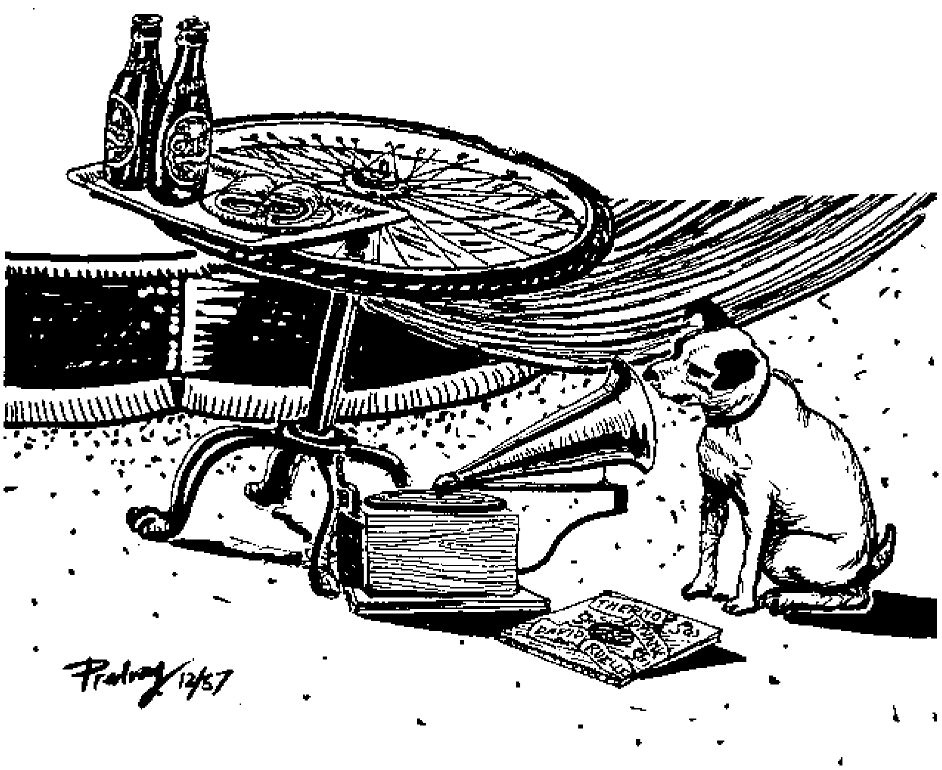
Classical and Quantum Chaos
Part I: Deterministic Chaos



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

version 10, Aug 26 2003 printed August 27, 2003
`www.nbi.dk/ChaosBook/ChaosBook` comments to: predrag@nbi.dk

Contents

Volume I: Classical chaos

Volume II: Quantum chaos

Volume www: Appendices on `www.nbi.dk/ChaosBook`

Contributors

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

This book is a result of collaborative labors of many people over a span of several decades. Coauthors of a chapter or a section are indicated in the byline to the chapter/section title. If you are referring to a specific coauthored section rather than the entire book, cite it as (for example):

> C. Chandre, F.K. Diakonos and P. Schmelcher, section "Discrete cyclist relaxation method", in P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner and G. Vattay, *Chaos: Classical and Quantum* (Niels Bohr Institute, Copenhagen 2003); `www.nbi.dk/ChaosBook`.

Chapters without a byline are written by Predrag Cvitanović. Friends whose contributions and ideas were invaluable to us but have not contributed written text to this book, are listed in the acknowledgements.

Roberto Artuso

Ronnie Mainieri

Gábor Vattay

Gregor Tanner

Acknowledgements

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

This book owes its existence to the Niels Bohr Institute's and Nordita's hospitable and nurturing environment, and the private, national and cross-national foundations that have supported the collaborators' research over a span of several decades. P.C. thanks M.J. Feigenbaum of Rockefeller University; D. Ruelle of I.H.E.S., Bures-sur-Yvette; I. Procaccia of the Weizmann Institute; P. Hemmer of University of Trondheim; The Max-Planck Institut für Mathematik, Bonn; J. Lowenstein of New York University; Edificio Celi, Milano; and Fundaçaõ de Faca, Porto Seguro, for the hospitality during various stages of this work, and the Carlsberg Foundation and Glen P. Robinson for support.

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

We thank Dorte Glass for typing parts of the manuscript; B. Lautrup and D. Viswanath for comments and corrections to the preliminary versions of this text; M.V. Berry for the quotation on page 603; H. Fogedby for the quotation on page 239; J. Greensite for the quotation on page 6; Ya.B. Pesin for the remarks quoted on page 612; E.A. Spiegel for quotations on page 1 and page 693.

Fritz Haake's heartfelt lament on page 206 was uttered at the end of the first conference presentation of cycle expansions, in 1988. G.P. Morriss advice to students as how to read the introduction to this book, page 4, was offerred during a 2002 graduate course in Dresden. Kerson Huang's interview of C.N. Yang quoted on page 111 is available on `www.nbi.dk/ChaosBook/extras`.

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

Profound thanks to all the unsung heroes - students and colleagues, too numerous to list here, who have supported this project over many years in many ways, by surviving pilot courses based on this book, by providing invaluable insights, by teaching us, by inspiring us.

Chapter 1

Overture

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

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

This chapter offers a quick survey of the main topics covered in the book. We start out by making promises - we will right wrongs, no longer shall you suffer the slings and arrows of outrageous Science of Perplexity. We relegate a historical overview of the development of chaotic dynamics to appendix A, and head straight to the starting line: A pinball game is used to motivate and illustrate most of the concepts to be developed in this book.

Throughout the book

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

fast track points you where to skip to

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

 indicates an exercise that might clarify a point in the text

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

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

1.1 Why 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 since 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. This book is a self-contained graduate textbook on classical and quantum chaos. 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 page 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. Nevertheless, a beautiful theory of deterministic chaos, of predictive quality comparable to that of the traditional perturbation expansions for nearly integrable systems, already exists.

In the traditional approach the integrable motions are used as zeroth-order approximations to physical systems, and weak nonlinearities are then accounted for perturbatively. For strongly nonlinear, non-integrable systems such expansions fail completely; 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 periodic motion of a harmonic oscillator.

So, what is chaos, and what is to be done about it? To get some feeling for how and why unstable cycles come about, we start by playing a game of pinball. The reminder of the chapter is a quick tour through the material covered in 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 The future as in a mirror

> All you need to know about chaos is contained in the introduction of the [Cvitanović *et al "Chaos: Classical and Quantum"*] book. However, in order to understand the introduction you will first have to read the rest of the book.
> Gary Morriss

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

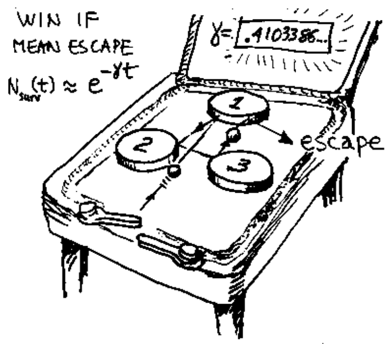


Figure 1.1: A physicist's bare bones game of pinball.

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. A physicist's pinball is free, frictionless, point-like, spin-less, perfectly elastic, and noiseless. Point-like pinballs are shot at the disks from random starting positions and angles; they spend some time bouncing between the disks and then escape.

At the beginning of 18th century Baron Gottfried Wilhelm Leibniz was confident that given the initial conditions one knew all what a deterministic system would do far into the future. He wrote [1.1], anticipating by century and a half the oft quoted Laplace's "Given for one instant an intelligence which could comprehend all the forces by which nature is animated...":

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

Leibniz chose to illustrate his faith in determinism precisely with the type of physical system that we shall use here as a paradigm of "chaos". His claim is wrong in a deep and subtle way: a state of a physical system can *never* be specified to infinite precision, 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.

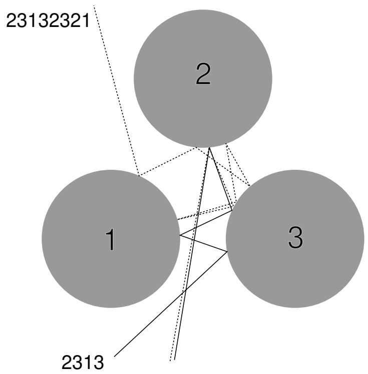


Figure 1.2: Sensitivity to initial conditions: two pinballs that start out very close to each other separate exponentially with time.

1.3.1 What is "chaos"?

> I accept chaos. I am not sure that it accepts me.
> Bob Dylan, *Bringing It All Back Home*

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

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

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

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

☞ sect. 8.3

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

Figure 1.3: Dynamics of a *chaotic* dynamical system is (a) everywhere locally unstable (positive Lyapunov exponent) and (b) globally mixing (positive entropy). (A. Johansen)

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. 10.1

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. 17.1

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) - fig. 1.3 - 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 Poincaré's vision of chaos as the interplay of local instability (unstable periodic orbits) and global mixing (intertwining of their stable and unstable manifolds). In a chaotic system any open ball of initial conditions, no matter how small, will in finite time overlap with any other finite region and in this sense spread over the extent of the entire asymptotically accessible 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 described by deterministic equations of motion - say, a bucket of boiling water described by the Navier-Stokes equations. But in practice the word "turbulence" tends to refer to messy dynamics which we understand poorly. As soon as a phenomenon is understood better, it is reclaimed and renamed: "a route to chaos", "spatiotemporal chaos", and so on.

☞ appendix B

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

$\Longrightarrow$ other swirls $\Longrightarrow$

For any finite spatial resolution, the system follows approximately for a finite time a pattern belonging to a finite alphabet of admissible patterns, and the long term dynamics can be thought of as a walk through the space of such patterns. Recasting this image into mathematics is what this book is about.

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

As in science popularizations too much has been made of the impact of the "chaos theory", a number of caveats are already needed 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. There are applications – such as spatially extended systems and statistical mechanics applications – where the few important degrees of freedom can be isolated and studied profitably by methods to be described here.

☞ sect. 2.4.1
☞ chapter 20

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.

1.4 A game of pinball

> Formulas hamper the understanding.
> S. Smale

We are now going to get down to the brasstacks. But first, a disclaimer: If you understand most of the rest of this chapter on the first reading, you either do not need this book, or you are delusional. If you do not understand it, is not because the people who wrote it are so much smarter than you: the most one can hope for at this stage is to give you a flavor of what lies ahead. If a statement in this chapter mystifies/intrigues, fast forward to a section indicated by ☞ on the margin, read only the parts that you feel you need. Of course, we think that you need to learn ALL of it, or otherwise we would not have written it up in the first place.

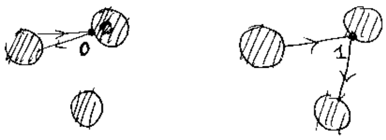


Figure 1.4: Binary labeling of the 3-disk pinball trajectories; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

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. 1.4 and 1.4.1. If successful, we can proceed with step III of sect. 1.5.1: investigate the *weights* of the different pieces of the system.

☞ sect. 2.4.1
☞ chapter 9
☞ chapter 10

We commence our analysis of the pinball game with steps I, II: diagnose, count. We shall return to step III – measure – in sect. 1.5.1.

☞ chapter 15

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
page 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.

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

☞ sect. 9.7

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. 30.2

1.4.1 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. 1.5 - but it is rather hard to perceive the systematics of orbits from their shapes. In mechanics a trajectory is fully and uniquely specified by its position and momentum at a given instant, and no two distinct phase space trajectories can intersect. Their projections on arbitrary subspaces, however, can and do intersect, in rather unilluminating ways. In the pinball example the problem is that we are looking at the projections of a 4-dimensional phase space trajectories onto a 2-dimensional subspace, the space coordinates. 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-

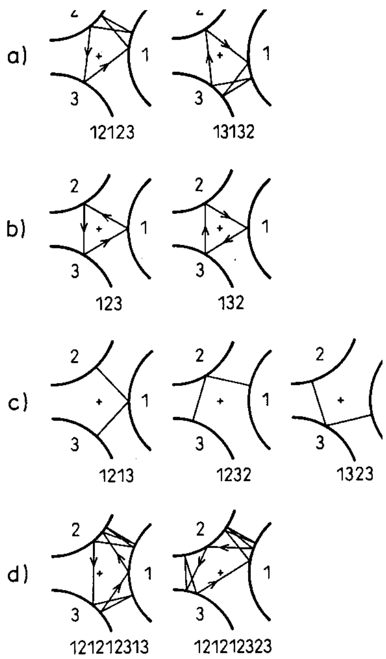


Figure 1.5: Some examples of 3-disk cycles: (a) $\overline{12123}$ and $\overline{13132}$ are mapped into each other by the flip across 1 axis. Similarly (b) $\overline{123}$ and $\overline{132}$ are related by flips, and (c) $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$ by rotations. (d) The cycles $\overline{121212313}$ and $\overline{121212323}$ are related only by time reversal. These symmetries are discussed in more detail in chapter 19. (from ref. [1.2])

dimensional map f that takes the coordinates of the pinball from one disk edge to another disk edge. Let us state this more precisely: the trajectory just after the moment of impact is defined by marking s_n, the arc-length position of the nth bounce along the billiard wall, and $p_n = p \sin \phi_n$ the momentum component parallel to the billiard wall at the point of impact, fig. 1.6. Such section of a flow is called a *Poincaré section*, and the particular choice of coordinates (due to Birkhoff) is particularly smart, as it conserves the phase-space volume. In terms of the Poincaré section, the dynamics is reduced to the *return map* $P : (s_n, p_n) \mapsto (s_{n+1}, p_{n+1})$ 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.

☞ sect. 5

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. 1.7 and 1.8. Provided that the disks are sufficiently separated, after n bounces the survivors are divided into 2^n distinct strips: the $\mathcal{M}_i$th strip consists of all points with itinerary $i = s_1 s_2 s_3 \ldots s_n$, $s = \{0, 1\}$. 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 nav-

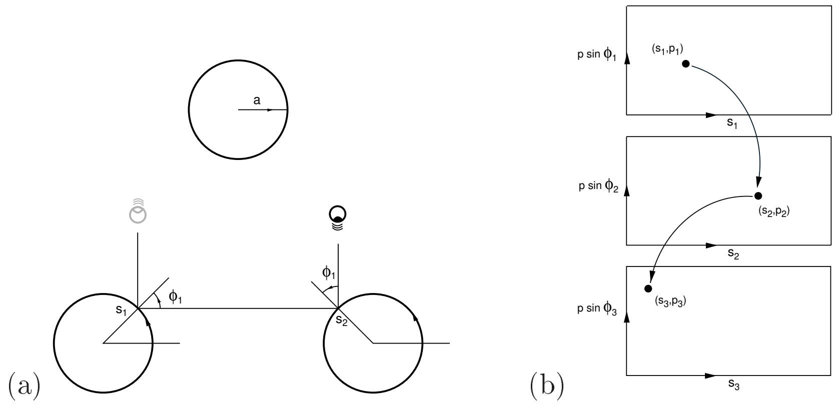
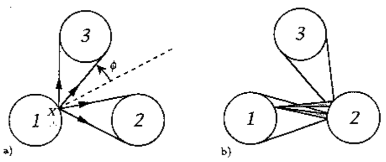


Figure 1.6: (a) The Poincaré section coordinates for the 3-disk game of pinball. (b) Collision sequence $(s_1, p_1) \mapsto (s_2, p_2) \mapsto (s_3, p_3)$ from the boundary of a disk to the boundary of the next disk presented in the Poincaré section coordinates.

Figure 1.7: (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 cones of initial conditions that hit more and more consecutive disks are nested within each other, as in fig. 1.8.



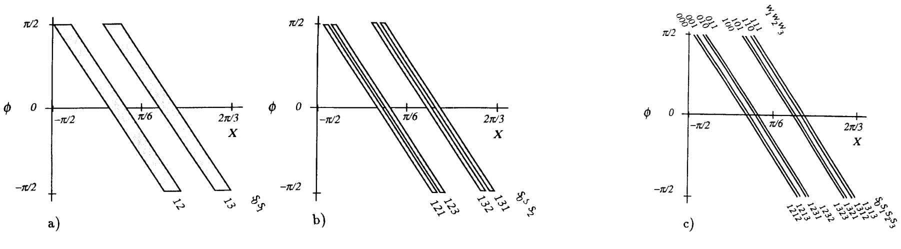


Figure 1.8: Ternary labeled regions of the 3-disk game of pinball phase space Poincaré section which correspond to trajectories that originate on disk 1 and remain confined for (a) one bounce, (b) two bounces, (c) three bounces. The Poincaré sections for trajectories originating on the other two disks are obtained by the appropriate relabeling of the strips (K.T. Hansen [1.3]).

igation 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.

1.4.2 Escape rate

☞ remark 7.1

What is a good physical quantity to compute for the game of pinball? Such system, for which almost any trajectory eventually leaves a finite region (the pinball table) never to return, is said to be open, or a *repeller*. The repeller *escape rate* is an eminently measurable quantity. An example of such 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
page 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. 1.8 again. In each bounce the initial conditions get thinned out, yielding twice as many thin strips as at the previous bounce. The total area that remains at a given time is the sum of the areas of the strips, so that the fraction of survivors after n bounces, or the *survival probability* is given by

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

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

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

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

1.5 Chaos for cyclists

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 $\gamma = 0.4103384077693464893384613078192\ldots$, you do not need this book. If you have no clue, hang on.

1.5.1 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

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

A trajectory of a pinball moving on a flat surface is specified by two position coordinates and the direction of motion, so in this case $d = 3$. Evaluation of a cycle Jacobian matrix is a 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 ☞ sect. 5.3

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 the beam of neighboring trajectories is defocused along the unstable eigendirection of the Jacobian matrix $\mathbf{J}$.

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

Each strip i in fig. 1.8 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 Jacobian matrix $\mathbf{J}^t(x_i)$ evaluated at the ith periodic point for $t = T_p$, the full period (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 be insensitive to smooth variations in the initial distribution.

☞ sect. 7.3

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

☞ sect. 11.1.1

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

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

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

Recall that for large n the 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:

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

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 (1.6), rather than a numerical extrapolation of a sequence of γ_n extracted from (1.3). As any finite truncation $n < n_{trunc}$ of (1.7) is a polynomial in z, convergent for any z, finding this pole requires that we know something about Γ_n for any n, and that might be a tall order. ☞ sect. 11.4

We could now proceed to estimate the location of the leading singularity of $\Gamma(z)$ from finite truncations of (1.7) by methods such as Padé approximants. However, as we shall now show, it pays to first perform a simple resummation that converts this divergence into a *zero* of a related function.

1.5.2 Dynamical zeta function

If a trajectory retraces a *prime* cycle r times, its expanding eigenvalue is 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 ✎ 10.5 page 195 ☞ sect. 4.4

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

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. 16.1
☞ sect. 12.4

Easy, you say: "Zeros of (1.9) can be read off the formula, a zero $z_p = |\Lambda_p|^{1/n_p}$ for each term in the product. What's the problem?" Dead wrong!

1.5.3 Cycle expansions

How are formulas such as (1.9) used? We start by computing the lengths and eigenvalues of the shortest cycles. This usually requires some numerical work, such as the Newton'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 14

☞ sect. 25.3

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

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

The virtue of the expansion is that the sum of all terms of the same total length n (grouped in brackets above) is a number that is exponentially smaller than a typical term in the sum, for geometrical reasons we explain in the next section.

☞ chapter 15

☞ sect. 15.1

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.

1.5.4 Shadowing

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

☞ sect. 15.1.3

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. 1.9; 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

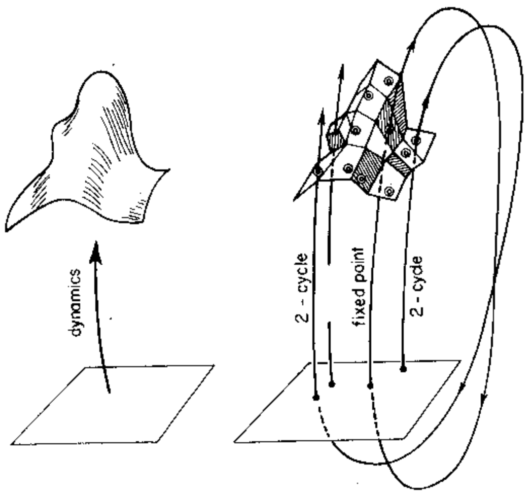


Figure 1.9: Approximation to (a) a smooth dynamics by (b) the skeleton of periodic points, together with their linearized neighborhoods. Indicated are segments of two 1-cycles and a 2-cycle that alternates between the neighborhoods of the two 1-cycles, shadowing first one of the two 1-cycles, and then the other.

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. 1.4, 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. 1.9.

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 13

1.6 Evolution

The above derivation of the dynamical zeta function formula for the escape rate has one shortcoming; it estimates the fraction of survivors as a function of the number of pinball bounces, but the physically interesting quantity is the escape rate measured in units of continuous time. For continuous time flows, the escape rate (1.2) is generalized as follows. Define a finite 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

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

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

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

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

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

☞ sect. 7.4.1

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

☞ chapter 26

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.

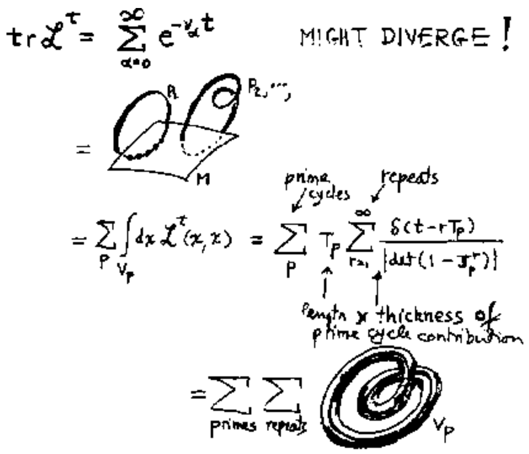


Figure 1.10: The trace of an evolution operator is concentrated in tubes around prime cycles, of length T_p and thickness $1/|\Lambda_p|^r$ for rth repeat of the prime cycle p.

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.

1.6.1 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. 11.3

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

This formula has a simple geometrical interpretation sketched in fig. 1.10. 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 $\operatorname{tr} e^{\mathcal{A}} = \sum e^{s_\alpha t}$, while the right-hand side equals zero everywhere except for the set $t = rT_p$. A Laplace transform smooths the sum over Dirac delta functions in cycle periods and yields the *trace formula* for the eigenspectrum $s_0, s_1, \cdots$ of the classical evolution operator:

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

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. 11.1

1.6.2 Spectral determinant

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

✎ 4.1 page 81

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

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 12

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

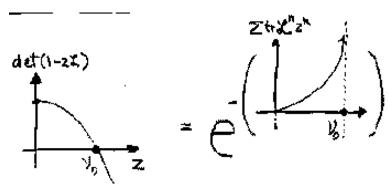


Figure 1.11: Spectral determinant is preferable to the trace as it vanishes smoothly at the leading eigenvalue, while the trace formula diverges.

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. 1.11; 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. 1.5.3.

☞ sect. 12.5.1

1.7 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 Boltzmann).

☞ chapter 20

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 18

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 [1.4] 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 [1.5], 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 [1.6] 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 [1.7], 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.

An introductory course should give students skills in qualitative and numerical analysis of dynamical systems for short times (trajectories, fixed points, bifurcations) and familiarize them with Cantor sets and symbolic dynamics for chaotic dynamics. With this, and a graduate level exposure to statistical mechanics, partial differential equations and quantum mechanics, the stage is set for any of the one-semester advanced courses based on this book. The courses taught so far start out with the introductory chapters on qualitative dynamics, symbolic dynamics and flows, and then continue in different directions:

Deterministic chaos. Chaotic averaging, evolution operators, trace formulas, zeta functions, cycle expansions, Lyapunov exponents, billiards, transport coefficients, thermodynamic formalism, period doubling, renormalization operators.

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

This book concentrates on the periodic orbit theory. The role of unstable periodic orbits was already fully appreciated by Poincaré [1.8, 1.9], 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 [1.10] 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 [1.11], an impressive compendium of modern dynamical systems theory. The fundamental papers in this field, all still valuable reading, are Smale [1.12], Bowen [1.13] and Sinai [1.14]. 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. Markov partitions were introduced by Sinai in ref. [1.15]. The classical text (though certainly not an easy read) on the subject of dynamical zeta functions is Ruelle's *Statistical Mechanics, Thermodynamic Formalism* [1.16]. 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 [1.17]. Further excellent mathematical references on thermodynamic formalism are Parry and Pollicott's monograph [1.18] with emphasis on the symbolic dynamics aspects of the formalism, and Baladi's clear and compact reviews of the theory dynamical zeta functions [1.19, 1.20].

A graduate level introduction to statistical mechanics from the dynamical point view is given by Dorfman [1.21]; the Gaspard monograph [1.4] covers the same ground in more depth. Driebe monograph [1.22] 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?"* [1.23].

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. [1.25, 1.26, 1.27].

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 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:

15.2
page 301

- **Dynamics**
  1. count prime cycles, exercise 1.1
  2. pinball simulator, exercise 5.1, exercise 14.4
  3. pinball stability, exercise 5.5, exercise 14.4
  4. pinball periodic orbits, exercise 14.6, exercise 14.5
  5. helium integrator, exercise 2.9, exercise 14.8
  6. helium periodic orbits, exercise 29.4, exercise 14.9

- **Averaging, numerical**
  1. pinball escape rate, exercise 12.12
  2. Lyapunov exponent, exercise 17.2
- **Averaging, periodic orbits**
  1. cycle expansions, exercise 15.1, exercise 15.2
  2. pinball escape rate, exercise 15.4, exercise 15.5
  3. cycle expansions for averages, exercise 15.1, exercise 16.3
  4. cycle expansions for diffusion, exercise 20.1
  5. desymmetrization exercise 19.1
  6. semiclassical quantization exercise 28.3
  7. ortho-, para-helium, lowest eigenenergies exercise 29.7

Solutions for some of the problems are included appendix O. 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] W. Parry and M. Pollicott, *Zeta Functions and the periodic Structure of Hyperbolic Dynamics*, *Astérisque* **187–188** (Société Mathématique de France, Paris 1990).

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

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

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

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

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

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

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

[1.26] F. Moss, "Chaos under control", *Nature* **370**, 615 (1994).

[1.27] J. Glanz,s *Science* **265**, 1174 (1994).

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

Exercises

Exercise 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. 10.4). 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.*

Exercise 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 15 you should be in position to make very accurate estimates).*

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 prerequisites 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 7.

fast track:
chapter 7, p. 105

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, we see the future in the present as in a mirror. The motion of the planets against the celestial firmament provides an example. Against the daily motion of the stars from East to West, the planets distinguish themselves by moving among the fixed stars. Ancients discovered that by

sect. 1.3

knowing a sequence of planet's positions - latitudes and longitudes - its future position could be predicted.

For the solar system, tracking the latitude and longitude in the celestial sphere suffices to completely specify the planet's apparent motion. All possible values for 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 tomorrow; or knowing the value of a stock market index today will not determine its value tomorrow. The phase space can be enlarged, in the hope that in a sufficiently 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 tomorrow. 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. We may think such sets a nuisance, but it is quite the contrary - they 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* $f^t : \mathcal{M} \to \mathcal{M}$ tells us 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. Note that time is not absolute, only

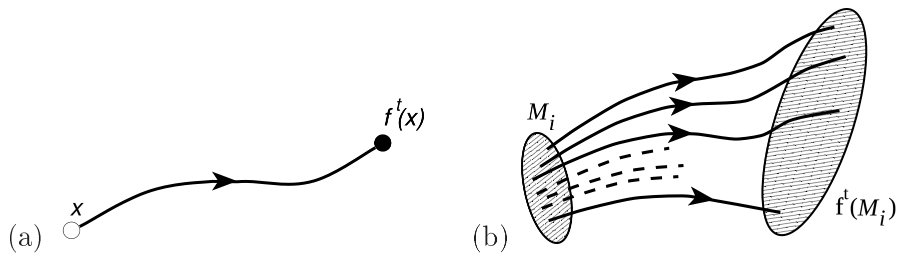


Figure 2.1: (a) A trajectory traced out by the evolution rule f^t. Starting from the phase space point x, after a time t, the point is at $f^t(x)$. (b) The evolution rule f^tcan be used to map a region $\mathcal{M}_i$ of the phase space into the region $f^t(\mathcal{M}_i)$.

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* of a *map*.

Nature provides us with innumerable dynamical systems. They manifest themselves through their trajectories: given an initial point x_0, the evolution rule traces out a sequence of points $x(t) = f^t(x_0)$, the *trajectory* through the point $x_0 = x(0)$. A trajectory is parameterized by the time t and thus belongs to $(f^t(x_0), t) \in \mathcal{M} \otimes \mathbb{R}$. 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 page 48

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

The subset of points in $\mathcal{M}$ that belong to the (possibly infinite) trajectory of a given point x_0 is called the *orbit* of x_0; we shall talk about forward orbits, backward orbits, periodic orbits. For a flow, an orbit is a continuous curve; for a map, a sequence of points.

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:

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

The ancient, no less than the contemporary quantum field theorists, tried to make sense of all dynamics in terms of periodic motions; epicycles, integrable systems. The embarassing truth is that for a generic dynamical systems 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

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

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

A *periodic orbit* corresponds to a trajectory that returns exactly to the initial point in a finite time. Periodic orbits form a very small subset of the phase space, in the same sense that rational numbers 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

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

In other words, the trajectory of a non-wandering point reenters the neighborhood $\mathcal{M}_0$ infinitely often. We shall denote by 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 (and you can prove that by the method of Lyapunov functionals, or any other method available in the literature), the flow is globally contracting onto a subset of $\mathcal{M}$ which we shall refer to as the *attractor*. The attractor may be unique, or there can coexist any number of distinct attracting sets, each with its own *basin of attraction*, the set of all points that fall into the attractor under foward evolution. The attractor can be a fixed point, a periodic orbit, aperiodic, or any combination of the above. The most interesting case is that of an aperiodic reccurent attractor, to which we shall refer loosely as a *strange attractor*.

☞ sect. 2.2

We say loosely, as will soon become apparent, that diagnosing and proving existence of a genuine, card carrying strange attractor is a highly nontrivial undertaking.

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 a 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. Quite to the contrary. As longer and longer cycles approximate more and more accurately finite segments of aperiodic trajectories, we shall establish control over non–wandering sets by defining them as the closures of the union of all periodic points.

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

2.2 Flows

> There is no beauty without some strangeness
> William Blake

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

2.2 page 48

(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. Why "semi-"group? It may fail to form a group if the dynamics is not reversible, and the rule $f^t(x)$ cannot be used to rerun the dynamics backwards in time, with negative t; with no reversibility, we cannot define the inverse $f^{-t}(f^t(x)) = f^0(x) = x$, in which case the family of mappings $f^t(x)$ does not form a group. In exceedingly many situations of interest - for times beyond the Lyapunov time, for asymptotic attractors, for infinite dimensional systems, for systems with noise, for non-invertible maps - the dynamics cannot be run backwards in time, hence, the circumspect sect. 2.4.1

emphasis on *semi*groups. On the other hand, there are many settings of physical interest, where dynamics is reversible (such as finite-dimensional Hamiltonian flows), and where the family of evolution maps f^t does form a group.

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

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

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

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

2.3 page 48

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

$$v(x) = \dot{x}(t)\,. \tag{2.5}$$

Newton's laws, Lagrange's method, or Hamilton's method are all familiar procedures for obtaining a set of differential equations for the vector field $v(x)$ that describes the evolution of a mechanical system. Equations of mechanics may appear different in form from (2.5), as they are often involve higher time derivatives, but an equation that is second or higher order in time can always be rewritten as a set of first order equations.

We are concerned here with a much larger world of general flows, mechanical or not, all defined by a time independent vector field (2.5). At each point of the phase space a vector indicates the local direction in which the orbit evolves. The length of the vector $|v(x)|$ is proportional to the speed at the point x, and the direction and length of $v(x)$ changes from point to point. When the phase space is a complicated manifold embedded in $\mathbb{R}^d$, one can no longer think of the vector field as being embedded in the 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. The vector field lives in the union of all these tangent planes, a space called the *tangent bundle* $\mathbf{T}\mathcal{M}$.

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

$$\begin{aligned}\dot{x}(t) &= y(t)\\ \dot{y}(t) &= 0.15\,y(t) - x(t) + x(t)^3\end{aligned} \tag{2.6}$$

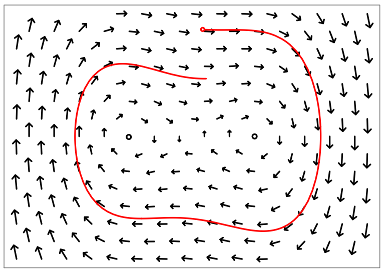
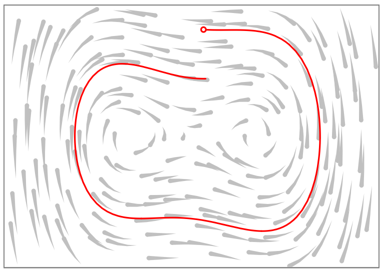
(a)  (b) 

Figure 2.2: (a) The two-dimensional vector field for the Duffing system (2.6), together with a short trajectory segment. (b) The flow lines. Each "comet" represents the same time interval of a trajectory, starting at the tail and ending at the head. The longer the comet, the faster the flow in that region.

plotted in fig. 2.2. The velocity 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 $\mathbf{T}\mathcal{M}$.

2.4 page 48

$$\text{If} \qquad v(x_q) = 0\,, \tag{2.7}$$

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 passing though x_0 at time $t = 0$ can be obtained by integrating the equations (2.5):

$$x(t) = f^t(x_0) = x_0 + \int_0^t d\tau\, v(x(\tau))\,, \qquad x(0) = x_0\,. \tag{2.8}$$

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. A non-autonomous system

$$\frac{dy}{d\tau} = w(y, \tau)\,, \tag{2.9}$$

can always be converted 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

$$v(x) = \left[\begin{array}{c} w(y,\tau) \\ 1 \end{array} \right]. \tag{2.10}$$

2.5 page 48

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

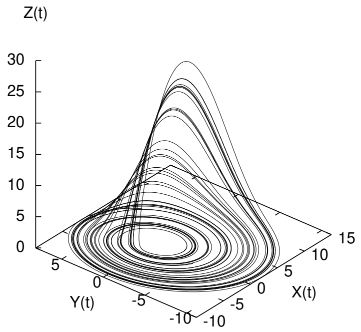


Figure 2.3: A trajectory of the Rössler flow at time $t = 250$. (G. Simon)

Example 2.2 A flow with a strange attractor: *A concrete example of an autonomous flow is the* Rössler system

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c)\,, \qquad a = b = 0.2\,, \quad c = 5.7\,. \end{aligned} \tag{2.11}$$

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. In order to get a feel for what typical 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 shrinks with time, so the flow is contracting. *All trajectories seem to converge to a* strange attractor. *We say "seem", as there exists no proof that such attractor is asymptotically aperiodic - it might well be that what we see is but a long transient on a way to an attractive periodic orbit. For now, accept that fig. 2.3 and similar figures in what follows are examples of "strange attractors".*

3.5
page 62

fast track:
chapter 2.3, p. 40

2.3 Computing trajectories

There is lots of strangeness without beauty
Benny Lautrup

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

$$x_i \to x_i + \delta\tau v_i(x)\,. \tag{2.12}$$

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 is neither the time nor the place to expand on them; how you learn them is your business. And if you have developed some nice routines for solving problems in this text or can point another student to some, let us know.

2.6 page 49
2.7 page 49
2.8 page 49
2.9 page 50

In the next section we dispose of the fear of "infinite-dimensional" dynamical systems - you might prefer to skip the section on first reading.

fast track:
chapter 3, p. 51

2.4 Infinite-dimensional flows

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

There is only one honorable cause that would justify sweating through so much formalism - this is but the sharpening of a pencil in order that we may attack the Navier-Stokes equation,

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t}+\mathbf{u}\cdot\nabla\mathbf{u}\right)=-\nabla p+\eta\left(\nabla^2\mathbf{u}+\frac{1}{3}\nabla(\operatorname{div}\mathbf{u})\right)+\mathbf{f}\,, \tag{2.13}$$

and solve the problem of turbulence. Being realistic, we are not so foolhardy to immediately plunge into *the* problem. Too many dimensions and indices - instead, we start small, in one spatial dimension, $\mathbf{u}\to u$, $\mathbf{u}\cdot\nabla\mathbf{u}\to\frac{1}{2}\partial_x u^2$, assume constant ρ, forget about the pressure p, and so on. This line of reasoning, as well as many other equally sensible threads of thought, such as the amplitude equations obtained via weakly nonlinear stability analysis of steady flows, lead to the essentially same nonlinear PDEs, like the one that we turn to in the next section.

Flows described by partial differential equations are considered infinite dimensional because if one writes them down as a set of ordinary differential

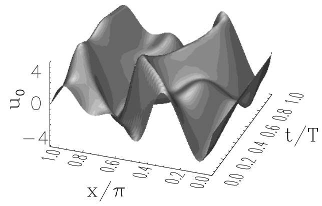


Figure 2.4: Spatiotemporally periodic solution $u_0(x, t)$. We have divided x by π and plotted only the $x > 0$ part, since we work in the subspace of the odd solutions, $u(x, t) = -u(-x, t)$. $N = 16$ Fourier modes truncation with $\nu = 0.029910$. (From ref. [2.7])

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.

in depth:
appendix B, p. 617

2.4.1 Fluttering flame front

The Kuramoto-Sivashinsky system, 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

$$u_t = (u^2)_x - u_{xx} - \nu u_{xxxx} \,. \tag{2.14}$$

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" (more correctly the "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. We take note, as in the Navier-Stokes equation (2.13), of the $u\partial_x u$ "inertial" term, the $\partial_x^2 u$ "diffusive" term (both with a "wrong" sign), etc.

The term $(u^2)_x$ makes this a *nonlinear system*. It is the simplest conceivable nonlinear PDE, playing the role in the theory of spatially extended systems analogous to the role that the x^2 nonlinearity plays in the dynamics of iterated mappings. The time evolution of a solution of the Kuramoto-Sivashinsky system is illustrated by fig. 2.4. How are such solutions computed? The salient feature of such partial differential equations is a theorem

☞ sect. 3.3

☞ remark 2.1

saying that for any finite value of the phase-space contraction parameter ν, the asymptotic dynamics is describable by a *finite* set of "inertial manifold" ordinary differential equations.

We are studying a "flame front" $u(x,t) = u(x+2\pi,t)$ 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:

$$u(x,t) = \sum_{k=-\infty}^{+\infty} b_k(t) e^{ikx} \,. \tag{2.15}$$

Since $u(x,t)$ is real, $b_k = b^*_{-k}$. Substituting (2.15) into (2.14) yields the infinite ladder of evolution equations for the Fourier coefficients b_k:

$$\dot{b}_k = (k^2 - \nu k^4) b_k + ik \sum_{m=-\infty}^{\infty} b_m b_{k-m} \,. \tag{2.16}$$

As it follows from this equation that $\dot{b}_0 = 0$, the solution integrated over space is constant in time. We shall consider only the case of this average - the mean value of u - equal to zero, $b_0 = \int dx\, u(x,t) = 0$.

The coefficients b_k are in general complex functions of time t. We can isolate a smaller subspace of the system (2.16) further by considering the case of b_k pure imaginary, $b_k = i a_k$, where a_k are real, with the evolution equations

$$\dot{a}_k = (k^2 - \nu k^4) a_k - k \sum_{m=-\infty}^{\infty} a_m a_{k-m} \,. \tag{2.17}$$

This picks out the subspace of odd solutions $u(x,t) = -u(-x,t)$, so $a_{-k} = -a_k$. By picking this subspace we eliminate the continuous translational symmetry from our consideration; that is probably not an option for an experimentalist, but will do for our purposes.

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 (2.14), but that is basically all we know as far as analytical solutions are concerned. You can integrate numerically the Fourier modes (2.17), 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 such truncation is called a

Galerkin truncation. For the parameter values explored below, $N \leq 16$ truncations were deemed sufficiently accurate. In other words, even though our starting point (2.14) is an infinite-dimensional dynamical system, the asymptotic dynamics unfolds on a finite-dimensional attracting manifold, and so we are back on the familiar territory of sect. 2.2: the theory of a finite number of ODEs applies to this infinite-dimensional PDE as well. 2.6 page 49

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 (2.15), as in fig. 2.4.

2.4.2 Fourier modes truncations

Thinking is extra price
Fernando Solla

Consider now the case of initial a_k sufficiently small that the bilinear $a_m a_{k-m}$ terms in (2.17) can be neglected. Then we have a set of decoupled linear equations for a_k whose solutions are exponentials, at most a finite number for which $k^2 > \nu k^4$ growing with time, and the infinity of modes for which $\nu k^4 > k^2$ decaying with time. The growth of the unstable long wavelengths (low $|k|$) excites the short wavelengths through the nonlinear term in (2.17). The excitations thus transferred are dissipated by the strongly damped short wavelengths, and a "chaotic equilibrium" can emerge. The very short wavelengths $|k| \gg 1/\sqrt{\nu}$ 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. As the damping parameter decreases, the solutions increasingly take on shock front character poorly represented by the Fourier basis, and many higher harmonics may need to be kept in truncations of (2.17).

Hence, while one may truncate the high modes in the expansion (2.17), 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 (2.17) is that the dynamics is difficult to visualize. The best we can do without much programming is to examine the trajectory's projections onto any three axes a_i, a_j, a_k, as in fig. 2.5.

☞ sect. 3.3

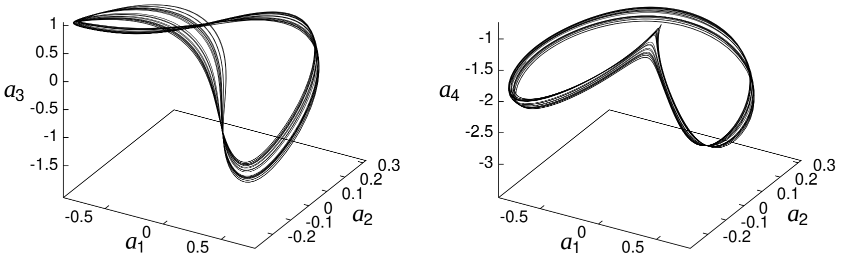


Figure 2.5: Projections of a typical 16-dimensional trajectory onto different 3-dimensional subspaces, coordinates (a) $\{a_1, a_2, a_3\}$, (b) $\{a_1, a_2, a_4\}$. $N = 16$ Fourier modes truncation with $\nu = 0.029910$. (From ref. [2.7].)

Examination of numerical plots such as fig. 2.5 suggests that a more thoughtful approach would be to find a coordinate transformation $y = h(x)$ to a "center manifold", such that in the new, curvilinear coordinates large scale dynamics takes place in (y_1, y_2) coordinates, with exponentially small dynamics in $y_3, y_4 \cdots$. But - thinking is extra price - we do not know how to actually accomplish that.

☞ chapter 6

We can now start to understand the remark on page 38 that for infinite dimensional systems time reversability is not an option: evolution forward in time strongly damps the higher Fourier modes. There is no turning back: if we reverse the time, the infinity of high modes that contract strongly forward in time now explodes, instantly rendering evolution backward in time meaningless. As everything in dynamics, this claim is also wrong, in a subtle way: if the initial $u(x, 0)$ is *in* the non–wandering set (2.2), the trajectory is well defined both forward and backward in time. For practical purposes, this subtlety is not of much use, as any time-reversed numerical trajectory in a finite-mode truncation will explode very quickly, unless special precautions are taken.

Commentary

Remark 2.1 Model ODE and PDE systems. Rössler system was introduced in ref. [2.2], as a simplified set of equations describing time evolution of concentrations of chemical reagents. The Duffing system (2.6) arises in study of electronic circuits. The theorem on finite dimenionality of inertial manifolds of phase-space contracting PDE flows is proven in ref. [2.4]. The Kuramoto-Sivashinsky equation was introduced in refs. [2.5, 2.6]; sect. 2.4 is based on V. Putkaradze's term project (see `wwwcb/extras`), and on the Christiansen *et al.* article [2.7]. 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. [2.8] - and this one is perhaps the simplest physically interesting spatially extended nonlinear system.

Remark 2.2 Diagnosing chaos. In sect. 1.3.1 we have stated that a deterministic system exhibits "chaos" if the dynamics is locally unstable (positive Lyapunov exponent) and globally mixing (positive entropy). In sect. 8.3 we shall define Lyapunov exponents, and discuss their evaluation, but already at this point it would be handy to have a few quick numerical methods to diagnose chaotic dynamics. Laskar's *frequency analysis* method [2.10] is useful for extracting quasi-periodic and weakly chaotic regions of phase space in Hamiltonian dynamics with many degrees of freedom. For references to several other numerical methods, see ref. [2.11].

Résumé

A *dynamical system* – a flow, or an iterated map – is defined by specifying a pair $(\mathcal{M}, f)$, where $\mathcal{M}$ is the phase 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 $\mathcal{M}$.

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

References

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

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

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

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

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

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

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

[2.8] 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).

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

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

[2.11] Ch. Skokos, *J. Phys* **A 34**, 10029 (2001).

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

Exercises

Exercise 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.

Exercise 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?

Exercise 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 $\dot{x} = x(t+1)$?*

Exercise 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.*

Exercise 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.*

Exercise 2.6 **Runge-Kutta integration.** *Implement the fourth-order Runge-Kutta integration formula (see, for example, ref. [2.9]) for $\dot{x} = v(x)$:*

$$
\begin{aligned}
x_{n+1} &= x_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + O(\delta\tau^5) \\
k_1 &= \delta\tau v(x_n)\,, \qquad k_2 = \delta\tau v(x_n + k_1/2) \\
k_3 &= \delta\tau v(x_n + k_2/2)\,, \quad k_4 = \delta\tau v(x_n + k_3)
\end{aligned}
\tag{2.18}
$$

or some other numerical integration routine.

Exercise 2.7 **Rössler system.** *Use the result of exercise 2.6 or some other integration routine to integrate numerically the Rössler system (2.11). Does the result look like a "strange attractor"?*

Exercise 2.8 **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 2.6. The equation $\ddot{x} = -x$ with initial conditions $x(0) = 2$ and $\dot{x} = 0$ has as solution $x(t) = e^{-t}(1 + e^{2\,t})$. 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.*

Exercise 2.9 **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 (5.3)*

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

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. 29.1. In the transformed coordinates* (x_1, x_2, p_1, p_2) *the Hamiltonian equations of motion take the form (6.9).*

(a) *Integrate the equations of motion by the fourth order Runge-Kutta computer routine of exercise 2.6 (or whatever integration routine you like). A convenient way to visualize the 3-*d *phase space orbit is by projecting it onto the 2-dimensional* $(r_1(t), r_2(t))$ *plane.*

(Gregor Tanner, Per Rosenqvist)

Exercise 2.10 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 go through a zero, or as complicated as having the flow cut through a curved hypersurface.

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

$$x_{n+1} = P(x_n)\,, \qquad x_n, x_{n+1} \in \mathcal{P}\,, \qquad n \in \mathbb{Z}^+ \tag{3.1}$$

(for economy of notation, the *maps* of this chapter will be taken d-dimensional, the associated *flows* $(d+1)$-dimensional). The choice of the section hypersurface $\mathcal{P}$ is altogether arbitrary. With a sufficiently clever choice of a Poincaré section or a set of sections, any orbit of interest intersects a section. Depending on the application, one might need to supplement the return map with the time of *first return function* $\tau(x_n)$ - sometimes referred 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

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

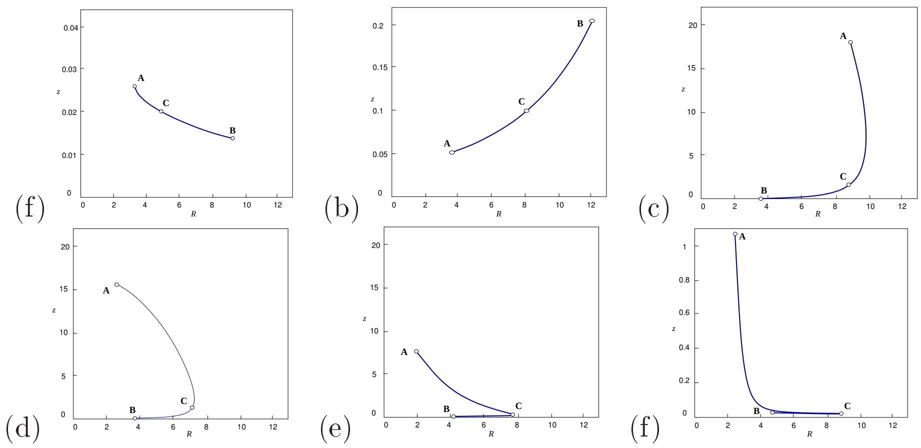


Figure 3.1: A sequence of Poincaré sections of the Rössler flow strange attractor, defined by planes through the z axis, oriented at angles (a) 270^o (b) 135^o, (c) 90^o, (d) 45^o, (e) 0^o, (f) 315^o, in the x-y plane. In order to guide the eye, a continuous line is drawn from a point A on the inner edge of the attractor to the point B on the outer edge (based on a calculation by G. Simon).

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 8

A few examples may help visualize this.

Example 3.1 Pendulum. *The phase space of a simple pendulum is 2-dimensional: momentum on the vertical axis and position on the horizontal axis. We choose the Poincaré section to be the positive horizontal axis. Now imagine what happens as a point traces a trajectory through this phase space. In the pendulum all orbits are loops, so any trajectory will periodically intersect the line, that is the Poincaré section, at one point. Consider next a pendulum with friction. Now every trajectory is an inward spiral, and the trajectory will intersect the Poincaré section at a series of points that get closer and closer to the origin.*

Pendulum dynamics is so simple that you can sketch it yourself on a piece of paper. Two next two examples offer a better illustration of the utility of visualization of dynamics by means of Poincaré sections.

Example 3.2 Rössler attractor. *Consider fig. 2.3, a typical trajectory of the 3-dimensional Rössler flow (2.11). It wraps around the z axis, so a good choice for a Poincaré section is a plane passing through the z axis. A sequence of such Poincaré sections placed radially at increasing angles with respect to the x axis, fig. 3.1, illustrates the "stretch & fold" action of the Rössler flow. To orient yourself, compare with fig. 2.3, and note the different z axis scales. A line segment $[A, B]$, traversing the width of the attractor, starts out close to the x-y plane, and after the stretching (b) $\to$ (c) $\to$ (d) followed by the folding (d) $\to$ (e) $\to$ (f), the folded segment is returned close to the*

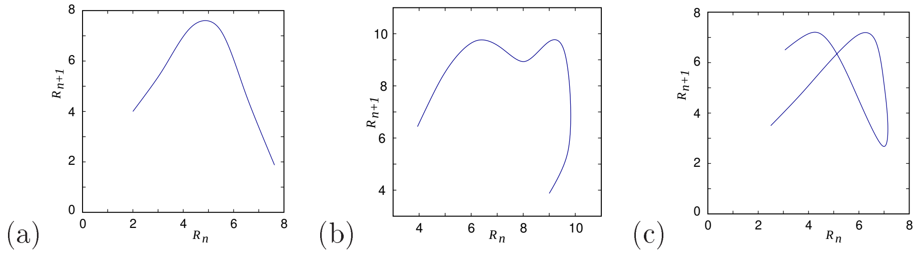


Figure 3.2: Return maps for the $R_n \to R_{n+1}$ radial distance for three distinct Poincaré sections for the Rössler flow, at angles (a) 0^o, (b) 90^o, (c) 45^o around the z-axis, see fig. 3.1. (G. Simon and A. Johansen)

x-y plane strongly compressed. In one Poincaré return the $[A, B]$ interval is stretched, folded and mapped onto itself, so the flow is expanding. *It is also* mixing, *as in one Poincaré return the point C from the interior of the attractor is mapped into the outer edge, while the edge point B lands in the interior.*

Once a particular Poincaré section is picked, we can also exhibit the return map (3.1), as in fig. 3.2. The case (a) is an example of a nice 1-to-1 return map. However, (b) and (c) appear multimodal and non-invertible, 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 $R_n \to R_{n+1}$.

The dynamics of high-dimensional flows, such as a truncation of the infinite tower of the Kuramoto-Sivashinsky modes (2.17) can be difficult to visualize. The question then is: how to look at such a flow? It is not clear that restricting the dynamics to a Poincaré section necessarily helps - after all, a section reduces a $(d+1)$-dimensional flow to a d-dimensional map, and how much is gained by replacing a continuous flow in 16 dimensionas by a set of points in 15 dimensions? Nevertheless, as we now show, much can be gleaned by examining trajectory's projections onto a subspace spanned by two or three coordinate axes.

Example 3.3 Kuramoto-Sivashinsky attractor. *Consider an N-mode truncation to the infinite tower of the Kuramoto-Sivashinsky coupled Fourier modes (2.17). We fix (arbitrarily) the Poincaré section to be the hyperplane $a_1 = 0$, and integrate (2.17) with the initial conditions $a_1 = 0$, and arbitrary values of the coordinates $a_2, \dots, a_N$ (where N is the truncation order). When a_1 becomes 0 the next time and the flow crosses the $a_1 = 0$ hyperplane in the same direction as initially, the coordinates $a_2, \dots, a_N$ are mapped into $(a'_2, \dots a'_N) = P(a_2, \dots, a_N)$, where P is the Poincaré mapping (3.1) of the $(N-1)$-dimensional $a_1 = 0$ hyperplane into itself. Fig. 3.3 is an example of a typical result. We have picked - arbitrarily - a subspace such as $a_6(n+1)$ vs. $a_6(n)$ in order to visualize the dynamics. While the topology of the attractor is still obscure, one thing is clear: even though the flow is infinite dimensional, the attractor is finite and thin, barely thicker than a line.*

fast track:
sect. 3.3, p. 55

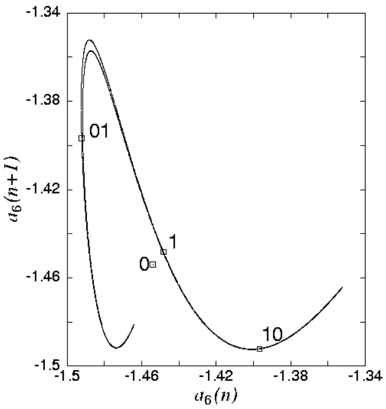


Figure 3.3: The attractor of the Kuramoto-Sivashinsky system (2.17), plotted as the a_6 component of the $a_1 = 0$ Poincaré section return map. Here 10,000 Poincaré section returns of a typical trajectory are plotted. Also indicated are the periodic points 0, 1, 01 and 10. $N = 16$ Fourier modes truncation with $\nu = 0.029910$. (From ref. [2.7].)

The above examples illustrate why a Poincaré section gives a more informative snapshot of the flow than the full flow portrait. For example, while the full flow portrait of the Rössler flow fig. 2.3 gives us no sense of the thickness of the attractor, we see clearly in the Rössler Poincaré sections fig. 3.1 that even though the return map is $2\text{-}d \to 2\text{-}d$, the flow contraction happens to be so strong that for all practical purposes it renders the return map 1-dimensional. Similarly, while no fine structure is discernable in the full flow portraits of Kuramoto-Sivashinsky dynamics fig. 2.5, the Poincaré return map fig. 3.3 reveals the fractal structure in the asymptotic attractor.

3.2 Constructing a Poincaré section

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

☞ remark 3.1

Consider the system (2.5) 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 in this plane if it satisfies the condition

3.2 page 61

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

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. 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

$$\frac{dt}{dx_1} = \frac{1}{v_1}\,,\ \cdots\,,\ \frac{dx_d}{dx_1} = \frac{v_d}{v_1}\,. \tag{3.6}$$

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 at the trajectory intersection point. If the section crossing is transverse, v_1 cannot vanish in the short segment bracketed by the integration step preceeding the section, and the point on the Poincaré section.

Of course, once you have determined the Poincaré section point, you use it as the initial point of the next segment of the trajectory, integrated again in the time variable, not in x_1, as globally v_1 vanishes at least twice for each complete Poincaré section return.

3.3 Do it again

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

$$f : \mathcal{M} \to \mathcal{M}\,, \tag{3.7}$$

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

$$\{f_A, f_B, \dots f_Z\} : \mathcal{M} \to \mathcal{M}\,. \tag{3.8}$$

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

$$f_Z(\cdots f_B(f_A(x))) \cdots) = f_Z \circ \cdots f_B \circ f_A(x)\,,$$

and the nth iterate of map f by

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

☞ sect. 2.1

The *trajectory* of x is the set of points

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

and the *orbit* of x is the subset of all points of $\mathcal{M}$ that can be reached by iterations of f.

The functional form of Poincaré return maps P such as figs. 3.2 and 3.3 can be approximated by tabulating the results of integration of the flow from x to the first Poincaré section return for many $x \in \mathcal{P}$, and constructing a function that interpolates through these points. If we find a good approximation to $P(x)$, we can get rid of numerical integration altogether, by replacing the continuous time trajectory $f^t(x)$ by iteration of the Poincaré return map $P(x)$. Multinomial approximations

$$P_k(x) = a_k + \sum_{j=1}^{d+1} b_{kj}x_j + \sum_{i,j=1}^{d+1} c_{kij}x_ix_j + \dots\,, \qquad x \in \mathcal{P} \tag{3.9}$$

to Poincaré return maps

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

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

Example 3.4 Hénon map. *The map*

$$\begin{aligned} x_{n+1} &= 1 - ax_n^2 + by_n \\ y_{n+1} &= x_n \end{aligned} \qquad (3.10)$$

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

$$x_{n+1} = 1 - ax_n^2 + bx_{n-1} \,. \qquad (3.11)$$

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 n 1-step recurrence relations.

The Hénon map is the simplest map that captures the "stretch & fold" dynamics of return maps such as the Rössler's, fig. 3.1. It can be obtained by a truncation of a polynomial approximation (3.9) to a Poincaré return map to second order.

A quick sketch of the long-time dynamics of such mapping, such as fig. 3.4, is obtained by picking an arbitrary starting point and iterating (3.10) on a computer. We plot here the dynamics in the (x_n, x_{n+1}) plane, rather than in the (x_n, y_n) plane, because we think of the Hénon map as a model return map $x_n \to x_{n+1}$. As we shall soon see, periodic orbits will be key to understanding the long-time dynamics, so we also plot a typical periodic orbit of such system, in this case an unstable period 7 cycle. Numerical determination of such cycles will be explained in sect. 25.1. 3.4 page 61

Example 3.5 Lozi map. *Another example frequently employed is the* Lozi map*, a linear, "tent map" version of the Hénon map given by*

$$\begin{aligned} x_{n+1} &= 1 - a|x_n| + by_n \\ y_{n+1} &= x_n \,. \end{aligned} \qquad (3.12)$$

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

What we get by iterating such maps is - at least qualitatively - not unlike what we get from Poincaré section of flows, figs. 3.2 and 3.3. 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-straddle methods and so on. 3.5 page 62

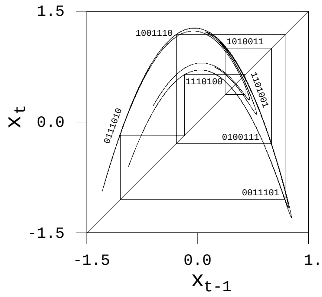


Figure 3.4: The strange attractor and an unstable period 7 cycle of the Hénon map (3.10) with $a = 1.4$, $b = 0.3$. The periodic points in the cycle are connected to guide the eye. (K.T. Hansen [1.3])

Example 3.6 Parabola. *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 (3.11). In fig. 3.4 b is rather large, $b = 0.3$, hence the attractor is rather thick, with the transverse fractal structure clearly visible. For vanishingly small b the Hénon map reduces to the 1-dimensional quadratic map*

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

3.6
page 62

By setting $b = 0$ we lose determinism, as on reals the inverse of map (3.13) has two preimages $\{x^+_{n-1}, x^-_{n-1}\}$ for most x_n. Still, this 1-dimensional approximation is very instructive.

As we shall see in sect. 9.4, 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.

appendix I.4

Commentary

Remark 3.1 Determining a Poincaré section. The idea of changing the integration variable from time to one of the coordinates, although simple, avoids the alternative of having to interpolate the numerical solution to determine the intersection. The trick described in sect. 3.2 was published by Hénon.

Remark 3.2 Hénon, Lozi maps. The Hénon map *per se* is of no particular physical import - 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 [2.1]. Y. Pomeau's studies of the Lorenz attractor on an analog computer, and his insights into its stretching and folding dynamics motivated Hénon [3.1] to introduce the Hénon mapping in 1976. Hénon's and Lorenz's original papers can be found in reprint collections refs. [3.2, 3.3]. They are a pleasure to read, and are still the best introduction to the physics motivating such models. Detailed description of the Hénon map dynamics was given by Mira and coworkers [3.4], as well as very many other authors.

The Lozi map [3.6] 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 [3.7], and the existence of the SRB measure was established by L.-S. Young.

☞ sect. 7.1

Résumé

Visualization of strange attractors is greatly facilitated by felicitous choice of Poincaré sections, and reduction of flows to Poincaré section return maps, which motivates study of discrete time dynamical systems generated by iteration of maps.

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

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

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

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

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

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

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

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

[3.13] G. Vattay, A. Wirzba and P.E. Rosenqvist, *Phys. Rev. Lett.* **73**, 2304 (1994).

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

[3.15] Ya.G. Sinai, *Usp. Mat. Nauk* **25**, 141 (1970).

[3.16] L.A. Bunimovich, *Funct. Anal. Appl.* **8**, 254 (1974).

[3.17] L.A. Bunimovich, *Comm. Math. Phys.***65**, 295 (1979).

[3.18] L. Bunimovich and Ya.G. Sinai, *Markov Partition for Dispersed Billiard, Comm. Math. Phys.* **78**, 247 (1980); **78**, 479 (1980); *Erratum, ibid.* **107**, 357 (1986).

[3.19] M. Benedicks and L.-S. Young, *Ergodic Theory & Dynamical Systems* **12**, 13–37 (1992).

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

Exercises

Exercise 3.1 **Rössler system** *(continuation of exercise 2.7) 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?*

Exercise 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*

$$\frac{dx_k}{ds} = \kappa f_k \,, \tag{3.14}$$

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*

$$x_{n+1} = g(x) \,. \tag{3.15}$$

How can this be used to find the Poincaré section?

Exercise 3.3 **Classical collinear helium dynamics.** *(continuation of exercise 2.9)*

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 (5.3). Compare your results with fig. 29.3(b).*

(Gregor Tanner, Per Rosenqvist)

Exercise 3.4 **Hénon map fixed points.** *Show that the two fixed points* (x_0, x_0), (x_1, x_1) *of the Hénon map (3.10) are given by*

$$\begin{aligned} x_0 &= \frac{-(1-b) - \sqrt{(1-b)^2 + 4a}}{2a} \,, \\ x_1 &= \frac{-(1-b) + \sqrt{(1-b)^2 + 4a}}{2a} \,. \end{aligned} \tag{3.16}$$

Exercise 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?*

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?*

Chapter 4

Local stability

(R. Mainieri and P. Cvitanović)

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

We now turn to the first, local class of such invariants, linear stability of flows and maps. This will give us metric information about local dynamics.

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 $x(x_0, t) = f^t(x_0)$. The system of linear *equations of variations* for the displacement of the infinitesimally close neighbor $x + \delta x$ follows from the flow equations (2.5) by Taylor expanding to linear order

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

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

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

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

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

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

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

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

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

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

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

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

which describes the deformation of an infinitesimal neighborhood at finite time t in the co-moving frame of $x(t)$, that is transformation of the initial, Lagrangian coordinate frame into the current, Eulerian coordinate frame.

As this is a deformation in the linear approximation, you can think of it as a linear deformation of an infinitesimal sphere envelopping x_0 into an ellipsoid around $x(t)$, described by the eigenvectors and eigenvalues of the Jacobian matrix of the linearized flow. Nearby trajectories separate along the *unstable directions*, approach each other along the *stable directions*, and maintain their distance along the *marginal directions*. In the literature adjectives *neutral* or *indifferent* are often used instead of "marginal".

One of the eigendirections is what one might expect, the direction of the flow itself. To see that, consider two points along a trajectory separated by infinitesimal flight time δt: $\delta x(0) = f^{\delta t}(x_0) - x_0 = v(x_0)\delta t$. At time t later

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

while from (4.5)

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

Dividing both equations by δt we observe that $\mathbf{J}^t(x_0)$ transports the velocity vector at x_0 to the velocity vector at $x(t)$ time t later:

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

$v(x(t))$ in general does not point in the same direction as $v(x_0)$, so this is not an eigenvalue condition for $\mathbf{J}^t$; Jacobian matrix computed for an arbitrary segment of an arbitrary trajectory has no invariant meaning. As the eigenvalues of $\mathbf{J}^t$ have invariant meaning only for periodic orbits, we postpone their interpretation to sect. 4.5. However, already at this stage we see that if the orbit is periodic, $x(T) = x(0)$, $v(x_0)$ is an eigenvector of the Jacobian matrix $\mathbf{J}^T(x_0)$ with a unit eigenvalue.

As we started by assuming that we know the equations of motion, from (4.3) we also know $\mathbf{A}$, the instantaneous rate of shear of an infinitesimal neighborhood $\delta x_i(t)$ of the trajectory $x(t)$. What we do not know is the finite time deformation matrix $\mathbf{J}^T$. Our next task is to relate the two.

We are interested in smooth, differentiable flows. If a flow is smooth, in a sufficiently small neighborhood it is essentially linear. Hence the next section, which might seem an embarrassment (what is a section on *linear* flows doing in a book on *non*linear dynamics?), offers a firm stepping stone on the way to understanding nonlinear flows. If you know your eigenvalues and eigenvectors, you may prefer to fast forward here.

fast track:
sect. 4.3, p. 69

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 equations of motion (2.5) are written in terms of a vector x and a constant matrix of variations $\mathbf{A}$ 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

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

In order to solve the matrix case, it is helpful to re-derive 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 the 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.

in depth:
appendix K.2, p. 699

fast track:
sect. 4.3, p. 69

4.2.1 Stability eigenvalues, diagonal case

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

$$\mathbf{J}^t = e^{t\mathbf{A}_D} = \begin{pmatrix} e^{t\lambda_1} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & e^{t\lambda_d} \end{pmatrix}\,. \tag{4.13}$$

Suppose next 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. Then $\mathbf{J}$ can also be brought to a diagonal form:

4.2
page 81

$$\mathbf{J}^t = e^{t\mathbf{A}} = \mathbf{U}^{-1}e^{t\mathbf{A}_D}\mathbf{U}\,. \tag{4.14}$$

In either case, the action of both $\mathbf{A}$ and $\mathbf{J}$ is very simple; the axes of orthonormal coordinate system where $\mathbf{A}$ is diagonal are also the eigendirections of both $\mathbf{A}$ and $\mathbf{J}^t$, and under the flow the neighborhood is deformed by a multiplication by an eigenvalue factor for each coordinate axis.

4.2.2 Complex stability eigenvalues

As $\mathbf{A}$ has only real entries, it will in general have either real eigenvalues, or complex conjugate pairs of eigenvalues. That is not surprising, but also the corresponding eigenvectors can be either real or complex. All coordinates used in defining the flow are real numbers, so what is the meaning of a *complex* eigenvector?

To develop some intuition about that, let us work out the behavior for the simplest nontrivial case, the case where $\mathbf{A}$ is a $[2\times2]$ matrix

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} . \tag{4.15}$$

The eigenvalues λ_1, λ_2 of $\mathbf{A}$ are the roots

$$\lambda_{1,2} = \frac{1}{2}\left(\mathrm{tr}\,\mathbf{A} \pm \sqrt{(\mathrm{tr}\,\mathbf{A})^2 - 4\det\mathbf{A}}\right) \tag{4.16}$$

of the characteristic equation

$$\begin{aligned} \det(\mathbf{A} - z\mathbf{1}) &= (\lambda_1 - z)(\lambda_2 - z) = 0\,, \qquad (4.17) \\ \left\|\begin{matrix} A_{11} - z & A_{12} \\ A_{21} & A_{22} - z \end{matrix}\right\| &= z^2 - (A_{11} + A_{22})\,z + (A_{11}A_{22} - A_{12}A_{21})\,. \end{aligned}$$

The qualitative behavior of the exponential of $\mathbf{A}$ for real eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ will differ from the case that they form a complex conjugate pair,

$$\lambda_1 = \lambda + i\theta\,, \quad \lambda_2 = \lambda_1^* = \lambda - i\theta\,.$$

These two possibilities are refined further into sub-cases depending on the signs of the real part. The matrix might have only one eigenvector, or two linearly independent eigenvectors, which may or may not be orthogonal. Along each of these directions the motion is of the form $\exp(t\lambda_k)x_k$. If the exponent λ_k is positive, then the component x_k will grow; if the exponent λ_k is negative, it will shrink.

We sketch the full set of possibilities in fig. 4.1(a), and work out in detail the case when $\mathbf{A}$ can be brought to the diagonal form. Then the solution (4.12) to the differential equation (4.7) can be written either as

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} e^{t\lambda_1} & 0 \\ 0 & e^{t\lambda_2} \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} , \tag{4.18}$$

or

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = e^{t\lambda} \begin{pmatrix} e^{it\theta} & 0 \\ 0 & e^{-it\theta} \end{pmatrix} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} . \tag{4.19}$$

In the case $\lambda_1 > 0$, $\lambda_2 < 0$, x_1 grows exponentially with time, and x_2 contracts exponentially. This behavior, called a *saddle*, is sketched in fig. 4.1(b), as are the remaining possibilities: in/out nodes, inward/outward spirals, and the center. Spirals arise from taking a real part of the action of $\mathbf{J}^t$ on a complex eigenvector. The magnitude of $|x(t)|$ diverges exponentially the $\lambda > 0$, and contracts toward 0 if the $\lambda < 0$, while the imaginary phase θ controls its oscillations.

4.2.3 General, nonsymmetric Jacobian matrix

In general $\mathbf{J}^t$ is neither diagonal, nor diagonalizable, nor constant along the trajectory. Still, any matrix, including $\mathbf{J}^t$, can be expressed in the singular value decomposition form

$$\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

where $\mathbf{D}$ is diagonal, and $\mathbf{U}$, $\mathbf{V}$ are orthogonal matrices. The diagonal elements $\Lambda_1, \Lambda_2, \ldots, \Lambda_d$ of $\mathbf{D}$ are called the *stability eigenvalues*.

Under the action of the flow an infinitesimally small ball of initial points is deformed into an ellipsoid: Λ_i is the relative stretching of the ith principal axis of the ellipsoid, the columns of the matrix $\mathbf{V}$ are the principal axes $\mathbf{e}_i$ of stretching in the Lagrangian coordinate frame, and the orthogonal matrix $\mathbf{U}$ gives the orientation of the ellipse in the Eulerian coordinates.

Now that we have some feeling for the qualitative behavior of eigenvectors and eigenvalues, we are ready to return to the general case: nonlinear flows.

4.3 Stability of 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)

$$x(t) = f^t(x_0) , \quad \delta x(x_0, t) = \mathbf{J}^t(x_0)\delta x(x_0, 0) . \tag{4.20}$$

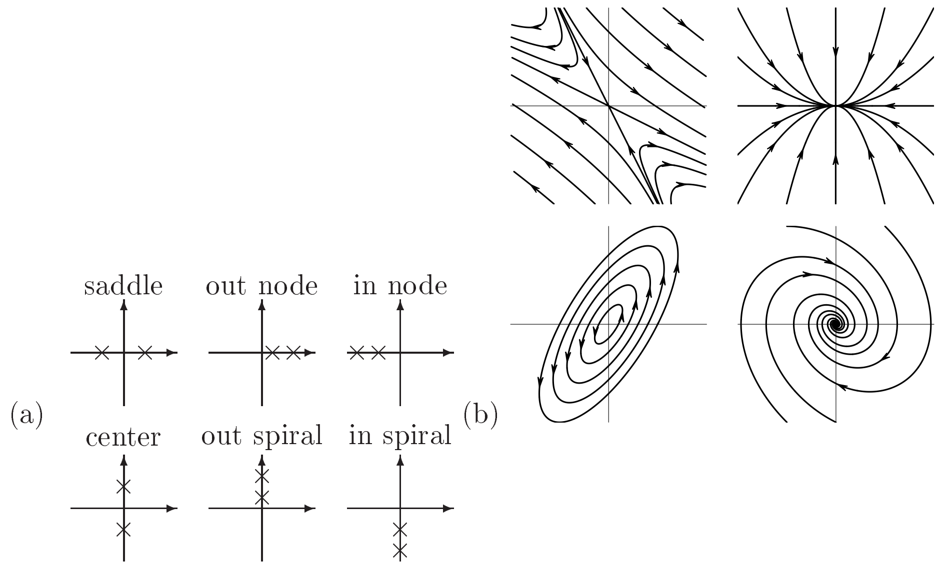


Figure 4.1: (a) Qualitatively distinct types of eigenvalues of a $[2\times2]$ stability matrix. (b) Streamlines for several typical 2-dimensional flows.

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

4.3.1 Stability of equilibria

For a start, consider the case where x_q is an equilibrium point (2.7). 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,

$$f^t(x) = x_q + e^{\mathbf{A}t}(x - x_q) + \cdots \,, \tag{4.21}$$

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

Example 4.1 Why does a flame front flutter? *The Kuramoto-Sivashinsky flat frame-front $u(x,t) = 0$ is an equilibrium point of (2.14). The matrix of variations (4.3) follows from (2.17)*

$$A_{kj}(a) = \frac{\partial v_k(x)}{\partial a_j} = (k^2 - \nu k^4)\delta_{kj} - 2ka_{k-j} \,. \tag{4.22}$$

For the $u(x,t) = 0$ equilibrium solution the matrix of variations is diagonal, and as in (4.13), so is the Jacobian matrix $\mathbf{J}^t(0)_{kj} = \delta_{kj} e^{(k^2-\nu k^4)t}$.

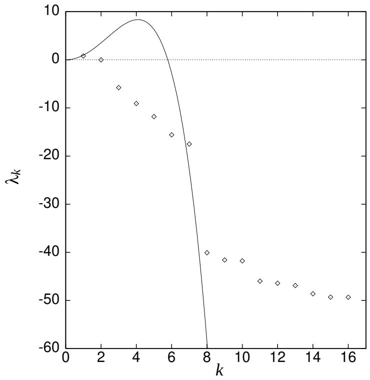


Figure 4.2: Lyapunov exponents $\lambda_{\overline{1},k}$ versus k for the least unstable spatio-temporally periodic orbit $\overline{1}$ of the Kuramoto-Sivashinsky system, compared with the stability exponents of the $u(x,t) = 0$ stationary solution, $\lambda_k = k^2 - \nu k^4$. $\lambda_{\overline{1},k}$ for $k \geq 8$ fall below the numerical accuracy of integration and are not meaningful. $N = 16$ Fourier modes, $\nu = 0.029924$, chaotic regime. The cycle $\overline{1}$ was computed using methods of chapter 14. (From ref. [2.7])

For $\nu > 1$, $u(x,t) = 0$ is the globally attractive stable equilibrium. With the "viscosity" $\nu = 1$ and below, the dynamics goes through a rich sequence of bifurcations on which we shall not dwell here. The $|k| < 1/\sqrt{\nu}$ long wavelength perturbations of the flat-front equilibrium are linearly unstable, while all $|k| > 1/\sqrt{\nu}$ short wavelength perturbations are strongly contractive. The high k eigenvalues, corresponding to rapid variations of the flame front, decay so fast that the corresponding eigendirections have no physical meaning. To illustrate the rapid contraction in the non-leading eigendirections we plot in fig. 4.2 the eigenvalues of the equilibrium in the unstable regime, for relatively low "viscosity" ν, and compare them with the stability eigenvalues of the least unstable cycle for the same value of ν. The equilibrium solution is very unstable, in 5 eigendirections, the least unstable cycle only in one, but for $k > 7$ the rate of contraction is so strong that higher eigendirections are numerically meaningless for either solution.

4.3.2 Stability of trajectories

Next, consider the case of a general, non-stationary trajectory $x(t)$. The exponential of a constant matrix can be defined either by its Taylor series expansion, or in terms of the Euler limit (4.12):

☞ appendix K.1

$$e^{t\mathbf{A}} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbf{A}^k \tag{4.23}$$

$$= \lim_{m\to\infty} \left(\mathbf{1} + \frac{t}{m}\mathbf{A}\right)^m . \tag{4.24}$$

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 linearized neighborhood is multiplicatively deformed along the flow, and the m discrete time steps approximation to $\mathbf{J}^t$ is therefore given by generalization of the Euler product (4.24) to

$$\mathbf{J}^t = \lim_{m\to\infty} \prod_{n=m}^{1} (\mathbf{1} + \Delta t \mathbf{A}(x_n)) = \lim_{m\to\infty} \prod_{n=m}^{1} e^{\Delta t\, \mathbf{A}(x_n)} \tag{4.25}$$

where $\Delta t = (t - t_0)/m$, $x_n = x(t_0 + n\Delta t)$. The two formulas for $\mathbf{J}^t$ agree to the leading order in Δt, and the $m \to \infty$ limit of this procedure is the integral

☞ appendix D

$$\mathbf{J}^t_{ij}(x_0) = \left[\mathbf{T} e^{\int_0^t d\tau \mathbf{A}(x(\tau))}\right]_{ij} . \tag{4.26}$$

☞ appendix H.1

where $\mathbf{T}$ stands for time-ordered integration. This formula for $\mathbf{J}$ is the main result of this chapter.

It is evident from the time-ordered product structure (4.25) that the Jacobian matrices are multiplicative along the flow,

$$\mathbf{J}^{t+t'}(x) = \mathbf{J}^{t'}(x')\mathbf{J}^t(x), \qquad \text{where } x' = f^t(x) . \tag{4.27}$$

The formula (4.25) is a matrix generalization of the crude Euler integrator (2.12), neither smart not accurate. Much better numerical accuracy is obtained by the following observation. To linear order in Δt, $(\mathbf{J}^{t+\Delta t} - \mathbf{J}^t)$ equals $\Delta t\, \mathbf{A}\mathbf{J}^t$, so the Jacobian matrix itself satisfies the linearized equation (4.1)

$$\frac{d}{dt}\mathbf{J}^t(x) = \mathbf{A}(x)\,\mathbf{J}^t(x) , \qquad \text{with the initial condition } \mathbf{J}^0(x) = \mathbf{1} . \tag{4.28}$$

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

"Simply" is perhaps too glib. Integration will work for short finite times, but for exponentially unstable flows one quickly runs into numerical over/underflow problems, and further thought will be required.

4.4 Stability of maps

The 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 evaluated at a discrete set of times $n = 1, 2, \ldots$,

$$\mathbf{J}^n_{ij}(x_0) = \left.\frac{\partial f^n_i(x)}{\partial x_j}\right|_{x=x_0} . \tag{4.29}$$

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

$$\Lambda_n = \frac{d}{dx} f^n(x) = \prod_{m=0}^{n-1} f'(x_m) , \quad x_m = f^m(x_0) . \tag{4.30}$$

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

$$\mathbf{J}^n(x_0) = \prod_{m=n-1}^{0} \mathbf{J}(x_m) , \quad \mathbf{J}(x)_{kl} = \frac{\partial}{\partial x_l} f_k(x) , \quad x_{m,j} = f^m(x_0)_j \tag{4.31}$$

follows from the chain rule for matrix derivatives

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

The $[d\times d]$ Jacobian matrix $\mathbf{J}^n$ for a map is evaluated by multiplication along the n points $x_0, x_1, x_2, \ldots x_{n-1}$ on the trajectory of x_0, with $\mathbf{J}(x)$ the single time step Jacobian matrix.

8.1
page 143

Example 4.2 Hénon map Jacobian matrix. *For the Hénon map (3.10) the Jacobian matrix for nth iterate of the map is*

$$\mathbf{J}^n(x_0) = \prod_{m=n}^{1} \begin{pmatrix} -2ax_m & b \\ 1 & 0 \end{pmatrix} , \qquad x_m = f^m_1(x_0, y_0) . \tag{4.32}$$

The determinant of the Hénon one time step Jacobian matrix (4.32) is constant,

$$\det \mathbf{J} = \Lambda_1 \Lambda_2 = -b \tag{4.33}$$

so in this case only one eigenvalue $\Lambda_1 = -b/\Lambda_2$ needs to be determined. This is not an accident; a constant jacobian was one of desiderata that led Hénon to constructing the map of this particular form.

4.5 Stability of periodic orbits

As noted on page 36, 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 re-parametrization 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. No less important, as we shall now show, is the fact that cycle stability eigenvalues are *metric* invariants: they determine the relative sizes of neighborhoods in a non–wandering set.

To prove this, we start by noting that due to the multiplicative structure (4.27) of Jacobian matrices, the stability of the rth repeat of a prime cycle p of period T_p is

$$\mathbf{J}^{rT_p}(x_0) = \mathbf{J}^{T_p}(f^{rT_p}(x_0)) \cdots \mathbf{J}^{T_p}(f^{T_p}(x_0))\mathbf{J}^{T_p}(x_0) = \mathbf{J}_p(x_0)^r \,, \qquad (4.34)$$

where $\mathbf{J}_p(x_0) = \mathbf{J}^{T_p}(x_0)$ is the stability matrix for a single traversal of the prime cycle p, $x_0 \in p$ is any point on the cycle, and $f^{rT_p}(x_0) = x_0$ as $f^t(x_0)$ returns to x_0 every multiple of the period T_p. Hence, it suffices to restrict our considerations to the stability of prime cycles.

4.5.1 Stability eigenvalues, stability exponents

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

$$\begin{aligned}
\text{expanding:} \quad & \{\Lambda_p\}_e &= \{\Lambda_{p,i} : |\Lambda_{p,i}| > 1\} \\
\text{marginal:} \quad & \{\Lambda_p\}_m &= \{\Lambda_{p,i} : |\Lambda_{p,i}| = 1\} \\
\text{contracting:} \quad & \{\Lambda_p\}_c &= \{\Lambda_{p,i} : |\Lambda_{p,i}| < 1\} \,.
\end{aligned} \qquad (4.35)$$

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

$$\Lambda_p = \prod_e \Lambda_{p,e} \,. \qquad (4.36)$$

As $\mathbf{J}_p$ is a real matrix, complex Λ_i always come in $\Lambda_{p,i}$, $\Lambda_{p,i+1} = \Lambda^*_{p,i}$ pairs, and Λ_p is always real.

Cycle *stability exponents* are the stretching/contraction rates per unit time

$$\lambda_{p,i} = \frac{1}{T_p} \ln |\Lambda_{p,i}| \ . \tag{4.37}$$

This definition is motivated by the form of the stability exponents for the linear case, for example (4.13), as well as the fact that exponents so defined can be interpreted as the Lyapunov exponents evaluated on the prime cycle p. As in the three cases of (4.35), we sort the stability exponents into three sets

☞ sect. 8.3

$$\begin{aligned} \text{expanding:} \quad & \{\lambda_p\}_e &= \{\lambda_{p,i} :\ \lambda_{p,i} > 0\} \\ \text{elliptic:} \quad & \{\lambda_p\}_m &= \{\lambda_{p,i} :\ \lambda_{p,i} = 0\} \\ \text{contracting:} \quad & \{\lambda_p\}_c &= \{\lambda_{p,i} : \lambda_{p,i} < 0\} . \end{aligned} \tag{4.38}$$

A periodic orbit p of a d-dimensional flow or a map is *stable* if all its stability exponents (other than the vanishing longitudinal stability exponent, to be explained in sect. 4.5.3 below) are strictly negative, $|\lambda_{p,i}| < 0$. The region of system parameter values for which a periodic orbit p is stable is called the *stability window* of p. The subset $\mathcal{M}$ of initial points that are asymptotically attracted to p (for a fixed set of system parameter velues) is called the *basin of attraction* of p.

If *all* stability exponents (other than the vanishing longitudinal stability exponent) of *all* periodic orbits of a flow are strictly bounded away from zero, $|\lambda_i| \geq \lambda_{min} > 0$, the flow is said to be *hyperbolic*. Otherwise the flow is said to be *nonhyperbolic*.

As we often do care about the sign of $\Lambda_{p,i}$ and, if $\Lambda_{p,i}$ is complex, its phase

$$\Lambda_{p,j} = \pm e^{T_p(\lambda_{p,j} \pm i\theta_{p,j})} \, , \tag{4.39}$$

☞ sect. 5.1.1

and keeping track of those by case-by-case enumeration is a self-inflicted, unnecessary nuisance (followed by much of the literature), almost all of our formulas will be stated in terms of the stability eigenvalues Λ rather than in the terms of the overall signs, stability exponents λ_i and phases θ_i.

Example 4.3 1-dimensional maps. *The simplest example of cycle stability is afforded by 1-dimensional maps. The stability of a prime cycle p follows from the chain rule (4.30) for stability of the n_pth iterate of the map*

$$\Lambda_p = \frac{d}{dx_0} f^{n_p}(x_0) = \prod_{m=0}^{n_p-1} f'(x_m)\,, \quad x_m = f^m(x_0)\,, \tag{4.40}$$

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

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

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

and superstable *if the orbit includes a critical point, so that the above product vanishes. 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. If $|\Lambda_p| > 1$, cycle p is* unstable.

4.5.2 Cycle stabilities are cycle invariants

The 1-dimensional map 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 form of the Jacobian matrix $\mathbf{J}_p(x_0)$ in (4.34) does depend on the choice of coordinates and the initial point $x_0 \in p$. Nevertheless, as we shall now show, the cycle stability *eigenvalues* are intrinsic property of a cycle also for multi-dimensional flows. Consider the ith eigenvalue, eigenvector pair $(\Lambda_{p,i}, \mathbf{e}_i)$ computed from $\mathbf{J}_p$ evaluated at a cycle point,

$$\mathbf{J}_p(x)\mathbf{e}_i(x) = \Lambda_{p,i}\mathbf{e}_i(x)\,, \quad x \in p\,. \tag{4.41}$$

Consider another point on the cycle a time t later, $x' = f^t(x)$ whose Jacobian matrix is $\mathbf{J}_p(x')$. By the group property (4.27), $\mathbf{J}^{T_p+t} = \mathbf{J}^{t+T_p}$, and the Jacobian matrix at x' can be written either as

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

Multiplying (4.41) by $\mathbf{J}^t(x)$, we find that the Jacobian matrix evaluated at x' has the same eigenvalue,

$$\mathbf{J}_p(x')\mathbf{e}_i(x') = \Lambda_{p,i}\mathbf{e}_i(x')\,, \quad \mathbf{e}_i(x') = \mathbf{J}^t(x)\mathbf{e}_i(x)\,, \tag{4.42}$$

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

$$\det\left(\mathbf{1} - \mathbf{J}_p^r(x)\right) = \det\left(\mathbf{1} - \mathbf{J}_p^r\right) . \tag{4.43}$$

We postpone the proof that the cycle stability eigenvalues are smooth conjugacy invariants of the flow to sect. 6.4.

4.5.3 Marginal eigenvalues

The 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 18

A periodic orbit of a continuous flow 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

$$\mathbf{J}_p(x)v(x) = v(x) , \tag{4.44}$$

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

$$\Lambda_{p,\|} = 1 . \tag{4.45}$$

✎ 6.2
page 104

The continuous invariance that gives rise to this marginal eigenvalues is the invariance of a cycle under a translation of its points along the cycle. As we shall see in sect. 11.3, this marginal stability direction can be eliminated by “fixing the gauge”, that is by cutting the cycle by a Poincaré section and eliminating the continuous flow Jacobian matrix in favor of the Jacobian matrix of the Poincaré section return map.

4.6 Neighborhood of a cycle

☞ sect. 8.3
☞ remark 8.3

The Jacobian of the flow (or the *sum* of stability exponents) is easily evaluated.

Consider $\det \mathbf{J}^t(x_0) = \prod_{i=1}^d \Lambda_i(x_0, t)$, the product of the stability eigenvalues. We shall refer to this determinant as the *Jacobian* of the flow. By means of the time-ordered product (4.25) and the identity $\ln \det M = \mathrm{tr} \ln M$ the Jacobian is given by

4.1 page 81

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

As the divergence $\partial_i v_i$ is a scalar quantity, the integral in the exponent needs no time ordering. All we need to do is to evaluate the time average

$$\langle \partial_i v_i \rangle_t = \frac{1}{t} \ln \left| \prod_{i=1}^d \Lambda_i(x_0, t) \right| = \sum_{i=1}^d \lambda_i(x_0, t) = \frac{1}{t} \int_0^t d\tau \sum_{i=1}^d A_{ii}(x(\tau)) \tag{4.47}$$

along the trajectory. If the flow is not singular (for example, the trajectory does not run head-on into the Coulomb $1/r$ singularity), the matrix of variations elements $|A_{ij}|$ are everywhere bounded in magnitude, and so is the trace $\sum_i A_{ii}$. The time integral in (4.47) grows at most linearly with t, hence $\langle \partial_i v_i \rangle_t$ is bounded, regardless of how large the time t might be, and numerical estimates of the $t \to \infty$ limit $\langle \partial_i v_i \rangle$ are not marred by any exponential blowups.

Even if we were to insist on extracting $\langle \partial_i v_i \rangle$ from (4.25) by first multiplying stability matrices along the flow, and then taking the logarithm, we can avoid exponential blowups in $\mathbf{J}^t$ by using the multiplicative structure (4.27), $\det \mathbf{J}^{t'+t}(x_0) = \det \mathbf{J}^{t'}(x') \det \mathbf{J}^t(x_0)$ to restart with $\mathbf{J}^0(x') = \mathbf{1}$ whenever the eigenvalues of $\mathbf{J}^t(x_0)$ start getting out of hand. In numerical evaluations of Lyapunov exponents, $\lambda_i = \lim_{t\to\infty} \lambda_i(x_0, t)$, sum rule (4.47) can serve as a helpful check on the accuracy of the computation.

☞ sect. 8.3

$\langle \partial_i v_i \rangle$ is also an important physical characterization of the flow - it tells us what is the behavior of a phase space volume in the infinitesimal neighborhood of the trajectory. If $\partial_i v_i < 0$, the flow is *locally contracting*, and the trajectory might be falling into an attractor. If $\partial_i v_i = 0$, 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 in sect. 5.1.1.

But before we can get to that, the alert student, pipes up. He does not like our definition of the Jacobian matrix in terms of the time-ordered exponential (4.26). Depending on the signs of stability eigenvalues, the left hand side of (4.46) can be either positive or negative. But the right hand side is an exponential of a real number, and that can only be positive. What gives? As we shall see much later on in this text, in discussion of

☞ chapter 24.1

topological topological indices arising in semiclassical quantization, this is not at all a dumb question.

in depth:
appendix K.1, p. 697

in depth:
appendix C.1, p. 621

4.6.1 There goes the neighborhood

In what follows, our task will be to determine the size of a *neighborhood* of $x(t)$, and that is why we care about the 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 sub-volume $|\mathcal{M}_i| = \prod_i^e \Delta x_i$ of the set of points which get no further away from $f^t(x_0)$ than L, the typical size of the system, is fixed by the condition that $\Delta x_i \Lambda_i = O(L)$ in each expanding direction i. Hence the neighborhood size scales as $\propto 1/|\Lambda_p|$ where Λ_p is the product of expanding eigenvalues (4.36) only; contracting ones play a secondary role. So secondary that even infinity of them (for example, the infinity of contracting eigendirections of the spatiotemporal dynamics of sect. 2.4.1) will not matter.

So the physically important information is carried by the expanding sub-volume, not the total volume computed so easily in (4.47). That is also the reason why the dissipative and the Hamiltonian chaotic flows are much more alike than one would have naively expected for "compressible" *vs.* "incompressible" flows. What matters are the expanding directions. Whether the contracting eigenvalues are inverses of the expanding ones or not is of secondary importance. As long as the number of unstable directions is finite, the same theory will describe finite-dimensional ODEs and infinite-dimensional PDEs.

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 a finite time t is described by the Jacobian matrix

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

where **T** stands for the *time-ordered* integration. Its eigenvalues and eigendirections describe deformation of an initial infinitesimal sphere of neighboring trajectories into an ellipsoid finite time t later. Nearby trajectories separate exponentially along unstable directions, approach each other along stable directions, and change slowly (algebraically) their distance along marginal directions. The Jacobian matrix $\mathbf{J}^t$ is in general neither symmetric, nor diagonalizable by a rotation, nor do its (left or right) eigenvectors define an orthonormal coordinate frame. Furthermore, while the stability matrices are multiplicative along the flow, in dimensions higher than one their eigenvalues in general are not. This lack of multiplicativity has important repercussion both for classical dynamics and quantum dynamics.

☞ appendix H.1

Periodic orbits play a central role in any invariant characterization of the dynamics, because (a) their existence and inter-relations are *topological*, coordinate choice independent property of the dynamics, and (b) their stability eigenvalues are metric invariants. We shall show in chapter 9 that extending their local stability eigendirections into stable and unstable manifolds yields also important global information about the topological organization of the phase space.

References

[4.1] J.-L. Thiffeault, "Derivatives and constraints in chaotic flows: Asymptotic behaviour and a numerical method", preprint (2001).

Exercises

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

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

for arbitrary finite dimensional matrix M.

Exercise 4.2 Stability, diagonal case. *Verify the relation (4.14)*

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

Exercise 4.3 A contracting baker's map. *Consider a contracting (or "dissipative") baker's map, on $[0,1]^2$, defined as*

$$\left(\begin{array}{c} x_{n+1} \\ y_{n+1} \end{array}\right) = \left(\begin{array}{c} x_n/3 \\ 2y_n \end{array}\right) \qquad y_n \le 1/2$$

$$\left(\begin{array}{c} x_{n+1} \\ y_{n+1} \end{array}\right) = \left(\begin{array}{c} x_n/3 + 1/2 \\ 2y_n - 1 \end{array}\right) \qquad y_n > 1/2$$

This map shrinks strips by factor 1/3 in the x direction, and stretches (and folds) by factor 2 in the y direction.

How fast does the phase space volume contract?

Chapter 5

Newtonian dynamics

You might think that the strangeness of contracting flows, flows such as the Rössler flow of fig. 2.3 is of concern only to chemists. Not at all - while it is easier to visualize aperiodic dynamics when the flow is contracting onto a lower-dimensional attracting set, there are plenty examples of chaotic flows that do preserve the full symplectic invariance of Hamiltonian dynamics.

5.1 Hamiltonian flows

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

☞ appendix C

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{5.1}$$

with the $2D$ phase space coordinates x split into the configuration space coordinates and the conjugate momenta of a Hamiltonian system with D degrees of freedom:

☞ sect. 26.1.1

$$x = (\mathbf{p}, \mathbf{q}), \qquad \mathbf{q} = (q_1, q_2, \ldots, q_D), \qquad \mathbf{p} = (p_1, p_2, \ldots, p_D). \tag{5.2}$$

Example 5.1 Collinear helium. *In chapter 29, 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 29

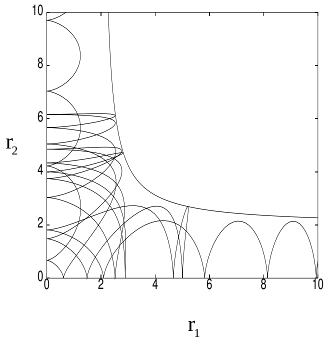


Figure 5.1: A typical colinear helium trajectory in the $r_1 - r_2$ plane; the trajectory enters here along the r_1 axis and then, like almost every other trajectory, after a few bounces escapes to infinity, in this case along the r_2 axis.

$$H = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_1 + r_2}\,. \tag{5.3}$$

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. 5.1. It looks messy, and, indeed, it will turn out to be no less chaotic than a pinball bouncing between three disks.

5.1.1 Stability of 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 (5.1) can be written as

$$\dot{x}_m = \omega_{mn}\frac{\partial H}{\partial x_n}, \qquad \omega = \begin{pmatrix} 0 & -\mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix}, \qquad m, n = 1, 2, \ldots, 2D \tag{5.4}$$

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

$$\omega_{mn} = -\omega_{nm}\,, \qquad \omega^2 = -\mathbf{1}\,. \tag{5.5}$$

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 (4.20). The matrix of variations in (4.28) takes form

$$A(x)_{mn} = \omega_{mk}H_{kn}(x)\,, \qquad \frac{d}{dt}\mathbf{J}^t(x) = A(x)\mathbf{J}^t(x)\,, \tag{5.6}$$

where $H_{kn} = \partial_k \partial_n H$ is the Hessian matrix of second derivatives. From (5.6) and the symmetry of H_{kn} it follows that

$$A^T \omega + \omega A = 0\,. \tag{5.7}$$

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):

$$\mathbf{J}^T \omega \mathbf{J} = \omega\,. \tag{5.8}$$

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

$$\mathbf{J}^{-1} = -\omega \mathbf{J}^T \omega\,, \tag{5.9}$$

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 $1/\Lambda^*$, so

5.8 page 94

$$\det \mathbf{J}^t(x_0) = 1 \qquad \text{for all } t \text{ and } x_0\text{'s}\,, \tag{5.10}$$

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

Example 5.2 2-dimensional symplectic flows *In the 2-dimensional case the eigenvalues (4.35) depend only on* $\mathrm{tr}\,\mathbf{J}^t$

$$\Lambda_{1,2} = \frac{1}{2}\left(\mathrm{tr}\,\mathbf{J}^t \pm \sqrt{(\mathrm{tr}\,\mathbf{J}^t - 2)(\mathrm{tr}\,\mathbf{J}^t + 2)}\right)\,. \tag{5.11}$$

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*)

$$\begin{array}{lll} \textit{either} \ \text{hyperbolic} & \Lambda_1 = e^{\lambda t}\,, \quad \Lambda_2 = e^{-\lambda t}\,, & (5.12) \\ \textit{or} \ \text{inverse hyperbolic} & \Lambda_1 = -e^{\lambda t}\,, \quad \Lambda_2 = -e^{-\lambda t}\,. & (5.13) \end{array}$$

in depth:
appendix C.1, p. 621

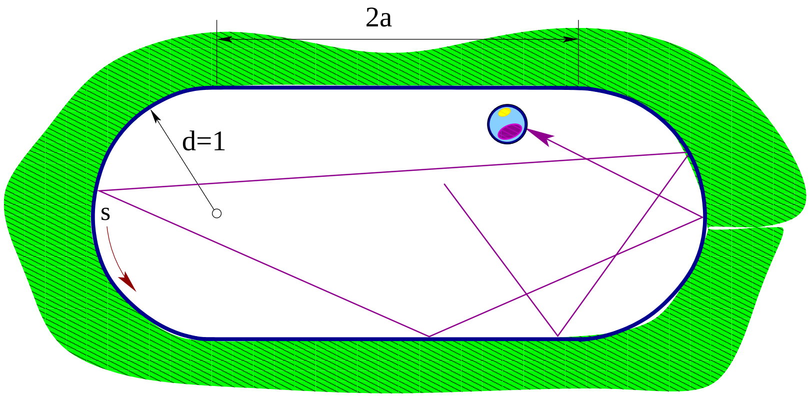


Figure 5.2: The stadium billiard is a 2-dimensional domain bounded by two semi-circles of radius $d = 1$ connected by two straight walls of length $2a$. At the points where the straight walls meet the semi-circles, the curvature of the border changes discontinuously; these are the only singular points of the flow. The length a is the only parameter.

5.2 Billiards

The dynamics that we have the best intuitive grasp on, and find easiest to grapple with both numerically and conceptually, is the dynamics of billiards. For billiards discrete time is altogether natural; a particle moving through a billiard suffers a sequence of instantaneous kicks, and executes a simple motion inbetween, so and there is no need to contrive a Poincaré section. We have already used this system in sect. 1.3 as the intuitively most accessible example of chaos. Here we define billiard dynamics more precisely, anticipating the applications to come.

A billiard is defined by a connected region $Q \subset \mathbb{R}^D$, with boundary $\partial Q \subset \mathbb{R}^{D-1}$ separating Q from its complement $\mathbb{R}^D \setminus Q$. Q can consist of one compact, finite volume component (in which case the billiard phase space is bounded, as for the stadium billiard fig. 5.2), or can be infinite in extent, with its complement $\mathbb{R}^D \setminus Q$ consisting of one or several finite or infinite volume components (in which case the phase space is open, as for the 3-disk pinball game fig. 1.1). In what follows we shall more often than not restrict our attention to *planar billiards*. A point particle ("pinball") of mass m and momentum $p_n = mv_n$ moves freely within the billiard, along a straight line, until it encounters the boundary. There it reflects specularly (*specular* = mirrorlike), with no change in tangential component of momentum, and instantaneous reversal of the momentum component normal to the boundary,

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

where $\hat{n}$ is a unit vector normal to the boundary ∂Q at the collision point.

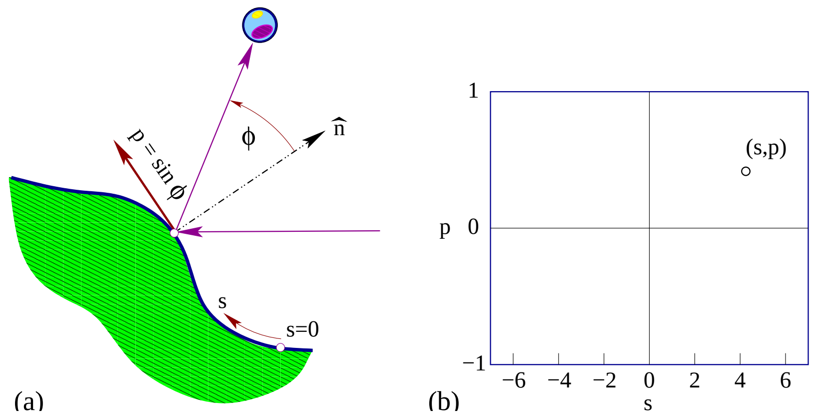


Figure 5.3: (a) A planar billiard trajectory is fixed by specifying the perimeter length parametrized by s and the outgoing trajectory angle ϕ, both measured anti-clockwise with respect to the outward normal $\hat{n}$. (b) The Birkhoff phase space coordinate pair (s, p) fully specifies the trajectory, with $p = \sin\phi$ is the momentum component tangential to the boundary.

The angle of incidence equals to the angle of reflection, fig. 5.3. A billiard is a Hamiltonian system with a $2D$-dimensional phase space $x = (p, q)$ and potential $V(q) = 0$ for $q \in Q$, $V(q) = \infty$ for $q \in \partial Q$.

A billiard flow has a natural Poincaré section defined by marking s_n, the arc length position of the nth bounce measured along the billiard boundary, and $p_n = p \sin\phi_n$, the momentum component parallel to the boundary, where ϕ_n is the angle between the outgoing trajectory and the normal to the boundary. We measure both the arc length s and the parallel momentum p anti-clockwise relative to the outward normal (see fig. 5.3 as well as fig. 1.6a). In $D = 2$, the Poincaré section is a cylinder (topologically an annulus), fig. 5.4, where the parallel momentum p ranges for -1 to 1, and the s coordinate is cyclic along each connected component of ∂Q. The volume in the full phase space is preserved by the Liouville theorem (5.10). The Birkhoff Poincaré section coordinates $x = (s, p) \in \mathcal{P}$, see fig. 5.3, are the natural choice (rather than, let's say, (s, ϕ)), because with them the the Poincaré return map preserves the phase space volume also in the (s, p) parametrized Poincaré section.

5.8
page 94
☞ sect. 5.3

Without loss of generality we will set $m = |v| = |p| = 1$ throughout. Poincaré section condition eliminates one dimension, and the energy conservation $|p| = \text{const.}$ another, hence the Poincaré section return map P is $(2D - 2)$-dimensional.

The dynamics is given by the Poincaré return map $P : (s_n, p_n) \mapsto (s_{n+1}, p_{n+1})$ from the nth collision to the $(n+1)$th collision. The discrete time dynamics map P is equivalent to the Hamiltonian flow (5.1) in the sense that both describe the same full trajectory. Let t_n be the instant of

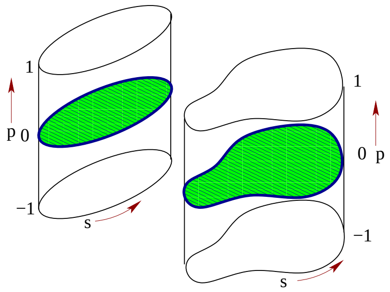


Figure 5.4: In $D = 2$ the billiard Poincaré section is a cylinder, with the parallel momentum p ranging over $p \in \{-1, 1\}$, and with the s coordinate is cyclic along each connected component of ∂Q. The rectangle fig. 5.3(b) is such cylinder unfolded, with periodic boundary conditions glueing together the left and the right edge of the rectangle.

nth collision. Then the position of the pinball $\in Q$ at time $t_n + \tau \le t_{n+1}$ is given by $2D-2$ Poincaré section coordinates $(s_n, p_n) \in \mathcal{P}$ together with τ, the distance reached by the pinball along the nth section of its trajectory.

5.1
page 93

Example 5.3 3-disk game of pinball *In case of bounces off a circular disk, the position coordinate $s = r\theta$ is given by angle $\theta \in [0, 2\pi]$. For example, for the 3-disk game of pinball of fig. 1.4 and fig. 1.6 we have two types of collisions:*

$$P_0 : \begin{cases} \phi' = -\phi + 2\arcsin p \\ p' = -p + \frac{a}{R}\sin\phi' \end{cases} \qquad \textit{back-reflection} \tag{5.14}$$

$$P_1 : \begin{cases} \phi' = \phi - 2\arcsin p + 2\pi/3 \\ p' = p - \frac{a}{R}\sin\phi' \end{cases} \qquad \textit{reflect to 3rd disk}\,. \tag{5.15}$$

Here a = radius of a disk, and R = center-to-center separation. Actually, as in this example we are computing intersections of circles and straight lines, nothing more than high-school geometry is required. There is no need to compute $\arcsin$'s either - one only needs to compute a square root per each reflection, and the simulations can be very fast.

5.2
page 93

Trajectory of the pinball in the 3-disk billiard is generated by a series of P_0's and P_1's. At each step on has to check whether the trajectory intersects the desired disk (and no disk inbetween). With minor modifications, the above formulas are valid for any smooth billiard as long as we replace a by the local curvature of the boundary at the point of collision.

5.3 Stability of 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 (4.27) 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

$$\delta z = \delta q_1 \cos\theta - \delta q_2 \sin\theta\,, \tag{5.16}$$

$\delta\theta$ 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

$$\frac{d}{dt}\delta\theta = 0, \quad \frac{d}{dt}\delta z = \delta\theta\,. \tag{5.17}$$

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

$$\begin{aligned} \delta z_k^- &= \delta z_{k-1} + \tau_k \delta\theta_{k-1}\,, \qquad \tau_k = t_k - t_{k-1} \\ \delta\theta_k^- &= \delta\theta_{k-1}\,, \end{aligned} \tag{5.18}$$

and the stability matrix (4.26) for the kth free flight segment is

$$\mathbf{J}_T(x_k) = \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix}. \tag{5.19}$$

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

$$\begin{aligned} \delta z_k &= -\delta z_k^- \\ \delta\theta_k &= -\,\delta\theta_k^- - \frac{2}{\rho_k \cos\phi_k}\delta z_k^-\,, \end{aligned} \tag{5.20}$$

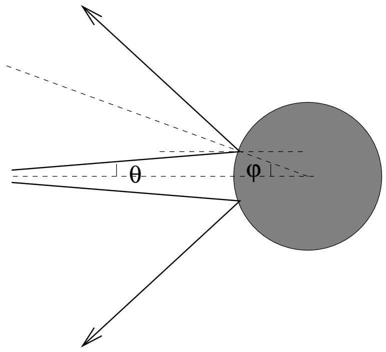


Figure 5.5: Defocusing of a beam of nearby trajectories at a billiard collision. (A. Wirzba)

so the Jacobian matrix associated with the reflection is

$$\mathbf{J}_R(x_k) = -\begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}, \qquad r_k = \frac{2}{\rho_k \cos\phi_k}. \tag{5.21}$$

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)

$$\mathbf{J}_p = (-1)^{n_p} \prod_{k=n_p}^{1} \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r_k & 1 \end{pmatrix}, \tag{5.22}$$

5.4 page 94

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 (5.11).

This is still another example of the Jacobian matrix chain rule (4.31) for discrete time systems, rather similar to the Hénon map stability (4.32). Stability of every flight segment or reflection taken alone is a shear with two unit eigenvalues, but acting in concert in the intervowen sequence (5.22) they can lead to a hyperbolic deformation of the infinitesimal neighborhood of a billiard trajectory.

5.5 page 94

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 14.

5.6 page 94
5.3 page 93
chapter 14

Commentary

Remark 5.1 Billiards. The 3-disk game of pinball is to chaotic dynamics what a pendulum is to integrable systems; the simplest physical example that captures the essence of chaos. Another contender for the title of the "harmonic oscillator of chaos" is the baker's map which is used as the red thread through Ott's introduction to chaotic dynamics [1.7]. 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 subtleties of cycle expansions curvature corrections that will be central to this treatise. ☞ chapter 15

That the 3-disk game of pinball is a quintessential example of deterministic chaos appears to have been first noted by B. Eckhardt [3.8]. The model was studied in depth classically, semiclassically and quantum mechanically by P. Gaspard and S.A. Rice [3.9], and used by P. Cvitanović and B. Eckhardt [3.10] 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 [3.11], construct semiclassical evolution operators and entire spectral determinants by P. Cvitanović and G. Vattay [3.12], and incorporate the diffraction effects into the periodic orbit theory by G. Vattay, A. Wirzba and P.E. Rosenqvist [3.13]. The full quantum mechanics and semiclassics of scattering systems is developed here in the 3-disk scattering context in chapter 28. Gaspard's monograph [1.4], 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 ☞ sect. 25.1 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.

Remark 5.2 Further reading. The chapter 1 of Gaspard monograph [1.4] is recommended reading if you are interested in Hamiltonian flows, and billiards in particular. A. Wirzba has generalized the stability analysis of sect. 5.3 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 [1.4], sect. 1.4.

Résumé

Visualization of strange attractors is greatly facilitated by felicitous choice of Poincaré sections, and reduction of flows to Poincaré section return maps. A particulary natural application of the Poincaré section method is the reduction of a billiard flow to a boundary-to-boundary return map.

Exercises

Exercise 5.1 **A pinball simulator.** *Implement the disk $\rightarrow$ disk maps to compute a trajectory of a pinball for a given starting point, and a given $R{:}a$ = (center-to-center distance):(disk radius) ratio for a 3-disk system. As this requires only computation of intersections of lines and circles together with specular reflections, implementation should be within reach of a high-school student. Please start working on this program now; it will be continually expanded in chapters to come, incorporating the Jacobian calculations, Newton root–finding, and so on.*

Fast code will use elementary geometry (only one $\sqrt{\cdots}$ per iteration, rest are multiplications) and eschew trigonometric functions. Provide a graphic display of the trajectories and of the Poincaré section iterates. To be able to compare with the numerical results of coming chapters, work with $R{:}a = 6$ and/or 2.5 values. Draw the correct versions of fig. 1.8 or fig. 9.4 for $R{:}a =$ 2.5 and/or 6.

Exercise 5.2 **Trapped orbits.** *Shoot 100,000 trajectories from one of the disks, and trace out the strips of fig. 1.8 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.*

Exercise 5.3 **Pinball stability.** *Add to your exercise 5.1 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.*

Exercise 5.4 **Stadium billiard.** *Consider the* Bunimovich stadium *[3.16, 3.17] defined in fig. 5.2. The Jacobian matrix associated with the reflection is given by (5.21). 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*

$$\mathbf{J}^k = (-1)^k\begin{pmatrix} -2k-1 & 2k\cos\phi \\ 2k/\cos\phi & 2k-1 \end{pmatrix} . \tag{5.23}$$

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

$$\mathbf{J}_p = (-1)^{\sum n_k} \prod_{k=1}^{n_p} \begin{pmatrix} 1 & \tau_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n_k r_k & 1 \end{pmatrix} . \tag{5.24}$$

Adopt your pinball simulator to the Bunimovich stadium.

Exercise 5.5 **Fundamental domain fixed points.** *Use the formula (5.22) 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
$\overline{0}$:	$R-2$	$R-1+R\sqrt{1-2/R}$
$\overline{1}$:	$R-\sqrt{3}$	$-\frac{2R}{\sqrt{3}}+1-\frac{2R}{\sqrt{3}}\sqrt{1-\sqrt{3}/R}$

(5.25)

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

Exercise 5.6 **Fundamental domain 2-cycle.** *Verify that for the $\overline{10}$-cycle the cycle length and the trace of the Jacobian matrix are given by*

$$\begin{aligned} L_{10} &= 2\sqrt{R^2-\sqrt{3}R+1}-2, \\ \operatorname{tr}\mathbf{J}_{10} &= 2L_{10}+2+\frac{1}{2}\frac{L_{10}(L_{10}+2)^2}{\sqrt{3}R/2-1} . \end{aligned} \tag{5.26}$$

The $\overline{10}$-cycle is drawn in fig. 9.5. The unstable eigenvalue Λ_{10} follows from (4.16).

Exercise 5.7 **A test of your pinball simulator.** *Test your exercise 5.3 pinball simulator by comparing what it yields with the analytic formulas of exercise 5.5 and 5.6.*

Exercise 5.8 **Birkhoff coordinates.** *Prove that the Birkhoff coordinates are phase-space volume preserving. Hint: compute the determinant of (5.22).*

Chapter 6

Get straight

A Hamiltonian system is said to be “integrable” if one can find a change of coordinates to an action-angle coordinate frame where the phase space dynamics is described by motion on circles, one circle for each degree of freedom. In the same spirit, a natural description of a hyperbolic, unstable flow would be attained if one found a change of coordinates into a frame where the stable/unstable manifolds are straight lines, and the flow is along hyperbolas. Achieving this globally for anything but a handful of contrived examples is too much to hope for. Still, as we shall now show, we can make some headway on straightening out the flow locally.

Even though such nonlinear coordinate transformations are very important, especially in celestial mechanics, we shall not use them much in what follows, so you can safely skip this chapter. Except, perhaps, you might want to convince yourself that cycle stabilities are indeed metric invariants of flows.

fast track:
chapter 6.4, p. 100

6.1 Changing coordinates

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

Denote by h the *conjugation function* which maps the coordinates of the initial phase space $\mathcal{M}$ into the reparametrized phase space $\mathcal{M}' = h(\mathcal{M})$, with a point $x \in \mathcal{M}$ related to a point $y \in \mathcal{M}'$ by $y = h(x)$. The change of coordinates must be one-to-one and span both $\mathcal{M}$ and $\mathcal{M}'$, so given any point y we can go back to $x = h^{-1}(y)$. For smooth flows the reparametrized dynamics should support the same number of derivatives as the initial one. If h is a (piecewise) analytic function, we refer to h as a *smooth conjugacy*.

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

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

Here "$\circ$" stands for functional composition $h \circ f(x) = h(f(x))$, so (6.1) is a shorthand for $y(t) = h(f^t(h^{-1}(y_0)))$.

The vector field $\dot{x} = v(x)$ in $\mathcal{M}$, locally tangent the flow f^t, is related to the flow by differentiation (2.4) along the trajectory. The vector field $\dot{y} = w(y)$ in $\mathcal{M}'$, locally tangent to g^t follows by the chain rule:

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

6.1
page 104

With the indices reinstated, this stands for

$$w_i(y) = \frac{\partial h_i(x)}{\partial x_j} v_j(x) \,, \qquad y_i = h_i(x) \,. \tag{6.3}$$

Imagine that the phase space is a rubber sheet with the flow lines drawn on it. A coordinate change corresponds to pulling and tugging on the rubber sheet, with h sufficiently smooth to preclude violent and irreversible acts such as cutting, glueing, or self-intersections of the distorted rubber sheet. Trajectories that are closed loops in $\mathcal{M}$ will remain closed loops in the new manifold $\mathcal{M}'$, by their shapes will change. Globally h deforms the rubber sheet in a highly nonlinear manner, but locally it simply rescales and shears the tangent field by $\partial_j h_i$, hence the simple transformation law (6.2) for the velocity fields.

The time itself is a parametrization of points along flow lines, and it can also be reparametrized, $s = s(t)$, with the attendent modification of (6.2). An example is the 2-body collision regularization of the helium Hamiltonian (5.3), to be undertaken in sect. 6.2 below.

6.2 Rectification of flows

A profitable way to exploit invariance of dynamics under smooth conjugacies is to use it to pick out the simplest possible representative of an equivalence class. In general and globally these are just words, as we have no clue how to pick such "canonical" representative, but for smooth flows we can always do it 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 (2.7), and points at which v is infinite. According to the theorem, in a small neighborhood of a non-singular point there exists a change of coordinates $y = h(x)$ such that $\dot{x} = v(x)$ in the new coordinates takes the standard form

$$\begin{aligned} &\dot{y}_1 = 1 \\ &\dot{y}_2 = \dot{y}_3 = \cdots = \dot{y}_d = 0\,, \end{aligned} \tag{6.4}$$

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

Example 6.1 Harmonic oscillator, rectified.

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

$$\dot{q} = \ p\,, \qquad \dot{p} = -q\,. \tag{6.5}$$

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 $y = (r, \theta)$:

$$h^{-1} : \left\{ \begin{array}{lllll} q & = & h_1^{-1}(r,\theta) & = & r\cos\theta \\ p & = & h_2^{-1}(r,\theta) & = & r\sin\theta \end{array} \right. . \tag{6.6}$$

The Jacobian matrix of the transformation is

$$h' = \left[\begin{array}{cc} \cos\theta & \sin\theta \\ -\dfrac{\sin\theta}{r} & -\dfrac{\cos\theta}{r} \end{array} \right] \tag{6.7}$$

resulting in (6.2) of rectified form

$$\dot{r} = 0\,, \qquad \dot{\theta} = -1\,. \tag{6.8}$$

In the new coordinates the radial coordinate r is constant, and the angular coordinate θ wraps around a cylinder with constant angular velocity. There is a subtle point in *this change of coordinates: the domain of the map h^{-1} is not the plane $\mathbb{R}^2$, but rather the plane minus the origin. We had mapped a plane into a cylinder, and coordinate transformations should not change the topology of the space in which the dynamics takes place; the coordinate transformation is not defined on the equilibrium point $x = (0,0)$, or $r = 0$.*

Example 6.2 Colinear helium, regularized.

Though very simple in form, the Hamiltonian (5.3) is not the most convenient for numerical investigations of the dynamics of the classical helium 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 29. For the purpose at hand it is sufficient to state the result: In the rescaled coordinates the equations of motion are

$$
\begin{aligned}
\dot{P}_1 &= 2Q_1 \left[2 - \frac{P_2^2}{8} - Q_2^2 \left(1 + \frac{Q_2^2}{R^4} \right) \right]; & \dot{Q}_1 &= \frac{1}{4} P_1 Q_2^2 \\
\dot{P}_2 &= 2Q_2 \left[2 - \frac{P_1^2}{8} - Q_1^2 \left(1 + \frac{Q_1^2}{R^4} \right) \right]; & \dot{Q}_2 &= \frac{1}{4} P_2 Q_1^2 \,.
\end{aligned}
\tag{6.9}
$$

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 29, p. 537

fast track:
chapter 6.4, p. 100

6.3 Rectification of maps

In sect. 6.2 we had argued that nonlinear coordinate transformations can be profitably employed to simplify the representation of a flow. We shall now apply the same idea to nonlinear maps, and determine a smooth nonlinear change of coordinates that flattens out the vicinity of a fixed point and makes the map *linear* in an open neighborhood. In its simplest form the idea can be implemented only for an isolated nondegenerate fixed point (otherwise are needed in the normal form expansion around the point), and only in a finite neigborhood of a point, as the conjugating function in general has a finite radius of convergence. In sect. 6.3.2 we will extend the method to periodic orbits.

6.3.1 Rectification of a fixed point in one dimension

6.2
page 104

Consider a 1-dimensional map $x_{n+1} = f(x_n)$ with a fixed point at $x = 0$, with stability $\Lambda = f'(0)$. If $|\Lambda| \neq 1$, one can determine term-by-term the power series for a smooth conjugation $h(x)$ centered at the fixed point, $h(0) = 0$, that flattens out the neighborhood of the fixed point

$$f(x) = h^{-1}(\Lambda h(x)) \tag{6.10}$$

and replaces the nonlinear map $f(x)$ by a *linear* map $y_{n+1} = \Lambda y_n$.

To compute the conjugation h we use the functional equation $h^{-1}(\Lambda x) = f(h^{-1}(x))$ and the expansions

$$\begin{aligned} f(x) &= \Lambda x + x^2 f_2 + x^3 f_3 + \dots \\ h^{-1}(x) &= x + x^2 h_2 + x^3 h_3 + \dots \quad . \end{aligned} \tag{6.11}$$

If $h(x)$ is a conjugation, so is any scaling $h(bx)$ of the function for a real number b. Hence the value of $h'(0)$ is not determined by the functional equation (6.10); it is convenient to set $h'(0) = 1$.

The algebra is not particularly illuminating and best left to computers. In any case, for the time being we will not use much beyond the first, linear term in these expansions.

Here we assume $\Lambda \neq 1$. If the fixed point has first $k-1$ derivatives vanishing, the conjugacy is to the kth *normal form.*

In several dimensions, Λ is replaced by the Jacobian matrix, and one has to check that the eigenvalues $\mathbf{J}$ are non-resonant, that is, there is no integer linear relation between the stability exponents (4.37).

6.3.2 Rectification of a 1-dimensional periodic orbit

Now that we have constructed the conjugation function for a fixed point, we turn to the problem of constructing it for periodic orbits. Each point around the cycle has a differently distorted neighborhood, with differing second and higher order derivatives, so we need to compute a different conjugation function h_a at each cycle point x_a. We expand the map f around each cycle point along the cycle,

$$y_a(\phi) = f_a(\phi) - x_{a+1} = \phi f_{a,1} + \phi^2 f_{a,2} + \dots$$

where x_a is a point on the cycle, $f_a(\phi) = f(x_a+\phi)$ is centered on the periodic orbit, and the index k in $f_{a,k}$ refers to the kth order in the expansion (6.11).

For a periodic orbit the conjugation formula (6.10) generalizes to

$$f_a(\phi) = h_{a+1}^{-1}(f_a'(0)h_a(\phi)) , \qquad a = 1, 2, \cdots, n ,$$

point by point. The conjugationg functions h_a are obtained in the same way as before, by equating coefficients of the expansion (6.11), and assuming that the cycle stability $\Lambda = \prod_{a=0}^{n-1} f'(x_a)$ is not marginal, $|\Lambda| \neq 1$. The explicit expressions for h_a in terms of f are obtained by iterating around the whole cycle,

$$f^n(x_a + \phi) = h_a^{-1}(\Lambda h_a(\phi)) + x_a . \tag{6.12}$$

☞ remark 6.1

evaluated at each cycle point a. Again we have the freedom to set $h_a'(0) = 1$ for all a.

6.3.3 Repeats of cycles

We have traded in our initial nonlinear map f for a (locally) linear map Λy and an equally complicated conjugation function h. What is gained by rewriting the map f in terms of the conjugacy function h? Once the neighborhood of a fixed point is linearized, the repeats of it are trivialized; from the conjugation formula (6.11) one can compute the derivatives of a function composed with itself r times:

$$f^r(x) = h^{-1}(\Lambda^r h(x)) .$$

One can already discern the form of the expansion for arbitrary repeats; the answer will depend on the conjugacy function $h(x)$ computed for a *single* repeat, and all the dependence on the repeat number will be carried by factors polynomial in Λ^r, a considerable simplification. The beauty of the idea is difficult to gauge at this stage - an appreciation only sets in when one starts computing perturbative corrections, be it in celestial mechanics (where the method was born), be it the quantum or stochastic corrections to “semiclassical” approximations.

6.4 Smooth conjugacies and cycle stability

In sect. 4.5.2 we have established that for a given flow the cycle stability eigenvalues are intrinsic to a given cycle, independent of the staring point along the cycle. Now we can prove a much stronger statement; cycle stability eigenvalues are *metric invariants* of the flow, the same 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. 6.1: If the same dynamics is given by a map f in x coordinates, and a map g in the $y = h(x)$ coordinates, then f and g (or any other good representation) are related by (6.2), 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,

$$
\begin{aligned}
g'(y) &= h'(f \circ h^{-1}(y)) f'(h^{-1}(y)) \frac{1}{h'(x)} \\
&= h'(x) f'(x) \frac{1}{h'(x)} = f'(x) \,,
\end{aligned}
\tag{6.13}
$$

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.

6.2
page 104

in depth:
appendix C.1, p. 621

Commentary

Remark 6.1 Rectification of maps. The methods outlined above are standard in the analysis of fixed points and construction of normal forms for bifurcations, see for example ref. [1.11, 6.2, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 3.5]. The geometry underlying such methods is pretty, and we enjoyed reading, for example, Percival and Richards [6.10], chaps. 2 and 4 of Ozorio de Almeida's monograph [6.11], and, as always, Arnol'd [6.1].

Recursive formulas for evaluation of derivatives needed to evaluate (6.11) are given, for example, in Appendix A of ref. [7.3]).

Résumé

Dynamics $(\mathcal{M}, f)$ is invariant under the group of all smooth conjugacies

$$(\mathcal{M}, f) \rightarrow (\mathcal{M}', g) = (h(\mathcal{M}), h \circ f \circ h^{-1}) .$$

This invariance can be used to (i) find a simplified representation for the flow and (ii) identify a set of invariants, numbers computed within a particular choice of $(\mathcal{M}, f)$, but invariant under all $\mathcal{M} \rightarrow h(\mathcal{M})$ smooth conjugacies.

The $2D$-dimensional phase space of an integrable Hamiltonian system of D degrees of freedom is fully foliated by D-tori. In the same spirit, for a uniformly hyperbolic, chaotic dynamical system one would like to change into a coordinate frame where the stable/unstable manifolds form a set of transversally interesecting hyper-planes, with the flow everywhere locally hyperbolic. That, cannot be done in general: Fully globally integrable and fully globally chaotic flows are a very small subset of all possible flows.

Nevertheless, we can profitably straighten out local neighborhoods of periodic orbits. Periodic orbits play a central role in any invariant characterization of the dynamics, as they form an infinite set of *metric invariants*: The stability eigenvalues of a periodic orbit remain invariant under any smooth nonlinear change of coordinates $f \rightarrow h \circ f \circ h^{-1}$.

What we *really* care about is developping invariant notions of what a given dynamical system is. The totality of smooth one-to-one nonlinear coordinate transformations h which map all trajectories of a given dynamical system $(\mathcal{M}, f^t)$ onto all trajectories of dynamical systems $(\mathcal{M}', g^t)$ gives us a huge equivalence class, much larger than the equivalence classes familiar from the theory of linear transformations, such as the rotation group $O(d)$ or the Galilean group of all rotations and translations in $\mathbb{R}^d$. In the theory of Lie groups, the full invariant specification of an object is given by a finite set of Casimir invariants. What a good full set of invariants for a group of general nonlinear smooth conjugacies might be is not known, but the set of all periodic orbits and their stability eigenvalues will turn out to be a good start.

References

[6.1] V.I. Arnol'd, *Ordinary Differential Equations* (Springer-Verlag, New York 1992).

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

[6.3] C. Simo, in *Dynamics and Mission Design Near Libration Points*, Vol. 1-4, (World Sci. Pub., Monograph Ser. Math., 2000-2001).

[6.4] C. L. Siegel. Iteration of analytic functions. *Ann. Math.*, 43:607–612, 1942.

[6.5] J. Moser. *Ann. Scuola Norm. Super. Pisa*, 20:265–315, 1966; 20:499–535, 1966.

[6.6] S. Sternberg. *Amer. J. Math.*, 79:809, 1957; 80:623, 1958; 81:578, 1959.

[6.7] K.-T. Chen. *Amer. J. Math.*, 85:693–722, 1963.

[6.8] G.R. Belitskiĭ. *Russian Math. Surveys*, 31:107–177, 1978.

[6.9] A.D. Brjuno. *Trans. Moscow Math. Soc.*, 25:131–288, 1971; 26:199–238, 1972.

[6.10] I. Percival and D. Richards, *Introduction to Dynamics* (Cambridge Univ. Press, Cambridge, 1982).

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

Exercises

Exercise 6.1 Coordinate transformations. *Changing coordinates is conceptually simple, but can become confusing when carried out in detail. The difficulty arises from confusing functional relationships, such as $x(t) = h^{-1}(y(t))$ with numerical relationships, such as $w(y) = h'(x)v(x)$. Working through an example will clear this up.*

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

(b) *Show that in the transformed space $\mathcal{M}'$, the differential equation is*

$$\frac{d}{dt}\left[\begin{array}{c} y_1 \\ y_2 \end{array}\right] = \frac{1}{3}\left[\begin{array}{c} 5y_1 + 2y_2 \\ y_1 + 4y_2 \end{array}\right] . \tag{6.14}$$

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

Exercise 6.2 Linearization for maps. *Let $f : C \to C$ be a map from the complex numbers into themselves, with a fixed point at the origin and analytic there. By manipulating power series, find the first few terms of the map h that conjugates f to 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)

Chapter 7

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, 3 and 5 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.

7.1 Measures

> Do I then measure, O my God, and know not what I measure?
> St. Augustine, *The confessions of Saint Augustine*

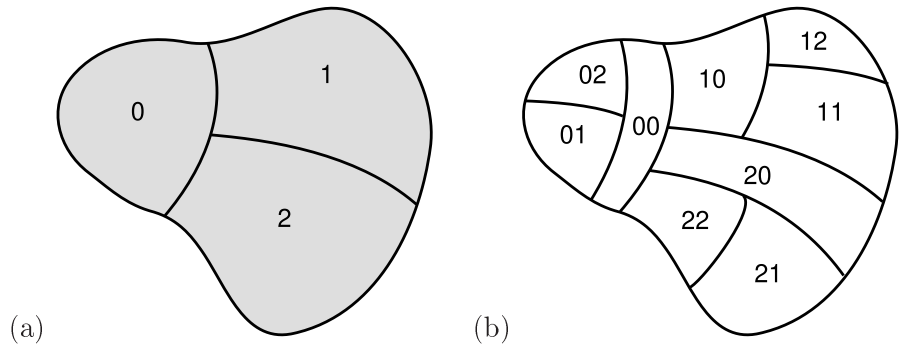


Figure 7.1: (a) First level of partitioning: A coarse partition of $\mathcal{M}$ into regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$. (b) $n = 2$ level of partitioning: A refinement of the above partition, with each region $\mathcal{M}_i$ subdivided into $\mathcal{M}_{i0}$, $\mathcal{M}_{i1}$, and $\mathcal{M}_{i2}$.

A fundamental concept in the description of dynamics of a chaotic system is that of *measure*, which we denote by $d\mu(x) = \rho(x)dx$. An intuitive way to define and construct a physically meaningful measure is by a process of *coarse-graining*. Consider a sequence 1, 2, ..., n, ... of more and more refined partitions of the phase space, fig. 7.1, into regions $\mathcal{M}_i$ defined by the characteristic function

$$\chi_i(x) = \begin{cases} 1 & \text{if } x \in \text{ region } \mathcal{M}_i \\ 0 & \text{otherwise} \end{cases} . \tag{7.1}$$

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:

$$\Delta\mu_i = \int_{\mathcal{M}} d\mu(x)\chi_i(x) = \int_{\mathcal{M}_i} d\mu(x) = \int_{\mathcal{M}_i} dx\, \rho(x) . \tag{7.2}$$

$\rho(x) = \rho(x,t)$ 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

$$\sum_i^{(n)} \Delta\mu_i = 1 , \tag{7.3}$$

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 $n \to \infty$ infinitely refined partition limit of $\Delta\mu_i = |\mathcal{M}_i|\rho(x_i)\,,\quad x_i \in \mathcal{M}_i$, with normalization

$$\int_{\mathcal{M}} dx\, \rho(x) = 1\,. \tag{7.4}$$

So far, any arbitrary sequence of partitions will do. What are intelligent ways of partitioning the phase space? We postpone the answer to chapter 9, after we have developed some intuition about how the dynamics transports densities.

☞ chapter 9

7.2 Perron-Frobenius operator

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*

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 (4.46)

$$\rho(x,t) = \frac{\rho(x_0,0)}{|\det \mathbf{J}^t(x_0)|}\,, \qquad x = f^t(x_0)\,, \tag{7.5}$$

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

$$\rho(x,t) = \mathcal{L}^t \rho(x) = \int_{\mathcal{M}} dx_0\, \delta\big(x - f^t(x_0)\big)\, \rho(x_0, 0)\,. \tag{7.6}$$

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:

$$\int dx\, \delta(h(x)) = \sum_{x \in \mathrm{Zero}\,[h]} \frac{1}{|h(x)'|}\,, \tag{7.7}$$

7.1 page 120

and in d dimensions the denominator is replaced by

$$\int dx\, \delta(h(x)) = \sum_{x \in \mathrm{Zero}\,[h]} \frac{1}{\left|\det \frac{\partial h(x)}{\partial x}\right|}\,. \tag{7.8}$$

7.2 page 120

Now you can check that (7.6) is just a rewrite of (7.5):

$$\begin{aligned} \mathcal{L}^t \rho(x) &= \sum_{x_0 = f^{-t}(x)} \frac{\rho(x_0)}{|f^t(x_0)'|} && \text{(1-dimensional)} \\ &= \sum_{x_0 = f^{-t}(x)} \frac{\rho(x_0)}{|\det \mathbf{J}^t(x_0)|} && (d\text{-dimensional})\,. \end{aligned} \tag{7.9}$$

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. 9.4.

We shall refer to the kernel of (7.6) as the *Perron-Frobenius operator*:

7.3 page 120

$$\mathcal{L}^t(x,y) = \delta\big(x - f^t(y)\big)\ . \tag{7.10}$$

sect. 13.3.1

remark 12.4

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. 627

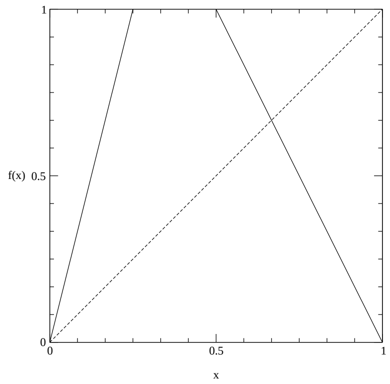


Figure 7.2: A piecewise-linear repeller (7.11): All trajectories that land in the gap between the f_0 and f_1 branches escape ($\Lambda_0 = 4$, $\Lambda_1 = -2$).

Example 7.1 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 the expanding 1-*d *map* $f(x)$ *of fig. 7.2, a piecewise-linear 2–branch repeller with slopes* $\Lambda_0 > 1$ *and* $\Lambda_1 < -1$ *:*

$$f(x) = \begin{cases} f_0 = \Lambda_0 x & \textit{if} \quad x \in \mathcal{M}_0 = [0, 1/\Lambda_0] \\ f_1 = \Lambda_1 (x-1) & \textit{if} \quad x \in \mathcal{M}_1 = [1 + 1/\Lambda_1, 1] \end{cases} . \tag{7.11}$$

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*

$$\rho(x) = \begin{cases} \rho_0 & \textit{if } x \in \mathcal{M}_0 \\ \rho_1 & \textit{if } x \in \mathcal{M}_1 \end{cases} . \tag{7.12}$$

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. 1.7, with* f_0 *and* f_1 *modelling its two strips of survivors.*

As can be easily checked by using (7.9), the Perron-Frobenius operator acts on this piecewise constant function as a [2×2] "transfer" matrix with matrix elements 7.1 page 120

$$\begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} \to \mathcal{L}\rho = \begin{pmatrix} \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \\ \frac{1}{|\Lambda_0|} & \frac{1}{|\Lambda_1|} \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \end{pmatrix} , \tag{7.13}$$

7.5 page 121

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 non-zero 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* $\mathcal{L}$*:*

$$\gamma = -s_0 = -\ln(1/|\Lambda_0| + 1/|\Lambda_1|) . \tag{7.14}$$

Voila! Here is the rationale for introducing operators – in one time step we have solved the problem of evaluating escape rate at infinite time. This 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. The example gives a foretaste of the enterprise upon which we are to embark in this book, but the full story is much subtler: 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 13

7.3 Invariant measures

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

$$\rho(x,t) = \rho(x,0) = \rho(x)\,. \tag{7.15}$$

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 (7.10), with the unit eigenvalue:

7.3 page 120

$$\mathcal{L}^t\rho(x) = \int_{\mathcal{M}} dy\, \delta(x - f^t(y))\rho(y) = \rho(x). \tag{7.16}$$

In general, depending on the choice of $f^t(x)$ and the function space for $\rho(x)$, there may be no, one, or many solutions of the eigenfunction condition (7.16). 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.

From a physical point of view, there is no way to prepare initial densities which are singular, so it makes sense to concentrate on measures which are limits of transformations which an initial smooth distribution $\rho(x)$ experiences under the action of f, rather than as a limit computed from a single trajectory,

$$\rho_0(x) = \lim_{t\to\infty} \int_{\mathcal{M}} dy\, \delta(x - f^t(y))\rho(y,0)\,, \qquad \int_{\mathcal{M}} dy\, \rho(y,0) = 1\,. \tag{7.17}$$

Intuitively, the "natural" measure should also be the one least sensitive to inevitable facts of life, such as noise, not matter how weak.

7.3.1 Natural measure

> **Huang:** Chen-Ning, do you think ergodic theory gives us useful insight into the foundation of statistical mechanics?
> **Yang:** I don't think so.
> Kerson Huang, *C.N. Yang interview*

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

$$\overline{\rho}_{x_0}(y) \quad = \quad \begin{cases} \lim_{t\to\infty} \frac{1}{t}\int_0^t d\tau\, \delta(y - f^\tau(x_0)) & \text{flows} \\ \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \delta\big(y - f^k(x_0)\big) & \text{maps} \end{cases}\,, \tag{7.18}$$

7.8 page 122
7.9 page 122

where x_0 is a generic inital point. Staring at an average over ∞ many Dirac deltas is not a prospect we cherish. Generated by the action of f, the natural measure satisfies the stationarity condition (7.16) and is thus invariant by construction. From the computational point of view, the natural measure is the visitation frequency defined by coarse-graining, integrating (7.18) over the $\mathcal{M}_i$ region

$$\Delta\overline{\mu}_i = \lim_{t\to\infty} \frac{t_i}{t}\,, \tag{7.19}$$

where t_i is the accumulated time that a trajectory of total duration t spends in the $\mathcal{M}_i$ region, with the initial point x_0 picked from some smooth density $\rho(x)$.

Let $a = a(x)$ be any *observable*. In mathematical literature $a(x)$ is a function belonging to some function space, for instance the space of integrable functions L^1, that associates to each point in phase space a number or a set of numbers. In physical application the observable $a(x)$ is of necessity a smooth function. The observable reports on some property of the dynamical system (several examples will be given in sect. 8.1).

The *space average* of the observable a with respect to measure ρ is given by the d-dimensional integral over the phase space $\mathcal{M}$:

$$\langle a \rangle_\rho = \frac{1}{|\rho_{\mathcal{M}}|}\int_{\mathcal{M}} dx\, \rho(x)a(x)\,, \quad |\rho_{\mathcal{M}}| = \int_{\mathcal{M}} dx\, \rho(x) = \text{mass in } \mathcal{M}\,. \tag{7.20}$$

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

Inserting the right hand side of (7.18) into (7.20) we see that the natural measure corresponds to *time average* of the observable a along a trajectory of the initial point x_0,

$$\overline{a(x_0)} = \lim_{t\to\infty} \frac{1}{t}\int_0^t d\tau\, a(f^\tau(x_0))\,. \tag{7.21}$$

☞ appendix A

Analysis of the above asymptotic 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 (7.21) exists for all initial x_0. As we shall not rely on this result in what follows we forgo a proof here. Furthermore, if the dynamical system is *ergodic*, the time average over almost any trajectory tends to the space average

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t d\tau\, a(f^\tau(x_0)) = \langle a \rangle \tag{7.22}$$

for "almost all" initial x_0. By "almost all" we mean that the time average is independent of the initial point apart from a set of 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,

$$\lim_{t\to\infty} \left\langle a(x)b(f^t(x))\right\rangle = \langle a \rangle \langle b \rangle\ , \tag{7.23}$$

☞ sect. 16.3

the dynamical system is said to be *mixing*.

An example of a numerical calculation of the natural measure (7.19) for the Hénon attractor (3.10) is given by the histogram of fig. 7.3. The phase space is partitioned into many equal size areas $\mathcal{M}_i$, and the coarse grained measure (7.19) computed by a long time iteration of the Hénon map, and represented by the height of the column over area $\mathcal{M}_i$. What you see is a typical invariant measure - a complicated, singular function concentrated

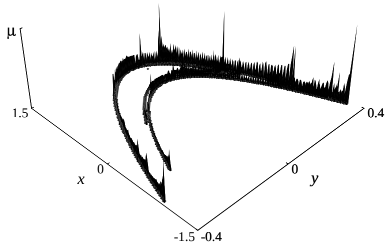


Figure 7.3: Natural measure (7.19) for the Hénon map (3.10) strange attractor at parameter values $(a, b) = (1.4, 0.3)$. See fig. 3.4 for a sketch of the attractor without the natural measure binning. (Courtesy of J.-P. Eckmann)

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 (7.18) is problematic and often hard to control, as generic dynamical systems are neither uniformly hyperbolic nor structurally stable: it is not known whether even the simplest model of a strange attractor, the Hénon attractor, is a strange attractor or merely a long stable cycle.

8.1
page 143

While dynamics can lead to very singular ρ's, in any physical setting we cannot do better than to measure it averaged over some region $\mathcal{M}_i$; the coarse-graining is not an approximation but a physical necessity. One is free to think of a measure as a probability density, as long as one keeps in mind the distinction between deterministic and stochastic flows. In deterministic evolution the evolution kernels are not probabilistic, the density of trajectories is transported *deterministically*. What this distinction means will became apparent later 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. 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, not in this book. 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 $\Delta\rho_i$.

☞ chapter 12

7.4 Density evolution for infinitesimal times

Consider the evolution of a smooth density $\rho(x) = \rho(x, 0)$ under an infinitesimal step $\delta\tau$, by expanding the action of $\mathcal{L}^{\delta\tau}$ to linear order in $\delta\tau$:

$$\begin{aligned}
\mathcal{L}^{\delta\tau}\rho(y) &= \int_{\mathcal{M}} dx\, \delta\Big(y - f^{\delta\tau}(x)\Big)\, \rho(x) = \int_{\mathcal{M}} dx\, \delta(y - x - \delta\tau v(x))\, \rho(x) \\
&= \frac{\rho(y - \delta\tau v(y))}{\left|\det\left(1 + \delta\tau \frac{\partial v(y)}{\partial x}\right)\right|} = \frac{\rho(y) - \delta\tau \sum_{i=1}^{d} v_i(y)\partial_i\rho(y)}{1 + \delta\tau \sum_{i=1}^{d} \partial_i v_i(y)}
\end{aligned}$$

$$\rho(x,\delta\tau) \;=\; \rho(x,0) - \delta\tau \sum_{i=1}^{d} \frac{\partial}{\partial x_i}(v_i(x)\rho(x,0))\,. \tag{7.24}$$

Here we have used the infinitesimal form of the flow (2.5), Dirac delta jacobian (7.9), and the $ln det = tr ln$ relation. Moving $\rho(y,0)$ to the left hand side and dividing by $\delta\tau$, we discover that the rate of the deformation of ρ under the infinitesimal action of the Perron-Frobenius operator is nothing but the *continuity equation* for the density:

$$\partial_t\rho + \partial\cdot(\rho v) \;=\; 0\,. \tag{7.25}$$

The family of Perron-Frobenius operators operators $\left\{\mathcal{L}^t\right\}_{t\in\mathbb{R}_+}$ forms a semi-group parametrized by time

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

(b) $\mathcal{L}^t\mathcal{L}^{t'} = \mathcal{L}^{t+t'} \qquad t,t'\geq 0 \qquad$ (semigroup property) .

The *generator* of the semigroup, the generator of infinitesimal time translation by an infinitesimal step $\delta\tau$ is determined by (7.24)

$$\mathcal{A}\rho(x) \;=\; +\lim_{\delta\tau\to 0^+}\frac{1}{\delta\tau}\left(\mathcal{L}^{\delta\tau} - I\right)\rho(x)\,. \;=\; -\partial_i(v_i(x)\rho(x))\,. \tag{7.26}$$

(If the flow is finite-dimensional and invertible, $\mathcal{A}$ is a generator of a group). Of course, the left hand side is the definition of the time derivative, so the evolution equation for $\rho(x)$ is

$$\left(\frac{\partial}{\partial t} - \mathcal{A}\right)\rho(x) \;=\; 0\,. \tag{7.27}$$

☞ appendix D.3

The above statements apply to any deterministic flow. If the density itself is a material invariant, combining (D.1) and (7.25) 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. 5.1.1. For incompressible flows the continuity equation (7.25) becomes a statement of conservation of the phase space volume (see sect. 5.1.1), or the *Liouville theorem*

$$\partial_t\rho + v_i\partial_i\rho = 0\,. \tag{7.28}$$

The finite time Perron-Frobenius operator (7.10) can be formally expressed by exponentiating the time evolution generator $\mathcal{A}$ as

$$\mathcal{L}^t = e^{t\mathcal{A}} . \tag{7.29}$$

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:

7.11 page 123

$$e^{-tv\frac{\partial}{\partial x}} a(x) = a(x - tv) . \tag{7.30}$$

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

appendix D.3

The Perron-Frobenius operator $\mathcal{L}$ acts multiplicatively in time, so it is reasonable to suppose that there exist constants $M > 0$, $\beta \geq 0$ such that $||\mathcal{L}^t|| \leq Me^{t\beta}$ for all $t \geq 0$. What does that mean? The operator norm is defined in the same spirit in which we defined the matrix norms in sect. K.2: We are assuming that no value of $\mathcal{L}^t\rho(x)$ grows faster than exponentially for any choice of function $\rho(x)$, so that the fastest possible growth can be bounded by $e^{t\beta}$, a reasonable expectation in the light of the simplest example studied so far, the exact escape rate (7.14). If that is so, multiplying $\mathcal{L}^t$ by $e^{-t\beta}$ we construct a new operator $e^{-t\beta}\mathcal{L}^t = e^{t(\mathcal{A}-\beta)}$ which decays exponentially for large t, $||e^{t(\mathcal{A}-\beta)}|| \leq M$. We say that $e^{-t\beta}\mathcal{L}^t$ is an element of a *bounded* semigroup with generator $\mathcal{A} - \beta I$. Given this bound, it follows by the Laplace transform

$$\int_0^\infty dt\, e^{-st}\mathcal{L}^t = \frac{1}{s - \mathcal{A}} , \qquad \mathrm{Re}\, s > \beta , \tag{7.31}$$

that the *resolvent* operator $(s - \mathcal{A})^{-1}$ is bounded ("resolvent" = able to cause separation into constituents)

sect. K.2

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

If one is interested in the spectrum of $\mathcal{L}$, 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 (7.29) into the resolvent form (7.31) and enables us to study its spectrum.

in depth:
appendix D.3, p. 631

7.4.1 Liouville operator

A case of special interest is the Hamiltonian or symplectic flow defined by the time-independent Hamilton's equations of motion (5.1). A reader versed in quantum mechanics will have observed by now that with replacement $\mathcal{A} \to -\frac{i}{\hbar}\hat{H}$, where $\hat{H}$ is the quantum Hamiltonian operator, (7.27) 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

$$\dot{q}_i = \frac{p_i}{m}, \qquad \dot{p}_i = -\frac{\partial V(q)}{\partial q_i}. \tag{7.32}$$

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

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial Q}{\partial p_i}\frac{dp_i}{dt} = \frac{\partial H}{\partial p_i}\frac{\partial Q}{\partial q_i} - \frac{\partial Q}{\partial p_i}\frac{\partial H}{\partial q_i}. \tag{7.33}$$

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

$$[A,B] = \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i} - \frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i}. \tag{7.34}$$

In terms of Poisson brackets the time evolution equation (7.33) takes the compact form

$$\frac{dQ}{dt} = [H,Q]. \tag{7.35}$$

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 (5.1) 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

$$\partial_t \rho + \partial_i(\rho v_i) = 0\,. \tag{7.36}$$

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 (5.1) we obtain the *Liouville equation*, a special case of (7.27):

$$\frac{\partial}{\partial t}\rho(q,p,t) = -\mathcal{A}\rho(q,p,t) = [H,\rho(q,p,t)]\,, \tag{7.37}$$

where [,] is the Poisson bracket (7.34). The generator of the flow (7.26) is now the generator of infinitesimal symplectic transformations,

$$\mathcal{A} = \dot{q}_i\frac{\partial}{\partial q_i} + \dot{p}_i\frac{\partial}{\partial p_i} = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i}\,. \tag{7.38}$$

or, by the Hamilton's equations for separable Hamiltonians

$$\mathcal{A} = -\frac{p_i}{m}\frac{\partial}{\partial q_i} + \partial_i V(q)\frac{\partial}{\partial p_i}\,. \tag{7.39}$$

7.12
page 123

This special case of the time evolution generator (7.26) for the case of symplectic flows is called the *Liouville operator*. You might have encountered it in statistical mechanics, while discussing what ergodicity means for 10^{23} hard balls, or along a 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 surprise learn that that suffices to get a grip on the foundations of the classical nonequilibrium statistical mechanics.

in depth:
sect. D.3, p. 631

Commentary

Remark 7.1 Ergodic theory. An overview of ergodic theory is outside the scope of this book: the interested reader may find it useful to consult [7.1]. The existence of time average (7.21) is the basic result of ergodic theory, known as the Birkhoff theorem, see for example refs. [7.1, 1.11], or the statement of the theorem 7.3.1 in ref. [D.1]. The natural measure (7.19) (more carefully defined than in the above sketch) is often referred to as the SBR or Sinai-Bowen-Ruelle measure [1.14, 1.13, 1.16].

Remark 7.2 Bounded semigroup. For a discussion of bounded semigroups of page 115 see, for example, Marsden and Hughes [7.2].

Résumé

In physically realistic setting initial state of a system can be specified only to a finite precision. Depending on the desired precision, and given a deterministic law of evolution, h, the system can then be tracked for some finite time, however, if dynamics is chaotic, it is not possible to calculate accurately the long time trajectory of a given initial point. Hence, the study of long time dynamics requires going into a trading in evolution of a single phase space point for 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. If the Lyapunov time (1.1), the time after which the notion of an individual deterministic trajectory loses meaning, is much shorter than the observation time, the "sharp" observables are dual to time, the eigenvalues of evolution operators. This is very much the same situation as in quantum mechanics, as atomic time scales are so short, what is accessible to measure is the energy spectrum.

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.

In what follows we shall find the second formulation more convenient, but the alternative is worth keeping in mind when posing and solving invariant density problems. However, as the classical evolution operators are not unitary, their eigenstates can be quite singular and difficult to work with. In what follows we shall learn how to avoid this altogether.

References

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

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

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

[7.4] We thank Laurette Tuckerman for providing us with the pseudospectral code.

[7.5] We have used the NAG library integration routine D02BHF. Our code and numerical results are available on `www.nbi.dk/ChaosBook/extras`.

Exercises

Exercise 7.1 **Integrating over Dirac delta functions.** *Let us verify a few of the properties of the delta function and check (7.9), as well as the formulas (7.7) and (7.8) 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.

Exercise 7.2 **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)
$$\int dx\, \delta^{(2)}\left(y\right) = \sum_{x:y(x)=0} \frac{1}{|y'|} \left\{ 3\frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3} \right\} \,. \tag{7.40}$$

(c)
$$\int dx\, b(x) \delta^{(2)}(y) = \sum_{x:y(x)=0} \frac{1}{|y'|} \left\{ \frac{b''}{(y')^2} - \frac{b'y''}{(y')^3} + b\left(3\frac{(y'')^2}{(y')^4} - \frac{y'''}{(y')^3}\right) \right\} . \tag{7.41}$$

These formulas are useful incomputing effects of weak noise on deterministic dynamics [7.3].

Exercise 7.3 **$\mathcal{L}^t$ generates a semigroup.** *Check that the Perron-Frobenius operator has the semigroup property,*

$$\int_M dz \mathcal{L}^{t_2}(y,z)\, \mathcal{L}^{t_1}(z,x) = \mathcal{L}^{t_2+t_1}(y,x) \,, \qquad t_1, t_2 \geq 0 \,. \tag{7.42}$$

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 $t_1, t_2 \in \mathbb{R}$.

Exercise 7.4 **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*

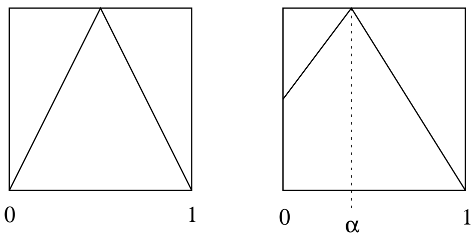
$$f(x) = a(1 - 2|x - 0.5|)$$

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).*

Exercise 7.5 **Invariant measure.** *We will compute the invariant measure for two different piecewise linear maps.*



(a) *Verify the matrix* $\mathcal{L}$ *representation (7.13).*

(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.*

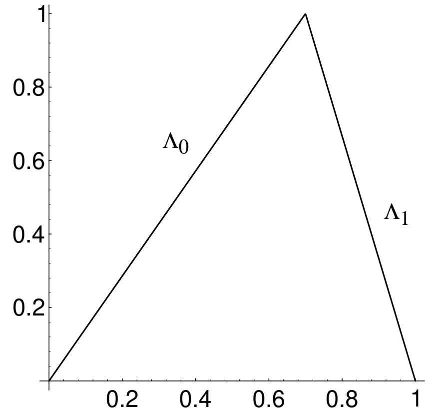
(d) *For this map there is an infinite number of invariant measures, but only one of them will be found when one carries out a numerical simulation. Determine that measure, and explain why your choice is the natural measure for this map.*

(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)

Exercise 7.6 **Escape rate for a flow conserving map.** *Adjust* Λ_0, Λ_1 *in (7.11) so that the gap between the intervals* $\mathcal{M}_0$, $\mathcal{M}_1$ *vanishes. Check that in that case the escape rate equals zero.*

Exercise 7.7 Eigenvalues of the skew Ulam tent map Perron-Frobenius operator. *Show that for the skew Ulam tent map*



$$f(x) = \begin{cases} f_0(x) = \Lambda_0 x\,, & x \in \mathcal{M}_0 = [0, 1/\Lambda_0) \\ f_1(x) = \frac{\Lambda_0}{\Lambda_0 - 1}(1-x)\,, & x \in \mathcal{M}_1 = (1/\Lambda_0, 1]\,. \end{cases} \tag{7.43}$$

the eigenvalues are available analytically, compute the first few.

Exercise 7.8 "Kissing disks"* *(continuation of exercises 5.1 and 5.2). 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 18.*

Exercise 7.9 Invariant measure for the Gauss map. *Consider the Gauss map (we shall need this map in chapter 21):*

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

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?*

Exercise 7.10 **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 $a(x)$.*

Exercise 7.11 **$\mathcal{A}$ as a generator of translations.** *Verify that for a constant velocity field the evolution generator $\mathcal{A}$ n (7.30) is the generator of translations,*

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

(hint: expand $a(x)$ in a Tylor series.)

Exercise 7.12 **Incompressible flows.** *Show that (7.9) 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 7.8.*

Chapter 8

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. 8.3 we try out the formalism on the first quantitative diagnosis that a system's got chaos, Lyapunov exponents.

8.1 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.

8.1.1 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,

$$A^t(x_0) = \int_0^t d\tau \, a(f^\tau(x_0)) \,. \tag{8.1}$$

If the dynamics is given by an iterated mapping and the time is discrete, $t \to n$, the integrated observable is given by

$$A^n(x_0) = \sum_{k=0}^{n-1} a(f^k(x_0)) \tag{8.2}$$

(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 27):

$$A^t(x_0) = \int_0^t d\tau \, \dot{\mathbf{q}}(\tau) \cdot \mathbf{p}(\tau) \,. \tag{8.3}$$

The *time average* of the observable along a trajectory is defined by

$$\overline{a(x_0)} = \lim_{t\to\infty} \frac{1}{t} A^t(x_0) \,. \tag{8.4}$$

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^oF and $+130^oF$ for all times – $A^t(x_0)$ is expected to grow not faster than t, and the limit (8.4) exists. For an example of a time average - the Lyapunov exponent - see sect. 8.3.

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

4.3
page 81

$$\begin{aligned}
\text{flows:} \quad A_p &= a_p T_p = \int_0^{T_p} a\left(f^\tau(x_0)\right) d\tau\,, \qquad x_0 \in p \\
\text{maps:} \quad &= a_p n_p = \sum_{i=0}^{n_p-1} a\left(f^i(x_0)\right)\,,
\end{aligned} \tag{8.5}$$

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 (8.4) requires therefore only a single traversal of the cycle:

$$a_p = \frac{1}{T_p} A_p\,. \tag{8.6}$$

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 (8.6) on any periodic orbit, that is, on a dense set of points (fig. 8.1(b)). For example, for an open

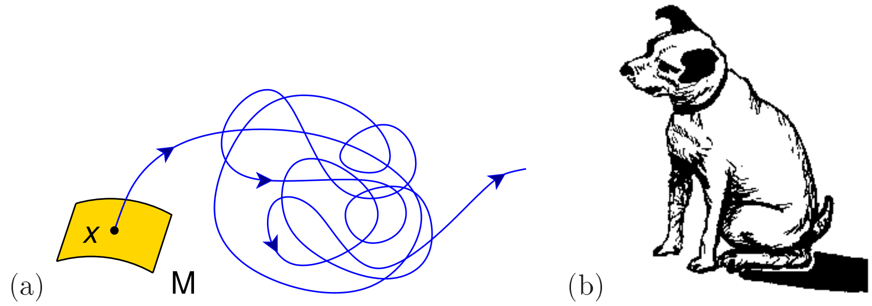


Figure 8.1: (a) A typical chaotic trajectory explores the phase space with the long time visitation frequency building up the natural measure $\rho_0(x)$. (b) time average evaluated along an atypical trajectory such as a periodic orbit fails to explore the entire accessible phase space. (A. Johansen)

system such as the Sinai gas of sect. 20.1 (an infinite 2-dimensional periodic array of scattering disks) the phase space is dense with initial points that correspond to periodic runaway trajectories. The mean distance squared traversed by any such trajectory grows as $x(t)^2 \sim t^2$, and its contribution to the diffusion rate $D \approx x(t)^2/t$, (8.4) evaluated with $a(x) = x(t)^2$, diverges. Seemingly there is a paradox; even though intuition says the typical motion should be diffusive, we have an infinity of ballistic trajectories.

☞ chapter 20

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.

8.1.2 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:

$$\begin{aligned} \langle a \rangle(t) &= \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\; a(x(t)) \\ |\mathcal{M}| &= \int_{\mathcal{M}} dx \; = \text{volume of } \mathcal{M}\,. \end{aligned} \tag{8.7}$$

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. 8.1.3).

What is it we *really* do in experiments? We cannot measure the time average (8.4), 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 (8.7), and consider instead

$$\langle a \rangle_\rho(t) = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx\, \rho(x) a(x(t))\,. \tag{8.8}$$

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 (8.7); 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 $\mathcal{M}$

$$\langle a \rangle = \lim_{t\to\infty} \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx\, \frac{1}{t} \int_0^t d\tau\, a(f^\tau(x))\,. \tag{8.9}$$

We use the same $\langle \cdots \rangle$ notation as for the space average (8.7), 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

$$\left\langle e^{\beta \cdot A^t} \right\rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dx\, e^{\beta \cdot A^t(x)}. \tag{8.10}$$

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 (8.4) 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

(8.10) 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 $\left\langle \exp(\beta \cdot A^t) \right\rangle$ 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

$$s(\beta) = \lim_{t\to\infty} \frac{1}{t} \ln \left\langle e^{\beta \cdot A^t} \right\rangle . \tag{8.11}$$

Now we understand one reason for why it is smarter to compute $\left\langle \exp(\beta \cdot A^t) \right\rangle$ rather than $\langle a \rangle$: the expectation value of the observable (8.9) and the moments of the integrated observable (8.1) can be computed by evaluating the derivatives of $s(\beta)$

$$\begin{aligned} \left.\frac{\partial s}{\partial \beta}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \left\langle A^t \right\rangle = \langle a \rangle , \\ \left.\frac{\partial^2 s}{\partial \beta^2}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \left(\left\langle A^t A^t \right\rangle - \left\langle A^t \right\rangle \left\langle A^t \right\rangle \right) \\ &= \lim_{t\to\infty} \frac{1}{t} \left\langle (A^t - t \langle a \rangle)^2 \right\rangle , \end{aligned} \tag{8.12}$$

8.2 page 144

and so forth. We have written out the formulas for a scalar observable; the vector case is worked out in the exercise 8.2. If we can compute the function $s(\beta)$, we have the desired expectation value without having to estimate any infinite time limits from finite time data.

Suppose we could evaluate $s(\beta)$ and its derivatives. What are such formulas good for? A typical application is to the problem of describing a particle scattering elastically off a 2-dimensional triangular array of disks. If the disks are sufficiently large to block any infinite length free flights, the particle will diffuse chaotically, and the transport coefficient of interest is the diffusion constant given by $\left\langle x(t)^2 \right\rangle \approx 4Dt$. In contrast to D estimated numerically from trajectories $x(t)$ for finite but large t, the above formulas yield the asymptotic D without any extrapolations to the $t \to \infty$ limit. For example, for $a_i = v_i$ and zero mean drift $\langle v_i \rangle = 0$, the diffusion constant is given by the curvature of $s(\beta)$ at $\beta = 0$,

$$D = \lim_{t\to\infty} \frac{1}{2dt} \left\langle x(t)^2 \right\rangle = \frac{1}{2d} \sum_{i=1}^{d} \left.\frac{\partial^2 s}{\partial \beta_i^2}\right|_{\beta=0} , \tag{8.13}$$

sect. 20.1

so if we can evaluate derivatives of $s(\beta)$, we can compute transport coefficients that characterize deterministic diffusion. As we shall see in chapter 20, periodic orbit theory yields an explicit closed form expression for D.

fast track:
sect. 8.2, p. 132

8.1.3 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, (8.9) 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

$$\int_{\mathcal{M}} dy\, \delta(y - f^t(x_0)) = 1\,, \quad t > 0\,, \qquad \text{iff } x_0 \in \text{ non–wandering set} \tag{8.14}$$

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

$$|\mathcal{M}(t)| = \int_{\mathcal{M}} dxdy\, \delta\big(y - f^t(x)\big) \ \sim \ |\mathcal{M}|e^{-\gamma t} \tag{8.15}$$

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 sect. 1.4.2

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. 16.1

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 (8.15), by redefining the space average (8.10) as

$$\left\langle e^{\beta\cdot A^t}\right\rangle = \int_{\mathcal{M}} dx\, \frac{1}{|\mathcal{M}(t)|} e^{\beta\cdot A^t(x)} \sim \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx\, e^{\beta\cdot A^t(x)+\gamma t}\,. \tag{8.16}$$

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 $\left\langle e^{\beta\cdot A^t}\right\rangle$.

8.2 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

$$\left\langle e^{\beta\cdot A^t}\right\rangle = \frac{1}{|\mathcal{M}|}\int_{\mathcal{M}} dx \int_{\mathcal{M}} dy\, \delta\big(y - f^t(x)\big)\, e^{\beta\cdot A^t(x)}\,. \tag{8.17}$$

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

$$1 = \int_{\mathcal{M}} dy\, \delta\big(y - f^t(x)\big)\ , \tag{8.18}$$

into (8.10) 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 $\mathcal{M}$.

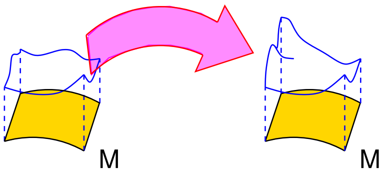


Figure 8.2: Space averaging pieces together the time average computed along the $t \to \infty$ trajectory of fig. 8.1 by a space average over infinitely many short t trajectory segments starting at all initial points at once. (A. Johansen)

As a matter of fact (and that is why we went to the trouble of defining the generator (7.26) 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 (8.17) as the *evolution operator*

$$\mathcal{L}^t(y, x) = \delta\big(y - f^t(x)\big)\, e^{\beta \cdot A^t(x)} . \tag{8.19}$$

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 (8.31). The evolution operator acts on scalar functions $a(x)$ as

$$\mathcal{L}^t a(y) = \int_{\mathcal{M}} dx\, \delta\big(y - f^t(x)\big)\, e^{\beta \cdot A^t(x)} a(x) . \tag{8.20}$$

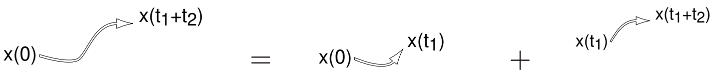
In terms of the evolution operator, the expectation value of the generating function (8.17) is given by

$$\left\langle e^{\beta \cdot A^t} \right\rangle = \left\langle \mathcal{L}^t \iota \right\rangle ,$$

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



$$\begin{aligned} A^{t_1+t_2}(x_0) &= \int_0^{t_1} d\tau\, a(x(\tau)) + \int_{t_1}^{t_1+t_2} d\tau\, a(x(\tau)) \\ &= A^{t_1}(x_0) + A^{t_2}(f^{t_1}(x_0)) . \end{aligned}$$

7.3
page 120

sect. 7.4

If $A^t(x)$ is additive along the trajectory, the evolution operator generates a semigroup

$$\mathcal{L}^{t_1+t_2}(y,x) = \int_{\mathcal{M}} dz\, \mathcal{L}^{t_2}(y,z)\mathcal{L}^{t_1}(z,x)\,, \tag{8.21}$$

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 (8.17) is preferable to (8.9) as a starting point for evaluation of dynamical averages: it recasts averaging in form of operators multiplicative along the flow.

8.3 Lyapunov exponents

(J. Mathiesen and P. Cvitanović)

sect. 1.3.1

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

$$x(t) = f^t(x_0) \quad \text{and} \quad x(t) + \delta x(t) = f^t(x_0 + \delta x(0)) \tag{8.22}$$

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

$$|\delta x(t)| \approx e^{\lambda t}|\delta x(0)| \tag{8.23}$$

where λ, the mean rate of separation of trajectories of the system, is called the *Lyapunov exponent*.

8.3.1 Lyapunov exponent as a time average

We can start out with a small δx and try to estimate λ from (8.23), but now that we have quantified the notion of linear stability in chapter 4 and defined the dynamical time averages in sect. 8.1.1, we can do better. The problem with measuring the growth rate of the distance between two points is that as the points separate, the measurement is less and less a local measurement. In study of experimental time series this might be the only option, but if we have the equations of motion, a better way is to measure the growth rate of vectors transverse to a given orbit.

The mean growth rate of the distance $|\delta x(t)|/|\delta x(0)|$ between neighboring trajectories (8.23) is given by the *Lyapunov exponent*

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln |\delta x(t)|/|\delta x(0)| \tag{8.24}$$

(For notational brevity we shall often suppress the dependence of quantities such as $\lambda = \lambda(x_0)$, $\delta x(t) = \delta x(x_0, t)$ on the initial point x_0 and the time t). For infinitesimal δx we know the $\delta x_i(t)/\delta x_j(0)$ ratio exactly, as this is by definition the Jacobian matrix (4.26)

$$\lim_{\delta x\to 0} \frac{\delta x_i(t)}{\delta x_j(0)} = \frac{\partial x_i(t)}{\partial x_j(0)} = \mathbf{J}^t_{ij}(x_0)\,,$$

so the leading Lyapunov exponent can be computed from the linear approximation (4.20)

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln \frac{\left|\mathbf{J}^t(x_0)\delta x(0)\right|}{|\delta x(0)|} = \lim_{t\to\infty} \frac{1}{2t} \ln \left| \hat{n}^T (\mathbf{J}^t)^T \mathbf{J}^t \hat{n} \right| \,. \tag{8.25}$$

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 positive eigenvalues $\{|\Lambda_1|^2 \geq \ldots \geq |\Lambda_d|^2\}$, and a complete orthonormal set of eigenvectors of $\{u_1, \ldots, u_d\}$. Expanding the initial orientation $\hat{n} = \sum (\hat{n}\cdot u_i) u_i$ in the $\mathbf{M}u_i = |\Lambda_i| u_i$ eigenbasis, we have

$$\hat{n}^T \mathbf{M} \hat{n} = \sum_{i=1}^{d} (\hat{n}\cdot u_i)^2 |\Lambda_i|^2 = (\hat{n}\cdot u_1)^2 e^{2\lambda_1 t} \left(1 + O(e^{-2(\lambda_1-\lambda_2)t})\right) \,, \tag{8.26}$$

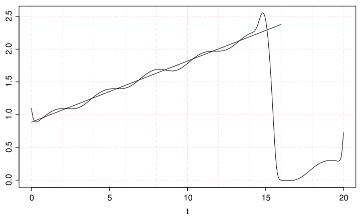


Figure 8.3: A numerical estimate of the leading Lyapunov exponent for the Rössler system (2.11) from the dominant expanding eigenvalue formula (8.25). The leading Lyapunov exponent $\lambda \approx 0.09$ is positive, so numerics supports the hypothesis that the Rössler attractor is strange. (J. Mathiesen)

where $t\lambda_i = \ln|\Lambda_i(x_0, t)|$, and we assume that $\lambda_1 > \lambda_2 \geq \lambda_3 \cdots$. For long times the largest Lyapunov exponent dominates exponentially (8.25), 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

$$\begin{aligned}\overline{\lambda(x_0)} &= \lim_{t\to\infty}\frac{1}{t}\left\{\ln|\hat{n}\cdot u_1| + \ln|\Lambda_1(x_0,t)| + O(e^{-2(\lambda_1-\lambda_2)t})\right\} \\ &= \lim_{t\to\infty}\frac{1}{t}\ln|\Lambda_1(x_0,t)|\,, \end{aligned} \tag{8.27}$$

where $\Lambda_1(x_0, t)$ is the leading eigenvalue of $\mathbf{J}^t(x_0)$. By choosing the initial displacement such that $\hat{n}$ is normal to the first (i-1) eigendirections we can define not only the leading, but all Lyapunov exponents as well:

$$\overline{\lambda_i(x_0)} = \lim_{t\to\infty}\frac{1}{t}\ln|\Lambda_i(x_0,t)|\,, \qquad i = 1, 2, \cdots, d\,. \tag{8.28}$$

The leading Lyapunov exponent now follows from the Jacobian matrix by numerical integration of (4.28). The equations can be integrated accurately for a finite time, hence the infinite time limit of (8.25) 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. 8.3 for the Rössler system (2.11). 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 (8.27) 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 $\ln|\Lambda_p|/T_p$, a local property of cycle p, not a global property of the dynamical system. Furthermore, even if x_0 happens to be a "generic" phase space point, it is still not obvious that $\ln|\Lambda(x_0,t)|/t$ should be converging to anything in particular. In a Hamiltonian system with coexisting elliptic islands and chaotic regions, a chaotic trajectory gets every so often captured in the neighborhood of an elliptic island and can stay there for arbitrarily long time; as there the orbit is nearly stable, during such episode $\ln|\Lambda(x_0,t)|/t$ can dip arbitrarily close to 0^+. For phase space volume non-preserving flows the trajectory can traverse locally contracting regions, and $\ln|\Lambda(x_0,t)|/t$ can occasionally go negative; even worse, one never knows whether the asymptotic attractor is periodic or "strange", so any finite estimate of $\overline{\lambda}$ might be dead wrong.

8.1 page 143

8.3.2 Evolution operator evaluation of Lyapunov exponents

A cure to these problems was offered in sect. 8.2. 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 (4.31) for the derivative of an iterated map, the stability of a 1-d mapping is multiplicative along the flow, so the integral (8.1) of the observable $a(x) = \ln|f'(x)|$, the local trajectory divergence rate, evaluated along the trajectory of x_0 is additive:

$$A^n(x_0) = \ln\left|f^{n\prime}(x_0)\right| = \sum_{k=0}^{n-1} \ln\left|f'(x_k)\right| \,. \tag{8.29}$$

The Lyapunov exponent is then the expectation value (8.9) given by a spatial integral (8.8) weighted by the natural measure

$$\lambda = \left\langle \ln|f'(x)| \right\rangle = \int_{\mathcal{M}} dx\, \rho_0(x) \ln|f'(x)| \,. \tag{8.30}$$

The associated (discrete time) evolution operator (8.19) is

$$\mathcal{L}(y,x) \;=\; \delta(y-f(x))\; e^{\beta \ln|f'(x)|} \,. \tag{8.31}$$

appendix H.1

Here we have restricted our considerations to $1-d$ maps, as for higher-dimensional flows only the Jacobian matrices are multiplicative, not the individual eigenvalues. Construction of the evolution operator for evaluation of the Lyapunov spectra 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.

in depth:
appendix H.1, p. 659

All that remains is to determine the value of the Lyapunov exponent

$$\lambda = \left\langle \ln |f'(x)| \right\rangle = \left.\frac{\partial s(\beta)}{\partial \beta}\right|_{\beta=1} = s'(1) \tag{8.32}$$

from (8.12), the derivative of the leading eigenvalue $s_0(\beta)$ of the evolution operator (8.31). The only question is: how?

sect. 15.2.4

8.4 Why not just run it on a computer?

(R. Artuso and P. Cvitanović)

All of the insight gained in this chapter and the preceding one was nothing but an elegant way of thinking of the evolution operator, $\mathcal{L}$, as a matrix (this point of view will be further elaborated in chapter 13). There are many textbook methods of approximating an operator $\mathcal{L}$ by sequences of finite matrix approximations $\mathcal{L}$, but in what follows the great achievement will be that we shall avoid constructing any matrix approximation to $\mathcal{L}$ altogether. Why a new method? Why not just run it on a computer, as many do with such relish in diagonalizing quantum Hamiltonians?

The simplest possible way of introducing a phase space discretization, fig. 8.4, is to partition the phase space $\mathcal{M}$ with a non-overlapping collection of sets $\mathcal{M}_i$, $i = 1, \ldots, N$, and to consider densities (7.2) that are locally constant on each $\mathcal{M}_i$:

$$\rho(x) = \sum_{i=1}^{N} \rho_i \frac{\chi_i(x)}{|\mathcal{M}_i|}$$

where $\chi_i(x)$ is the characteristic function (7.1) of the set $\mathcal{M}_i$. The density ρ_i at a given instant is related to the densities at the previous step in time

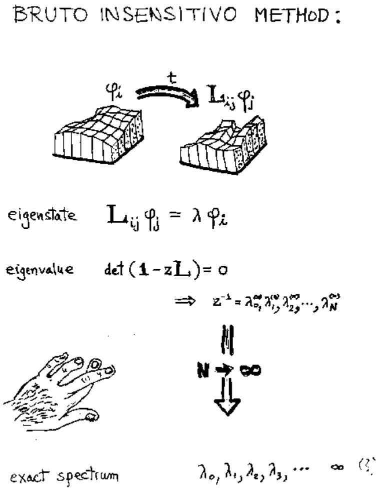


Figure 8.4: Phase space discretization approach to computing averages.

by the action of the Perron-Frobenius operator, as in (7.6):

$$\begin{aligned}
\rho'_j &= \int_{\mathcal{M}} dy\, \chi_j(y)\rho'(y) = \int_{\mathcal{M}} dx\, dy\, \chi_j(y)\, \delta(y - f(x))\, \rho(x) \\
&= \sum_{i=1}^{N} \rho_i \frac{|\mathcal{M}_i \cap f^{-1}(\mathcal{M}_j)|}{|\mathcal{M}_i|} .
\end{aligned}$$

In this way

$$\mathbf{L}_{ij} = \frac{|\mathcal{M}_i \cap f^{-1}(\mathcal{M}_j)|}{|\mathcal{M}_i|} , \quad \rho' = \rho\mathbf{L} \tag{8.33}$$

is a matrix approximation to the Perron-Frobenius operator, and its leading left eigenvector is a piecewise constant approximation to the invariant measure. It is an old idea of Ulam that such an approximation for the Perron-Frobenius operator is a meaningful one.

The problem with such phase space discretization approaches is that they are blind, the grid knows not what parts of the phase space are more or less important. This observation motivates the next step in developing the theory of long-time dynamics of chaotic systems: in chapter 9 we shall exploit the intrinsic topology of the flow to give us both an invariant partition of the phase space and a measure of the partition volumes, in the spirit of fig. 1.9.

Furthermore, a piecewise constant ρ belongs to an unphysical function space, and with such approximations one is plagued by numerical artifacts such as spurious eigenvalues. In chapter 13 we shall employ a more refined approach to extracting spectra, by expanding the initial and final densities

ρ, ρ' in some basis $\varphi_0, \varphi_1, \varphi_2, \cdots$ (orthogonal polynomials, let us say), and replacing $\mathcal{L}(y,x)$ by its φ_α basis representation $\mathbf{L}_{\alpha\beta} = \langle\varphi_\alpha|\mathcal{L}|\varphi_\beta\rangle$. The art is then the subtle art of finding a "good" basis for which finite truncations of $\mathbf{L}_{\alpha\beta}$ give accurate estimates of the eigenvalues of $\mathcal{L}$.

☞ chapter 13

Regardless of how sophisticated the choice of basis might be, the basic problem cannot be avoided - as illustrated by the natural measure for the Hénon map (3.10) sketched in fig. 7.3, eigenfunctions of $\mathcal{L}$ are complicated, singular functions concentrated on fractal sets, and in general cannot be represented by a nice basis set of smooth functions. We shall resort to matrix representations of $\mathcal{L}$ and the φ_α basis approach only insofar this helps us prove that the spectrum that we compute is indeed the correct one, and that finite periodic orbit truncations do converge.

in depth:
chapter 1, p. 1

Commentary

Remark 8.1 "Pressure". The quantity $\langle\exp(\beta\cdot A^t)\rangle$ is called a "partition function" by Ruelle [12.1]. Mathematicians decorate it with considerably more Greek and Gothic letters than is the case in this treatise. Ruelle [12.2] and Bowen [8.1] had given name "pressure" $P(a)$ to $s(\beta)$ (where a is the observable introduced here in sect. 8.1.1), defined by the "large system" limit (8.11). As we shall apply the theory also to computation of the physical gas pressure exerted on the walls of a container by a bouncing particle, we prefer to $s(\beta)$ as simply the leading eigenvalue of the evolution operator introduced in sect. 7.4. The "convexity" properties such as $P(a) \leq P(|a|)$ will be pretty obvious consequence of the definition (8.11). In the case that $\mathcal{L}$ is the Perron-Frobenius operator (7.10), the eigenvalues $\{s_0(\beta), s_1(\beta), \cdots\}$ are called the *Ruelle-Pollicott resonances* [8.2, 8.3, 8.4], 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 8.2 Microcanonical ensemble. In statistical mechanics the space average (8.7) performed over the Hamiltonian system constant energy surface invariant measure $\rho(x)dx = dqdp\,\delta(H(q,p)-E)$

of volume $|\mathcal{M}| = \int_{\mathcal{M}} dqdp\, \delta(H(q,p) - E)$

$$\langle a(t) \rangle = \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} dqdp\, \delta(H(q,p) - E) a(q,p,t) \tag{8.34}$$

is called the *microcanonical ensemble average.*

Remark 8.3 Lyapunov exponents. The Multiplicative Ergodic Theorem of Oseledec [8.5] states that the limit (8.28) 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 [8.6] $\lambda_i(x_0)$.

There is much literature on numerical computation of the Lyapunov exponents, see for example refs. [4.1, 8.12, 8.13, 8.14].

Résumé

The expectation value $\langle a \rangle$ of an observable $a(x)$ measured $A^t(x) = \int_0^t d\tau a(x(\tau))$ and averaged along the flow $x \to f^t(x)$ is given by the derivative $\partial s / \partial \beta$ of the leading eigenvalue $e^{ts(\beta)}$ of the corresponding evolution operator $\mathcal{L}^t$.

Using the Perron-Frobenius operator (7.10) whose leading eigenfunction, the natural measure, once computed, yields expectation value (7.20) of any observable $a(x)$ a separate evolution operator $\mathbf{L}$ has to be constructed for each and every observable. However, by the time the scaffolding is removed both $\mathcal{L}$'s and their eigenfunctions will be gone, and only the formulas for expectation value of observables will remain.

The next question is: how do we evaluate the eigenvalues of $\mathcal{L}$? We saw in sect. 7.1, 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. In chapter 9 we take the first step, and replace the *ad hoc* partitioning (8.33) by the intrinsic, topologically invariant partitioning. In chapter 10 we apply this information to our first application of the evolution operator formalism, evaluation of the topological entropy, the growth rate of the number of topologically distinct orbits. This small victory will then be refashioned in chapters 11 and 12 into a systematic method for computing eigenvalues of evolution operators in terms of periodic orbits.

☞ chapter 15

References

[8.1] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer Lect. Notes on Math. **470** (1975).

[8.2] D. Ruelle, "The thermodynamical formalism for expanding maps" *J. Diff. Geo.* **25**, 117 (1987).

[8.3] M. Pollicott, "On the rate of mixing of Axiom A flows", *Invent. Math.* **81**, 413 (1985).

[8.4] D. Ruelle, *J. Diff. Geo.* **25**, 99 (1987).

[8.5] V.I. Oseledec, Trans. Moscow Math.Soc. 19, 197 (1968).

[8.6] A.M. Lyapunov, *General problem of stability of motion*, *Ann. of Math. Studies* **17** (Princeton Univ. Press).

[8.7] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)]

[8.8] Ya.B. Pesin, *Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties*, *Ergodic Theory and Dynamical Systems,* **12**, 123 (1992).

[8.9] Ya.B. Pesin, *Func. Anal. Applic.* **8**, 263 (1974).

[8.10] A. Katok, *Liapunov exponents, entropy and periodic orbits for diffeomorphisms*, *Publ. Math. IHES* **51**, 137 (1980).

[8.11] D. Bessis, G. Paladin, G. Turchetti and S. Vaienti, *Generalized Dimensions, Entropies and Lyapunov Exponents from the Pressure Function for Strange Sets*, *J. Stat. Phys.* **51**, 109 (1988).

[8.12] A. Wolf, J.B. Swift, et al., "Determining Lyapunov Exponents from a Time Series", *Physica D* **16**, 285 (1985).

[8.13] J.-P. Eckmann, S.O. Kamphorst, et al., "Liapunov exponents from time series", *Phys. Rev. A* **34**, 4971 (1986).

[8.14] G. Bennettin, L. Galgani and J.-M. Strelcyn, *Meccanica* **15**, 9 (1980).

Exercises

Exercise 8.1 **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?*

Exercise 8.2 **Expectation value of a vector observable and its moments.** *Check and extend the expectation value formulas (8.12) 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)

$$\left.\frac{\partial s}{\partial \beta_i}\right|_{\beta=0} = \lim_{t\to\infty} \frac{1}{t} \left\langle A_i^t \right\rangle = \langle a_i \rangle \,, \tag{8.35}$$

(b)

$$\begin{aligned} \left.\frac{\partial^2 s}{\partial \beta_i \partial \beta_j}\right|_{\beta=0} &= \lim_{t\to\infty} \frac{1}{t} \left(\left\langle A_i^t A_j^t \right\rangle - \left\langle A_i^t \right\rangle \left\langle A_j^t \right\rangle \right) \\ &= \lim_{t\to\infty} \frac{1}{t} \left\langle (A_i^t - t \langle a_i \rangle)(A_j^t - t \langle a_j \rangle) \right\rangle \,. \end{aligned} \tag{8.36}$$

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 (8.35) vanishes,* $\langle a_i \rangle = 0$. *The 4-th order moment formula*

$$K(t) = \frac{\left\langle x^4(t) \right\rangle}{\left\langle x^2(t) \right\rangle^2} - 3 \tag{8.37}$$

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 (20.22) for a concrete example). If the observable is a vector, the kurtosis is given by*

$$K(t) = \frac{\sum_{ij} [\langle A_i A_i A_j A_j \rangle + 2 (\langle A_i A_j \rangle \langle A_j A_i \rangle - \langle A_i A_i \rangle \langle A_j A_j \rangle)]}{\left(\sum_i \langle A_i A_i \rangle\right)^2} \quad (8.38)$$

Chapter 9

Qualitative dynamics, for pedestrians

In this chapter we begin to learn how to use qualitative properties of a flow in order to *partition* the phase space in a topologically invariant way, and *name* topologically distinct orbits. This will enable us (in the next chapter) to *count* the distinct orbits, and in the process touch upon all the main themes of this book, going the whole distance from diagnosing chaotic dynamics to computing zeta functions.

We start by a simple physical example, symbolic dynamics of a 3-disk game of pinball, and then show that also for smooth flows the qualitative dynamics of stretching and folding flows enables us to partition the phase space and assign symbolic dynamics itineraries to trajectories. Here we illustrate the method on a $1-d$ approximation to Rössler flow. We then turn this topological dynamics into a multiplicative operation on the phase space partition 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.

9.1 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

$$\left\{y \in \mathcal{M}_0 : \quad y = f^t(x_0), \quad t > t_0\right\} \tag{9.1}$$

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*. As the subject can get quite technical, a summary of the basic notions and definitions of symbolic dynamics is relegated to sect. 9.7; check that section 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. 9.1, 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 the itinerary for the trajectory sketched in the figure is $0 \mapsto 2 \mapsto 1 \mapsto 0 \mapsto 1 \mapsto 2 \mapsto \cdots$.

If there is no way to reach partition $\mathcal{M}_i$ from partition $\mathcal{M}_j$, and conversely, partition $\mathcal{M}_j$ from partition $\mathcal{M}_i$, the phase space consists of at least two disconnected pieces, and we can analyze it piece by piece. 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 said to be *metrically indecomposable*.

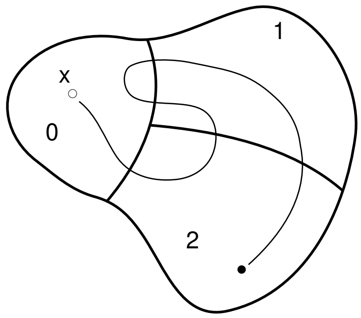


Figure 9.1: A trajectory with itinerary 021012.

In general one also encounters transient regions - regions to which the dynamics does not return to once they are exited. Hence we have to distinguish between (for us uninteresting) wandering trajectories that never return to the initial neighborhood, and the non–wandering set (2.2) of the *recurrent* trajectories.

The allowed transitions between the regions of a partition are encoded in the $[N \times N]$-dimensional *transition matrix* whose elements take values

$$T_{ij} \quad = \quad \begin{cases} 1 & \text{if a transition region } \mathcal{M}_j \to \text{region } \mathcal{M}_i \text{ is possible} \\ 0 & \text{otherwise}\,. \end{cases} \tag{9.2}$$

An example is the complete N-ary dynamics for which all transition matrix entries equal unity (one can reach any region from any other region in one step)

$$T_c = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} . \tag{9.3}$$

Further examples of transition matrices, such as the 3-disk transition matrix (9.5) and the 1-step memory sparse matrix (9.14), 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.

Knowing that a point from $\mathcal{M}_i$ reaches $\mathcal{M}_j$ in one step is not quite good enough. We would be happier if we knew that *any* point in $\mathcal{M}_i$ reaches $\mathcal{M}_j$; otherwise we have to subpartition $\mathcal{M}_i$ into the points which land in $\mathcal{M}_j$, and those which do not, and often we will find ourselves partitioning *ad infinitum*.

Such considerations motivate the notion of a *Markov partition*, a partition for which no memory of preceding steps is required to fix the transitions

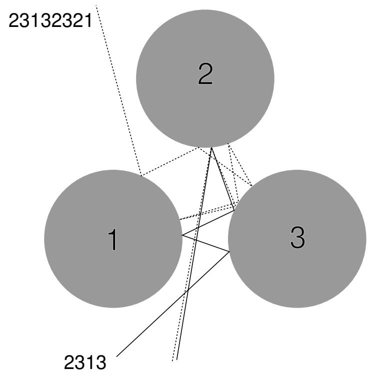


Figure 9.2: Two pinballs that start out very close to each other exhibit the same qualitative dynamics _2313_ for the first three bounces, but due to the exponentially growing separation of trajectories with time, follow different itineraries thereafter: one escapes after _2313_, the other one escapes after _23132321_.

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,

$$\text{either} \quad \mathcal{M}_j \cap f(\mathcal{M}_i) = \emptyset \quad \text{or} \quad \mathcal{M}_j \subset f(\mathcal{M}_i)\,. \tag{9.4}$$

An example is the 1-dimensional expanding mapping sketched in fig. 9.8, and more examples are worked out in sect. 20.2.

Example 9.1 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. 9.2 have itineraries _2313_, _23132321_ respectively. The 3-disk prime cycles given in figs. 1.5 and 9.5 are further examples of such itineraries.

1.1
page 32

At each bounce a cone of initially nearby trajectories defocuses (see fig. 1.7), and in order to attain a desired longer and longer itinerary of bounces the initial point $x_0 = (s_0, p_0)$ has to be specified with a larger and larger precision, and lie within initial phase space strips drawn in fig. 9.3. Similarly, it is intuitively clear that as we go backward in time (in this case, simply reverse the velocity vector), we also need increasingly precise specification of $x_0 = (s_0, p_0)$ in order to follow a given past itinerary. Another way to look at the survivors after two bounces is to plot $\mathcal{M}_{s_1.s_2}$, the intersection of $\mathcal{M}_{.s_2}$ with the strips $\mathcal{M}_{s_1.}$ obtained by time reversal (the velocity changes sign $\sin\phi \to -\sin\phi$). $\mathcal{M}_{s_1.s_2}$ 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. More generally, for hyperbolic flows the intersection of the future and past itineraries, the bi-infinite itinerary $S^{-}.S^{+} = \cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots$ specifies a unique trajectory. This is intuitively clear for our 3-disk game of pinball,

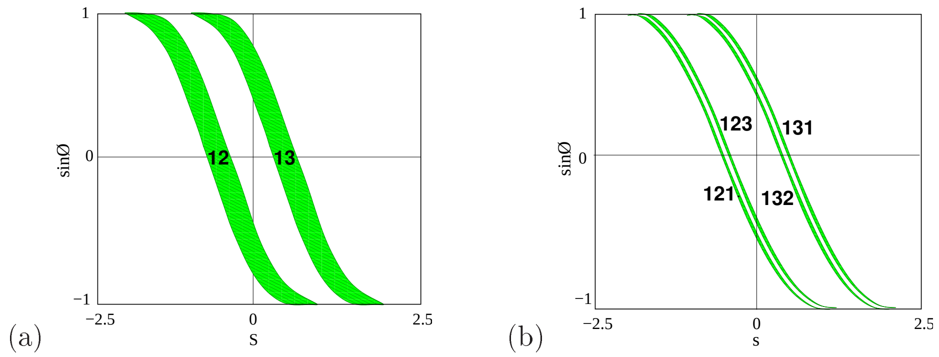


Figure 9.3: The 3-disk game of pinball Poincaré section, trajectories emanating from the disk 1 with $x_0 =$ (arclength, parallel momentum) $= (s_0, p_0)$, disk radius : center separation ratio a:R = 1:2.5. (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in one bounce, respectively. (b) Strips of initial points $\mathcal{M}_{121}$, $\mathcal{M}_{131}$ $\mathcal{M}_{132}$ and $\mathcal{M}_{123}$ which reach disks 1, 2, 3 in two bounces, respectively. (Y. Lan)

and is stated more formally in the definition (9.4) of a Markov partition. The definition requires that the dynamics be expanding forward in time in order to ensure that the cone of trajectories with a given itinerary becomes sharper and sharper as the number of specified symbols is increased.

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, with the transition matrix (9.2) given by*

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} . \tag{9.5}$$

For convex disks the separation between nearby trajectories increases at every reflection, implying that the stability matrix has an expanding eigenvalue. By the Liouville phase-space volume conservation (5.10), the other transverse eigenvalue is contracting. This example demonstrates that finite Markov partitions can be constructed for hyperbolic *dynamical systems which are expanding in some directions, contracting in others.*

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. 9.2, p. 153

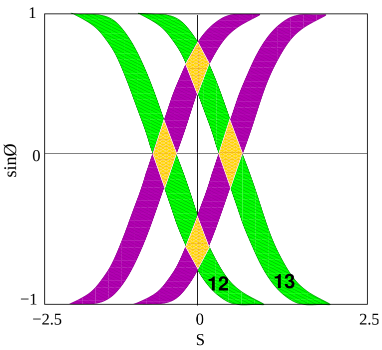


Figure 9.4: The Poincaré section of the phase space for the binary labelled pinball, see also fig. 9.5(b). For definitiveness, this set is generated by starting from disk 1, preceded by disk 2. Indicated are the fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle periodic points $\overline{01}$, $\overline{10}$, together with strips which survive 1, 2, ... bounces. Iteration corresponds to the decimal point shift; for example, all points in the rectangle [01.01] map into the rectangle [010.1] in one iteration.
PC: do this figure right, in terms of strips!

9.1.1 A brief detour; recoding, 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 19, 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. 1.5. 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. 1.4:

9.1
page 169

$$
\begin{array}{lll}
0 & : & \text{pinball returns to the disk it came from} \\
1 & : & \text{pinball continues to the third disk}\,.
\end{array}
\tag{9.6}
$$

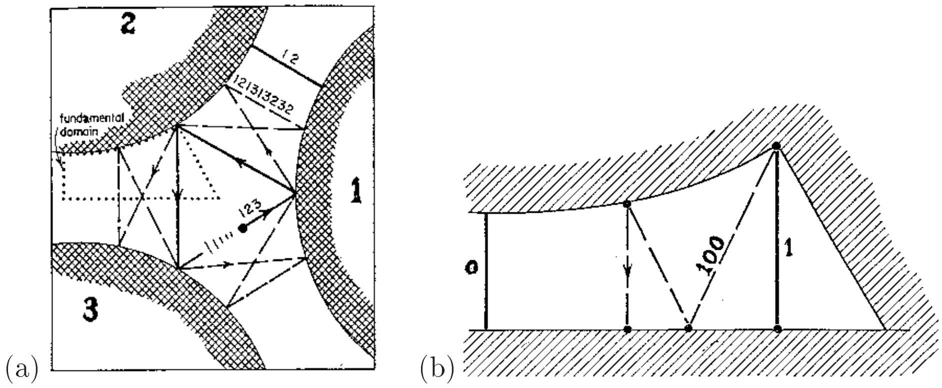


Figure 9.5: The 3-disk game of pinball with the disk radius : center separation ratio a:R = 1:2.5. (a) The three disks, with $\overline{12}$, $\overline{123}$ and $\overline{121232313}$ cycles indicated. (b) The fundamental domain, that is the small 1/6th wedge indicated in (a), consisting of a section 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.

9.2
page 169

This *binary* symbolic dynamics has two immediate advantages over the ternary one; the prohibition of self-bounces is automatic, and the coding utilizes the symmetry of the 3-disk pinball game in elegant manner. If the disks are sufficiently far apart there are no further restrictions on symbols, the symbolic dynamics is complete, and *all* binary sequences are admissible itineraries. As this type of symbolic dynamics pops up frequently, we list the shortest binary prime cycles in table 9.2.

9.3
page 169

Given a ternary sequence and labels of 2 preceding disks these rules fix the corresponding binary symbol. For example

$$\begin{array}{lcccccccccccccccc} \text{ternary:} & 3 & 1 & 2 & 1 & 3 & 1 & 2 & 3 & 2 & 1 & 2 & 3 & 1 & 3 & 2 & 3 \\ \text{binary:} & & . & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \end{array} \tag{9.7}$$

The first 2 disks initialize the trajectory and its direction. Due to the 3-disk symmetry the six 3-disk sequences initialized by 12, 13, 21, 23, 31, 32 respectively have the same weights, the same size partitions, and are coded by a single binary sequence. For periodic orbits, the equivalent ternary cycles reduce to binary cycles of 1/3, 1/2 or the same length. How this works is best understood by inspection of table 9.1, fig. 9.5 and fig. 19.3.

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. 9.5b. 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

$\tilde{p}$	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 9.1: 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. 19.2.2. 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 $\overline{001101}$.

trajectory, a repeating binary symbols block corresponds either to the full periodic orbit or to an irreducible segment (examples are shown in fig. 9.5 and table 9.1). An irreducible segment corresponds to a periodic orbit in the fundamental domain. Table 9.1 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 19, life is much simpler in the fundamental domain than in the full system, so whenever possible our computations will be carried out in the fundamental domain.

9.5
page 170

Symbolic dynamics for N-disk game of pinball is so straightforward that one may altogether fail to see the connection between the topology of hyperbolic flows and their symbolic dynamics. This is brought out more clearly by the 1-d visualization of "stretch & fold" flows to which we turn now. Inspecting the fig. 9.3b we see that the relative ordering of regions with differing finite itineraries is a qualitative, topological property of the flow, so it makes sense to define a simple "canonical" representative partition which in a simple manner exhibits spatial ordering common to an entire class of topologically similar nonlinear flows.

in depth:
chapter 19, p. 365

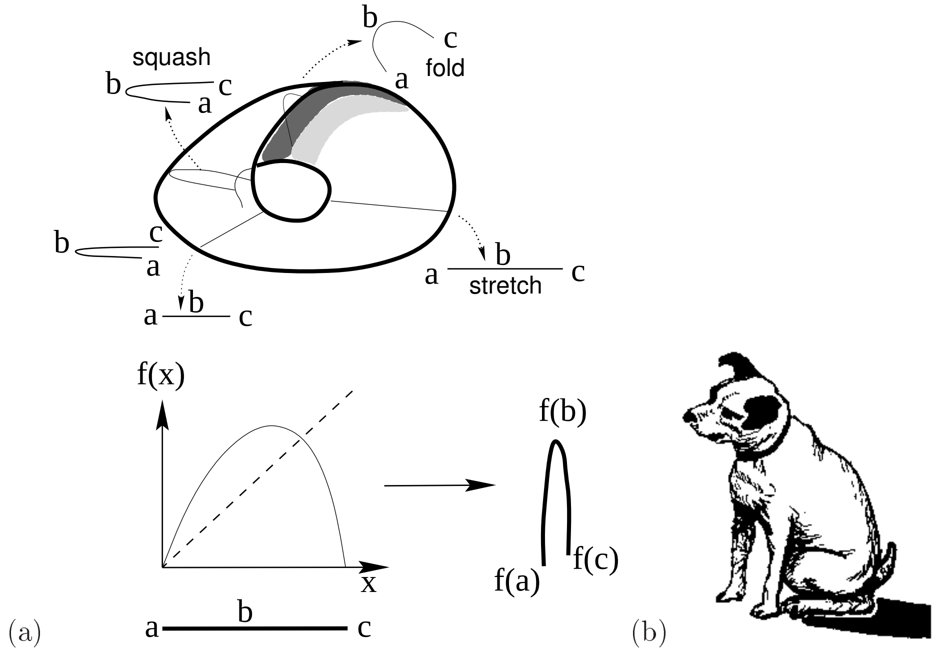


Figure 9.6: (a) A recurrent flow that stretches and folds. (b) The "stretch & fold" return map on the Poincaré section.

9.2 Stretch and fold

Suppose concentrations of certain chemical reactants worry you, or the variations in the Chicago temperature, humidity, pressure and winds affect your mood. All such properties vary within some fixed range, and so do their rates of change. Even if we are studying an operator system such as the 3-disk pinball game, we tend to be interested in a finite region around the disks and ignore the escapees. So a typical dynamical system that we care about is *bounded.* If the price for keeping going is high - for example, we try to stir up some tar, and observe it come to a dead stop the moment we cease our labors - the dynamics tends to settle into a simple limiting state. However, as the resistance to change decreases - the tar is heated up and we are more vigorous in our stirring - the dynamics becomes unstable. 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.

9.3 Going global: Stable/unstable manifolds

In sects. 7.1 and 9.1 we introduced the concept of partitioning the phase space, in any way you please. Here we show that the dynamics itself generates a natural partition.

In sect. 4.5 we established that stability eigenvalues of periodic orbits are invariants of a given flow. The invariance of stabilities of a periodic orbit is a local property of the flow. Now we show that every periodic orbit carries with it stable and unstable manifolds which provide a topologically invariant *global* foliation of the phase space.

A neighborhood of a trajectory deforms as it is transported by the flow. In the linear approximation, the matrix of variations 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.

The fixed or periodic point x^* stability matrix $\mathbf{J}_p(x^*)$ eigenvectors (4.42) 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

$$\begin{aligned} W^s &= \left\{x \in \mathcal{M} : f^t(x) - x^* \to 0 \text{ as } t \to \infty\right\} \\ W^u &= \left\{x \in \mathcal{M} : f^{-t}(x) - x^* \to 0 \text{ as } t \to \infty\right\} . \end{aligned} \tag{9.8}$$

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

$$\begin{aligned} W^s &= \{x \in \mathcal{P} : f^n(x) - x^* \to 0 \text{ as } n \to \infty\} \\ W^u &= \left\{x \in \mathcal{P} : f^{-n}(x) - x^* \to 0 \text{ as } n \to \infty\right\} . \end{aligned} \tag{9.9}$$

For $n \to \infty$ any finite segment of W^s, respectively W^u converges to the linearized map eigenvector $\mathbf{e}^{(e)}$, respectively $\mathbf{e}^{(c)}$. In this sense each eigenvector defines a (curvilinear) axis of the stable, respectively unstable manifold. Conversely, we can use an arbitrarily small segment of a fixed point

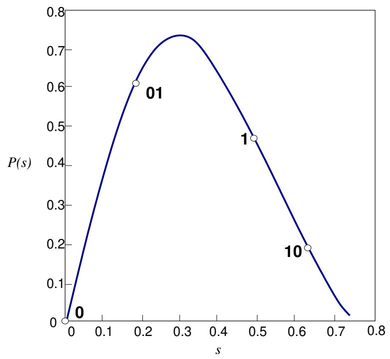


Figure 9.7: The Poincaré return map of the Kuramoto-Sivashinsky system (2.17) fig. 3.3, from the unstable manifold of the $\overline{1}$ fixed point to the (neighborhood of) the unstable manifold. Also indicated are the periodic points $\overline{0}$ and $\overline{01}$.

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?

9.4 Temporal ordering: itineraries

In this section we learn how to *name* and *count* periodic orbits for the simplest, and nevertheless very instructive case, for 1-d maps of an interval.

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. 9.5 eliminates *all* itineraries that cannot occur for a given unimodal map.

Suppose that the compression of the folded interval in fig. 9.6 is so fierce that we can neglect the thickness of the attractor. For example, the Rössler flow (2.11) 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.

Example 9.2 Tent map. *The simplest examples of unimodal maps are the* complete tent map *fig. 9.8(a),*

$$f(\gamma) = 1 - 2|\gamma - 1/2|\,, \tag{9.10}$$

and the quadratic map (sometimes also called the logistic map)

$$x_{t+1} = 1 - a x_t^2\,, \tag{9.11}$$

with the one critical point at $x_c = 0$. Another example is the repelling unimodal map of fig. 9.8b. We refer to (9.10) 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, they may nevertheless serve as effective descriptions of hyperbolic flows.

For the unimodal maps of fig. 9.8 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 (9.3). 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 (9.16) for a given trajectory,

$$s_n = \left\{ \begin{array}{ll} 1 & \text{if } x_n > x_c \\ 0 & \text{if } x_n < x_c \end{array} \right. . \tag{9.12}$$

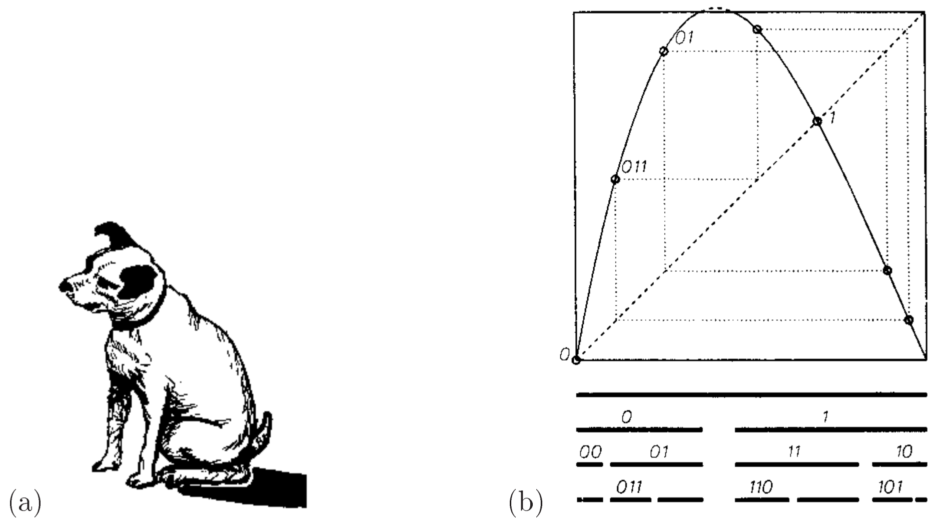


Figure 9.8: (a) The complete tent map together with intervals that follow the indicated itinerary for n steps. (b) A unimodal repeller with the remaining intervals after 1, 2 and 3 iterations. Intervals marked $s_1s_2\cdots s_n$ are unions of all points that do not escape in n iterations, and follow the itinerary $S^+ = s_1s_2\cdots s_n$. 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 $\overline{011}$.

We shall refer to $S^+(x_0) = .s_1s_2s_3\cdots$ 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?

As binary type symbolic dynamics pops up frequently, we list the shortest binary prime cycles in table 9.2.

9.3
page 169

9.5 Spatial ordering

The tent map (9.10) consists of two straight segments joined at $x = 1/2$. The symbol s_n defined in (9.12) 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

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
	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
	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
	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
	001001111
	001010111
	001011011
	001011101
	001100111
	001101011
	001101101
	001110101
	010101011
	000111111
	001011111
	001101111
	001110111
	001111011
	001111101
	010101111
	010110111
	010111011
	001111111
	010111111
	011011111
	011101111
	011111111

Table 9.2: Prime cycles for the binary symbolic dynamics up to length 9.

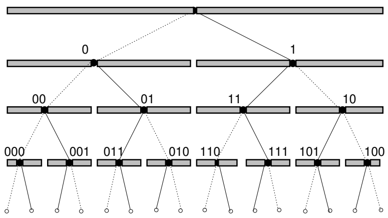


Figure 9.9: Alternating binary tree relates the itinerary labelling of the unimodal map fig. 9.8 intervals to their spatial ordering. Dotted line stands for 0, full line for 1; the binary sub-tree whose root is a full line (symbol 1) reverses the orientation, due to the orientation reversing fold in figs. 9.8 and 9.6.

the sequence of s_n's into a binary number by the following algorithm:

$$
\begin{aligned}
w_{n+1} &= \begin{cases} w_n & \text{if } s_n = 0 \\ 1 - w_n & \text{if } s_n = 1 \end{cases} , \qquad w_1 = s_1 \\
\gamma(S^+) &= 0.w_1 w_2 w_3 \ldots = \sum_{n=1}^{\infty} w_n / 2^n .
\end{aligned}
\tag{9.13}
$$

This follows by inspection from the binary tree of fig. 9.9. 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.. 9.6 page 170

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 (9.10). 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 (9.10) 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 $\overline{\gamma} > \gamma$.

9.6 Topological dynamics

9.6.1 Finite memory

In the complete N-ary symbolic dynamics case (see example (9.3)) the choice of the next symbol requires no memory of the previous ones. However, any further refinement of the partition requires finite memory.

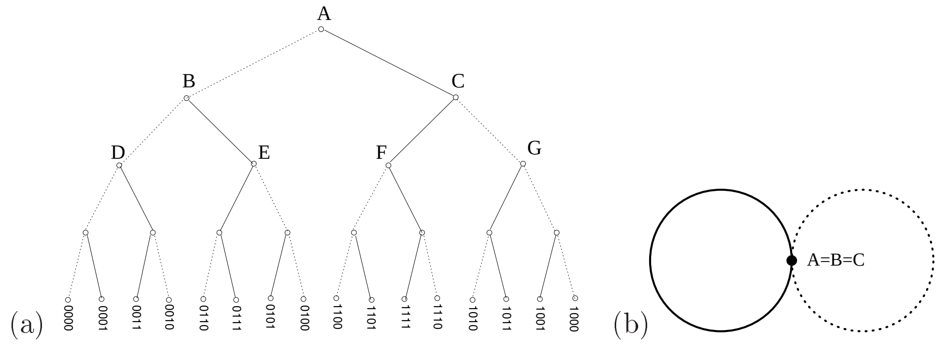


Figure 9.10: (a) The self-similarity of the complete binary symbolic dynamics represented by a binary tree (b) identification of nodes $B = A$, $C = A$ leads to the finite 1-node, 2-links Markov graph. All admissible itineraries are generated as walks on this finite Markov graph.

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 fig. 9.4, as well as fig. 1.8. Topologically f acts as a left shift, 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

9.8
page 170

$$\phi' = T\phi \;=\; \begin{pmatrix} T_{00,00} & 0 & T_{00,10} & 0 \\ T_{01,00} & 0 & T_{01,10} & 0 \\ 0 & T_{10,01} & 0 & T_{10,11} \\ 0 & T_{11,01} & 0 & T_{11,11} \end{pmatrix} \begin{pmatrix} \phi_{00} \\ \phi_{01} \\ \phi_{10} \\ \phi_{11} \end{pmatrix} . \qquad (9.14)$$

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.

10.1
page 194

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.

10.4
page 195
10.1
page 194

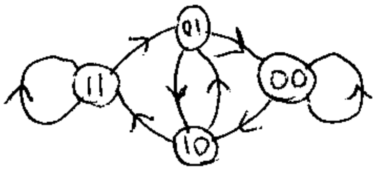
(b)  (a) 

Figure 9.11: (a) The 2-step memory Markov graph, links version obtained by identifying nodes $A = D = E = F = G$ in fig. 9.10(a). Links of this graph correspond to the matrix entries in the transition matrix (9.14). (b) the 2-step memory Markov graph, node version.

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.

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. 9.10(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. 9.10(a) correspond to the 16 distinct binary strings of length 4, and so on. By habit we have drawn the tree as the alternating binary tree of fig. 9.9, but that has no significance as far as enumeration of itineraries is concerned - an ordinary binary tree would serve just as well.

The trouble with an infinite tree is that it does not fit on a piece of paper. On the other hand, we are not doing much - at each node we are turning either left or right. Hence all nodes are equivalent, and can be identified. To say it in other words, the tree is self-similar; the trees originating in nodes B and C are themselves copies of the entire tree. The result of identifying $B = A$, $C = A$ is a single node, 2-link Markov graph of fig. 9.10(b): any itinerary generated by the binary tree fig. 9.10(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

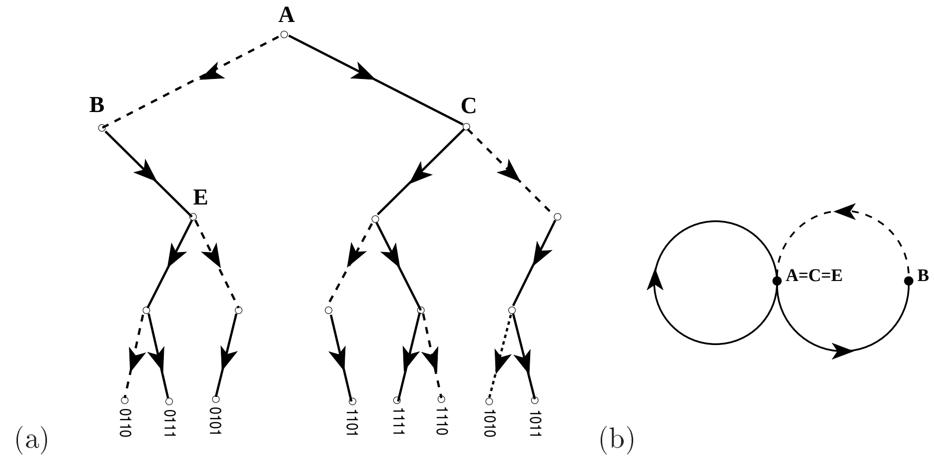


Figure 9.12: (a) The self-similarity of the _00_ pruned binary tree: trees originating from nodes C and E are the same as the entire tree. (b) Identification of nodes $A = C = E$ leads to the finite 2-node, 3-links Markov graph; as 0 is always followed by 1, the walks on this graph generate only the admissible itineraries.

trees originating in D, E, F and G with the entire tree leads to the 2-step memory Markov graph of fig. 9.11a. The corresponding transition matrix is given by (9.14).

fast track:
chapter 10, p. 171

9.7 Symbolic dynamics, basic notions

In this section we collect the basic notions and definitions of symbolic dynamics. The reader might prefer to skim through this material on first reading, return to it later as the need arises.

Shifts.

We associate with every initial point $x_0 \in \mathcal{M}$ the *future itinerary*, a sequence of symbols $S^+(x_0) = s_1 s_2 s_3 \cdots$ which indicates the order in which the regions are visited. If the trajectory $x_1, x_2, x_3, \ldots$ of the initial point x_0 is generated by

$$x_{n+1} = f(x_n)\,, \tag{9.15}$$

then the itinerary is given by the symbol sequence

$$s_n = s \qquad \text{if} \qquad x_n \in \mathcal{M}_s \,. \tag{9.16}$$

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

$$S(x_0) = (s_k)_{k\in\mathbb{Z}} = S^-.S^+ = \cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots \,. \tag{9.17}$$

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*

$$\mathcal{A}^{\mathbb{Z}} = \{(s_k)_{k\in\mathbb{Z}} : s_k \in \mathcal{A} \text{ for all k } \in \mathbb{Z}\} \,. \tag{9.18}$$

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 (9.16), 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,

$$\sigma(\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots) = \cdots s_{-2}s_{-1}s_0s_1.s_2s_3\cdots \,, \tag{9.19}$$

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 $b = s_ks_{k+1}\cdots s_{k+n_b-1}$ 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*

$$p = \overline{s_1s_2\cdots s_{n_p}} = \{x_{s_1s_2\cdots s_{n_p}}, x_{s_2\cdots s_{n_p}s_1}, \cdots, x_{s_{n_p}s_1\cdots s_{n_p-1}}\} \,. \tag{9.20}$$

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

$$x_{s_2 \cdots s_{n_p} s_1} = x_{\overline{s_2 \cdots s_{n_p} s_1} \cdot \overline{s_2 \cdots s_{n_p} s_1}} .$$

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 *prime*; we shall refer to it as "prime" throughout this text).

Partitions.

A partition is called *generating* if every infinite symbol sequence corresponds to a distinct point in the phase space. Finite Markov partition (9.4) 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*

$$\Sigma = \{(s_k)_{k \in \mathbb{Z}}\} , \tag{9.21}$$

is the set of all *admissible* infinite itineraries, and $\sigma : \Sigma \to \Sigma$ is the shift operator (9.19). The designation "subshift" comes form the fact that $\Sigma \subset \mathcal{A}^{\mathbb{Z}}$ is the subset of the full shift (9.18). 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 $\cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots$ to each distinct trajectory, but there might exist full shift symbol sequences (9.18) which are not realized as trajectories; such sequences are called *inadmissible*, and we say that the symbolic dynamics is *pruned.* The word is suggested by "pruning" of branches corresponding to forbidden sequences for symbolic dynamics organized hierarchically into a tree structure, as explained in sect. 9.6.

Pruning.

If the dynamics is pruned, the alphabet must be supplemented by a *grammar*, a set of pruning rules. After the inadmissible sequences have been pruned, it is often convenient to parse the symbolic strings into words of variable length - this is called *coding*. Suppose that the grammar can be stated as a finite number of pruning rules, each forbidding a block of finite length,

$$\mathcal{G} = \{b_1, b_2, \cdots b_k\} \ , \tag{9.22}$$

where a *pruning block* b is a sequence of symbols $b = s_1 s_2 \cdots s_{n_b}$, $s \in \mathcal{A}$, of finite length n_b. In this case we can always construct a finite Markov partition (9.4) by replacing finite length words of the original partition by letters of a new alphabet. In particular, if the longest forbidden block is of length $M+1$, we say that the symbolic dynamics is a shift of finite type with M-step memory. In that case we can *recode* the symbolic dynamics in terms of a new alphabet, with each new letter given by an admissible block of at most length M. In the new alphabet the grammar rules are implemented by setting $T_{ij} = 0$ in (9.3) for forbidden transitions.

A topological dynamical system (Σ, σ) for which all admissible itineraries are generated by a finite transition matrix

$$\Sigma = \left\{(s_k)_{k\in\mathbb{Z}} \, : \, T_{s_k s_{k+1}} = 1 \quad \text{for all } k\right\} \tag{9.23}$$

is called a subshift of *finite type*. Such systems are particularly easy to handle; the topology can be converted into symbolic dynamics by representing the transition matrix by a finite directed *Markov graph*, a convenient visualization of topological dynamics.

Markov graphs.

A Markov graph describes compactly the ways in which the 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 (9.2) takes value $T_{ij} = 1$.

$$\text{(a)} \quad T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \text{(b)}$$

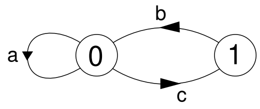


Figure 9.13: (a) The transition matrix for a simple subshift on two-state partition $\mathcal{A} = \{0, 1\}$, with grammar $\mathcal{G}$ given by a single pruning block $b = 11$ (consecutive repeat of symbol 1 is inadmissible): the state $\mathcal{M}_0$ maps both onto $\mathcal{M}_0$ and $\mathcal{M}_1$, 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.

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.

The simplest example is given in fig. 9.13.

Résumé

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.

In chapters 11 and 12 we will establish that spectra of evolution operators can be extracted from periodic orbit sums:

$$\sum (\text{spectral eigenvalues}) = \sum (\text{periodic orbits}) \ .$$

In order to implement this theory we need to know what periodic orbits can exist, and the symbolic dynamics developed above is an invaluable tool toward this end.

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

[9.1] B.P. Kitchens, Symbolic dynamics: one-sided, two-sided, and countable state Markov shifts (Springer, Berlin 1998).

[9.2] D.A. Lind and B. Marcus, *An introduction to symbolic dynamics and coding* (Cambridge Univ. Press, Cambridge 1995).

[9.3] 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).

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

[9.5] R.L. Devaney, *A First Course in Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1992).

[9.6] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1987).

[9.7] J. Guckenheimer and P. Holmes, *Non-linear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, New York, 1986).

[9.8] A. Salomaa, *Formal Languages* (Academic Press, San Diego, 1973).

[9.9] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading MA, 1979).

[9.10] D.M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs* (Academic Press, New York, 1980).

[9.11] T. Bedford, M.S. Keane and C. Series, eds., *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces* (Oxford University Press, Oxford, 1991).

[9.12] M.S. Keane, *Ergodic theory and subshifts of finite type*, in ref. [9.11].

[9.13] P.J. Myrberg, *Ann. Acad. Sc. Fenn., Ser. A,* **256**, 1 (1958); **259**, 1 (1958).

[9.14] 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).

[9.15] A.N. Sarkovskii, *Ukrainian Math. J.* **16**, 61 (1964).

[9.16] 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).

[9.17] W. Thurston, *"On the geometry and dynamics of diffeomorphisms of surfaces", Bull. Amer. Math. Soc. (N.S.)* **19**, 417 (1988).

[9.18] P. Collet and J.P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhauser, Boston, 1980).

[9.19] 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).

[9.20] V. Baladi, *"Infinite kneading matrices and weighted zeta functions of interval maps"*, *J. Functional Analysis* **128**, 226 (1995).

[9.21] D. Ruelle, *"Sharp determinants for smooth interval maps"*, Proceedings of Montevideo Conference 1995, IHES preprint (March 1995).

[9.22] V. Baladi and D. Ruelle, *"Sharp determinants"*, *Invent. Math.* **123**, 553 (1996).

[9.23] P. Cvitanović, G.H. Gunaratne and I. Procaccia, *Phys. Rev.* **A 38**, 1503 (1988).

[9.24] G. D'Alessandro, P. Grassberger, S. Isola and A. Politi, "On the topology of the Hénon Map", *J. Phys.* **A 23**, 5285 (1990).

[9.25] G. D'Alessandro, S. Isola and A. Politi, "Geometric properties of the pruning front", *Prog. Theor. Phys.* **86**, 1149 (1991).

[9.26] Y. Ishii, "Towards the kneading theory for Lozi attractors. I. Critical sets and pruning fronts", Kyoto Univ. Math. Dept. preprint (Feb. 1994).

[9.27] 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.

[9.28] A. de Carvalho, Ph.D. thesis, CUNY New York 1995; "Pruning fronts and the formation of horseshoes", preprint (1997).

[9.29] K.T. Hansen, *CHAOS* **2**, 71 (1992).

[9.30] K.T. Hansen, *Nonlinearity* **5**

[9.31] K.T. Hansen, *Nonlinearity* **5**

[9.32] K.T. Hansen, *Symbolic dynamics III, The stadium billiard*, to be submitted to *Nonlinearity*

[9.33] K.T. Hansen, *Symbolic dynamics IV; a unique partition of maps of Hénon type*, in preparation.

[9.34] E. Aurell, *"Convergence of dynamical zeta functions"*, *J. Stat. Phys.* **58**, 967 (1990).

[9.35] M. Benedicks and L. Carleson, *Ann. of Math.*, **122**, 1 (1985).

[9.36] P. Dahlqvist, *On the effect of pruning on the singularity structure of zeta functions*, J. Math. Phys. **38**, 4273 (1997).

[9.37] E. Hille, *Analytic function theory II* (Ginn and Co. 1959).

[9.38] P.H. Richter, H-J. Scholz and A. Wittek, "A Breathing Chaos", *Nonlinearity* **1**, 45 (1990).

Exercises

Exercise 9.1 **Binary symbolic dynamics.** *Verify that the shortest prime binary cycles of the unimodal repeller of fig. 9.8 are* $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$. *Compare with table 9.2. Try to sketch them in the graph of the unimodal function* $f(x)$*; compare ordering of the periodic points with fig. 9.9. The point is that while overlayed on each other the longer cycles look like a hopeless jumble, the cycle points are clearly and logically ordered by the alternating binary tree.*

Exercise 9.2 **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. 9.5, and interpret the symbols* $\{0, 1\}$ *by relating them to topologically distinct types of collisions. Compare with table 9.1. 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.*

Exercise 9.3 **Generating prime cycles.** *Write a program that generates all binary prime cycles up to given finite length.*

Exercise 9.4 **A contracting baker's map.** *Consider a contracting (or "dissipative") baker's defined in exercise 4.3.*

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 $(\epsilon_1 \ldots \epsilon_{n_p})$, $\epsilon_i \in \{0, 1\}$

$$A_p = \frac{3}{4} \sum_{j=1}^{n_p} \delta_{j,1} .$$

Exercise 9.5 **Reduction of 3-disk symbolic dynamics to binary.**

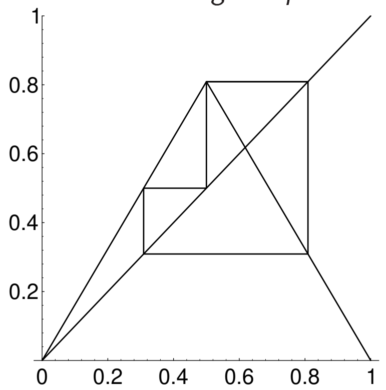
(a) *Verify that the 3-disk cycles*
$\{\overline{1\,2}, \overline{1\,3}, \overline{2\,3}\}$, $\{\overline{1\,2\,3}, \overline{1\,3\,2}\}$, $\{\overline{12\,13}$ + *2 perms.*$\}$,
$\{\overline{121\,232\,313}$ + *5 perms.*$\}$, $\{\overline{121\,323}$+ *2 perms.*$\}$, $\cdots$,
correspond to the fundamental domain cycles $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$ *respectively.*

(b) *Check the reduction for short cycles in table 9.1 by drawing them both in the full 3-disk system and in the fundamental domain, as in fig. 9.5.*

(c) *Optional: Can you see how the group elements listed in table 9.1 relate irreducible segments to the fundamental domain periodic orbits?*

Exercise 9.6 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 (9.13). This follows by inspection from the binary tree of fig. 9.9.*

Exercise 9.7 "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 10.6 and exercise 10.8. See also exercise 10.7 and exercise 10.9.

Exercise 9.8 Binary 3-step transition matrix. *Construct [8×8] binary 3-step transition matrix analogous to the 2-step transition matrix (9.14). 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 $m \to \infty$.*

Chapter 10

Counting, for pedestrians

> 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 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 flows with finite Markov graphs this determinant is a finite polynomial which can be read off the Markov graph.

The method goes well beyond the problem at hand, and forms the core of the entire treatise, making tangible a rather abstract notion of "spectral determinants" yet to come.

10.1 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 disks are too close and some part of trajectories is pruned, 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*:

$$h = \lim_{n\to\infty} \frac{1}{n} \ln K_n \ . \tag{10.1}$$

We shall now relate this quantity to the leading eigenvalue of the transition matrix.

The transition matrix element $T_{ij} \in \{0,1\}$ in (9.2) 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

10.1
page 194

$$(T^n)_{ij} = \sum_{k_1,k_2,\ldots,k_{n-1}} T_{ik_1}T_{k_1k_2}\ldots T_{k_{n-1}j}$$

receives a contribution 1 from every admissible sequence of transitions, so $(T^n)_{ij}$ is the number of admissible n symbol itineraries starting with j and ending with i.

Example 10.1 3-disk itinerary counting

The $(T^2)_{13} = 1$ element of T^2 for the 3-disk transition matrix (9.5)

$$\begin{pmatrix} c & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} . \tag{10.2}$$

corresponds to $3 \to 2 \to 1$, the only 2-step path from 3 to 1, while $(T^2)_{33} = 2$ counts the two itineraries 313 and 323.

The total number of admissible itineraries of n symbols is

$$K_n = \sum_{ij} (T^n)_{ij} = (\,1,1,\ldots,1\,)\ T^n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} . \tag{10.3}$$

We can also count the number of prime cycles and pruned periodic points, but in order not to break up the flow of the main argument, we relegate these pretty results to sects. 10.5.2 and 10.5.3. 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 (10.3) 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

$$K_n = \sum_{\alpha=0}^{N-1} c_\alpha \lambda_\alpha^n \,.$$

10.2 page 194

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 (10.1) is given by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} \ln c_0 \lambda_0^n \left[1 + \frac{c_1}{c_0} \left(\frac{\lambda_1}{\lambda_0} \right)^n + \cdots \right] \\ &= \ln \lambda_0 + \lim_{n\to\infty} \left[\frac{\ln c_0}{n} + \frac{1}{n} \frac{c_1}{c_0} \left(\frac{\lambda_1}{\lambda_0} \right)^n + \cdots \right] \\ &= \ln \lambda_0 \,. \end{aligned} \qquad (10.4)$$

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 (10.1) 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 interactions: this analogy with statistical mechanics will be developed further in chapter 17.

10.2 Topological trace formula

There are two standard ways of getting at eigenvalues of a matrix - by evaluating the trace $\mathrm{tr}\, T^n = \sum \lambda_\alpha^n$, or by evaluating the determinant $\det(1 - zT)$. We start by evaluating the trace of transition matrices.

Consider an M-step memory transition matrix, like the 1-step memory example (9.14). 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 (9.14) and exercise 9.8). 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 string by one symbol; 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 (9.14), 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 9.2 and 10.1)

9.8 page 170

$$
\begin{aligned}
z\, \mathrm{tr}\, T &= t_0 + t_1, \\
z^2 \mathrm{tr}\, T^2 &= t_0^2 + t_1^2 + 2t_{10}, \\
z^3 \mathrm{tr}\, T^3 &= t_0^3 + t_1^3 + 3t_{100} + 3t_{101}, \\
z^4 \mathrm{tr}\, T^4 &= t_0^4 + t_1^4 + 2t_{10}^2 + 4t_{1000} + 4t_{1001} + 4t_{1011}.
\end{aligned} \tag{10.5}
$$

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 10.1: 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 9.2 and 10.2. (L. Rondoni)

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 (10.5) picks up contributions from all repeats of prime cycles, with each cycle contributing n_p periodic points, so the total number of periodic points of period n is given by

$$N_n = \mathrm{tr}\, T^n = \sum_{n_p|n} n_p t_p^{n/n_p} = \sum_p n_p \sum_{r=1}^{\infty} \delta_{n,n_p r} t_p^r \,. \tag{10.6}$$

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

$$\sum_{n=1}^{\infty} z^n N_n = \mathrm{tr}\, \frac{zT}{1 - zT} \,. \tag{10.7}$$

Substituting (10.6) into the left hand side, and replacing the right hand side by the eigenvalue sum $\mathrm{tr}\, T^n = \sum \lambda_\alpha^n$, we obtain our first example of a trace formula, the *topological trace formula*

$$\sum_{\alpha=0} \frac{z\lambda_\alpha}{1 - z\lambda_\alpha} = \sum_p \frac{n_p t_p}{1 - t_p} \,. \tag{10.8}$$

A trace formula relates the spectrum of eigenvalues of an operator - in this case the transition matrix - to the spectrum of periodic orbits of the dynamical system. The z^n sum in (10.7) is a discrete version of the Laplace transform, (see chapter 11), and the resolvent on the left hand side is the antecedent of the more sophisticated trace formulas (11.9), (11.19) and (27.3). We shall now use this result to compute the spectral determinant of the transition matrix.

10.3 Determinant of a graph

Our next task is to determine the zeros of the *spectral determinant* of an $[M \times M]$ transition matrix

$$\det(1 - zT) = \prod_{\alpha=0}^{M-1} (1 - z\lambda_\alpha) \ . \tag{10.9}$$

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 15 for arbitrary dynamical averages. For T a finite matrix (10.9) is just the characteristic equation for T. However, we shall be able to compute this object even when the dimension of T and other such operators goes to ∞, and for that reason we prefer to refer to (10.9) 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:

4.1
page 81

$$\begin{aligned}\det(1 - zT) &= \exp(\mathrm{tr}\ \ln(1 - zT)) = \exp\left(-\sum_{n=1} \frac{z^n}{n} \mathrm{tr}\, T^n\right) \\ &= 1 - z\, \mathrm{tr}\, T - \frac{z^2}{2}\left((\mathrm{tr}\ T)^2 - \mathrm{tr}\,(T^2)\right) - \dots \end{aligned} \tag{10.10}$$

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 (10.10) to the corresponding Markov graph of chapter 9: to this end we start by the usual algebra textbook expression

$$\det(1-zT) = \sum_{\{\pi\}} (-1)^{P_\pi} (1-zT)_{1,\pi_1} \cdot (1-zT)_{2,\pi_2} \cdots (1-zT)_{M,\pi_M} \qquad (10.11)$$

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 (10.11) yields a polynomial of order M in z: a contribution of order n in z picks up $M-n$ unit factors along the diagonal, the remaining matrix elements yielding

$$(-z)^n (-1)^{P_{\tilde{\pi}}} T_{\eta_1,\tilde{\pi}_{\eta_1}} \cdots T_{\eta_n,\tilde{\pi}_{\eta_n}} \qquad (10.12)$$

where $\tilde{\pi}$ is the permutation of the subset of n distinct symbols $\eta_1 \ldots \eta_n$ indexing T matrix elements. As in (10.5), 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 (10.12) may be factored in terms of *local traces* $t_{c_1} t_{c_2} \cdots t_{c_k}$, that is loops on the Markov graph. These loops are non-intersecting, as each node may only be reached by *one* link, and they are indeed loops, as if a node is reached by a link, it has to be the starting point of another *single* link, as each 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 (10.12) that is, only the diagonal elements of T are picked up. We have $k = n$ loops and an even permutation $\tilde{\pi}$ so the sign is given by $(-1)^k$, k being the number of loops. Now take the case in which we have i single loops and j loops of length 2 (we must thus have $n = 2j + i$). The parity of the permutation gives $(-1)^j$ and the first factor in (10.12) gives $(-1)^n = (-1)^{2j+i}$. So once again these terms combine into $(-1)^k$, where $k = i + j$ is the number of loops. We may summarize our findings as follows: 10.3 page 194

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:

$$\det(1-zT) = \sum_{k=0}^{f} \sum_{\pi}{}' (-1)^k t_{p_1} \cdots t_{p_k} \qquad (10.13)$$

Any self-intersecting loop is *shadowed* by a product of two loops that share the intersection point. As both the long loop t_{ab} and its shadow $t_a t_b$ in the case at hand carry the same weight $z^{n_a+n_b}$, the cancellation is exact, and the loop expansion (10.13) is finite, with f the maximal number of non-intersecting loops.

We refer to the set of all non-self-intersecting loops $\{t_{p_1}, t_{p_2}, \cdots t_{p_f}\}$ as the 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. 10.4, p. 181

10.3.1 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 graph of fig. 9.10(b) is a little bit too simple, but anyway, let us start humbly; there are only two non-intersecting loops, yielding

$$\det(1 - zT) = 1 - t_0 - t_1 = 1 - 2z\,. \tag{10.14}$$

The leading (and only) zero of this characteristic polynomial yields the topological entropy $e^h = 2$. As we know that there are $K_n = 2^n$ binary strings of length N, we are not surprised. Similarly, for complete symbolic dynamics of N symbols the Markov graph has one node and N links, yielding

$$\det(1 - zT) = 1 - Nz\,, \tag{10.15}$$

whence the topological entropy $h = \ln N$.

A more interesting example is the "golden mean" pruning of fig. 10.1. There is only one grammar rule, that a repeat of symbol 0 is forbidden.

10.4
page 195

The non-intersecting loops are of length 1 and 2, so the topological polynomial is given by

$$\det(1 - zT) = 1 - t_1 - t_{01} = 1 - z - z^2\,. \tag{10.16}$$

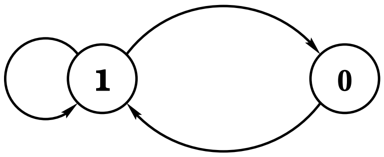


Figure 10.1: The golden mean pruning rule Markov graph, see also fig. 9.12.

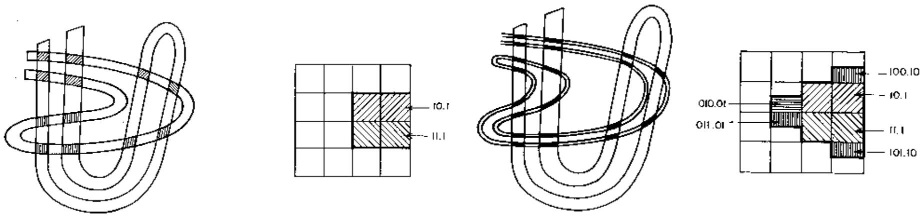


Figure 10.2: (a) An incomplete Smale horseshoe: the inner forward fold does not intersect the two rightmost backward folds. (b) The primary pruned region in the symbol square and the corresponding forbidden binary blocks. (c) An incomplete Smale horseshoe which illustrates (d) the monotonicity of the pruning front: the thick line which delineates the left border of the primary pruned region is monotone on each half of the symbol square. The backward folding in figures (a) and (c) is only schematic - in invertible mappings there are further missing intersections, all obtained by the forward and backward iterations of the primary pruned region.

The leading root of this polynomial is the golden mean, so the entropy (10.4) 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. 10.3(d) are indicated in fig. 10.3(e). The determinant can be written down by inspection, as the sum of all possible partitions of the graph into products of non-intersecting loops, with each loop carrying a minus sign:

$$\det(1-T) = 1 - t_0 - t_{0011} - t_{0001} - t_{00011} + t_0 t_{0011} + t_{0011} t_{0001} \tag{10.17}$$

With $t_p = z^{n_p}$, where n_p is the length of the p-cycle, the smallest root of

10.10
page 196

$$0 = 1 - z - 2z^4 + z^8 \tag{10.18}$$

yields the topological entropy $h = -\ln z$, $z = 0.658779\ldots$, $h = 0.417367\ldots$, significantly smaller than the entropy of the covering symbolic dynamics, the complete binary shift $h = \ln 2 = 0.693\ldots$

in depth:
sect. P.1, p. 791

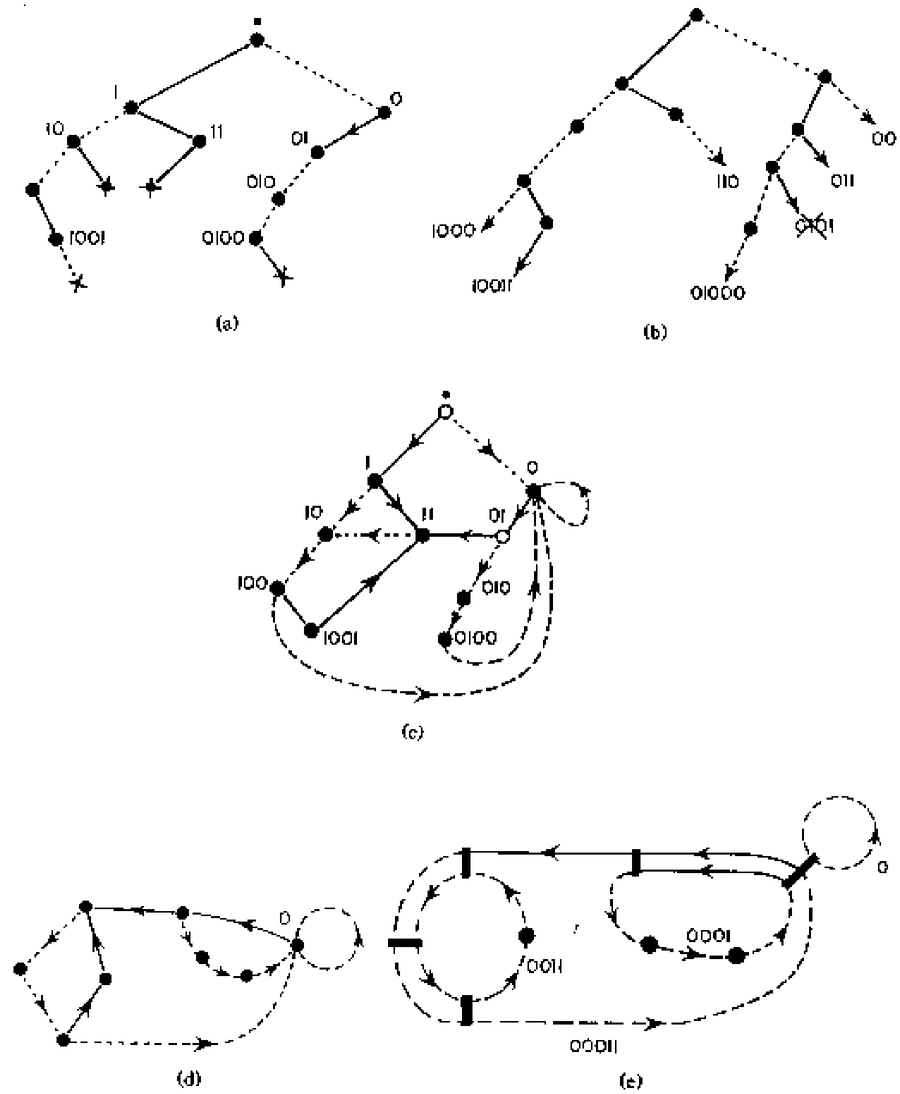


Figure 10.3: Conversion of the pruning front of fig. 10.2d into a finite Markov graph. (a) Starting with the start node ".", delineate all pruning blocks on the binary tree. A solid line stands for "1" and a dashed line for "0". Ends of forbidden strings are marked with $\times$. Label all internal nodes 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 (10.17).

10.4 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 (10.10)

4.1
page 81

$$\det(1 - zT) = 1 - \sum_{n=1}^{\infty} \hat{c}_n z^n . \tag{10.19}$$

For finite dimensional matrices the expansion is a finite polynomial, and (10.19) is an identity; however, for infinite dimensional operators the cumulant expansion coefficients $\hat{c}_n$ *define* the determinant.

Let us now evaluate the determinant in terms of traces for an arbitrary transition matrix. In order to obtain an expression for the spectral determinant (10.9) in terms of cycles, substitute (10.6) into (10.19) and sum over the repeats of prime cycles

$$\det(1 - zT) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{t_p^r}{r} \right) = \prod_p (1 - t_p) . \tag{10.20}$$

where for the topological entropy the weight assigned to a prime cycle p of length n_p is $t_p = z^{n_p}$ if the cycle is admissible, or $t_p = 0$ if it is pruned. This determinant is called the *topological* or the *Artin-Mazur* zeta function, conventionally denoted by

$$1/\zeta_{\mathrm{top}} = \prod_p (1 - z^{n_p}) = 1 - \sum_{n=1} \hat{c}_n z^n . \tag{10.21}$$

Counting cycles amounts to giving each admissible prime cycle p weight $t_p = z^{n_p}$ and expanding the Euler product (10.21) as a power series in z. As the precise expression for coefficients $\hat{c}_n$ in terms of local traces t_p is more general than the current application to counting, we shall postpone its derivation to chapter 15.

The topological entropy h can now be determined from the leading zero $z = e^{-h}$ of the topological zeta function. For a finite $[N \times N]$ transition matrix, the number of terms in the characteristic equation (10.13) 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. 10.6, and many more later on.

fast track:
sect. 10.6, p. 186

10.4.1 Topological zeta function for flows

We now apply the method we used in deriving (11.19) to the problem of deriving the topological zeta functions for flows. By analogy to (11.17), the time-weighted density of prime cycles of period t is

$$\Gamma(t) = \sum_p \sum_{r=1} T_p \delta(t - rT_p)\,. \tag{10.22}$$

A Laplace transform smooths the sum over Dirac delta spikes and yields the *topological trace formula*

$$\sum_p \sum_{r=1} T_p \int_{0_+}^{\infty} dt\, e^{-st}\, \delta(t - rT_p) = \sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} \tag{10.23}$$

and the *topological zeta function* for flows:

$$\begin{aligned} 1/\zeta_{\mathrm{top}}(s) &= \prod_p \left(1 - e^{-sT_p}\right) \\ \sum_p T_p \sum_{r=1}^{\infty} e^{-sT_p r} &= -\frac{\partial}{\partial s} \ln 1/\zeta_{\mathrm{top}}(s)\,. \end{aligned} \tag{10.24}$$

This is the continuous time version of the discrete time topological zeta function (10.21) for maps; its leading zero $s = -h$ yields the topological entropy for a flow.

10.5 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.

10.5.1 Counting periodic points

N_n, the number of periodic points of period n can be computed from (10.19) and (10.7) as a logarithmic derivative of the topological zeta function

$$\begin{aligned}\sum_{n=1} N_n z^n &= \mathrm{tr}\left(-z\frac{d}{dz}\ln(1-zT)\right) = -z\frac{d}{dz}\ln\det\,(1-zT) \\ &= \frac{-z\frac{d}{dz}1/\zeta_{\mathrm{top}}}{1/\zeta_{\mathrm{top}}}\,. \end{aligned} \tag{10.25}$$

We see that the trace formula (10.8) diverges at $z \to e^{-h}$, as the denominator has a simple zero there.

As a check of formula (10.19) in the finite grammar context, consider the complete N-ary dynamics (9.3) 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 (10.19) we verify (10.15). The logarithmic derivative formula (10.25) 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. 10.3(e). Substituting (10.18) we obtain

$$\sum_{n=1} N_n z^n = \frac{z+8z^4-8z^8}{1-z-2z^4+z^8}\,. \tag{10.26}$$

Now the topological zeta function is not merely a tool for extracting the asymptotic growth of N_n; it actually yields the exact and not entirely trivial recursion relation for the numbers of periodic points: $N_1 = N_2 = N_3 = 1$, $N_n = 2n+1$ for $n = 4,5,6,7,8$, and $N_n = N_{n-1} + 2N_{n-4} - N_{n-8}$ for $n > 8$.

10.5.2 Counting prime cycles

Having calculated the number of periodic points, our next objective is to evaluate the number of *prime* cycles M_n for a dynamical system whose symbolic dynamics is built from N symbols. The problem of finding M_n is classical in combinatorics (counting necklaces made out of n beads out of N different kinds) and is easily solved. There are N^n possible distinct strings of length n composed of N letters. These N^n strings include all M_d prime d-cycles whose period d equals or divides n. A prime cycle is a non-repeating symbol string: for example, $p = \overline{011} = \overline{101} = \overline{110} = \ldots 011011 \ldots$ is prime, but $\overline{0101} = 010101 \ldots = \overline{01}$ is not. A prime d-cycle contributes d strings to the sum of all possible strings, one for each cyclic permutation. The total number of possible periodic symbol sequences of length n is therefore related to the number of prime cycles by

$$N_n = \sum_{d|n} d M_d \,, \tag{10.27}$$

where N_n equals $\operatorname{tr} T^n$. The number of prime cycles can be computed recursively

$$M_n = \frac{1}{n}\left(N_n - \sum_{d|n}^{d<n} d M_d\right) \,,$$

10.11 page 197

or by the *Möbius inversion formula*

$$M_n = n^{-1} \sum_{d|n} \mu\left(\frac{n}{d}\right) N_d \,. \tag{10.28}$$

where the Möbius function $\mu(1) = 1$, $\mu(n) = 0$ if n has a squared factor, and $\mu(p_1 p_2 \ldots p_k) = (-1)^k$ if all prime factors are different.

10.12 page 197

We list the number of prime cycles up to length 10 for 2-, 3- and 4-letter complete symbolic dynamics in table 10.2. The number of *prime* cycles follows by Möbius inversion (10.28).

10.5.3 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 the diagonal entries, $T_{N-disk} = T_c - \mathbf{1}$, so the number of the N-disk periodic points is

$$N_n = \operatorname{tr} T^n_{N-disk} = (N-1)^n + (-1)^n (N-1) \tag{10.29}$$

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 10.2: Number of prime cycles for various alphabets and grammars up to length 10. The first column gives the cycle length, the second the formula (10.28) for the number of prime cycles for complete N-symbol dynamics, columns three through five give the numbers for $N = 2, 3$ and 4.

(here T_c is the complete symbolic dynamics transition matrix (9.3)). For the N-disk pruned case (10.29) Möbius inversion (10.28) yields

$$\begin{aligned} M_n^{N-disk} &= \frac{1}{n}\sum_{d|n} \mu\left(\frac{n}{d}\right)(N-1)^d + \frac{N-1}{n}\sum_{d|n}\mu\left(\frac{n}{d}\right)(-1)^d \\ &= M_n^{(N-1)} \quad \text{for} \quad n > 2\,. \end{aligned} \tag{10.30}$$

There are no fixed points, $M_1^{N-disk} = 0$. The number of periodic points of period 2 is $N^2 - N$, hence there are $M_2^{N-disk} = N(N-1)/2$ prime cycles of length 2; for lengths $n > 2$, the number of prime cycles is the same as for the complete $(N-1)$-ary dynamics of table 10.2.

10.5.4 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 (10.30)). To obtain the topological zeta function, just divide out the binary 1- and 2-cycles $(1-zt_0)(1-zt_1)(1-z^2t_{01})$ and multiply with the correct 3-disk 2-cycles $(1-z^2t_{12})(1-z^2t_{13})(1-z^2t_{23})$:

10.15 page 198
10.16 page 198

$$\begin{aligned} 1/\zeta_{3-disk} &= (1-2z)\frac{(1-z^2)^3}{(1-z)^2(1-z^2)} \\ &= (1-2z)(1+z)^2 = 1 - 3z^2 - 2z^3\,. \end{aligned} \tag{10.31}$$

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 10.3: List of the 3-disk prime cycles up to length 10. Here n is the cycle length, M_n the number of prime cycles, N_n the number of periodic points and S_n the number of distinct prime cycles under the C_{3v} symmetry (see chapter 19 for further details). Column 3 also indicates the splitting of N_n into contributions from orbits of lengths that divide n. The prefactors in the fifth column indicate the degeneracy m_p of the cycle; for example, 3·12 stands for the three prime cycles $\overline{12}$, $\overline{13}$ and $\overline{23}$ related by $2\pi/3$ rotations. Among symmetry related cycles, a representative $\hat{p}$ which is lexically lowest was chosen. The cycles of length 9 grouped by parenthesis are related by time reversal symmetry, but not by any other C_{3v} transformation.

The factorization reflects the underlying 3-disk symmetry; we shall rederive it in (19.25). As we shall see in chapter 19, symmetries lead to factorizations of topological polynomials and topological zeta functions.

The example of exercise 10.17 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 $(1 - t_b) = (1 - z)$:

10.17 page 198

$$1/\zeta_{\text{top}} = (1 - z)\left(1 - z - \frac{z}{1 - z}\right) = 1 - 3z + z^2$$

10.20 page 199

10.6 Topological zeta function for an infinite partition

(K.T. Hansen and P. Cvitanović)

Now consider an example of a dynamical system which (as far as we know - there is no proof) has an infinite partition, or an infinity of longer

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 10.4: List of the 4-disk prime cycles up to length 8. The meaning of the symbols is the same as in table 10.3. 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.

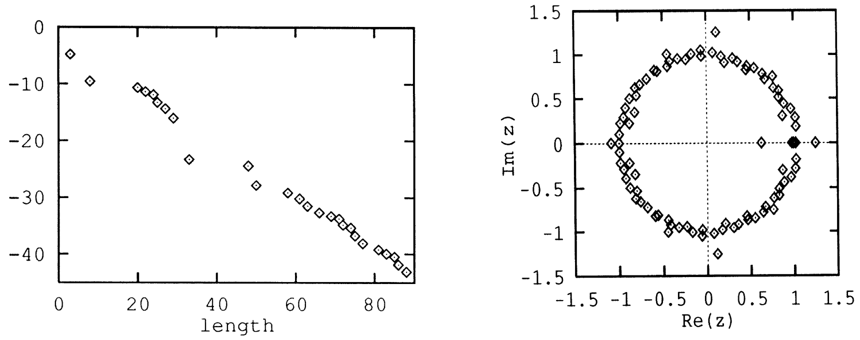


Figure 10.4: (a) The logarithm of the difference between the leading zero of the finite polynomial approximations to topological zeta function and our best estimate, as a function of the length for the quadratic map $A = 3.8$. (b) The 90 zeroes of the characteristic polynomial for the quadratic map $A = 3.8$ approximated by symbolic strings up to length 90. (from ref. [1.3])

and longer pruning rules. Take the 1-d quadratic map

$$f(x) = Ax(1-x)$$

with $A = 3.8$. It is easy to check numerically that the itinerary or the "kneading sequence" of the critical point $x = 1/2$ is

$$K = 1011011110110111101011110111110\ldots$$

where the symbolic dynamics is defined by the partition of fig. 9.8. 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

$$\begin{aligned} 1/\zeta_{\mathrm{top}}^{(29)} &= 1 - z^1 - z^2 + z^3 - z^4 - z^5 + z^6 - z^7 + z^8 - z^9 - z^{10} \\ &\quad + z^{11} - z^{12} - z^{13} + z^{14} - z^{15} + z^{16} - z^{17} - z^{18} + z^{19} + z^{20} \\ &\quad - z^{21} + z^{22} - z^{23} + z^{24} + z^{25} - z^{26} + z^{27} - z^{28} \,. \end{aligned} \tag{10.32}$$

The smallest real root of this approximate topological zeta function is

$$z = 0.62616120\ldots \tag{10.33}$$

Constructing finite Markov graphs of increasing length corresponding to $A \to 3.8$ we find polynomials with better and better estimates for the topological entropy. For the closest stable period 90 orbit we obtain our best estimate of the topological entropy of the repeller:

$$h = -\ln 0.62616130424685\ldots = 0.46814726655867\ldots. \tag{10.34}$$

Fig. 10.4 illustrates the convergence of the truncation approximations to the topological zeta function as a plot of the logarithm of the difference between the zero of a polynomial and our best estimate (10.34), plotted as a function of the length of the stable periodic orbit. The error of the estimate (10.33) is expected to be of order $z^{29} \approx e^{-14}$ because going from length 28 to a longer truncation yields typically combinations of loops with 29 and more nodes giving terms $\pm z^{29}$ and of higher order in the polynomial.

Hence the convergence is exponential, with exponent of $-0.47 = -h$, the topological entropy itself.

In fig. 10.4(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 $\lambda_1/\lambda_0 = 1/e^h = e^{-h}$.

10.7 Shadowing

The topological zeta function is a pretty function, but the infinite product (10.20) 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 (15.5) of chapter 15; 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 12.2. As we have already argued in sect. 1.5.4 and appendix J.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 (10.13), 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. 10.4(b).

One striking aspect of the pruned cycle expansion (10.32) compared to the trace formulas such as (10.7) is that coefficients are not growing exponentially - indeed they all remain of order 1, so instead having a radius of convergence e^{-h}, in the example at hand the topological zeta function has the unit circle as the radius of convergence. In other words, exponentiating the spectral problem from a trace formula to a spectral determinant as in (10.19) increases the *analyticity domain*: the pole in the trace (10.8) 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, 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 20.

You might well ask: What is wrong with computing an entropy from (10.1)? 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. 9.12, or the cycle expansion of (10.26) - and the finite estimates of $h_n = \ln K_n/n$ converge nonuniformly to h, and on top of that with a slow rate of convergence, $|h - h_n| \approx O(1/n)$ as in (10.4). The determinant (10.9) is much smarter, as by construction it encodes the self-similarity of the dynamics, and yields the asymptotic value of h with no need for any finite n extrapolations.

So, the main lesson of learning how to count well, a lesson that will be affirmed over and over, is that while the trace formulas are a conceptually essential step in deriving and understanding periodic orbit theory, the spectral determinant is the right object to use in actual computations. Instead of resumming all of the exponentially many periodic points required by trace formulas at each level of truncation, spectral determinants incorporate only the small incremental corrections to what is already known - and that makes them more convergent and economical to use.

Commentary

Remark 10.1 "Entropy". The ease with which the topological entropy can be motivated obscures the fact that our definition does not lead to an invariant characterization of the dynamics, as the choice of symbolic dynamics is largely arbitrary: the same caveat applies to other entropies to be discussed in chapter 17, and to get proper invariants one needs to evaluate a supremum over all possible partitions. The key mathematical point that eliminates the need of such search is the existence of *generators*, *i.e.* partitions that under dynamics are able to probe the whole phase space on arbitrarily small scales: more precisely a generator is a finite partition $\Omega = \omega_1 \ldots \omega_N$, with the following property: take $\mathcal{M}$ the subalgebra of the phase space generated by Ω, and consider the partition built upon all possible intersections 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. [10.1] and chapter 17.

Remark 10.2 Perron-Frobenius matrices. For a proof of Perron theorem on the leading eigenvalue see ref. [1.11]. Ref. [10.2], sect. A4.1 contains a clear discussion of the spectrum of the transition matrix.

Remark 10.3 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. [10.3, 10.4, 10.5].

Remark 10.4 T is not trace class. Note to the erudite reader: the transition matrix T (in the infinite partition limit (10.19)) is *not* trace class in the sense of appendix K. Still the trace is well defined in the $n \to \infty$ limit.

Remark 10.5 Artin-Mazur zeta functions. Motivated by A. Weil's zeta function for the Frobenius map [10.6], Artin and Mazur [12.13] introduced the zeta function (10.21) that counts periodic points for diffeomorphisms (see also ref. [10.7] for their evaluation for maps of the interval). Smale [10.8] conjectured rationality of the zeta functions for Axiom A diffeomorphisms, later proved by Guckenheimer [10.9] and Manning [10.10]. See remark 12.4 on page 229 for more zeta function history.

Remark 10.6 Ordering periodic orbit expansions. In sect. 15.4 we will introduce an alternative way of hierarchically organizing cumulant expansions, in which the order is dictated by stability rather than cycle length: such a procedure may be better suited to perform computations when the symbolic dynamics is not well understood.

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 (10.21) of the topological zeta function (that is, the spectral determinant of the transition matrix):

$$1/\zeta_{\mathrm{top}}(z) \;=\; 1 - \sum_{k=1} \hat{c}_k z^k \,.$$

For subshifts of finite type, the transition matrix is finite, and the topological zeta function is a finite polynomial evaluated by the loop expansion (10.13) of $\det(1 - zT)$. For infinite grammars the topological zeta function is defined by its cycle expansion. The topological entropy h is given by the smallest zero $z = e^{-h}$. This expression for the entropy is *exact*; in contrast to the definition (10.1), 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 11, 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. 10.6, and fig. 10.4 in particular.

References

[10.1] V.I. Arnold and A. Avez, “Ergodic Problems of Classical Mechanics”, *Addison-Wesley, Redwood City* (1989).

[10.2] J. Zinn-Justin, “Quantum Field Theory and Critical Phenomena”, *Clarendon Press, Oxford* (1996).

[10.3] A. Salomaa, “Formal Languages”, *Academic Press, San Diego* (1973).

[10.4] J.E. Hopcroft and J.D. Ullman, "Introduction to Automata Theory, Languages and Computation", *Addison-Wesley, Reading Ma* (1979).

[10.5] D.M. Cvektović, M. Doob and H. Sachs, "Spectra of Graphs", *Academic Press, New York* (1980).

[10.6] A. Weil, *Bull.Am.Math.Soc.* **55**, 497 (1949).

[10.7] 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).

[10.8] S. Smale, *Ann. Math.*, **74**, 199 (1961).

[10.9] J. Guckenheimer, *Invent. Math.* **39**, 165 (1977).

[10.10] A. Manning, *Bull. London Math. Soc.* **3**, 215 (1971).

Exercises

Exercise 10.1 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 10.3.*

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* $\{0, 1\}$*.*

Exercise 10.2 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.*

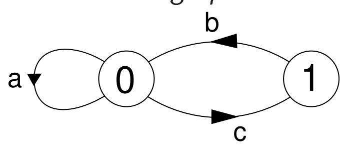
(b) *Do eigenvalues* λ_k *need to be distinct,* $\lambda_k \neq \lambda_l$ *for* $k \neq l$*?*

Exercise 10.3 Loop expansions. *Prove by induction the sign rule in the determinant expansion (10.13):*

$$\det(1 - z\mathbf{T}) = \sum_{k \geq 0} \sum_{p_1 + \cdots + p_k} (-1)^k t_{p_1} t_{p_2} \cdots t_{p_k} \,.$$

Exercise 10.4 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?*

Exercise 10.5 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, $\cdots$.

Exercise 10.6 <u>"Golden mean" pruned map.</u> *Continuation of exercise 9.7: 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 10.8. See also exercise 10.9.

Exercise 10.7 Alphabet {0,1}, prune $\_00\_$. *The Markov diagram fig. 9.12(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 (10.13) becomes*

$$\begin{aligned} 1/\zeta &= (1-t_1)(1-t_2)(1-t_{12})(1-t_{112})\ldots \\ &= 1 - t_1 - t_2 - (t_{12} - t_1 t_2) - (t_{112} - t_{12} t_1) - (t_{122} - t_{12} t_2)\ldots \end{aligned} \quad (10.35)$$

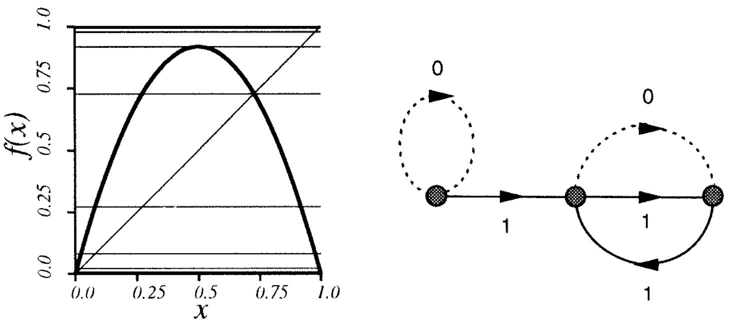


Figure 10.5: (a) A unimodal map for which the critical point maps into the right hand fixed point in three iterations, and (b) the corresponding Markov graph (Kai T. Hansen).

In the original binary alphabet this corresponds to:

$$
\begin{aligned}
1/\zeta &= 1 - t_1 - t_{10} - (t_{110} - t_1 t_{10}) \\
&\quad - (t_{1110} - t_{110} t_1) - (t_{11010} - t_{110} t_{10}) \ldots
\end{aligned}
\tag{10.36}
$$

This symbolic dynamics describes, for example, circle maps with the golden mean winding number, see chapter 21. For unimodal maps this symbolic dynamics is realized by the tent map of exercise 10.6.

Exercise 10.8 **Spectrum of the "golden mean" pruned map.** *(medium - Exercise 10.6 continued)*

(a) *Determine an expression for* $\mathrm{tr}\,\mathcal{L}^n$*, the trace of powers of the Perron-Frobenius operator (7.10) for the tent map of exercise 10.6.*

(b) *Show that the spectral determinant for the Perron-Frobenius operator is*

$$
\det(1 - z\mathcal{L}) = \prod_{k \text{ even}} \left(1 + \frac{z}{\Lambda^{k+1}} - \frac{z^2}{\Lambda^{2k+2}}\right) \prod_{k \text{ odd}} \left(1 + \frac{z}{\Lambda^{k+1}} + \frac{z^2}{\Lambda^{2k+2}}\right). \tag{10.37}
$$

Exercise 10.9 **A unimodal map example.** *Consider a unimodal map of fig. 10.5(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. 10.5(b).*

(Kai T. Hansen)

Exercise 10.10 **Glitches in shadowing.**** *Note that the combination* t_{00011} *minus the "shadow"* $t_0 t_{0011}$ *in (10.17) cancels exactly, and does not contribute to the topological polynomial (10.18). Are you able to construct a smaller Markov graph than fig. 10.3(e)?*

Exercise 10.11 **Whence Möbius function?** *To understand where the Möbius function comes from consider the function*

$$f(n) = \sum_{d|n} g(d) \tag{10.38}$$

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

$$g(n) = \sum_{d|n} \mu(n/d) f(d) \tag{10.39}$$

Exercise 10.12 **Counting prime binary cycles.** *In order to get comfortable with Möbius inversion reproduce the results of the second column of table 10.2.*

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.

Exercise 10.13 **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 (9.10) 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 (10.5) collapse to $tr(T^k) = (t_0 + t_1)^k$, and $1/\zeta$ is simply*

$$\prod_p (1 - t_p) = 1 - t_0 - t_1 \tag{10.40}$$

Substituting (10.40) into the identity

$$\prod_p (1 + t_p) = \prod_p \frac{1 - {t_p}^2}{1 - t_p}$$

we obtain

$$\begin{aligned} \prod_p (1 + t_p) &= \frac{1 - t_0^2 - t_1^2}{1 - t_0 - t_1} = 1 + t_0 + t_1 + \frac{2t_0 t_1}{1 - t_0 - t_1} \\ &= 1 + t_0 + t_1 + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} 2 \binom{n-2}{k-1} t_0^k t_1^{n-k} . \end{aligned} \tag{10.41}$$

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 $M_{n,k}$ $\quad (n = 1, 2, \dots ; k = \{0, 1\}$ *for* $n = 1;\ k = 1, \dots , n-1$ *for* $n \geq 2)$ *the number of prime* n*-cycles whose labels contain* k *zeros. Show that*

$$\begin{aligned} M_{1,0} &= M_{1,1} = 1 \\ nM_{n,k} &= \sum_{m \left| \frac{n}{k} \right.} \mu(m) \binom{n/m}{k/m}, \quad n \geq 2\,, k = 1, \dots , n-1 \end{aligned}$$

where the sum is over all m *which divide both* n *and* k*.*

Exercise 10.14 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?*

Exercise 10.15 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*

$$\begin{aligned} 1/\zeta_{4-disk} &= (1-3z)\frac{(1-z^2)^6}{(1-z)^3(1-z^2)^3} \\ &= (1-3z)(1+z)^3 = 1 - 6z^2 - 8z^3 - 3z^4\ . \end{aligned} \tag{10.42}$$

Exercise 10.16 N-disk pinball topological polynominal. *Show that for an* N*-disk pinball, the topological polynominal is given by*

$$\begin{aligned} 1/\zeta_{N-disk} &= (1-(N-1)z)\,\frac{(1-z^2)^{N(N-1)/2}}{(1-z)^{N-1}(1-z^2)^{(N-1)(N-2)/2}} \\ &= (1-(N-1)z)\,(1+z)^{N-1}\ . \end{aligned} \tag{10.43}$$

The topological polynomial has a root $z^{-1} = N-1$*, as we already know it should from (10.29) or (10.15). We shall see in sect. 19.4 that the other roots reflect the symmetry factorizations of zeta functions.*

Exercise 10.17 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 (10.13) becomes*

$$\begin{aligned} 1/\zeta &= (1-t_a)(1-t_b)(1-t_c)(1-t_{cb})(1-t_{ac})(1-t_{cbb})\dots \\ &= 1 - t_a - t_b - t_c + t_a t_b - (t_{cb} - t_c t_b) - (t_{ac} - t_a t_c) - (t_{cbb} - t_{cb} t_b)\dots \end{aligned}$$

The effect of the $\_ab\_$ *pruning is essentially to unbalance the 2 cycle curvature* $t_{ab} - t_a t_b$*; the remainder of the cycle expansion retains the curvature form.*

Exercise 10.18 Alphabet $\{0,1\}$, prune n repeats. *of "0" $\_000\ldots00\_$.*

This is equivalent to the n symbol alphabet $\{1, 2, \ldots, n\}$ unrestricted symbolic dynamics, with symbols corresponding to the possible 10...00 block lengths: 2=10, 3=100, ..., n=100...00. The cycle expansion (10.13) becomes

$$1/\zeta = 1 - t_1 - t_2 \ldots - t_n - (t_{12} - t_1 t_2) \ldots - (t_{1n} - t_1 t_n) \ldots \tag{10.44}$$

.

Exercise 10.19 Alphabet $\{0,1\}$, prune $\_1000\_$, $\_00100\_$, $\_01100\_$.

Show that the topological zeta function is given by

$$1/\zeta = (1 - t_0)(1 - t_1 - t_2 - t_{23} - t_{113}) \tag{10.45}$$

with the unrestricted 4-letter alphabet $\{1, 2, \underline{23}, \underline{113}\}$. Here 2, 3, refer to 10, 100 respectively, as in exercise 10.18.

Exercise 10.20 Alphabet $\{0,1\}$, prune $\_1000\_$, $\_00100\_$, $\_01100\_$, $\_10011\_$. *The first three pruning rules were incorporated in the preceeding exercise.*

(a) Show that the last pruning rule $\_10011\_$ leads (in a way similar to exercise 10.19) to the alphabet $\{\underline{21^k}, \underline{23}, \underline{21^k113}; \overline{1}, \overline{0}\}$, and the cycle expansion

$$1/\zeta = (1 - t_0)(1 - t_1 - t_2 - t_{23} + t_1 t_{23} - t_{2113}) \tag{10.46}$$

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

$$1/\zeta_{top} = (1 - z)(1 - z - z^2 - z^5 + z^6 - z^7) \tag{10.47}$$

and check that it yields the exact value of the entropy $h = 0.522737642\ldots$.

Chapter 11

Trace formulas

> The trace formula is not a formula, it is an idea.
> Martin Gutzwiller

Dynamics is posed in terms of local equations, but the ergodic averages require global information. How can we use a local description of a flow to learn something about the global behavior? We have given a quick sketch of this program in sects. 1.5 and 1.6; now we redo the same material in greater depth. In chapter 8 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.

11.1 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 (8.17) 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

10.2
page 194

$$\mathrm{tr}\,\mathcal{L}^t = \int dx\, \mathcal{L}^t(x,x) = \int dx\, \delta\big(x - f^t(x)\big)\, e^{\beta \cdot A^t(x)}\,. \tag{11.1}$$

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),

$$\mathrm{tr}\,\mathcal{L}^t = \sum_{\alpha=0}^{\infty} e^{s_\alpha t}\,. \tag{11.2}$$

On the other hand, we have learned in sect. 7.2 how to evaluate the delta-function integral (11.1).

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 (11.19) covers both cases.

11.1.1 Hyperbolicity assumption

According to (7.8) the trace (11.1) 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. 11.3, 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,

$$\mathrm{tr}\,_p\mathcal{L}^{n_p} = \int_{\mathcal{M}_p} dx\, \delta(x - f^{n_p}(x)) = \frac{n_p}{\left|\det\,\left(\mathbf{1} - \mathbf{J}_p\right)\right|} = n_p \prod_{i=1}^{d} \frac{1}{|1 - \Lambda_{p,i}|} \tag{11.3}$$

(in (7.9) 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 (4.35). As the integral (11.3) 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. 11.3, 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 (11.3) into a product over the expanding and the contracting eigenvalues

$$\left|\det\left(\mathbf{1} - \mathbf{J}_p\right)\right|^{-1} = \frac{1}{|\Lambda_p|}\prod_e \frac{1}{1 - 1/\Lambda_{p,e}}\prod_c \frac{1}{1-\Lambda_{p,c}}\,, \tag{11.4}$$

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:

$$\begin{array}{lcll} |\Lambda_{p,e}| & > & e^{\lambda_e T_p} & \text{any } p \text{, any expanding eigenvalue } |\Lambda_{p,e}| > 1 \\ |\Lambda_{p,c}| & < & e^{-\lambda_c T_p} & \text{any } p \text{, any contracting eigenvalue } |\Lambda_{p,c}| < 1 \end{array} \tag{11.5}$$

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 18.

It follows from (11.4) 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. 12.3. However, for evaluation of the full spectrum the exact cycle weight (11.3) has to be kept.

11.2 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:

$$\mathrm{tr}\,\mathcal{L}^n = \int dx\,\mathcal{L}^n(x,x) = \sum_{x_i\in\mathrm{Fix}f^n} \frac{e^{\beta\cdot A_i}}{|\det\,(\mathbf{1}-\mathbf{J}^n(x_i))|}\,. \tag{11.6}$$

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 (8.5) 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 (7.8) 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 (8.5), the integrated observable A^n is additive along the cycle: If a prime cycle p trajectory retraces itself r times, $n = rn_p$, we obtain A_p repeated r times, $A_i = A^n(x_i) = rA_p$, $x_i \in p$.

A prime cycle is a single traversal of the orbit, and its label is a non-repeating symbol string. There is only one prime cycle for each cyclic permutation class. For example, the four cycle points $\overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ belong to the same prime cycle $p = 0011$ of length 4. As both the stability of a cycle and the weight A_p are the same everywhere along the orbit, each prime cycle of length n_p contributes n_p terms to the sum, one for each cycle point. Hence (11.6) can be rewritten as a sum over all prime cycles and their repeats

☞ chapter 9

$$\mathrm{tr}\,\mathcal{L}^n = \sum_p n_p \sum_{r=1}^{\infty} \frac{e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\delta_{n,n_p r}\,, \tag{11.7}$$

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 (11.17). 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

$$\sum_{n=1}^{\infty} z^n \mathrm{tr}\,\mathcal{L}^n = \mathrm{tr}\,\frac{z\mathcal{L}}{1-z\mathcal{L}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r} e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\,. \tag{11.8}$$

Expressing the trace as in (11.2), in terms of the sum of the eigenvalues of $\mathcal{L}$, we obtain the *trace formula for maps*:

$$\sum_{\alpha=0}^{\infty} \frac{z e^{s_\alpha}}{1 - z e^{s_\alpha}} = \sum_p n_p \sum_{r=1}^{\infty} \frac{z^{n_p r}\, e^{r\beta \cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\,. \tag{11.9}$$

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. 11.3, p. 206

11.2.1 A trace formula for transfer operators

For a piecewise-linear map (7.11), we can explicitly evaluate the trace formula. By the piecewise linearity and the chain rule $\Lambda_p = \Lambda_0^{n_0}\Lambda_1^{n_1}$, where the cycle p contains n_0 symbols 0 and n_1 symbols 1, the trace (11.6) reduces to

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{m=0}^{n} \binom{n}{m} \frac{1}{|1 - \Lambda_0^m \Lambda_1^{n-m}|} = \sum_{k=0}^{\infty} \left(\frac{1}{|\Lambda_0|\Lambda_0^k} + \frac{1}{|\Lambda_1|\Lambda_1^k} \right)^n . \tag{11.10}$$

The eigenvalues are simply

$$e^{s_k} = \frac{1}{|\Lambda_0|\Lambda_0^k} + \frac{1}{|\Lambda_1|\Lambda_1^k}\,. \tag{11.11}$$

For $k = 0$ this is in agreement with the explicit transfer matrix (7.13) eigenvalues (7.14).

Alert reader should experience anxiety at this point. Is it not true that we have already written down explicitly the transfer operator in (7.13), 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 (7.13) 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. 13.1, the full spectrum (11.11) 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 (7.11) follows from (11.8)

$$\operatorname{tr} \frac{z\mathcal{L}}{1-z\mathcal{L}} = \frac{z\left(\frac{1}{|\Lambda_0-1|}+\frac{1}{|\Lambda_1-1|}\right)}{1-z\left(\frac{1}{|\Lambda_0-1|}+\frac{1}{|\Lambda_1-1|}\right)}\,, \tag{11.12}$$

verifying the trace formula (11.9).

11.3 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 $\operatorname{tr}\mathcal{L}^t$ 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. 1.10), and choose a local coordinate system with a longitudinal coordinate $dx_\parallel$ along the direction of the flow, and d transverse coordinates $x_\perp$

$$\operatorname{tr}_p\mathcal{L}^t = \int_{\mathcal{M}_p} dx_\perp dx_\parallel\, \delta\big(x_\perp - f^t_\perp(x)\big)\,\delta\Big(x_\parallel - f^t_\parallel(x)\Big)\ . \tag{11.13}$$

(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

$$\int_0^{L_p} dx_\| \, \delta\Big(x_\| - f_\|^t(x)\Big) \;=\; \int_0^{T_p} d\tau \, v(\tau) \, \delta\left(\int_\tau^{t+\tau} d\sigma \, v(\sigma) \Big|_{mod\, L_p} \right) . \tag{11.14}$$

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 (11.14) as

$$\int_0^{L_p} dx_\| \, \delta\Big(x_\| - f_\|^t(x)\Big) \;=\; \sum_{r=1}^{\infty} \delta(t - rT_p) \int_0^{T_p} d\tau \, v(\tau) \frac{1}{v(\tau + t)},$$

having used (7.7). 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

$$\oint_p dx_\| \, \delta\Big(x_\| - f_\|^t(x)\Big) \;=\; T_p \sum_{r=1}^{\infty} \delta(t - rT_p) \;\; . \tag{11.15}$$

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 27. For the remaining transverse integration variables the Jacobian is defined in a reduced Poincaré surface of section $\mathcal{P}$ of constant $x_\|$. Linearization of the periodic flow transverse to the orbit yields

$$\int_{\mathcal{P}} dx_\perp \delta\Big(x_\perp - f_\perp^{rT_p}(x)\Big) = \frac{1}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} \;, \tag{11.16}$$

where $\mathbf{J}_p$ is the p-cycle $[d \times d]$ *transverse* Jacobian matrix, and as in (11.5) we have to assume hyperbolicity, that is that the magnitudes of all transverse eigenvalues are bounded away from unity.

Substituting (11.15), (11.16) into (11.13), we obtain an expression for $\mathrm{tr}\,\mathcal{L}^t$ as a sum over all prime cycles p and their repetitions

$$\mathrm{tr}\,\mathcal{L}^t = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r\beta \cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} \delta(t - rT_p) \; . \tag{11.17}$$

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

$$\begin{aligned} \int_\epsilon^\infty dt\, e^{-st}\, \mathrm{tr}\, \mathcal{L}^t &= \mathrm{tr}\, \frac{e^{-(s-\mathcal{A})\epsilon}}{s-\mathcal{A}} = \sum_{\alpha=0}^{\infty} \frac{e^{-(s-s_\alpha)\epsilon}}{s-s_\alpha} \\ &= \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\,, \end{aligned} \tag{11.18}$$

where $\mathcal{A}$ is the generator of the semigroup of dynamical evolution, sect. 7.4. The *classical trace formula for flows* is the $\epsilon \to 0$ limit of the above expression:

$$\sum_{\alpha=0}^{\infty} \frac{1}{s-s_\alpha} = \sum_p T_p \sum_{r=1}^{\infty} \frac{e^{r(\beta\cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1}-\mathbf{J}_p^r\right)\right|}\,. \tag{11.19}$$

11.1
page 213

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 (11.9) is a special case of the trace formula for flows. The relation between the continuous and discrete time cases can be summarized as follows:

$$\begin{aligned} T_p &\leftrightarrow n_p \\ e^{-s} &\leftrightarrow z \\ e^{t\mathcal{A}} &\leftrightarrow \mathcal{L}^n\,. \end{aligned} \tag{11.20}$$

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 (11.19) 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. 12.2, after we complete our offering of trace formulas.

11.4 An asymptotic trace formula

In order to illuminate the manipulations of sect. 11.2 and relate them to something we already possess intuition about, we now rederive the heuristic sum of sect. 1.5.1 from the exact trace formula (11.9). The Laplace transforms (11.9) or (11.19) are designed to capture the time $\to \infty$ asymptotic behavior of the trace sums. By the hyperbolicity assumption (11.5) for $t = T_p r$ large the cycle weight approaches

$$\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right| \to |\Lambda_p|^r \,, \tag{11.21}$$

where Λ_p is the product of the expanding eigenvalues of $\mathbf{J}_p$. Denote the corresponding approximation to the nth trace (11.6) by

$$\Gamma_n = \sum_i^{(n)} \frac{1}{|\Lambda_i|} \,, \tag{11.22}$$

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 (11.9) by $\Gamma(z)$. Equivalently, think of this as a replacement of the evolution operator (8.19) by a transfer operator (as in sect. 11.2.1). For concreteness consider a dynamical system whose symbolic dynamics is complete binary, for example the 3-disk system fig. 1.4. In this case distinct periodic points that contribute to the nth periodic points sum (11.7) are labelled by all admissible itineraries composed of sequences of letters $s_i \in \{0, 1\}$:

$$\begin{aligned} \Gamma(z) &= \sum_{n=1}^{\infty} z^n \Gamma_n = \sum_{n=1}^{\infty} z^n \sum_{x_i \in \mathrm{Fix} f^n} \frac{e^{\beta \cdot A^n(x_i)}}{|\Lambda_i|} \\ &= z \left\{ \frac{e^{\beta \cdot A_0}}{|\Lambda_0|} + \frac{e^{\beta \cdot A_1}}{|\Lambda_1|} \right\} + z^2 \left\{ \frac{e^{2\beta \cdot A_0}}{|\Lambda_0|^2} + \frac{e^{\beta \cdot A_{01}}}{|\Lambda_{01}|} + \frac{e^{\beta \cdot A_{10}}}{|\Lambda_{10}|} + \frac{e^{2\beta \cdot A_1}}{|\Lambda_1|^2} \right\} \\ &\quad + z^3 \left\{ \frac{e^{3\beta \cdot A_0}}{|\Lambda_0|^3} + \frac{e^{\beta \cdot A_{001}}}{|\Lambda_{001}|} + \frac{e^{\beta \cdot A_{010}}}{|\Lambda_{010}|} + \frac{e^{\beta \cdot A_{100}}}{|\Lambda_{100}|} + \ldots \right\} \end{aligned} \tag{11.23}$$

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

$$\Gamma(z) = \sum_p \frac{n_p t_p}{1 - t_p} \,, \qquad t_p = z^{n_p} e^{\beta \cdot A_p} / |\Lambda_p| \,. \tag{11.24}$$

This is precisely our initial heuristic estimate (1.8). Note that we could not perform such sum over r in the exact trace formula (11.9) 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 (12.9).

☞ sect. 12.2

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 (11.6) tends to the leading $\mathcal{L}^n$ eigenvalue e^{ns_0}. Summing this asymptotic estimate level by level

$$\Gamma(z) \approx \sum_{n=1}^{\infty} (ze^{s_0})^n = \frac{ze^{s_0}}{1 - ze^{s_0}} \tag{11.25}$$

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 (11.24) 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 (12.11), and in sect. 12.5.1 we shall give the exact relation between the exact and the approximate trace formulas.

Commentary

Remark 11.1 Who's dunne it? Continuous time flow traces weighted by the cycle periods were introduced by Bowen [11.1] who treated them as Poincaré section suspensions weighted by the "time ceiling" function (3.2). They were used by Parry and Pollicott [11.2]. The derivation presented here [11.3] was designed to parallel as closely as possible the derivation of the Gutzwiller semiclassical trace formula, chapters 26 and 27.

Remark 11.2 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 (11.1) 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 (11.6) are called *flat* traces (see the review [11.4] and chapter 13). Other names for traces appear as well: for instance, in the context of $1{-}d$ mappings, *sharp* traces refer to generalizations of (11.6) 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 [12.12]. Such traces are connected to the theory of kneading invariants (see ref. [11.4] 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. [11.5, 15.6, 11.6].

Résumé

The description of a chaotic dynamical system in terms of cycles can be visualized as a tessellation of the dynamical system, fig. 1.9, 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 13). In order to implement such formulas, we have to determine "all" prime cycles. This task we postpone to chapter 14.

References

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

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

[11.3] P. Cvitanović and B. Eckhardt, *J. Phys.* **A 24**, L237 (1991).

[11.4] V. Baladi and D. Ruelle, *Ergodic Theory Dynamical Systems* **14**, 621 (1994).

[11.5] R. Artuso, E. Aurell and P. Cvitanović, *Nonlinearity* **3**, 325 (1990); *ibidem* 361 (1990)

[11.6] M. Pollicott, *J. Stat. Phys.* **62**, 257 (1991).

Exercises

Exercise 11.1 $t \to 0_+$ **regularization of eigenvalue sums****. *In taking the Laplace transform (11.19) 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. 27.1.1). Regularize the divergent sum in (11.19) following (for example) the prescription of appendix K.6 and assign to such volume term some interesting role in the theory of classical resonance spectra. E-mail the solution to the authors.*

Exercise 11.2 **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 12

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 (11.9), (11.19) and (11.24) 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 27 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 13). We shall now cure this problem by going from trace formulas to determinants. The idea is illustrated by fig. 1.11: Determinants tend to have larger analyticity domains because if $\mathrm{tr}\,\mathcal{L}/(1 - z\mathcal{L}) = -\frac{d}{dz} \ln \det(1 - z\mathcal{L})$ diverges at a particular value of z, then $\det(1 - z\mathcal{L})$ might have an isolated zero there, and a zero of a function is easier to determine 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 $\log\det = \mathrm{tr}\,\log$. Traces of evolution operators can be evaluated as integrals over Dirac delta functions, and in this way the spectra of evolution operators become related to periodic orbits.

12.1 Spectral determinants for maps

The eigenvalues z_k of a linear operator are given by the zeros of the determinant

$$\det(1 - z\mathcal{L}) = \prod_k (1 - z/z_k)\,. \tag{12.1}$$

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

4.1
page 81

$$\ln\det(1 - M) = \mathrm{tr}\,\ln(1 - M) = -\sum_{n=1}^{\infty} \frac{1}{n}\mathrm{tr}\,M^n\,, \tag{12.2}$$

to relate the spectral determinant of an evolution operator for a map to its traces (11.7), that is, periodic orbits:

$$\begin{aligned} \det(1 - z\mathcal{L}) &= \exp\left(-\sum_n^{\infty} \frac{z^n}{n}\mathrm{tr}\,\mathcal{L}^n\right) \\ &= \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{z^{n_p r} e^{r\beta\cdot A_p}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\right)\,. \end{aligned} \tag{12.3}$$

Going the other way, the trace formula (11.9) can be recovered from the spectral determinant by taking a derivative

$$\mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}} = -z\frac{d}{dz}\ln\det(1 - z\mathcal{L})\,. \tag{12.4}$$

fast track:
sect. 12.2, p. 217

Example 12.1 Spectral determinants of transfer operators

For a piecewise-linear map (7.11) with a finite Markov partition, an explicit formula for the spectral determinant follows by substituting the trace formula (11.10) into (12.3):

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty} \left(1 - \frac{t_0}{\Lambda_0^k} - \frac{t_1}{\Lambda_1^k}\right) , \tag{12.5}$$

where $t_s = z/|\Lambda_s|$. The eigenvalues are - as they should be - (11.11), the ones that we already determined from the trace formula (11.9).

The exponential spacing of eigenvalues guarantees that the spectral determinant (12.5) is an entire function. It is this property that will generalize to piecewise smooth flows with finite Markov paritions, and single out spectral determinants rather than the trace formulas or dynamical zeta functions as the tool of choice for evaluation of spectra.

12.2 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. $[\cdots]$ define $l(\gamma)$ to be the minimal period of γ $[\cdots]$ 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 (12.3)

$$\det(s - \mathcal{A}) = \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} \frac{e^{r(\beta \cdot A_p - sT_p)}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\right) , \tag{12.6}$$

and then check that the trace formula (11.19) is the logarithmic derivative of the spectral determinant so defined

$$\operatorname{tr} \frac{1}{s - \mathcal{A}} = \frac{d}{ds} \ln \det(s - \mathcal{A}) . \tag{12.7}$$

To recover $\det(s - \mathcal{A})$ integrate both sides $\int_{s_0}^{s} ds$. With z set to $z = e^{-s}$ as in (11.20), the spectral determinant (12.6) has the same form for both maps and flows. We shall refer to (12.6) as *spectral determinant*, as the spectrum of the operator $\mathcal{A}$ is given by the zeros of

$$\det(s-\mathcal{A}) = 0\,. \tag{12.8}$$

We now note that the r sum in (12.6) 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 (11.4) as geometric series, substituting back into (12.6), and resumming the logarithms, we find that the spectral determinant is formally given by the infinite product

$$\begin{aligned}
\det(s-\mathcal{A}) &= \prod_{k_1=0}^{\infty}\cdots\prod_{l_c=0}^{\infty}\frac{1}{\zeta_{k_1\cdots l_c}} \\
1/\zeta_{k_1\cdots l_c} &= \prod_p\left(1-t_p\frac{\Lambda_{p,e+1}^{l_1}\Lambda_{p,e+2}^{l_2}\cdots\Lambda_{p,d}^{l_c}}{\Lambda_{p,1}^{k_1}\Lambda_{p,2}^{k_2}\cdots\Lambda_{p,e}^{k_e}}\right) && (12.9) \\
t_p &= t_p(z,s,\beta) = \frac{1}{|\Lambda_p|}e^{\beta\cdot A_p - sT_p}z^{n_p}\,. && (12.10)
\end{aligned}$$

Here we have inserted a topological cycle length weigth z^{n_p} for reasons which will become apparent in chapter 15; 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 (8.1), 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 (8.5) 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 11; how do we actually evaluate the averages introduced in sect. 8.1? The eigenvalues of the dynamical averaging evolution operator are given by the values of s for which the spectral determinant (12.6) of the evolution operator (8.19) vanishes. If we can compute the leading eigenvalue $s_0(\beta)$ and its derivatives, we are done. Unfortunately, the infinite product formula (12.9) 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 15, but this point in the narrative is a natural point to introduce a still another variant of a determinant, the dynamical zeta function.

12.3 Dynamical zeta functions

It follows from sect. 11.1.1 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 (12.6) is replaced by the *dynamical zeta function*

$$1/\zeta \;=\; \exp\left(-\sum_p \sum_{r=1}^{\infty} \frac{1}{r} t_p^r\right) \tag{12.11}$$

that we have already derived heuristically in sect. 1.5.2. Resumming the logarithms using $\sum_r t_p^r/r = -\ln(1-t_p)$ we obtain the *Euler product rep.* of the dynamical zeta function:

$$1/\zeta \;=\; \prod_p (1-t_p) \;. \tag{12.12}$$

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 (11.24) plays the same role vis-a-vis the dynamical zeta function

$$\Gamma(s) = \frac{d}{ds}\ln\zeta^{-1} = \sum_p \frac{T_p t_p}{1-t_p}\,, \tag{12.13}$$

as the exact trace formula (11.19) plays vis-a-vis the spectral determinant (12.6), see (12.7). The heuristically derived dynamical zeta function of sect. 1.5.2 now re-emerges as the $1/\zeta_{0\cdots 0}(z)$ part of the *exact* spectral determinant; other factors in the infinite product (12.9) affect the non-leading eigenvalues of $\mathcal{L}$.

To summarize: the dynamical zeta function (12.12) 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 (8.5)), 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

$$1/\zeta(s) = 0 \tag{12.14}$$

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 15, p. 279

12.3.1 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 18): Γ_n, the trace sum (11.22), can be expressed in terms of the dynamical zeta function (12.12)

$$1/\zeta(z) = \prod_p \left(1 - \frac{z^{n_p}}{|\Lambda_p|}\right) \quad . \tag{12.15}$$

as a contour integral

$$\Gamma_n = \frac{1}{2\pi i} \oint_{\gamma_r^-} z^{-n} \left(\frac{d}{dz} \log \zeta^{-1}(z)\right) dz \quad , \tag{12.16}$$

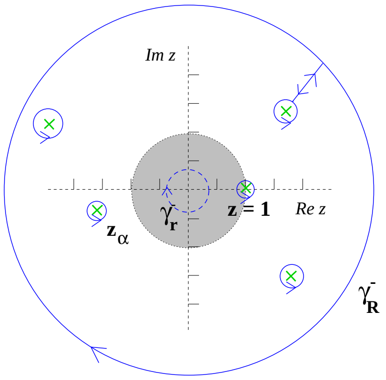


Figure 12.1: The survival probability Γ_n can be split into contributions from poles (x) and zeros (o) between the small and the large circle and a contribution from the large circle.

12.7
page 233

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 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 (11.22) 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. 12.1. Residue calculus turns this into a sum over the zeros z_α and poles z_β of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha|<R}^{\text{zeros}} \frac{1}{z_\alpha^n} - \sum_{|z_\beta|<R}^{\text{poles}} \frac{1}{z_\beta^n} + \frac{1}{2\pi i} \oint_{\gamma_R^-} dz\, z^{-n} \frac{d}{dz} \log \zeta^{-1}, \qquad (12.17)$$

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 $\zeta^{-1}(z)$.

12.3.2 Dynamical zeta functions for transfer operators

Ruelle's original dynamical zeta function was a generalization of the topological zeta function (10.21) that we shall discuss in chapter 10 to a function that assigns different weights to different cycles:

$$\zeta(z) = \exp \sum_{n=1}^{\infty} \frac{z^n}{n} \left(\sum_{x_i \in \mathrm{Fix} f^n} \operatorname{tr} \prod_{j=0}^{n-1} g(f^j(x_i)) \right).$$

11.2
page 213

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 (4.30) the stability of any n-cycle of a 1-d map is given by $\Lambda_p = \prod_{j=1}^{n} f'(x_i)$, so the 1-$d$ map cycle stability is the simplest example of a multiplicative cycle weight $g(x_i) = f'(x_i)$, and indeed - via the Perron-Frobenius evolution operator (7.9) - 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 (9.2). As explained in sect. 10.3, 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 (7.11).

Example 12.2 Piecewise linear repeller

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 (11.24) are of a particularly simple form

$$\mathrm{tr}\, T^n = \Gamma_n = \left(\frac{1}{|\Lambda_0|} + \frac{1}{|\Lambda_1|}\right)^n .$$

The dynamical zeta function (12.11) evaluated by resumming the traces

12.2 page 231

$$1/\zeta(z) = 1 - z/|\Lambda_0| - z/|\Lambda_1| \qquad (12.18)$$

is indeed the determinant $\det(1-zT)$ of the transfer operator (7.13), almost as simple as the topological zeta function (10.25).

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. 9.7).

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 15 how to exploit systematically this redundancy.

12.4 False zeros

Compare (12.18) with the Euler product (12.12). 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 (12.12) vanishes whenever one of its factors vanishes. If that were true, each factor $(1 - z^{n_p}/|\Lambda_p|)$ would yield

$$0 = 1 - e^{n_p(\gamma - \lambda_p)}, \tag{12.19}$$

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 (12.18)

$$0 = 1 - e^{\gamma - \lambda + h}, \qquad h = \ln 2. \tag{12.20}$$

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 (12.12) to the true zero $z = e^{\lambda - h}$.

12.5 More examples of spectral determinants

For expanding 1-d mappings the spectral determinant (12.9) takes form

$$\det(s - \mathcal{A}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k\right), \qquad t_p = \frac{e^{\beta A_p - sT_p}}{|\Lambda_p|} z^{n_p}. \tag{12.21}$$

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 (11.4) is expanded as follows:

$$\frac{1}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} = \frac{1}{|\Lambda|^r (1 - 1/\Lambda_p^r)^2} = \frac{1}{|\Lambda|^r} \sum_{k=0}^{\infty} \frac{k+1}{\Lambda_p^{kr}}. \tag{12.22}$$

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

$$\det\,(s - \mathcal{A}) = \prod_p \prod_{k=0}^{\infty} \left(1 - t_p/\Lambda_p^k\right)^{k+1} \,. \tag{12.23}$$

13.4
page 260

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.

12.5.1 Spectral determinants *vs.* dynamical zeta functions

In sect. 12.3 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

$$1 = \frac{1}{\det\,(1 - \mathbf{J})} \sum_{k=0}^{d} (-1)^k \mathrm{tr}\, \left(\wedge^k \mathbf{J}\right) \,, \tag{12.24}$$

inserted into the exponential representation (12.11) 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 (12.11) yields an expression for the dynamical zeta function for 1-d maps as a ratio of two spectral determinants

$$1/\zeta = \frac{\det\,(1 - \mathcal{L})}{\det\,(1 - \mathcal{L}_{(1)})} \tag{12.25}$$

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 13, 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. 12.3 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

$$1/\zeta = \frac{\det(1-\mathcal{L})\det(1-\mathcal{L}_{(2)})}{\det(1-\mathcal{L}_{(1)})}\,. \tag{12.26}$$

This establishes that for nice hyperbolic flows dynamical zeta function is meromorphic in 2-d.

Example 12.3 Dynamical zeta functions for 2-d Hamiltonian flows

The relation (12.26) 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 (12.11) of $1/\zeta_k$*, and obtain*

$$1/\zeta_k = \frac{F_k F_{k+2}}{F_{k+1}^2}\,. \tag{12.27}$$

Even though we have no guarantee that F_k *are entire, we do know 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 (12.23) 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.*

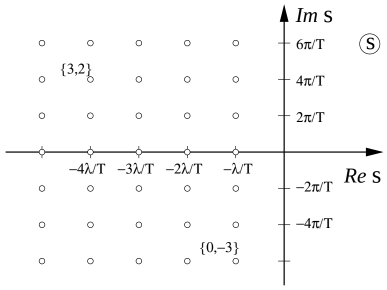


Figure 12.2: The classical resonances $\alpha = \{k, n\}$ for a 2-disk game of pinball, equation (12.28).

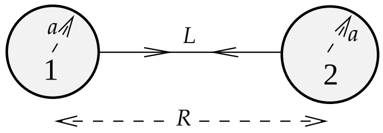


Figure 12.3: A game of pinball consisting of two disks of equal size in a plane, with its only periodic orbit. (A. Wirzba)

12.6 All too many eigenvalues?

What does the 2-dimensional hyperbolic Hamiltonian flow spectral determinant (12.23) tell us? Consider one of the simplest conceivable hyperbolic flows: the game of pinball of fig. 12.3 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 5.5), and the resonances $\det(s_\alpha - \mathcal{A}) = 0$, $\alpha = \{k, n\}$ plotted in fig. 12.2

$$s_\alpha = -(k{+}1)\lambda + n\frac{2\pi i}{T}\,, \quad n \in \mathbb{Z}\,, k \in \mathbb{Z}_+\,, \quad \text{multiplicity } k{+}1\,, \tag{12.28}$$

can be read off the spectral determinant (12.23) for a single unstable cycle:

$$\det(s - \mathcal{A}) = \prod_{k=0}^{\infty} \left(1 - e^{-sT}/|\Lambda|\Lambda^k\right)^{k+1} . \tag{12.29}$$

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 (11.19), the trace formula for flows:

$$\begin{aligned}
\sum_{\alpha=0}^{\infty} e^{s_\alpha t} &= \sum_{k=0}^{\infty}\sum_{n=-\infty}^{\infty}(k+1)e^{(k+1)\lambda t+i2\pi nt/T} \\
&= \sum_{k=0}^{\infty}(k+1)\left(\frac{1}{|\Lambda|\Lambda^k}\right)^{t/T}\sum_{n=-\infty}^{\infty} e^{i2\pi n/T} \\
&= \sum_{k=0}^{\infty}\frac{k+1}{|\Lambda|^r\Lambda^{kr}}\sum_{r=-\infty}^{\infty}\delta(r-t/T) \\
&= T\sum_{r=-\infty}^{\infty}\frac{\delta(t-rT)}{|\Lambda|(1-1/\Lambda^r)^2} \qquad (12.30)
\end{aligned}$$

So the two sides of the trace formula (11.19) 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 (11.7). 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. 15.1.

Commentary

Remark 12.1 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 [1.13]: expanding holomorphic case, see Ruelle [13.9]; piecewise monotone maps of the interval, see Hofbauer and Keller [12.15] and Baladi and Keller [12.18].

Remark 12.2 Smale's wild idea. Smale's wild idea quoted on page 217 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 (12.9) simplify to $s^{-(s+k)T_p}$, as intuited in Smale's wild idea quoted on page 217 (where $l(\gamma)$ is the cycle period denoted here by T_p). Nevertheless, Smale's intuition was remarkably on the target.

Remark 12.3 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. [12.4], a very readable introduction to the Selberg Zeta function.

Remark 12.4 Zeta functions, antecedents. For a function to be deserving of the appellation "zeta function", one expects it to have an Euler product (12.12) representation, and perhaps also satisfy a functional equation. Various kinds of zeta functions are reviewed in refs. [12.8, 12.9, 12.10]. Historical antecedents of the dynamical zeta function are the fixed-point counting functions introduced by Weil [12.11], Lefschetz [12.12] and Artin and Mazur [12.13], and the determinants of transfer operators of statistical mechanics [1.14].

In his review article Smale [1.12] 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 (12.3) between spectral determinants and dynamical zeta functions, but as he was motivated by the Artin-Mazur zeta function (10.21) 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" [9.22, 13.27].

The nomenclature has not settled down yet; what we call evolution operators here is called transfer operators [1.16], Perron-Frobenius operators [12.6] and/or Ruelle-Araki operators elsewhere. Here we refer to kernels such as (8.19) as evolution operators. We follow Ruelle in usage of the term "dynamical zeta function", but elsewhere in the literature function (12.12) is often called the Ruelle zeta function. Ruelle [1.17] 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 [1.4], functional determinants, Fredholm determinants, or even - to maximize confusion - dynamical zeta functions [12.14]. 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 13 and is illustrated, for example, by table 15.2). 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 15) is that the infinite time behavior of an unstable system turns out to be as easy to determine as its short time behavior.

References

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

[12.2] D. Ruelle, *Bull. Amer. Math. Soc.* **78**, 988 (1972)

[12.3] M. Pollicott, *Invent. Math.* **85**, 147 (1986).

[12.4] H.P. McKean, *Comm. Pure and Appl. Math.* **25** , 225 (1972); **27**, 134 (1974).

[12.5] W. Parry and M. Pollicott, *Ann. Math.* **118**, 573 (1983).

[12.6] 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).

[12.7] P. Cvitanović, P.E. Rosenqvist, H.H. Rugh, and G. Vattay, *CHAOS* **3**, 619 (1993).

[12.8] 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.

[12.9] Kiyosi Itô, ed., *Encyclopedic Dictionary of Mathematics*, (MIT Press, Cambridge, 1987).

[12.10] N.E. Hurt, "Zeta functions and periodic orbit theory: A review", *Results in Mathematics* **23**, 55 (Birkhäuser, Basel 1993).

[12.11] A. Weil, *"Numbers of solutions of equations in finite fields"*, *Bull. Am. Math. Soc.* **55**, 497 (1949).

[12.12] D. Fried, "Lefschetz formula for flows", *The Lefschetz centennial conference*, *Contemp. Math.* **58**, 19 (1987).

[12.13] E. Artin and B. Mazur, *Annals. Math.* **81**, 82 (1965)

[12.14] M. Sieber and F. Steiner, *Phys. Lett.* **A 148**, 415 (1990).

[12.15] F. Hofbauer and G. Keller, *"Ergodic properties of invariant measures for piecewise monotonic transformations"*, *Math. Z.* **180**, 119 (1982).

[12.16] G. Keller, *"On the rate of convergence to equilibrium in one-dimensional systems"*, *Comm. Math. Phys.* **96**, 181 (1984).

[12.17] F. Hofbauer and G. Keller, *"Zeta-functions and transfer-operators for piecewise linear transformations"*, *J. reine angew. Math.* **352**, 100 (1984).

[12.18] V. Baladi and G. Keller, *"Zeta functions and transfer operators for piecewise monotone transformations"*, *Comm. Math. Phys.* **127**, 459 (1990).

Exercises

Exercise 12.1 **Escape rate for a 1-d repeller, numerically.** *Consider the quadratic map*

$$f(x) = Ax(1-x) \qquad (12.31)$$

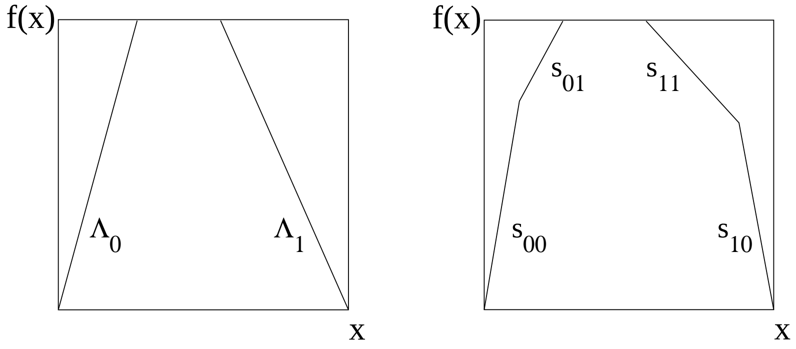
on the unit interval. The trajectory of a point starting in the unit interval either stays in the interval forever or after some iterate leaves the interval and diverges to minus infinity. Estimate numerically the escape rate (16.8), the rate of exponential decay of the measure of points remaining in 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 15.2.

Exercise 12.2 **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 (7.11) 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?*

Exercise 12.3 **Dynamical zeta functions from Markov graphs.** *Extend sect. 10.3 to evaluation of dynamical zeta functions for piecewise linear maps with finite Markov graphs. This generalizes the results of exercise 12.2.*

Exercise 12.4 **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_{root} is a root of the function F, show that the infinite product diverges when evaluated at z_{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)

Exercise 12.5 **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 $\det(1 - z\mathcal{L}_{(s)}) = \prod_{p,k}(1 - z^{n_p}/|\Lambda_p|\Lambda_p^{k+s})$.

Exercise 12.6 **Escape rate for the Ulam map.** *(medium) We will try and compute the escape rate for the Ulam map (14.21)*

$$f(x) = 4x(1 - x),$$

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.

Exercise 12.7 Contour integral for survival probability. *Perform explicitly the contour integral appearing in (12.16).*

Exercise 12.8 Dynamical zeta function for maps. *In this problem we will compare the dynamical zeta function and the spectral determinant. Compute the exact dynamical zeta function for the skew Ulam tent map (7.43)*

$$1/\zeta(z) = \prod_{p \in P} \left(1 - \frac{z^{n_p}}{|\Lambda_p|} \right) .$$

What are its roots? Do they agree with those computed in exercise 7.7?

Exercise 12.9 Dynamical zeta functions for Hamiltonian maps. *Starting from*

$$1/\zeta(s) = \exp \left(- \sum_p \sum_{r=1}^{\infty} \frac{1}{r} t_p^r \right)$$

for a 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.

Exercise 12.10 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 12.12 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?*

Exercise 12.11 **Finite truncations.** *(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 (7.13), 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?

Exercise 12.12 **Pinball escape rate from numerical simulation*** *Estimate the escape rate for $R : a = 6$ 3-disk pinball by shooting 100,000 randomly initiated pinballs into the 3-disk system and plotting the logarithm of the number of trapped orbits as function of time. For comparison, a numerical simulation of ref. [3.9] yields $\gamma = .410\ldots$.*

Chapter 13

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 11 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. 13.1 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 12. The program is enunciated in sect. 13.2, with the focus on expanding maps. In sect. 13.3 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 spectral determinants or dynamical zeta functions (see remark 13.1), 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 16, p. 305

13.1 The simplest of spectral determinants: A single fixed point

In order to get some feeling for the determinants defined so formally in sect. 12.2, 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 (7.10) is

$$\mathcal{L}\phi(y) = \int dx\, \delta(y - \Lambda x)\, \phi(x) = \frac{1}{|\Lambda|}\phi(y/\Lambda)\,. \tag{13.1}$$

From this one immediately gets that the monomials y^n are eigenfunctions:

$$\mathcal{L}y^n = \frac{1}{|\Lambda|\Lambda^n} y^n \,, \qquad n = 0, 1, 2, \ldots \tag{13.2}$$

We note that the eigenvalues Λ^{-n-1} fall off exponentially with n, and that the trace of $\mathcal{L}$ is

$$\mathrm{tr}\,\mathcal{L} = \frac{1}{|\Lambda|}\sum_{n=0}^{\infty} \Lambda^{-n} = \frac{1}{|\Lambda|(1 - \Lambda^{-1})} = \frac{1}{|f(0)' - 1|}\,,$$

in agreement with (11.6). A similar result is easily obtained for powers of $\mathcal{L}$, and for the spectral determinant (12.3) one obtains:

$$\det(1 - z\mathcal{L}) = \prod_{k=0}^{\infty}\left(1 - \frac{z}{|\Lambda|\Lambda^k}\right) = \sum_{k=0}^{\infty} Q_k t^k\,, \qquad t = -z/|\Lambda|\,, \tag{13.3}$$

where the coefficients Q_k are given explicitly by the *Euler formula*

13.3
page 260

$$Q_k = \frac{1}{1-\Lambda^{-1}}\,\frac{\Lambda^{-1}}{1-\Lambda^{-2}}\cdots\frac{\Lambda^{-k+1}}{1-\Lambda^{-k}} \quad . \tag{13.4}$$

(if you cannot figure out exercise 13.3 check the solutions on 771 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. 13.3.1, 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 (13.3) is *entire* in the complex z plane.

What is the meaning of (13.3)? It gives us an interpretation of the index k in the Selberg product representation of the spectral determinant (12.9): k labels the kth local fixed-point eigenvalue $1/|\Lambda|\Lambda^k$.

Now if the spectral determinant is entire, on the basis of (12.25) 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 15 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

$$h(z) \;= \sum_{k=0}^{\infty} \sigma_k z^k$$

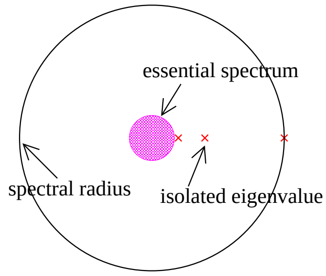


Figure 13.1: Spectrum for Perron-Frobenius operator in an extended function space: only a few isolated eigenvalues remain between the spectral radius and the essential spectral radius, bounding continuous spectrum

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

$$h_N(z) = \sum_{k=0}^{N} \sigma_k z^k = \frac{a}{b} + \frac{z(a-b)(1-z^N/b^N)}{b^2(1-z/b)}.$$

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

13.1 page 260

- 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

$$\mathcal{L}y^\eta = \frac{1}{|\Lambda|\Lambda^\eta} y^\eta$$

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. 13.1.

In sect. 13.2 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.

13.2
page 260

- A deceptively innocent assumption hides behind many features discussed so far: that (13.1) 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. 13.3, 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.

13.2 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.

13.2.1 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

$$\mathcal{K}\varphi(x) \;=\; \int_{\mathcal{Q}} dy\, \mathcal{K}(x,y)\varphi(y)\,, \tag{13.5}$$

and one is tempted to resort to the classical Fredholm theory in order to establish analyticity properties of spectral determinants. This path to enlightenment is blocked by the singular nature of the kernel, which is a distribution, whereas the standard theory of integral equations usually concerns itself with regular kernels $\mathcal{K}(x,y) \in L^2(\mathcal{Q}^2)$. Here we briefly recall some steps of the Fredholm theory, before going to our major example in sect. 13.2.2.

The general form of Fredholm integral equations of the second kind is

$$\varphi(x) \;=\; \int_{\mathcal{Q}} dy\, \mathcal{K}(x,y)\varphi(y) \;+\; \xi(x) \tag{13.6}$$

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 (13.5), 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 K). 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. 13.1) that much work is needed to bring Perron-Frobenius operators into this picture.

We rewrite (13.6) in the form

$$\mathcal{T}\varphi \;=\; \xi\,, \mathcal{T} \;=\; \mathbb{1} - \mathcal{K}\,. \tag{13.7}$$

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

$$(\mathbb{1} - \lambda\mathcal{K})^{-1} = \mathbb{1} + \lambda\mathcal{K} + \lambda^2\mathcal{K}^2 + \cdots = \mathbb{1} + \lambda\mathcal{W}, \tag{13.8}$$

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 (13.8) 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

$$\mathcal{K}\begin{pmatrix} x_1 \ldots x_n \\ y_1 \ldots y_n \end{pmatrix} = \begin{vmatrix} \mathcal{K}(x_1,y_1) & \ldots & \mathcal{K}(x_1,y_n) \\ \ldots & \ldots & \ldots \\ \mathcal{K}(x_n,y_1) & \ldots & \mathcal{K}(x_n,y_n) \end{vmatrix}$$

we may write the explicit expressions

$$\begin{aligned} D(\lambda) &= 1 + \sum_{n=1}^{\infty} (-1)^n \frac{\lambda^n}{n!} \int_{\mathcal{Q}^n} dz_1 \ldots dz_n\, \mathcal{K}\begin{pmatrix} z_1 \ldots z_n \\ z_1 \ldots z_n \end{pmatrix} \\ &= \exp - \sum_{m=1}^{\infty} \frac{\lambda^m}{m} \operatorname{tr} \mathcal{K}^m \end{aligned} \tag{13.9}$$

and

$$\mathcal{D}(x,y;\lambda) = \mathcal{K}\begin{pmatrix} x \\ y \end{pmatrix} + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} \int_{\mathcal{Q}^n} dz_1 \ldots dz_n\, \mathcal{K}\begin{pmatrix} x & z_1 & \ldots & z_n \\ y & z_1 & \ldots & z_n \end{pmatrix}$$

$D(\lambda)$ is known as the Fredholm determinant (see (12.24) and appendix K): 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.

13.2.2 Bernoulli shift

Consider now the Bernoulli shift

$$x \mapsto 2x \bmod 1 \qquad\qquad x \in [0,1] \tag{13.10}$$

and look at spectral properties in appropriate function spaces. The Perron-Frobenius operator associated with this map is given by

$$\mathcal{L}h(y) = \frac{1}{2}h\left(\frac{y}{2}\right) + \frac{1}{2}h\left(\frac{y+1}{2}\right) . \tag{13.11}$$

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

$$h_\theta(y) = \sum_{k \neq 0} \exp(2\pi i k y) \frac{1}{|k|^\theta} \tag{13.12}$$

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. 13.1).

13.5 page 260

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. 13.1. 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 (13.11) 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 13.3.

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 (13.10) 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 $y \in \mathrm{Cl}\, D$.

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:

$$\mathrm{tr}\, \mathcal{L}_F = \oint \frac{sF'(w)}{w - F(w)} dw.$$

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

13.6
page 260

$$\mathrm{tr}\, \mathcal{L}_F = \frac{sF'(x^*)}{1 - F'(x^*)} = \frac{1}{|f'(x^*) - 1|} ,$$

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 27). 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 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. H.1.

Property 2 is violated by the 1-d tent map (see fig. 13.2)

$$f(x) = \alpha(1 \ - \ |1-2x|) \ , \quad 1/2 < \alpha < 1 \ .$$

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. In practice this means that while the leading eigenvalue of $\mathcal{L}$ might be computable, the rest of the

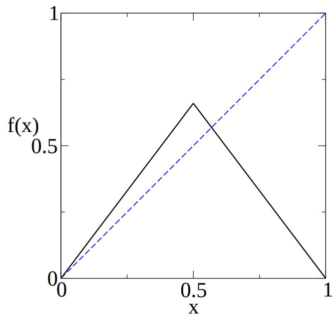


Figure 13.2: A (hyperbolic) tent map without a finite Markov partition.

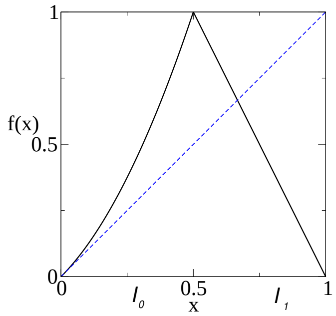


Figure 13.3: A Markov map with a marginal fixed point.

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. 13.3)

$$f(x) = \begin{cases} x + 2x^2 & , \quad x \in I_0 = [0, \frac{1}{2}] \\ 2 - 2x & , \quad x \in I_1 = [\frac{1}{2}, 1] \end{cases} .$$

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 18.

The property 4 is required as the heuristic approach of chapter 11 faces two major hurdles:

1. The trace (11.7) is not well defined since the integral kernel is singular.

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:

$$\det(1 - z\mathcal{L}) = \exp\left(-\sum_{n=1} \frac{z^n}{n} \frac{2^n - 1}{2^n - 1}\right) = 1 - z. \tag{13.13}$$

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.

13.3 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 ($0 < \lambda_s < 1$, $\Lambda_u > 1$):

$$f(z) = (f_1(z_1, z_2), f_2(z_1, z_2)) = (\lambda_s z_1, \Lambda_u z_2) \tag{13.14}$$

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:

$$\mathcal{L}h(z_1, z_2) = \frac{1}{\lambda_s \cdot \Lambda_u} h(z_1/\lambda_s, z_2/\Lambda_u) \tag{13.15}$$

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):

$$\mathcal{L}h(z_1, z_2) = \oint \oint \frac{s\, h(w_1, w_2)}{(z_1 - f_1(w_1, w_2)(f_2(w_1, w_2) - z_2)} \frac{dw_1}{2\pi i} \frac{dw_2}{2\pi i}. \tag{13.16}$$

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 (13.15). 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 13.11) 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 $f(\mathcal{M}_i) \cap \mathrm{Int}(\mathcal{M}_j) = \emptyset$, or for each pair $w_h \in \mathrm{Cl}(D_i^h)$, $z_v \in \mathrm{Cl}(D_j^v)$ there exist unique analytic functions of w_h, z_v: $w_v = w_v(w_h, z_v) \in \mathrm{Int}(D_i^v)$, $z_h = z_h(w_h, z_v) \in \mathrm{Int}(D_j^h)$, such that $f(w_h, w_v) = (z_h, z_v)$. Furthermore, if $w_h \in I_i^h$ and $z_v \in I_j^v$, then $w_v \in I_i^v$ and $z_h \in I_j^h$ (see fig. 13.4).

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 (13.16) 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

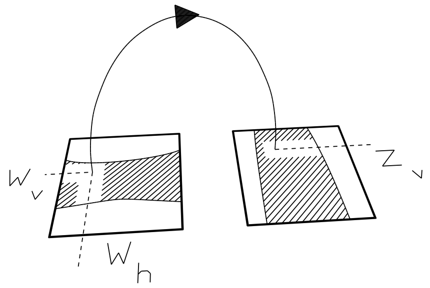


Figure 13.4: For an analytic hyperbolic map, specifying the contracting coordinate w_h at the initial rectangle and the expanding coordinate z_v at the image rectangle defines a unique trajectory between the two rectangles. In particular, w_v and z_h (not shown) are uniquely specified.

and the theorem states that the associated spectral determinant is entire, and that the trace formula (11.7) is correct.

13.3.1 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 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. 13.1 we start with a map with a single fixed point, but this time with a nonlinear map f with a nonlinear inverse $F = f^{-1}$, $s = \mathrm{sgn}(F')$

$$\mathcal{L}\circ\phi(z) = \int dx\ \delta(z - f(x))\,\phi(x) = s\ F'(z)\ \phi(F(z))\,.$$

Assume that F is a contraction of the unit disk, that is

$$|F(z)| < \theta < 1 \quad \text{and} \quad |F'(z)| < C < \infty \quad \text{for} \quad |z| < 1\,, \tag{13.17}$$

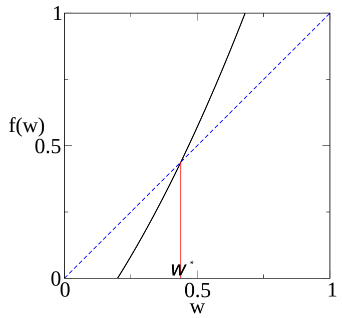


Figure 13.5: A nonlinear one-branch repeller with a single fixed point w^*.

and expand ϕ in a polynomial basis by means of the Cauchy formula

$$\phi(z) = \sum_{n\geq 0} z^n \phi_n = \oint \frac{dw}{2\pi i} \, \frac{\phi(w)}{w-z} \,, \qquad \phi_n = \oint \frac{dw}{2\pi i} \, \frac{\phi(w)}{w^{n+1}}$$

In this basis, $\mathcal{L}$ is a represented by the matrix

$$\mathcal{L} \circ \phi(w) = \sum_{m,n} w^m L_{mn} \phi_n \,, \qquad L_{mn} = \oint \frac{dw}{2\pi i} \, \frac{s\, F'(w)(F(w))^n}{w^{m+1}} \,. \tag{13.18}$$

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

$$\mathrm{tr}\ \mathcal{L} = \frac{s\, F'(w^*)}{1 - F'(w^*)} = \frac{1}{|f'(w^*) - 1|} \,.$$

13.6
page 260

We recognize this result as a generalization of the single piecewise-linear fixed-point example (13.2), $\phi_n = y^n$, 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 (13.17) into the contour integral (13.18) 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 $1/\Lambda^{k^{1/d}}$.

13.4 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 8) 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 7) 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 $\lambda_1, \lambda_2 \ldots \lambda_M$ 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

$$\mathcal{L}\varphi = \sum_{i=1}^{M} \lambda_i \psi_i L_i \psi_i^* \varphi + \mathbf{P}\mathcal{L}\varphi \tag{13.19}$$

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 (13.19) becomes

$$\mathcal{L}^n\varphi = \sum_{i=1}^{M} \lambda_i^n \psi_i L_i^n \psi_i^* \varphi + \mathbf{P}\mathcal{L}^n\varphi \tag{13.20}$$

If we now consider expressions like

$$C(n)_{\xi,\varphi} = \int_{\mathcal{M}} dy\, \xi(y)\, (\mathcal{L}^n\varphi)\,(y) = \int_{\mathcal{M}} dw\, (\xi \circ f^n)(w)\varphi(w) \tag{13.21}$$

we have

$$C(n)_{\xi,\varphi} = \lambda_1^n \omega_1(\xi,\varphi) + \sum_{i=2}^{L} \lambda_i^n \omega(n)_i(\xi,\varphi) + \mathcal{O}(\theta^n) \tag{13.22}$$

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 (13.21) defines a correlation function: as each term in (13.22) 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). 13.7 page 260

So let us come back the Bernoulli shift example (13.10), 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 (13.22) we have that the asymptotic decay of correlation function is (unless also $\omega_1(\eta, \eta) = 0$)

$$C_{\eta,\eta}(n) \sim \exp(-n \log 2) \tag{13.23}$$

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

$$\eta(z) = \int_0^1 dw\, \eta(w) + \sum_{m=1}^{\infty} \frac{\eta^{(m-1)}(1) - \eta^{(m-1)}(0)}{m!} B_m(z)\,. \tag{13.24}$$

As we are considering zero–average functions, we have from (13.21), 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 (13.24) 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

$$\Psi_i(w) = \frac{(-1)^{i-1}}{i!} \left(\delta^{(i-1)}(w-1) - \delta^{(i-1)}(w) \right) \tag{13.25}$$

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.

Commentary

Remark 13.1 Surveys of rigorous theory. For a physicist, Driebe's monograph [1.22] might be the most accessible introduction into main theories touched upon in this chapter. We recommend references listed in sect. 1.8 for an introduction into the mathematical 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. [13.1, 13.2]. An alternative approach to spectral properties of the Perron-Frobenius operator is illustrated in ref. [13.3].

The ergodic theory, as presented by Sinai [13.15] 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 [13.16]; more advanced and more old fashioned presentations are Walters [13.17] and Denker, Grillenberger and Sigmund [13.18]; and a more formal Peterson [13.19].

Remark 13.2 Fredholm theory. Our brief summary of Fredholm theory is based on the exposition in ref. [13.4]. A technical introduction of the theory from an operatorial point of view is contained in ref. [13.5]. The theory has been generalized in ref. [13.6].

Remark 13.3 Bernoulli shift. For a more detailed discussion, consult chapter 17.1 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 [13.7]: a concise and readable statement of the results is contained in ref. [13.8].

Remark 13.4 Higher dimensions and generalized Fredholm theory. When extending Bernoulli shift to higher dimensions. Extensions of Fredholm theory [13.6], which avoid problems with multi-dimensional residue calculus, may be used: see ref. [13.9].

Remark 13.5 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. [13.3].

Remark 13.6 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 (12.23). Proofs are getting too hard for the purposes of this chapter; details are discussed in ref.(?).

Remark 13.7 Examples. Examples of analytic hyperbolic maps are provided by small analytic perturbations of the cat map (where the Markov partitioning is non-trivial [13.10]), the 3-disk repeller, and the 2-d baker's map.

Remark 13.8 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. [13.11, 13.12]). 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 13.9 Perron-Frobenius theorem. A proof of the Perron-Frobenius theorem may be found in ref. [13.13]. For positive transfer operators such theorem has been generalized by Ruelle [13.14].

Remark 13.10 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 [13.20] for the spectral determinants of d-dimensional expanding flows.

Remark 13.11 Axiom A systems. Proofs outlined in sect. 13.3 follow the thesis work of H.H. Rugh [13.9, 13.20, 13.21]. For mathematical introduction to the subject, consult the excellent review by V. Baladi [13.1]. Rigorous treatment is given in refs. [13.9, 13.20, 13.21]. It would take us too far to give and explain the definition of the Axiom A systems (see refs. [13.22, 13.23]). 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 13.12 Exponential mixing speed of the Bernoulli shift. We see from (13.23) 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 13.13 Left eigenfunctions. We shall never use explicit form of left eigenfunctions, corresponding to highly singular kernels like (13.25). Many details have been elaborated in a number of papers, like ref. [13.24], with a daring physical interpretation.

Remark 13.14 Ulam's idea. The approximation of Perron-Frobenius operator defined by (8.33) has been shown to reproduce correctly the spectrum for expanding maps, once finer and finer Markov partitions are used [13.25]. The subtle point of choosing a phase space partitioning for a "generic case" is discussed in ref. [13.26].

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

[13.1] 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).

[13.2] 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.*

[13.3] M. Viana, *Stochastic dynamics of deterministic systems*, (Col. Bras. de Matemática, Rio de Janeiro,1997)

[13.4] A.N. Kolmogorov and S.V. Fomin, *Elements of the theory of functions and functional analysis* (Dover,1999).

[13.5] R.G. Douglas, *Banach algebra techniques in operator theory* (Springer, New York,1998).

[13.6] A. Grothendieck, *La théorie de Fredholm*, *Bull. Soc. Math. France* **84**, 319 (1956).

[13.7] D. Ruelle, *Inst. Hautes Études Sci. Publ. Math.* **72**, 175-193 (1990).

[13.8] 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)

[13.9] D. Ruelle, *Inv. Math.* **34**, 231-242 (1976).

[13.10] R.L. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1987).

[13.11] F. Christiansen, P. Cvitanović and H.H. Rugh, *J. Phys* **A 23**, L713 (1990).

[13.12] D. Alonso, D. MacKernan, P. Gaspard and G. Nicolis, *Phys. Rev.* **E54**, 2474 (1996).

[13.13] P. Walters, *An introduction to ergodic theory.* (Springer, New York 1982).

[13.14] D. Ruelle, *Commun. Math. Phys.* **9**, 267 (1968).

[13.15] Ya.G. Sinai, *Topics in ergodic theory.* (Princeton Univ. Press, Princeton 1994).

[13.16] I. Kornfeld, S. Fomin and Ya. Sinai, *Ergodic Theory* (Springer, 1982).

[13.17] P. Walters, *An introduction to ergodic theory*, Springer Graduate Texts in Math. Vol **79** (Springer, New York, 1982).

[13.18] M. Denker, C. Grillenberger and K. Sigmund, *Ergodic theory on compact spaces*, (Springer Lecture Notes in Math. **470**, 1975).

[13.19] K. Peterson, *Ergodic theory* (Cambridge Univ. Press, Cambridge 1983).

[13.20] D. Fried, *Ann. Scient. Éc. Norm. Sup.* **19**, 491 (1986).

[13.21] H.H. Rugh, *Nonlinearity* **5**, 1237 (1992).

[13.22] S. Smale, *Bull. Amer. Math. Soc.* **73**, 747 (1967).

[13.23] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, *Springer Lect. Notes in Math.* **470**, 1975.

[13.24] H.H. Hasegawa and W.C. Saphir, *Phys. Rev.* **A46**, 7401 (1992).

[13.25] G. Froyland, *Commun. Math. Phys.* **189**, 237 (1997)

[13.26] G. Froyland, *Extracting dynamical behaviour via markov models*, in A. Mees (ed.) *Nonlinear dynamics and statistics: Proceedings Newton Institute, Cambridge 1998* (Birkhauser,2000).

[13.27] V. Baladi, A. Kitaev, D. Ruelle, and S. Semmes, *"Sharp determinants and kneading operators for holomorphic maps"*, IHES preprint (1995).

[13.28] A. Zygmund, *Trigonometric series* (Cambridge Univ. Press, Cambridge 1959).

Exercises

Exercise 13.1 **What space does $\mathcal{L}$ act on?** *Show that (13.2) is a complete basis on the space of analytic functions on a disk (and thus that we found the* complete *set of eigenvalues).*

Exercise 13.2 **What space does $\mathcal{L}$ act on?** *What can be said about the spectrum of (13.1) on $L^1[0,1]$? Compare the result with fig. 13.1.*

Exercise 13.3 **Euler formula.** *Derive the Euler formula (13.4)*

$$\begin{aligned}\prod_{k=0}^{\infty}(1+tu^k) &= 1+\frac{t}{1-u}+\frac{t^2u}{(1-u)(1-u^2)}+\frac{t^3u^3}{(1-u)(1-u^2)(1-u^3)}\cdots \\ &= \sum_{k=0}^{\infty} t^k \frac{u^{\frac{k(k-1)}{2}}}{(1-u)\cdots(1-u^k)}\,, \qquad |u|<1. \end{aligned} \tag{13.26}$$

Exercise 13.4 **2-d product expansion**.** *We conjecture that the expansion corresponding to (13.26) is in this case*

$$\begin{aligned}\prod_{k=0}^{\infty}(1+tu^k)^{k+1} &= \sum_{k=0}^{\infty}\frac{F_k(u)}{(1-u)^2(1-u^2)^2\cdots(1-u^k)^2}t^k \\ &= 1+\frac{1}{(1-u)^2}t+\frac{2u}{(1-u)^2(1-u^2)^2}t^2 \\ &\quad +\frac{u^2(1+4u+u^2)}{(1-u)^2(1-u^2)^2(1-u^3)^2}t^3+\cdots \end{aligned} \tag{13.27}$$

$F_k(u)$ 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 13.3).

Exercise 13.5 **Bernoulli shift on L spaces.** *Check that the family (13.12) belongs to $L^1([0,1])$. What can be said about the essential spectral radius on $L^2([0,1])$? A useful reference is [13.28].*

Exercise 13.6 **Cauchy integrals.** *Rework all complex analysis steps used in the Bernoulli shift example on analytic functions on a disk.*

Exercise 13.7 **Escape rate.** *Consider the escape rate from a strange repeller: find a choice of trial functions ξ and φ such that (13.21) 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 14

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*

$$f^{t+T}(x) = f^t(x) , \qquad T > 0 \tag{14.1}$$

for a given flow or mapping.

We know from chapter 11 that cycles are the necessary ingredient for evaluation of spectra of evolution operators. In chapter 9 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. A serious cyclist might want to also learn about the variational methods to find cycles, chapter 25. They are particularly useful when little is understood about the topology of a flow, such as in high-dimensional periodic orbit searches. ☞ chapter 25

fast track:
chapter 15, p. 279

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_p for which (14.1) has a solution. A cycle point of a flow f^t which crosses a Poincaré section n_p times is a fixed point of the P^{n_p} iterate of the Poincaré section return map P, hence we shall refer to all cycles as "fixed points" in this chapter. By cyclic invariance, stability eigenvalues and the period of the cycle are independent of the choice of the initial point, so it will suffice to solve (14.1) at a single cycle point.

☞ sect. 4.5.2

If the cycle is an attracting limit cycle with a sizable basin of attraction, it can be found by integrating the flow for sufficiently long time. If the cycle is unstable, simple integration forward in time will not reveal it, and methods to be described here need to be deployed. In essence, any method for finding a cycle is based on devising a new dynamical system which possesses the same cycle, but for which this cycle is attractive. Beyond that, there is a great freedom in constructing such systems, and many different methods are used in practice.

Due to the exponential divergence of nearby trajectories in chaotic dynamical systems, fixed point searches based on direct solution of the fixed-point condition (14.1) as an initial value problem can be numerically very unstable. Methods that start with initial guesses for a number of points along the cycle, such as the multipoint shooting method described here in sect. 14.3, and the variational) methods of chapter 25, are considerably more robust and safer.

☞ chapter 25

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 9.2. The relations between the temporal symbol sequences and the spatial layout of the topologically distinct regions of the phase space discussed in chapter 9 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.

14.1 Where are the cycles?

Ergodic exploration of recurrences that we turn to now sometimes performs admirably well in getting us started.

Both in the example of the Rössler flow and of the Kuramoto-Sivashinsky system we sketched the attractors by running a long chaotic trajectory, and noted that the attractors are very thin, but otherwise the return maps that we plotted were disquieting – neither fig. 3.2 nor fig. 3.3 appeared to be 1-to-1 maps. In this section we show how to use such information to approximately locate cycles. In the remainder of this chapter and in chapter 25 we shall learn how to turn such guesses into highly accurate cycles.

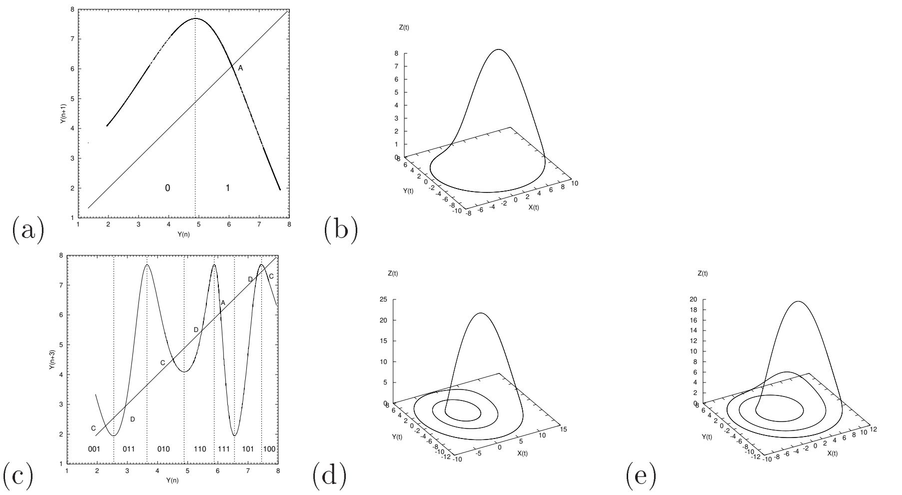


Figure 14.1: (a) $y \to P_1(y, z)$ return map for $x = 0$, $y > 0$ Poincaré section of the Rössler flow fig. 2.3. (b) The $\overline{1}$-cycle found by taking the fixed point $y_{k+n} = y_k$ together with the fixed point of the $z \to z$ return map (not shown) an initial guess $(0, y^(0), z^(0))$ for the Newton-Raphson search. (c) $y_{k+3} = P_1^3(x_k, y_k)$, the third iterate of Poincaré return map (3.1) together with the corresponding plot for $z_{k+3} = P_2^3(x_k, y_k)$, is used to pick starting guesses for the Newton-Raphson searches for the two 3-cycles: (d) the $\overline{001}$ cycle, and (e) the $\overline{011}$ cycle. (G. Simon)

Example 14.1 Rössler attractor *(G. Simon and P. Cvitanović)*

Run a long simulation of the Rössler flow f^t, plot a Poincaré section, as in fig. 3.1, and extract the corresponding Poincaré return map P, as in fig. 3.2. Luck is with us; the fig. 14.1(a) return map $y \to P_1(y, z)$ looks much like a parabola, so we take the unimodal map symbolic dynamics, sect. 9.4, as our guess for the covering dynamics. Strictly speaking, the attractor is "fractal", but for all practical purposes the return map is 1-dimensional; your printer will need a resolution better than 10^{14} dots per inch to start resolving its structure.

Periodic points of a prime cycle p of cycle length n_p for the $x = 0$, $y > 0$ Poincaré section of the Rössler flow fig. 2.3 are fixed points $(y, z) = Pn(y, z)$ of the nth Poincaré return map.

Using the fixed point $y_{k+1} = y_k$ in fig. 14.1(a) together with the simultaneous fixed point of the $z \to P_1(y, z)$ return map (not shown) as a starting guess $(0, y^{(0)}, z^{(0)})$ for the Newton-Raphson search for the cycle p with symbolic dynamics label $\overline{1}$, we find the cycle fig. 14.1(b) with the Poincaré section point $(0, y_p, z_p)$, period T_p, expanding, marginal, contracting stability eigenvalues $(\Lambda_{p,e}, \Lambda_{p,m}, \Lambda_{p,c})$, and Lyapunov exponents $(\lambda_{p,e}, \lambda_{p,m}, \lambda_{p,c})$:

$$
\begin{array}{lrcl}
\overline{1}\text{-cycle:} & (x, y, z) & = & (0, 6.09176832, 1.2997319) \\
 & T_1 & = & 5.88108845586 \\
 & (\Lambda_{1,e}, \Lambda_{1,m}, \Lambda_{1,c}) & = & (-2.40395353, 1 + 10^{-14}, -1.29 \times 10^{-14}) \\
 & (\lambda_{1,e}, \lambda_{1,m}, \lambda_{1,c}) & = & (0.149141556, 10^{-14}, -5.44) \,.
\end{array}
\tag{14.2}
$$

The Newton-Raphson method that we used is described in sect. 14.5.

14.7
page 276

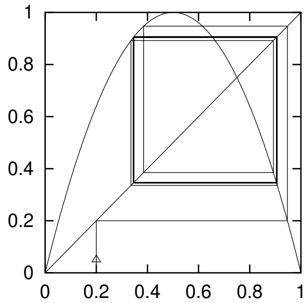


Figure 14.2: The inverse time path to the $\overline{01}$-cycle of the logistic map f(x)=4x(1-x) from an initial guess of x=0.2. At each inverse iteration we chose the 0, respectively 1 branch.

As an example of a search for longer cycles, we use $y_{k+3} = P^3(y_k, z_k)$, the third iterate of Poincaré return map (3.1) plotted in fig. 14.1(c), together with a corresponding plot for $z_{k+3} = f^3(y_k, z_k)$, to pick starting guesses for the Newton-Raphson searches for the two 3-cycles plotted in fig. 14.1(d), (e). For a listing of the short cycles of the Rössler flow, consult table 14.1.

The numerical evidence suggests (but a proof is lacking) that all cycles that comprise the strange attractor of the Rössler system are hyperbolic, each with an expanding eigenvalue $|\Lambda_e| > 1$, a contracting eigenvalue $|\Lambda_c| < 1$, and a marginal eigenvalue $|\Lambda_m| = 1$ corresponding to displacements along the direction of the flow.

For the Rössler system the contracting eigenvalues turn out to be insanely contracting, a factor of e^{-32} per one par-course of the attractor, so their numerical determination is quite difficult. Fortunately, they are irrelevant; for all practical purposes the strange attractor of the Rössler system is 1-dimensional, a very good realization of a horseshoe template.

14.2 One-dimensional mappings

14.2.1 Inverse iteration

Let us first consider a very simple method to find unstable cycles of a 1-dimensional map such as the logistic map. Unstable cycles of 1-d maps are attracting cycles of the inverse map. The inverse map is not single valued, so at each backward iteration we have a choice of branch to make. By choosing branch according to the symbolic dynamics of the cycle we are trying to find, we will automatically converge to the desired cycle. The rate of convergence is given by the stability of the cycle, i.e. the convergence is exponentially fast. Fig. 14.2 shows such path to the $\overline{01}$-cycle of the logistic map.

14.11
page 278

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 14.11. 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

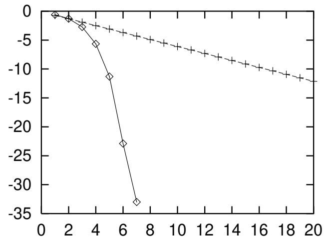


Figure 14.3: Convergence of Newton's method ($\Diamond$) vs. inverse iteration ($+$). The error after n iterations searching for the $\overline{01}$-cycle of the logistic map $f(x) = 4x(1-x)$ with an initial starting guess of $x_1 = 0.2, x_2 = 0.8$. y-axis is $\log_{10}$ of the error. The difference between the exponential convergence of the inverse iteration method and the super-exponential convergence of Newton's method is dramatic.

still be left with both stable and unstable directions. The best strategy is to directly attack the problem of finding solutions of $f^T(x) = x$.

14.2.2 Newton's method

Newton's method for determining a zero x^* of a function $F(x)$ of one variable is based on a linearization around a starting guess x_0:

$$F(x) \approx F(x_0) + F'(x_0)(x - x_0). \tag{14.3}$$

An approximate solution x_1 of $F(x) = 0$ is

$$x_1 = x_0 - F(x_0)/F'(x_0). \tag{14.4}$$

The approximate solution can then be used as a new starting guess in an iterative process. A fixed point of a map f is a solution to $F(x) = x - f(x) = 0$. We determine x by iterating

$$\begin{aligned} x_m &= g(x_{m-1}) = x_{m-1} - F(x_{m-1})/F'(x_{m-1}) \\ &= x_{m-1} - \frac{1}{1 - f'(x_{m-1})}(x_{m-1} - f(x_{m-1}))\,. \end{aligned} \tag{14.5}$$

Provided that the fixed point is not marginally stable, $f'(x) \neq 1$ at the fixed point x, a fixed point of f is a super-stable fixed point of the Newton-Raphson map g, $g'(x) = 0$, and with a sufficiently good initial guess, the Newton-Raphson iteration will converge super-exponentially fast.

To illustrate the efficiency of the Newton's method we compare it to the inverse iteration method in fig. 14.3. Newton's method wins hands down: the number of significant digits of the accuracy of x estimate doubles with each iteration.

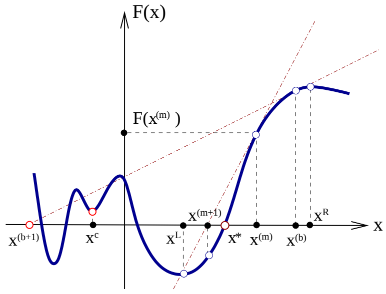


Figure 14.4: Newton method: bad initial guess $x^{(b)}$ leads to the Newton estimate $x^{(b+1)}$ far away from the desired zero of $F(x)$. Sequence $\cdots$, $x^{(m)}$, $x^{(m+1)}$, $\cdots$, starting with a good guess converges super-exponentially to x^*. The method diverges if it iterates into the basin of attraction of a local minimum x^c.

In order to avoid jumping too far from the desired x^* (see fig. 14.4), one often initiates the search by the *damped Newton method*,

$$\Delta x_m = x_{m+1} - x_m = -\frac{F(x_m)}{F'(x_m)}\,\Delta\tau\,, \qquad 0 < \Delta\tau \leq 1\,,$$

takes small $\Delta\tau$ steps at the beginning, reinstating to the full $\Delta\tau = 1$ jumps only when sufficiently close to the desired x^*.

14.3 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 9 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

$$\frac{d}{dx}F(x)(x'-x) = -F(x), \tag{14.6}$$

where $\frac{d}{dx}F(x)$ is an $[n \times n]$ matrix:

$$\frac{d}{dx}F(x) \quad = \quad \left(\begin{array}{cccccc} 1 & & & & & -f'(x_n) \\ -f'(x_1) & 1 & & & & \\ & \cdots & 1 & & & \\ & & \cdots & 1 & & \\ & & & -f'(x_{n-1}) & 1 & \end{array}\right). \tag{14.7}$$

This matrix can easily be inverted numerically by first eliminating the elements below the diagonal. This creates non-zero elements in the nth 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:

$$\left(\begin{array}{ccc} 1 & 0 & -f'(x_3) \\ -f'(x_1) & 1 & 0 \\ 0 & -f'(x_2) & 1 \end{array}\right)\left(\begin{array}{c} \delta_1 \\ \delta_2 \\ \delta_3 \end{array}\right) = \left(\begin{array}{c} -F_1 \\ -F_2 \\ -F_3 \end{array}\right), \tag{14.8}$$

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

$$\begin{array}{c} \left(\begin{array}{ccc} 1 & 0 & -f'(x_3) \\ 0 & 1 & -f'(x_1)f'(x_3) \\ 0 & 0 & 1-f'(x_2)f'(x_1)f'(x_3) \end{array}\right)\left(\begin{array}{c} \delta_1 \\ \delta_2 \\ \delta_3 \end{array}\right) = \\ \left(\begin{array}{c} -F_1 \\ -F_2 - f'(x_1)F_1 \\ -F_3 - f'(x_2)F_2 - f'(x_2)f'(x_1)F_1 \end{array}\right) \end{array}. \tag{14.9}$$

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

$$\begin{pmatrix} 1 & 0 & -f'(x_3) \\ 0 & 1 & -f'(x_1)f'(x_3) \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 - f'(x_1)F_1 \\ \frac{-F_3 - f'(x_2)F_2 - f'(x_2)f'(x_1)F_1}{1 - f'(x_2)f'(x_1)f'(x_3)} \end{pmatrix} . \tag{14.10}$$

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.

14.1
page 276

14.4 d-dimensional mappings

(F. Christiansen)

Armed with symbolic dynamics informed initial guesses we can utilize the Newton-Raphson iteration in d-dimensions as well.

14.4.1 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 (3.10)) can be written as 1-dimensional time delay mappings of the form

$$f(x_i) = f(x_{i-1}, x_{i-2}, \dots, x_{i-d}). \tag{14.11}$$

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.

14.5 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. ☞ sect. 4.5.3

14.5.1 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 (2.5) is given as

$$\dot{x} = v(x), \tag{14.12}$$

The corresponding Jacobian matrix $\mathbf{J}$ (4.26) is obtained by integrating the linearized equation (4.28)

$$\dot{\mathbf{J}} = \mathbf{AJ}\,, \qquad A_{ij}(x) = \frac{\partial v_i(x)}{\partial x_j}$$

along the trajectory. The flow and the corresponding Jacobian matrix are integrated simultaneously, by the same numerical routine. Integrating an initial condition on the Poincaré surface until a later crossing of the same and linearizing around the flow we can write

$$f(x') \approx f(x) + \mathbf{J}(x' - x). \tag{14.13}$$

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's method described below.

To find a fixed point of the flow near a starting guess x we must solve the linearized equation

$$(1 - \mathbf{J})(x' - x) = -(x - f(x)) = -F(x) \tag{14.14}$$

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 (14.14) 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

$$(x - x_0) \cdot a = 0, \tag{14.15}$$

where a is a vector normal to the Poincaré section and x_0 is any point in the Poincaré section. (14.14) then becomes

$$\begin{pmatrix} 1 - \mathbf{J} & v(x) \\ a & 0 \end{pmatrix} \begin{pmatrix} x' - x \\ \delta T \end{pmatrix} = \begin{pmatrix} -F(x) \\ 0 \end{pmatrix}. \tag{14.16}$$

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

$$\begin{pmatrix} 1-\Lambda & 0 & 0 & 0 \\ 0 & 1-\Lambda_s & 0 & 0 \\ 0 & 0 & 0 & v_z \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \delta_x \\ \delta_y \\ \delta_z \\ \delta t \end{pmatrix} = \begin{pmatrix} -F_x \\ -F_y \\ -F_z \\ 0 \end{pmatrix} . \qquad (14.17)$$

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 $[4 \times 4]$ 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. (27.2).

For periodic orbits (14.16) generalizes in the same way as (14.7), but with n additional equations – one for each point on the Poincaré surface. The Newton setup looks like this

$$\left(\begin{array}{ccccc|ccc} 1 & & & & -J_n & & & \\ -J_1 & 1 & & & & v_1 & & \\ & \cdots & 1 & & & & \ddots & \\ & & \cdots & 1 & & & & v_n \\ & & & -J_{n-1} & 1 & & & \\ \hline & a & & & & 0 & & \\ & & \ddots & & & & \ddots & \\ & & & a & & & & 0 \end{array}\right) \begin{pmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \delta_n \\ \delta t_1 \\ \cdot \\ \delta t_n \end{pmatrix} = \begin{pmatrix} -F_1 \\ -F_2 \\ \cdot \\ \cdot \\ -F_n \\ 0 \\ \cdot \\ 0 \end{pmatrix} .$$

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

$$\begin{pmatrix} 1-\mathbf{J} & v(x) & \nabla H(x) \\ a & 0 & 0 \end{pmatrix} \begin{pmatrix} x'-x \\ \delta t \\ \delta E \end{pmatrix} = \begin{pmatrix} -(x - f(x)) \\ 0 \\ 0 \end{pmatrix} \qquad (14.18)$$

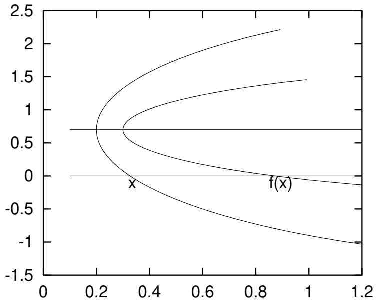


Figure 14.5: Illustration of the optimal Poincaré surface. The original surface $y = 0$ yields a large distance $x - f(x)$ for the Newton iteration. A much better choice is $y = 0.7$.

simultaneously with

$$H(x') - H(x) = 0. \tag{14.19}$$

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 (14.18) and (14.19) together. One might be tempted to linearize (14.19) and put it into (14.18) 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.

14.5.2 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. 14.5 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

$$(x' - x) \cdot v(x) = 0. \tag{14.20}$$

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.

14.6 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 (4.31). For example, the Jacobian matrix $\mathbf{J}_p$ for a Hénon map (3.10) prime cycle p of length n_p is given by (4.32), and the Jacobian matrix $\mathbf{J}_p$ for a 2-dimensional billiard prime cycle p of length n_p follows from (5.22). As explained on page 72, evaluation of the Jacobian matrix for a flow will require an integration along the prime cycle.

Commentary

Remark 14.1 Piece-wise linear maps. The Lozi map (3.12) is linear, and 100,000's of cycles can be easily computed by [2x2] matrix multiplication and inversion.

Résumé

There is no general computational algorithm that is guaranteed to find all solutions (up to a given period T_{max}) to the periodic orbit condition

$$f^{t+T}(x) = f^t(x)\,, \qquad T > 0$$

for a general flow or mapping. Due to the exponential divergence of nearby trajectories in chaotic dynamical systems, direct solution of the periodic orbit condition can be numerically very unstable.

A prerequisite for a systematic and complete cycle search is a good (but hard to come by) understanding of the topology of the flow. Usually one starts by - possibly analytic - determination of the equilibria of the flow. Their locations, stabilities, stability eigenvectors and invariant manifolds offer skeletal information about the topology of the flow. Next step is numerical long-time evolution of "typical" trajectories of the dynamical system under investigation. Such numerical experiments build up the "natural measure", and reveal regions most frequently visited. The periodic orbit searches can then be initialized by taking nearly recurring orbit segments and deforming them into a closed orbits. With a sufficiently good initial guess the Newton-Raphson formula (14.16)

☞ sect. 7.3.1

$$\left(\begin{array}{cc} 1-\mathbf{J} & v(x) \\ a & 0 \end{array}\right)\left(\begin{array}{c} \delta x \\ \delta T \end{array}\right) = \left(\begin{array}{c} f(x)-x \\ 0 \end{array}\right)$$

yields improved estimate $x' = x + \delta x, T' = T + \delta T$. Iteration then yields the period T and the location of a periodic point x_p in the Poincaré surface $(x_p - x_0)\cdot a = 0$, where a is a vector normal to the Poincaré section at x_0.

The problem one faces with high-dimensional flows is that their topology is hard to visualize, and that even with a decent starting guess for a point on a periodic orbit, methods like the Newton-Raphson method are likely to fail. Methods that start with initial guesses for a number of points along the cycle, such as the multipoint shooting method of sect. 14.3, are more robust.

☞ chapter 25

The relaxation (or variational) methods take this strategy to its logical extreme, and start by a guess of not a few points along a periodic orbit, but a guess of the entire orbit. As these methods are intimately related to variational principles and path integrals, we postpone their introduction to chapter 25.

References

[14.1] D.W. Moore and E.A. Spiegel, "A thermally excited nonlinear oscillator", *Astrophys. J.,* **143**, 871 (1966).

[14.2] N.H. Baker, D.W. Moore and E.A. Spiegel, *Quar. J. Mech. and Appl. Math.* **24**, 391 (1971).

[14.3] E.A. Spiegel, *Chaos: a mixed metaphor for turbulence*, *Proc. Roy. Soc.* **A413**, 87 (1987).

[14.4] M. Baranger and K.T.R. Davies *Ann. Physics* **177**, 330 (1987).

[14.5] B.D. Mestel and I. Percival, *Physica D* **24**, 172 (1987); Q. Chen, J.D. Meiss and I. Percival, *Physica D* **29**, 143 (1987).

[14.6] find Helleman et all Fourier series methods

[14.7] J.M. Greene, *J. Math. Phys.* **20**, 1183 (1979)

[14.8] H.E. Nusse and J. Yorke, "A procedure for finding numerical trajectories on chaotic saddles" *Physica* **D 36**, 137 (1989).

[14.9] D.P. Lathrop and E.J. Kostelich, "Characterization of an experimental strange attractor by periodic orbits"

[14.10] T. E. Huston, K.T.R. Davies and M. Baranger *Chaos* **2**, 215 (1991).

[14.11] M. Brack, R. K. Bhaduri, J. Law and M. V. N. Murthy, *Phys. Rev. Lett.* **70**, 568 (1993).

[14.12] 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).

Exercises

Exercise 14.1 Cycles of the Ulam map. *Test your cycle-searching routines by computing a bunch of short cycles and their stabilities for the Ulam map*

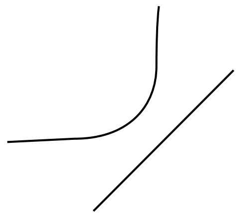
$$f(x) = 4x(1-x)\,. \tag{14.21}$$

Exercise 14.2 Cycles stabilities for the Ulam map, exact. *In exercise 14.1 you should have observed that the numerical results for the cycle stability eigenvalues (4.30) 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 (9.10). This problem is perhaps too hard, but give it a try - the answer is in many introductory books on nolinear dynamics.)*

Exercise 14.3 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 $y = 1/x$.*



Exercise 14.4 Cycle stability. *Add to the pinball simulator of exercise 5.1 a routine that evaluates the expanding eigenvalue for a given cycle.*

Exercise 14.5 Pinball cycles. *Determine the stability and length of all fundamental domain prime cycles of the binary symbol string lengths up to 5 (*or longer*) for $R : a = 6$ 3-disk pinball.*

Exercise 14.6 Newton-Raphson method. *Implement the Newton-Raphson method in 2-d and apply it to determination of pinball cycles.*

Exercise 14.7 Rössler system cycles. *(continuation of exercise 2.7, and exercise 3.1) Determine all cycles up to 5 Poincaré sections returns for the Rössler system (2.11), as well as their stabilities. (Hint: implement (14.16), the multipoint shooting methods for flows; you can cross-check your shortest cycles against the ones listed in table 14.1.)*

n_p	p	y_p	z_p	Λ_e
1	1	6.091768319056803	1.299731937639821	-2.4039535318268
2	01	3.915804049621049	3.692833386542665	-3.5120069815161
3	001	2.278281031720258	7.416480984019008	-2.3419235232340
	011	2.932877559129124	5.670805943881501	5.3449081538885
4	0111	3.466758713211455	4.506217531477667	-16.6967406980700
5	01011	4.162798782914948	3.303903338609633	-23.1995830097831
	01111	3.278914359770783	4.890452922955567	36.8863297988981
6	001011	2.122093931936202	7.886172854283211	-6.8576654190825
	010111	4.059210605826523	3.462265228606606	61.6490940089068
	011111	3.361494458061049	4.718206217035575	-92.0825560711089
7	0101011	3.842769382372052	3.815493592299824	77.7611048852412
	0110111	3.025956697151134	5.451444475664179	-95.1838846735358
	0101111	4.102255295518855	3.395643547170646	-142.2379888163439
	0111111	3.327986189581191	4.787462810306583	218.0283602810993

Table 14.1: The Rössler system (2.11): The itinerary p, a periodic point $x_p = (0, y_p, z_p)$ and the expanding eigenvalue Λ_p for all cycles up to the topological length 7. (Joachim Mathiesen)

Exercise 14.8 **Cycle stability, helium.** *Add to the helium integrator of exercise 2.9 a routine that evaluates the expanding eigenvalue for a given cycle.*

Exercise 14.9 **Colinear helium cycles.** *Determine the stability and length of all fundamental domain prime cycles up to symbol sequence length 5 or longer for collinear helium of fig. 29.5.*

Exercise 14.10 **Uniqueness of unstable cycles***.** *Prove that there exists only one 3-disk prime cycle for a given finite admissible prime cycle symbol string. Hints: look at the Poincaré section mappings; can you show that there is exponential contraction to a unique periodic point with a given itinerary? Exercise 25.1 might be helpful in this effort.*

Exercise 14.11 **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 (3.10), the coordinates of a periodic orbit of length n_p satisfy the equation*

$$x_{p,i+1} + x_{p,i-1} = 1 - a x_{p,i}^2 \,, \quad i = 1, ..., n_p \,, \tag{14.22}$$

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 (14.22)

$$x''_{p,i} = S_{p,i} \sqrt{\frac{1 - x'_{p,i+1} - x'_{p,i-1}}{a}} \,, \quad i = 1, ..., n_p \tag{14.23}$$

p	Λ_p	$\sum x_{p,i}$
0	0.71516752438$\times 10^1$	-0.6076252185107
1	-0.29528463259$\times 10^1$	0.2742918851774
10	-0.98989794855$\times 10^1$	0.3333333333333
100	-0.13190727397$\times 10^3$	-0.2060113295833
110	0.55896964996$\times 10^2$	0.5393446629166
1000	-0.10443010730$\times 10^4$	-0.8164965809277
1100	0.57799826989$\times 10^4$	0.0000000000000
1110	-0.10368832509$\times 10^3$	0.8164965809277
10000	-0.76065343718$\times 10^4$	-1.4260322065792
11000	0.44455240007$\times 10^4$	-0.6066540777738
10100	0.77020248597$\times 10^3$	0.1513755016405
11100	-0.71068835616$\times 10^3$	0.2484632276044
11010	-0.58949885284$\times 10^3$	0.8706954728949
11110	0.39099424812$\times 10^3$	1.0954854155465
100000	-0.54574527060$\times 10^5$	-2.0341342556665
110000	0.32222060985$\times 10^5$	-1.2152504370215
101000	0.51376165109$\times 10^4$	-0.4506624359329
111000	-0.47846146631$\times 10^4$	-0.3660254037844
110100	-0.63939998436$\times 10^4$	0.3333333333333
101100	-0.63939998436$\times 10^4$	0.3333333333333
111100	0.39019387269$\times 10^4$	0.5485837703548
111010	0.10949094597$\times 10^4$	1.1514633582661
111110	-0.10433841694$\times 10^4$	1.3660254037844

Table 14.2: All periodic orbits up to 6 bounces for the Hamiltonian Hénon mapping (14.22) 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.)

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 (4.32). Verify that the times and the stabilities of the short periodic orbits for the Hénon repeller (3.10) at $a = 6$ are listed in table 14.2; in actual calculations all prime cycles up to topological length $n = 20$ have been computed.

(G. Vattay)

Exercise 14.12 "Center of mass" puzzle****.** *Why is the "center of mass", listed in table 14.2, a simple rational every so often?*

Chapter 15

Cycle expansions

Recycle... It's the Law!
Poster, New York City Department of Sanitation

The Euler product representations of spectral determinants (12.9) and dynamical zeta functions (12.12) 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 (12.9), (12.12) 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).

15.1 Pseudocycles and shadowing

How are periodic orbit formulas such as (12.12) 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 25.3 and 14.2. It is important not to miss any short cycles, as the calculation is as accurate as the shortest cycle dropped - including cycles longer than the shortest omitted does not improve the accuracy (more precisely, improves it, but painfully slowly).

Expand the dynamical zeta function (12.12) as a formal power series,

$$1/\zeta = \prod_p (1 - t_p) = 1 - \sum_{\{p_1 p_2 \dots p_k\}}' (-1)^{k+1} t_{p_1} t_{p_2} \dots t_{p_k} \tag{15.1}$$

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 (15.1) is now compactly written as

$$1/\zeta = 1 - \sum_\pi' t_\pi \,. \tag{15.2}$$

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

$$t_\pi = (-1)^{k+1} \frac{1}{|\Lambda_\pi|} e^{\beta A_\pi - s T_\pi} z^{n_\pi} \,. \tag{15.3}$$

depends on the pseudocycle topological length, integrated observable, period, and stability

$$\begin{aligned} n_\pi &= n_{p_1} + \dots + n_{p_k} \,, \qquad T_\pi = T_{p_1} + \dots + T_{p_k} \\ A_\pi &= A_{p_1} + \dots + A_{p_k} \,, \qquad \Lambda_\pi = \Lambda_{p_1} \Lambda_{p_2} \cdots \Lambda_{p_k} \,. \end{aligned} \tag{15.4}$$

15.1.1 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 (12.12) is given by

$$\begin{aligned} 1/\zeta = \; & (1 - t_0)(1 - t_1)(1 - t_{01})(1 - t_{001})(1 - t_{011}) \\ & (1 - t_{0001})(1 - t_{0011})(1 - t_{0111})(1 - t_{00001})(1 - t_{00011}) \\ & (1 - t_{00101})(1 - t_{00111})(1 - t_{01011})(1 - t_{01111}) \dots \end{aligned}$$

(see table 9.2), and the first few terms of the expansion (15.2) 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

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_1 - [(t_{01} - t_1 t_0)] - [(t_{001} - t_{01} t_0) + (t_{011} - t_{01} t_1)] \\ &\quad - [(t_{0001} - t_0 t_{001}) + (t_{0111} - t_{011} t_1) \\ &\qquad + (t_{0011} - t_{001} t_1 - t_0 t_{011} + t_0 t_{01} t_1)] - \dots \\ &= 1 - \sum_f t_f - \sum_n \hat{c}_n \,. \end{aligned} \tag{15.5}$$

All terms in this expansion up to length $n_p = 6$ are given in table 15.1. 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. 10.3 and sect. 15.3). 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 (10.21). If the flow is continuous and smooth, orbits of similar symbolic dynamics will traverse the same neighborhoods and will have similar weights, and the weights in such combinations will almost cancel. The utility of cycle expansions of dynamical zeta functions and spectral determinants, lies precisely in this organization into nearly cancelling combinations: cycle expansions are dominated by short cycles, with long cycles giving exponentially decaying corrections.

$-t_0$				
$-t_1$				
$-t_{10}$	$+\ t_1 t_0$			
$-t_{100}$	$+\ t_{10} t_0$			
$-t_{101}$	$+\ t_{10} t_1$			
$-t_{1000}$	$+\ t_{100} t_0$			
$-t_{1001}$	$+\ t_{100} t_1$	$+\ t_{101} t_0$	$-\ t_1 t_{10} t_0$	
$-t_{1011}$	$+\ t_{101} t_1$			
$-t_{10000}$	$+\ t_{1000} t_0$			
$-t_{10001}$	$+\ t_{1001} t_0$	$+\ t_{1000} t_1$	$-\ t_0 t_{100} t_1$	
$-t_{10010}$	$+\ t_{100} t_{10}$			
$-t_{10101}$	$+\ t_{101} t_{10}$			
$-t_{10011}$	$+\ t_{1011} t_0$	$+\ t_{1001} t_1$	$-\ t_0 t_{101} t_1$	
$-t_{10111}$	$+\ t_{1011} t_1$			
$-t_{100000}$	$+\ t_{10000} t_0$			
$-t_{100001}$	$+\ t_{10001} t_0$	$+\ t_{10000} t_1$	$-\ t_0 t_{1000} t_1$	
$-t_{100010}$	$+\ t_{10010} t_0$	$+\ t_{1000} t_{10}$	$-\ t_0 t_{100} t_{10}$	
$-t_{100011}$	$+\ t_{10011} t_0$	$+\ t_{10001} t_1$	$-\ t_0 t_{1001} t_1$	
$-t_{100101}$	$-t_{100110}$	$+\ t_{10010} t_1$	$+\ t_{10110} t_0$	
	$+\ t_{10} t_{1001}$	$+\ t_{100} t_{101}$	$-\ t_0 t_{10} t_{101}$	$-\ t_1 t_{10} t_{100}$
$-t_{101110}$	$+\ t_{10110} t_1$	$+\ t_{1011} t_{10}$	$-\ t_1 t_{101} t_{10}$	
$-t_{100111}$	$+\ t_{10011} t_1$	$+\ t_{10111} t_0$	$-\ t_0 t_{1011} t_1$	
$-t_{101111}$	$+\ t_{10111} t_1$			

Table 15.1: The binary curvature expansion (15.5) 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.

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 sect. 15.4. 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.

15.1.2 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, $n \le N$. The result is the Nth order polynomial approximation

$$1/\zeta_N = 1 - \sum_{n=1}^{N} \hat{c}_n z^n \,. \tag{15.6}$$

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 (15.6) 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 (15.6) as function of s.

15.1.3 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 (12.9) 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 (11.9) truncated to all prime cycles p and their repeats such that $n_p r \le N$:

$$\begin{aligned}
\left.\mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}}\right|_{(i)} &= \left.\mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}}\right|_{(i-1)} + n_{(i)} \sum_{r=1}^{n_{(i)} r \le N} \frac{e^{(\beta\cdot A_{(i)} - s T_{(i)})r}}{\left|\prod \left(1 - \Lambda^r_{(i),j}\right)\right|} z^{n_{(i)} r} \\
\left.\mathrm{tr}\,\frac{z\mathcal{L}}{1 - z\mathcal{L}}\right|_{N} &= \sum_{n=1}^{N} C_n z^n \,, \qquad C_n = \mathrm{tr}\,\mathcal{L}^n \,.
\end{aligned} \tag{15.7}$$

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 $n_p \le N$. 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 (11.8) as a power series, we compute the Nth order approximation to the spectral determinant (12.3)

$$\det(1-z\mathcal{L})|_N = 1 - \sum_{n=1}^{N} Q_n z^n \,, \qquad Q_n = Q_n(\mathcal{L}) = n\text{th cumulant} \tag{15.8}$$

as follows. The logarithmic derivative relation (12.4) yields

$$\begin{aligned}\left(\operatorname{tr}\frac{z\mathcal{L}}{1-z\mathcal{L}}\right)\det(1-z\mathcal{L}) &= -z\frac{d}{dz}\det(1-z\mathcal{L}) \\ (C_1 z + C_2 z^2 + \cdots)(1 - Q_1 z - Q_2 z^2 - \cdots) &= Q_1 z + 2Q_2 z^2 + 3Q_3 z^3 \cdots\end{aligned}$$

so the nth order term of the spectral determinant cycle (or in this case, the cumulant) expansion is given recursively by the trace formula expansion coefficients

$$Q_n = \frac{1}{n}\left(C_n - C_{n-1}Q_1 - \cdots C_1 Q_{n-1}\right) . \tag{15.9}$$

Given the trace formula (15.7) 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 (15.6), by replacing $\prod\left(1-\Lambda^r_{(i),j}\right)$ in (15.7) by the product of expanding eigenvalues $\Lambda_{(i)} = \prod_e \Lambda_{(i),e}$, as in sect. 12.3.

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 (15.8) is the cycle expansion for (12.6), 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 (15.8) 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 (15.5) and (15.8) to different maximal cycle lengths, table 15.2.

☞ chapter 13

The pleasant surprise is that the coefficients in these expansions can be proven to fall off exponentially or even faster , due to analyticity of $\det(s-\mathcal{A})$ or $1/\zeta(s)$ for s values well beyond those for which the corresponding trace formula diverges.

$R{:}a$	N	. $\det(s-\mathcal{A})$	$1/\zeta(s)$	$1/\zeta(s)_{\text{3-disk}}$
	1	0.39	0.407	
	2	0.4105	0.41028	0.435
	3	0.410338	0.410336	0.4049
6	4	0.4103384074	0.4103383	0.40945
	5	0.4103384077696	0.4103384	0.410367
	6	0.410338407769346482	0.4103383	0.410338
	7	0.4103384077693464892		0.4103396
	8	0.410338407769346489338468		
	9	0.4103384077693464893384613074		
	10	0.4103384077693464893384613078192		
	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 15.2: 3-disk repeller escape rates computed from the cycle expansions of the spectral determinant (12.6) and the dynamical zeta function (12.12), 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 (15.5), and the fifth from the full 3-disk cycle expansion (15.32). 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.)

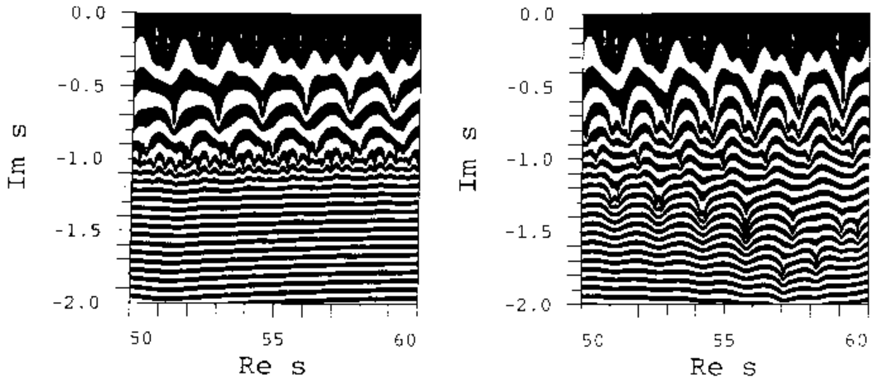


Figure 15.1: Examples of the complex s plane scans: contour plots of the logarithm of the absolute values of (a) $1/\zeta(s)$, (b) spectral determinant $\det(s-\mathcal{A})$ for the 3-disk system, separation $a:R=6$, A_1 subspace are evaluated numerically. The eigenvalues of the evolution operator $\mathcal{L}$ 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)

15.1.4 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. 15.1, 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. 15.1), a zero $1/\zeta(s)=0$ can now be easily determined by standard numerical methods, such as the iterative Newton algorithm (14.4)

$$s_{n+1} = s_n - \left(\zeta(s_n)\frac{\partial}{\partial s}\zeta^{-1}(s_n)\right)^{-1} = s_n - \frac{1/\zeta(s_n)}{\langle T \rangle_\zeta} \,. \qquad (15.10)$$

The derivative of $1/\zeta(s)$ required for the Newton iteration is given by the

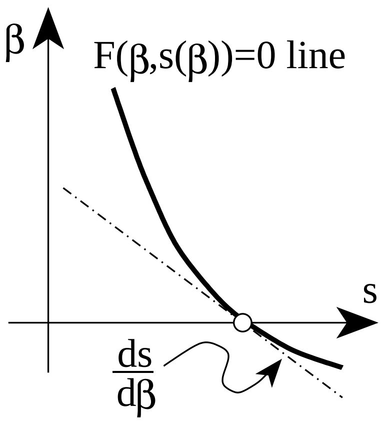


Figure 15.2: The eigenvalue condition is satisfied on the curve $F = 0$ the (β, s) plane. The expectation value of the observable (8.12) is given by the slope of the curve.

cycle expansion (15.18) that we need to evaluate anyhow, as $\langle T \rangle_\zeta$ enters our cycle averaging formulas.

15.2 Cycle formulas for dynamical averages

The eigenvalue condition in any of the three forms that we have given so far - the level sum (16.18), the dynamical zeta function (15.2), the spectral determinant (15.8):

$$
\begin{aligned}
1 &= \sum_i^{(n)} t_i \,, \qquad t_i = t_i(\beta, s(\beta)) = \frac{1}{|\Lambda_i|} e^{\beta \cdot A_i - s(\beta) T_i} && (15.11) \\
0 &= 1 - \sum_\pi{}' t_\pi \,, \qquad t_\pi = t_\pi(z, \beta, s(\beta)) && (15.12) \\
0 &= 1 - \sum_{n=1}^\infty Q_n \,, \qquad Q_n = Q_n(\beta, s(\beta)) \,, && (15.13)
\end{aligned}
$$

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. 15.2; 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

$$
\begin{aligned}
0 &= \frac{d}{d\beta} F(\beta, s(\beta)) \\
&= \frac{\partial F}{\partial \beta} + \frac{ds}{d\beta} \left. \frac{\partial F}{\partial s} \right|_{s=s(\beta)} \implies \frac{ds}{d\beta} = -\frac{\partial F}{\partial \beta} \Big/ \frac{\partial F}{\partial s} \,, && (15.14)
\end{aligned}
$$

and the second derivative of $F(\beta, s(\beta)) = 0$ yields

$$\frac{d^2 s}{d\beta^2} = -\left[\frac{\partial^2 F}{\partial \beta^2} + 2\frac{ds}{d\beta}\frac{\partial^2 F}{\partial\beta\partial s} + \left(\frac{ds}{d\beta}\right)^2 \frac{\partial^2 F}{\partial s^2}\right] / \frac{\partial F}{\partial s} . \tag{15.15}$$

Denoting by

$$\begin{aligned}
\langle A \rangle_F &= -\left.\frac{\partial F}{\partial \beta}\right|_{\beta, s=s(\beta)} , \qquad \langle T \rangle_F = \left.\frac{\partial F}{\partial s}\right|_{\beta, s=s(\beta)} \\
\left\langle (A - \langle A \rangle)^2 \right\rangle_F &= \left.\frac{\partial^2 F}{\partial \beta^2}\right|_{\beta, s=s(\beta)}
\end{aligned} \tag{15.16}$$

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 (8.12) and its variance

$$\begin{aligned}
\langle a \rangle &= \frac{\langle A \rangle_F}{\langle T \rangle_F} \\
\left\langle (a - \langle a \rangle)^2 \right\rangle &= \frac{1}{\langle T \rangle_F} \left\langle (A - \langle A \rangle)^2 \right\rangle_F .
\end{aligned} \tag{15.17}$$

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 (16.18), (15.2) and (15.8), the above quantities have explicit cycle expansions.

15.2.1 Dynamical zeta function cycle expansions

For the dynamical zeta function condition (15.12), the cycle averaging formulas (15.14), (15.17) require evaluation of the derivatives of dynamical zeta function at a given eigenvalue. Substituting the cycle expansion (15.2) for dynamical zeta function we obtain

$$\begin{aligned}
\langle A \rangle_\zeta &:= -\frac{\partial}{\partial \beta}\frac{1}{\zeta} = \sum{}' A_\pi t_\pi \\
\langle T \rangle_\zeta &:= \frac{\partial}{\partial s}\frac{1}{\zeta} = \sum{}' T_\pi t_\pi , \quad \langle n \rangle_\zeta := -z\frac{\partial}{\partial z}\frac{1}{\zeta} = \sum{}' n_\pi t_\pi ,
\end{aligned} \tag{15.18}$$

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 (15.4), 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

$$\langle A \rangle_\zeta = \sum_\pi{}' (-1)^{k+1} \frac{A_{p_1} + A_{p_2} \cdots + A_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} , \tag{15.19}$$

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

$$\begin{aligned} \langle T \rangle_\zeta &= \frac{T_0}{|\Lambda_0|} + \frac{T_1}{|\Lambda_1|} + \left(\frac{T_{01}}{|\Lambda_{01}|} - \frac{T_0 + T_1}{|\Lambda_0 \Lambda_1|} \right) \\ &\quad + \left(\frac{T_{001}}{|\Lambda_{001}|} - \frac{T_{01} + T_0}{|\Lambda_{01} \Lambda_0|} \right) + \left(\frac{T_{011}}{|\Lambda_{011}|} - \frac{T_{01} + T_1}{|\Lambda_{01} \Lambda_1|} \right) + \ldots \end{aligned} \tag{15.20}$$

Note that the cycle expansions for averages are grouped into the same shadowing combinations as the dynamical zeta function cycle expansion (15.5), 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 (15.17). 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

$$\left\langle A^2 \right\rangle_\zeta = \sum{}' (-1)^{k+1} \frac{\left(A_{p_1} + A_{p_2} \cdots + A_{p_k}\right)^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} . \tag{15.21}$$

15.2.2 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 15.2. 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 13, 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 (15.13) 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

(15.9) and (15.7). The cycle averaging formulas formulas are exact, and highly convergent for nice hyperbolic dynamical systems. An example of its utility is the cycle expansion formula for the Lyapunov exponent of sect. 15.2.4. Further applications of cycle expansions will be discussed in chapter 16.

15.2.3 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 $\langle a \rangle_{\mathrm{dscr}}$ measured in discrete time given by the number of reflections off billiard walls, the two averages are related by

$$\langle a \rangle_{\mathrm{dscr}} = \langle a \rangle \langle T \rangle_\zeta \, / \, \langle n \rangle_\zeta \ , \tag{15.22}$$

where $\langle n \rangle_\zeta$ is the average of the number of bounces n_p along the cycle p.

15.2.4 Cycle expansion formula for Lyapunov exponents

In sect. 8.3 we defined the Lyapunov exponent for a 1-d mapping, related it to the leading eigenvalue of an evolution operator and promised to evaluate it. Now we are finally in position to deliver on our promise.

The cycle averaging formula (15.19) yields a *closed* expression for the Lyapunov exponent in terms of prime cycles:

$$\lambda = \frac{1}{\langle n \rangle_\zeta} \sum{}' (-1)^{k+1} \frac{\log |\Lambda_{p_1}| + \cdots + \log |\Lambda_{p_k}|}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} \ . \tag{15.23}$$

For a repeller, the $1/|\Lambda_p|$ weights are replaced by normalized measure (16.10) $\exp(\gamma n_p)/|\Lambda_p|$, where γ is the escape rate.

We mention here without proof that for 2-d Hamiltonian flows such as our game of pinball there is only one expanding eigenvalue and (15.23) applies as it stands.

in depth:
chapter H.1, p. 659

15.3 Cycle expansions for finite alphabets

A finite Markov graph like the one given in fig. 10.3(d) is a compact encoding of the transition or the Markov matrix for a given subshift. It is a sparse matrix, and the associated determinant (10.17) can be written down by inspection: it is the sum of all possible partitions of the graph into products of non-intersecting loops, with each loop carrying a minus sign:

$$\det(1-T) = 1 - t_0 - t_{0011} - t_{0001} - t_{00011} + t_0 t_{0011} + t_{0011} t_{0001} \qquad (15.24)$$

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 (10.18).

The determinant (15.24) is exact for the finite graph fig. 10.3(e), as well as for the associated transfer operator of sect. 7.1. For the associated (infinite dimensional) evolution operator, it is the beginning of the cycle expansion of the corresponding dynamical zeta function:

$$\begin{aligned} 1/\zeta &= 1 - t_0 - t_{0011} - t_{0001} + t_{0001} t_{0011} \\ &\qquad - (t_{00011} - t_0 t_{0011} + \ldots \text{curvatures}) \ldots \end{aligned} \qquad (15.25)$$

The cycles 0, 0001 and 0011 are the *fundamental* cycles introduced in (15.5); they are not shadowed by any combinations of shorter cycles, and are the basic building blocks of the dynamics. All other cycles appear together with their shadows (for example, $t_{00011} - t_0 t_{0011}$ 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 (15.2) 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 (10.21) offers a quick way of checking the fundamental part of a cycle expansion.

Since for finite grammars the topological zeta functions reduce to polynomials, we are assured that there are just a few fundamental cycles and that all long cycles can be grouped into curvature combinations. For example, the fundamental cycles in exercise 9.5 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.

15.4 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 18, 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

$$|\Lambda_{p_1} \cdots \Lambda_{p_k}| \leq \Lambda_{\max} \tag{15.26}$$

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.

15.4.1 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 (15.5) 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 (16.9)? The argument has essentially already been laid out in sect. 10.7: 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 (16.9) error, where h is the entropy and T typical cycle length for cycles of stability Λ.

15.4.2 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.

15.4.3 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 18, the simplest of which is the Farey map

$$f(x) = \begin{cases} x/(1-x) & 0 \le x \le 1/2 \quad L \\ (1-x)/x & 1/2 \le x \le 1 \quad R\,, \end{cases} \tag{15.27}$$

a map which will reappear in the intermittency chapter 18, and in chapter 21, in context 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. 10.5.2 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. 15.3).

As the Farey map maps the unit interval onto itself, the leading eigenvalue of the Perron-Frobenius operator should equal $s_0 = 0$, so $1/\zeta(0) = 0$.

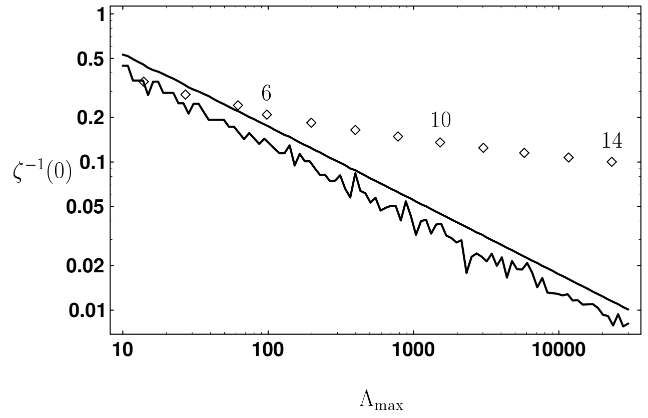


Figure 15.3: Comparison of cycle expansion truncation schemes for the Farey map (15.27); the deviation of the truncated cycles expansion for $|1/\zeta_N(0)|$ from the exact flow conservation value $1/\zeta(0) = 0$ is a measure of the accuracy of the truncation. The jagged line is logarithm of the stability ordering truncation error; the smooth line is smoothed according to sect. 15.4.2; 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.

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. 15.3. We see that topological length truncation schemes are hopelessly bad in this case; stability length truncations are somewhat better, but still rather bad. In simple cases like this one, where intermittency is caused by a single marginal fixed point, the convergence can be improved by going to infinite alphabets.

15.5 Dirichlet series

A Dirichlet series is defined as

$$f(s) = \sum_{j=1}^{\infty} a_j e^{-\lambda_j s} \tag{15.28}$$

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 (15.2) ordered by the increasing pseudocycle periods are often Dirichlet series. For example, for the pseudocycle weight (15.3), 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*

$$\sigma_a = \lim_{N\to\infty} \sup \frac{1}{\lambda_N} \ln \sum_{j=1}^{N} |a_j| \,, \qquad (15.29)$$

and σ_c is the *abscissa of conditional convergence*

$$\sigma_c = \lim_{N\to\infty} \sup \frac{1}{\lambda_N} \ln \left| \sum_{j=1}^{N} a_j \right| \,. \qquad (15.30)$$

We shall encounter another example of a Dirichlet series in the semiclassical quantization chapter 26, 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 15.1 Pseudocycle expansions. Bowen's introduction of shadowing ϵ-pseudoorbits [1.13] was a significant contribution to Smale's theory. Expression "pseudoorbits" seems to have been introduced in the Parry and Pollicott's 1983 paper [12.5]. Following them M. Berry [15.9] 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. [15.10, 15.2]. Some literature [12.14] refers to the pseudoorbits as "composite orbits", and to the cycle expansions as "Dirichlet series" (see also remark 15.6 and sect. 15.5).

Remark 15.2 Cumulant expansion. To statistical mechanician the curvature expansions are very reminiscent of cumulant expansions. Indeed, (15.9) is the standard Plemelj-Smithies cumulant formula (K.28) for the Fredholm determinant, discussed in more detail in appendix K.

Remark 15.3 Exponential growth of the number of cycles. Going from $N_n \approx N^n$ periodic points of length n to M_n prime cycles reduces the number of computations from N_n to $M_n \approx N^{n-1}/n$. Use of discrete symmetries (chapter 19) reduces the number of nth level terms by another factor. While the formulation of the theory from the trace (11.24) to the cycle expansion (15.5) 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 15.4 Shadowing cycle-by-cycle. A glance at the low order curvatures in the table 15.1 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. [15.3], table 1). However, splitting $\hat{c}_n$ into individual cycle curvatures is not possible in general [9.34]; 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 15.1 is shared by the two cycles.

Remark 15.5 Stability ordering. The stability ordering was introduced by Dahlqvist and Russberg [15.12] in a study of chaotic dynamics for the $(x^2y^2)^{1/a}$ potential. The presentation here runs along the lines of Dettmann and Morriss [15.13] for the Lorentz gas which is hyperbolic but the symbolic dynamics is highly pruned, and Dettmann and Cvitanović [15.14] for a family of intermittent maps. In the applications discussed in the above papers, the stability ordering yields a considerable improvement over the topological length ordering.

Remark 15.6 Are cycle expansions Dirichlet series? Even though some literature [12.14] 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 (12.12), (27.18) 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

[15.1] P. Cvitanović, *Phys. Rev. Lett.* **61**, 2729 (1988).

[15.2] R. Artuso, E. Aurell and P. Cvitanović, "Recycling of strange sets I: Cycle expansions", *Nonlinearity* **3**, 325 (1990).

[15.3] R. Artuso, E. Aurell and P. Cvitanović, "Recycling of strange sets II: Applications", *Nonlinearity* **3**, 361 (1990).

[15.4] S. Grossmann and S. Thomae, *Z. Naturforsch.* **32 a**, 1353 (1977); reprinted in ref. [15.5].

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

[15.6] F. Christiansen, P. Cvitanović and H.H. Rugh, *J. Phys* **A 23**, L713 (1990).

[15.7] J. Plemelj, *"Zur Theorie der Fredholmschen Funktionalgleichung"*, *Monat. Math. Phys.* **15**, 93 (1909).

[15.8] F. Smithies, *"The Fredholm theory of integral equations"*, *Duke Math.* **8**, 107 (1941).

[15.9] M.V. Berry, in *Quantum Chaos and Statistical Nuclear Physics*, ed. T.H. Seligman and H. Nishioka, *Lecture Notes in Physics* **263**, 1 (Springer, Berlin, 1986).

[15.10] P. Cvitanović, *"Invariant measurements of strange sets in terms of cycles"*, *Phys. Rev. Lett.* **61**, 2729 (1988).

[15.11] B. Eckhardt and G. Russberg, *Phys. Rev.* **E 47**, 1578 (1993).

[15.12] P. Dahlqvist and G. Russberg, *"Periodic orbit quantization of bound chaotic systems"*, *J. Phys.* **A 24**, 4763 (1991); P. Dahlqvist *J. Phys.* **A 27**, 763 (1994).

[15.13] C. P. Dettmann and G. P. Morriss, *Phys. Rev. Lett.* **78**, 4201 (1997).

[15.14] C. P. Dettmann and P. Cvitanović, *Cycle expansions for intermittent diffusion Phys. Rev. E* **56**, 6687 (1997); chao-dyn/9708011.

Exercises

Exercise 15.1 **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.*

Exercise 15.2 **Escape rate for a 1-d repeller.** *(Continuation of exercise 12.1 - easy, but long)*
Consider again the quadratic map (12.31)

$$f(x) = Ax(1-x)$$

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. 14.2.1).*

(c) *Determine the leading zero of the zeta function (12.12) 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 15.2. Compare the coefficients of the spectral determinant and the zeta function cycle expansions. Which expansion converges faster?*

(Per Rosenqvist)

Exercise 15.3 **Escape rate for the Ulam map.** *Check that the escape rate for the Ulam map, $A = 4$ in (12.31)*

$$f(x) = 4x(1-x),$$

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.

Exercise 15.4 **Pinball escape rate, semi-analytical.** *Estimate the 3-disk pinball escape rate for $R : a = 6$ by substituting analytical cycle stabilities and periods (exercise 5.5 and exercise 5.6) into the appropriate binary cycle expansion. Compare with the numerical estimate exercise 12.12*

Exercise 15.5 **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 14.5 into the binary cycle expansion.*

Exercise 15.6 **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 (15.5) 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?*

Exercise 15.7 **Counting the 3-disk pinball counterterms.** *Verify that the number of terms in the 3-disk pinball curvature expansion (15.31) is given by*

$$\begin{aligned}\prod_p (1+t_p) &= \frac{1-3z^4-2z^6}{1-3z^2-2z^3} = 1+3z^2+2z^3+\frac{z^4(6+12z+2z^2)}{1-3z^2-2z^3} \\ &= 1+3z^2+2z^3+6z^4+12z^5+20z^6+48z^7+84z^8+184z^9+\dots\end{aligned}$$

This means that, for example, c_6 has a total of 20 terms, in agreement with the explicit 3-disk cycle expansion (15.32).

Exercise 15.8 **3–disk unfactorized zeta cycle expansions.** *Check that the curvature expansion (15.2) for the 3-disk pinball, assuming no symmetries between disks, is given by*

$$\begin{aligned}1/\zeta &= (1-z^2t_{12})(1-z^2t_{13})(1-z^2t_{23})(1-z^3t_{123})(1-z^3t_{132}) \\ &\quad (1-z^4t_{1213})(1-z^4t_{1232})(1-z^4t_{1323})(1-z^5t_{12123})\cdots \\ &= 1-z^2t_{12}-z^2t_{23}-z^2t_{31}-z^3t_{123}-z^3t_{132} \\ &\quad -z^4[(t_{1213}-t_{12}t_{13})+(t_{1232}-t_{12}t_{23})+(t_{1323}-t_{13}t_{23})] \\ &\quad -z^5[(t_{12123}-t_{12}t_{123})+\cdots]-\cdots \end{aligned} \tag{15.31}$$

The symmetrically arranged 3-disk pinball cycle expansion of the Euler product (15.2) (see table 10.4 and fig. 19.2) is given by:

$$\begin{aligned}1/\zeta &= (1-z^2t_{12})^3(1-z^3t_{123})^2(1-z^4t_{1213})^3 \\ &\quad (1-z^5t_{12123})^6(1-z^6t_{121213})^6(1-z^6t_{121323})^3\dots \\ &= 1-3z^2\,t_{12}-2z^3\,t_{123}-3z^4\,(t_{1213}-t_{12}^2)-6z^5\,(t_{12123}-t_{12}t_{123}) \\ &\quad -z^6\,(6\,t_{121213}+3\,t_{121323}+t_{12}^3-9\,t_{12}t_{1213}-t_{123}^2) \\ &\quad -6z^7\,(t_{1212123}+t_{1212313}+t_{1213123}+t_{12}^2t_{123}-3\,t_{12}t_{12123}-t_{123}t_{1213}) \\ &\quad -3z^8\,(2\,t_{12121213}+t_{12121313}+2\,t_{12121323}+2\,t_{12123123} \\ &\qquad +2\,t_{12123213}+t_{12132123}+3\,t_{12}^2t_{1213}+t_{12}t_{123}^2 \\ &\qquad -6\,t_{12}t_{121213}-3\,t_{12}t_{121323}-4\,t_{123}t_{12123}-t_{1213}^2)-\cdots \end{aligned} \tag{15.32}$$

Remark 15.7 Unsymmetrized cycle expansions. The above 3-disk cycle expansions might be useful for cross-checking purposes, but, as we shall see in chapter 19, they are not recommended for actual computations, as the factorized zeta functions yield much better convergence.

Exercise 15.9 4–disk unfactorized dynamical zeta function cycle expansions *For the symmetriclly arranged 4-disk pinball the symmetry group is C_{4v}, of order 8. The degenerate cycles can have multiplicities 2, 4 or 8 (see table 10.2):*

$$\begin{aligned}1/\zeta &= (1-z^2t_{12})^4(1-z^2t_{13})^2(1-z^3t_{123})^8(1-z^4t_{1213})^8(1-z^4t_{1214})^4\\ &\quad(1-z^4t_{1234})^2(1-z^4t_{1243})^4(1-z^5t_{12123})^8(1-z^5t_{12124})^8(1-z^5t_{12134})^8\\ &\quad(1-z^5t_{12143})^8(1-z^5t_{12313})^8(1-z^5t_{12413})^8\cdots \end{aligned} \tag{15.33}$$

and the cycle expansion is given by

$$\begin{aligned}1/\zeta &= 1-z^2(4\,t_{12}+2\,t_{13})-8z^3\,t_{123}\\ &\quad -z^4(8\,t_{1213}+4\,t_{1214}+2\,t_{1234}+4\,t_{1243}-6\,t_{12}^2-t_{13}^2-8\,t_{12}t_{13})\\ &\quad -8z^5(t_{12123}+t_{12124}+t_{12134}+t_{12143}+t_{12313}+t_{12413}-4\,t_{12}t_{123}-2\,t_{13}t_{123})\\ &\quad -4z^6(2\,S_8+\,S_4+\,t_{12}^3+3\,t_{12}^2\,t_{13}+\,t_{12}t_{13}^2-8\,t_{12}t_{1213}-4\,t_{12}t_{1214}\\ &\quad -2\,t_{12}t_{1234}-4\,t_{12}t_{1243}-4\,t_{13}t_{1213}-2\,t_{13}t_{1214}-\,t_{13}t_{1234}\\ &\quad -2\,t_{13}t_{1243}-7\,t_{123}^2)-\cdots \end{aligned} \tag{15.34}$$

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 10.4.

Exercise 15.10 Tail resummations. *A simple illustration of such tail resummation is the ζ function for the Ulam map (14.21) 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, (12.18) 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*

$$1/\zeta(z) = \frac{(1-2t)(1-t_0)}{(1-t)} \tag{15.35}$$

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 (15.5) unexpanded, and noted that all curvature combinations in (15.5) which include the t_0 factor are unbalanced, so that the cycle expansion is an infinite series:

$$\prod_p (1-t_p) = (1-t_0)(1-t-t^2-t^3-t^4-\ldots) \tag{15.36}$$

(we shall return to such infinite series in chapter 18). The geometric series in the brackets sums up to (15.35). 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 (15.35).

Chapter 16

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.

16.1 Escape rates

We start by verifying the claim (8.11) 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

$$\hat{\Gamma}_{\mathcal{M}}(t) = \frac{1}{|\mathcal{M}|} \int \int_{\mathcal{M}} dx dy\, \delta\big(y - f^t(x)\big) \ . \tag{16.1}$$

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. 1.8. 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

$$\hat{\Gamma}_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i| \,, \tag{16.2}$$

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. 74) 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 11 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

$$\hat{\Gamma}_1 \sim \frac{1}{\Lambda} + \frac{1}{\Lambda}, \cdots \qquad , \hat{\Gamma}_n \sim \frac{2^n}{\Lambda^n} = e^{-n(\lambda - h)} := e^{-n\gamma} \,, \tag{16.3}$$

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 (17.16), with correctly defined "entropy" this result is exact.

As at each bounce one loses routinely the same fraction of trajectories, one expects the sum (16.2) to fall off exponentially with n. More precisely, by the hyperbolicity assumption of sect. 11.1.1 the expanding eigenvalue of the Jacobian matrix of the flow is exponentially bounded from both above and below,

$$1 < |\Lambda_{min}| \leq |\Lambda(x)| \leq |\Lambda_{max}| \,, \tag{16.4}$$

and the area of each strip in (16.2) is bounded by $|\Lambda_{max}^{-n}| \leq |\mathcal{M}_i| \leq |\Lambda_{min}^{-n}|$. Replacing $|\mathcal{M}_i|$ in (16.2) 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

$$\ln |\Lambda_{max}| \geq -\frac{1}{n} \ln \hat{\Gamma}_n + \ln 2 \geq \ln |\Lambda_{min}| \,. \tag{16.5}$$

The argument based on (16.5) 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 (16.2) can be replaced by the sum over periodic orbit stabilities. By (16.4) the size of $\mathcal{M}_i$ strip can be bounded by the stability Λ_i of ith periodic point:

$$\mathcal{C}_1 \frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \mathcal{C}_2 \frac{1}{|\Lambda_i|} \,, \tag{16.6}$$

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

$$\hat{\mathcal{C}}_1 \, \Gamma_n < \sum_i^{(n)} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \hat{\mathcal{C}}_2 \, \Gamma_n \quad , \tag{16.7}$$

where $\Gamma_n = \sum_i^{(n)} 1/|\Lambda_i|$ is the asymptotic trace sum (11.22). In this way we have established that for hyperbolic systems the survival probability sum (16.2) can be replaced by the periodic orbit sum (11.22).

We conclude that for hyperbolic, locally unstable flows the fraction (16.1) 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

$$\gamma = -\lim_{t\to\infty} \frac{1}{t} \ln \Gamma_{\mathcal{M}}(t) \,. \tag{16.8}$$

16.1 page 317
7.4 page 121

16.1.1 Periodic orbit averages

We now refine the reasoning of sect. 16.1. Consider the trace (11.6) in the asymptotic limit (11.21):

$$\mathrm{tr}\,\mathcal{L}^n = \int dx\, \delta(x - f^n(x))\, e^{\beta A^n(x)} \approx \sum_i^{(n)} \frac{e^{\beta A^n(x_i)}}{|\Lambda_i|} \,.$$

The factor $1/|\Lambda_i|$ was interpreted in (16.2) 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. 1.9, 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 $[s_{-m} \cdots s_0 . s_1 s_2 \cdots s_n]$ 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 $\langle 1 \rangle = 1$:

$$\left\langle e^{\beta A^n} \right\rangle_n \approx \frac{\sum_i^{(n)} e^{\beta A^n(x_i)}/|\Lambda_i|}{\sum_i^{(n)} 1/|\Lambda_i|} = \sum_i^{(n)} \mu_i\, e^{\beta A^n(x_i)} \,. \tag{16.9}$$

Here μ_i is the *normalized natural measure*

$$\sum_i^{(n)} \mu_i = 1 \,, \qquad \mu_i = e^{n\gamma}/|\Lambda_i| \,, \tag{16.10}$$

correct both for the closed systems as well as the open systems of sect. 8.1.3.

Unlike brute numerical slicing of the integration space into an arbitrary lattice (for a critique, see sect. 8.4), 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.

16.1.2 Unstable periodic orbits are dense

(L. Rondoni and P. Cvitanović)

Our goal in sect. 8.1 was to evaluate the space and time averaged expectation value (8.9). 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.

16.2 Flow conservation sum rules

If the dynamical system is bounded, all trajectories remain confined for all times, escape rate (16.8) equals $\gamma = -s_0 = 0$, and the leading eigenvalue (J.2) of the Perron-Frobenius operator (7.10) 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:

$$\begin{aligned} 1/\zeta(0,0) &= 1 + \sum_\pi{}' \frac{(-1)^k}{|\Lambda_{p_1}\cdots\Lambda_{p_k}|} = 0 \\ F(0,0) &= 1 - \sum_{n=1}^{\infty} c_n(0,0) = 0 \end{aligned} \tag{16.11}$$

obtained by setting $s = 0$ in (15.12), (15.13) 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 (15.11) yields

$$\mathrm{tr}\,\mathcal{L}^n = \sum_{i \in \mathrm{Fix} f^n} \frac{1}{|\det\,(\mathbf{1} - \mathbf{J}^n(x_i))\,|} = 1 + e^{s_1 n} + \dots\,. \tag{16.12}$$

For flows one can apply this rule by grouping together cycles from $t = T$ to $t = T + \Delta T$

$$\begin{aligned} \frac{1}{\Delta T} \sum_{p,r}^{T \le rT_p \le T + \Delta T} \frac{T_p}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|} &= \frac{1}{\Delta T} \int_T^{T+\Delta T} dt \left(1 + e^{s_1 t} + \dots\right) \\ &= 1 + \frac{1}{\Delta T} \sum_{\alpha=1}^{\infty} \frac{e^{s_\alpha T}}{s_\alpha} \left(e^{s_\alpha \Delta T} - 1\right) \approx 1 + e^{s_1 T} \end{aligned} \tag{16.13}$$

As is usual for the the fixed level trace sums, the convergence of (16.12) is controled by the gap between the leading and the next-to-leading eigenvalues of the evolution operator.

16.3 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

$$C_{AB}(t; x_0) = \lim_{T \to \infty} \frac{1}{T} \int_0^T d\tau A(x(\tau + t)) B(x(\tau))\,, \qquad x_0 = x(0) \tag{16.14}$$

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

$$C_{AB}(t) = \int_{\mathcal{M}} dx_0\, \varrho(x_0) A(f^t(x_0)) B(x_0)\,. \tag{16.15}$$

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

$$\hat{C}_{AB} := C_{AB} - \langle A\rangle \langle B\rangle \tag{16.16}$$

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 (16.15) 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. 7.1 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

$$\tilde{C}_{AB}(t) = \sum_{\alpha \neq 0} e^{-s_\alpha t} c_\alpha \int_{\mathcal{M}} dy\, A(y)\varphi_\alpha(y). \tag{16.17}$$

16.2
page 317

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 (16.17) 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.

16.4 Trace formulas *vs.* level sums

Trace formulas (11.9) and (11.19) diverge precisely where one would like to use them, at s equal to eigenvalues s_α. Instead, one can proceed as follows; according to (11.23) 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

$$1 = \sum_{i \in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)} e^{-s_{(n)} n}}{|\Lambda_i|} \tag{16.18}$$

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 (16.2) 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 (16.8): $s_{(n)} \to s_{(n)} - \ln \alpha / n$. This can be partially fixed by defining a level average

$$\left\langle e^{\beta A(s)} \right\rangle_{(n)} := \sum_{i \in \mathrm{Fix} f^n} \frac{e^{\beta A^n(x_i)} e^{sn}}{|\Lambda_i|} \tag{16.19}$$

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 (16.18), 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 (15.5) 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.

16.4.1 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. 10.1) and the equipartition measure for the ith region $\mathcal{M}_i$ is simply

$$\Delta\mu_i = 1/K_n \to e^{-nh} . \tag{16.20}$$

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, or even a scaling function describing a non–wandering set in the parameter space (sect. 21.3).

Commentary

Remark 16.1 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 18), the measure develops a square-root singularity on the $\overline{0}$ cycle:

$$\mu_0 = \frac{1}{|\Lambda_0|^{1/2}} \,. \tag{16.21}$$

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 [H.19, H.15, H.12, 9.23].

Remark 16.2 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 (16.13) seems to have be discussed for the first time by Hannay and Ozorio de Almeida [H.1, 6.11]. Another novelty of the cycle averaging formulas and one of their main virtues, in contrast to the explicit analytic results such as those of ref. [15.4], 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 16.3 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 [H.21], and a more mathematical argument is presented in [H.22].

Remark 16.4 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 16.5 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 (16.2) to the spectral determinants and dynamical zeta functions. One might be tempted to believe that given a deterministic rule, a sum like (16.2) could be evaluated to any desired precision. For short finite times this is indeed true: every region $\mathcal{M}_i$ in (16.2) 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 (16.2) 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 (16.2) 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

[16.1] F. Christiansen, G. Paladin and H.H. Rugh, *Phys. Rev. Lett.* **65**, 2087 (1990).

Exercises

Exercise 16.1 **Escape rate of the logistic map.**

(a) *Calculate the fraction of trajectories remaining trapped in the interval* $[0,1]$ *for the logistic map*

$$f(x) = a(1 - 4(x - 0.5)^2), \tag{16.22}$$

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* $a_c = 1$*?*

Exercise 16.2 **Four scale map decay.** *Compute the second largest eigenvalue of the Perron-Frobenius operator for the four scale map*

$$f(x) = \begin{cases} a_1 x & \text{if } 0 < x < b/a_1, \\ (1-b)((x - b/a_1)/(b - b/a_1)) + b & \text{if } b/a_1 < x < b, \\ a_2(x - b) & \text{if } b < x < b + b/a_2, \\ (1-b)((x - b - b/a_2)/(1 - b - b/a_2)) + b & \text{if } b + b/a_2 < x < 1. \end{cases} \tag{16.23}$$

Exercise 16.3 **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 15.2).*

(e) *for the 2-branch flow conserving map*

$$f_0(x) = \frac{h-p+\sqrt{(h-p)^2+4hx}}{2h}, \qquad x \in [0,p] \tag{16.24}$$
$$f_1(x) = \frac{h+p-1+\sqrt{(h+p-1)^2+4h(x-p)}}{2h}, \qquad x \in [p,1]$$

This is a nonlinear perturbation of ($h = 0$) *Bernoulli map (13.10); the first 15 eigenvalues of the Perron-Frobenius operator are listed in ref. [16.1] 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 (8.23) (hint: rescale* δx *every so often, to avoid numerical overflows),*

(g) *iterated stability (8.27).*

How good is the numerical accuracy compared with the periodic orbit theory predictions?

Chapter 17

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

(G. Vattay)

In the preceding chapters we characterized chaotic systems via global quantities such as averages. It turned out that these are closely related to very fine details of the dynamics like stabilities and time periods of individual periodic orbits. In statistical mechanics a similar duality exists. Macroscopic systems are characterized with thermodynamic quantities (pressure, temperature and chemical potential) which are averages over fine details of the system called microstates. One of the greatest achievements of the theory of dynamical systems was when in the sixties and seventies Bowen, Ruelle and Sinai made the analogy between these two subjects explicit. Later this "Thermodynamic Formalism" of dynamical systems became widely used 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.

17.1 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 $s_1 s_2 ... s_n$. We can associate many different quantities to these sets. There are geometric measures such as the volume $V(s_1 s_2 ... s_n)$, the area $A(s_1 s_2 ... s_n)$ or the length $l(s_1 s_2 ... s_n)$ of this set. Or in general we can have some measure $\mu(\mathcal{M}_{s_1 s_2 ... s_n}) = \mu(s_1 s_2 ... s_n)$ of this set. As we have seen in (16.10) the most important is the natural measure, which is the probability that a non-periodic trajectory visits the set $\mu(s_1 s_2 ... s_n) = P(s_1 s_2 ... s_n)$. The natural measure is additive. Summed up for all possible symbol sequences of length n it gives the measure of the whole phase space:

$$\sum_{s_1 s_2 ... s_n} \mu(s_1 s_2 ... s_n) = 1 \tag{17.1}$$

expresses probability conservation. Also, summing up for the last symbol we get the measure of a one step shorter sequence

$$\sum_{s_n} \mu(s_1 s_2 ... s_n) = \mu(s_1 s_2 ... s_{n-1}).$$

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

$$\lambda(s_1 s_2 ... s_n) = -\frac{1}{n} \log \mu(s_1 s_2 ... s_n). \tag{17.2}$$

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

$$Z_n(\beta) = \sum_{s_1 s_2 ... s_n} \mu^\beta(s_1 s_2 ... s_n), \tag{17.3}$$

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 $E(s_1 s_2 ... s_n) = -\log \mu(s1s_2 ... s_n)$ 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 [H.6] defined by the growth rate of the partition sum

$$K_\beta = \lim_{n\to\infty} \frac{1}{n} \frac{1}{1-\beta} \log \left(\sum_{s_1 s_2 ... s_n} \mu^\beta (s_1 s_2 ... s_n) \right). \tag{17.4}$$

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 (17.2). 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

$$g'(\lambda^*) = \beta. \tag{17.5}$$

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 (17.4) we can now connect the Rényi entropies and $g(\lambda)$

$$(\beta - 1)K_\beta = \lambda^*(\beta)\beta - g(\lambda^*(\beta)). \tag{17.6}$$

Equations (17.5) and (17.6) 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

$$g(\lambda) = \lambda\beta^*(\lambda) - (\beta^*(\lambda) - 1)K_{\beta^*(\lambda)}, \tag{17.7}$$

where now $\beta^*(\lambda)$ can be determined from

$$\frac{d}{d\beta^*}[(\beta^* - 1)K_{\beta^*}] = \lambda. \tag{17.8}$$

Obviously, if we can determine the Rényi entropies we can recover the distribution of probabilities from (17.7) and (17.8).

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 (16.10)

$$\mu(s_1, ..., s_n) \approx \frac{e^{n\gamma}}{|\Lambda_{\overline{s_1 s_2 \ldots s_n}}|}. \tag{17.9}$$

The partition sum (17.3) now reads

$$Z_n(\beta) \approx \sum_i \frac{e^{n\beta\gamma}}{|\Lambda_i|^\beta}, \tag{17.10}$$

where the summation goes for periodic orbits of length n. We can define the characteristic function

$$\Omega(z, \beta) = \exp\left(-\sum_n \frac{z^n}{n} Z_n(\beta)\right). \tag{17.11}$$

According to (17.4) for large n the partition sum behaves as

$$Z_n(\beta) \sim e^{-n(\beta-1)K_\beta}. \tag{17.12}$$

Substituting this into (17.11) 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 (17.10) into (17.11) and introducing prime and repeated periodic orbits as usual we get

$$\Omega(z,\beta) = \exp\left(-\sum_{p,r}\frac{z^{n_p r} e^{\beta\gamma n_p r}}{r|\Lambda_p^r|^\beta}\right).$$

We can see that the characteristic function is the same as the zeta function we introduced for Lyapunov exponents (H.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

$$P(\beta) = (\beta-1)K_\beta + \beta\gamma, \tag{17.13}$$

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 (H.4).

From (17.13) we can get all the important quantities of the thermodynamic formalism. For $\beta = 0$ we get the topological entropy

$$P(0) = -K_0 = -h_{top}. \tag{17.14}$$

For $\beta = 1$ we get the escape rate

$$P(1) = \gamma. \tag{17.15}$$

Taking the derivative of (17.13) in $\beta = 1$ we get Pesin's formula [H.2] connecting Kolmogorov's entropy and the Lyapunov exponent

$$P'(1) = \overline{\lambda} = K_1 + \gamma. \tag{17.16}$$

17.1 page 330

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. 17.2 we show a typical pressure and $g(\lambda)$ curve computed for the two scale tent map of Exercise 17.4. 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.

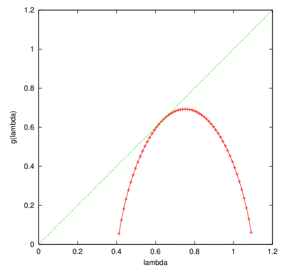


Figure 17.1:

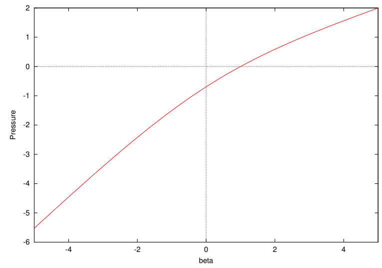


Figure 17.2: $g(\lambda)$ and $P(\beta)$ for the map of Exercise 17.4 at $a = 3$ and $b = 3/2$. See Solutions O for calculation details.

17.2 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 ith 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 [H.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

$$Z_\epsilon(q) = \sum_i \mu_i^q. \tag{17.17}$$

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 [H.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 (17.17) 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 [H.17] of the distribution, hence its name is *information dimension*.

Using equisize grids is impractical in most of the applications. Instead, we can rewrite (17.17) into the more convenient form

$$\sum_i \frac{\mu_i^q}{\epsilon^\tau(q)} \sim 1. \tag{17.18}$$

If we cover the ith branch of the fractal with a grid of size l_i instead of ϵ we can use the relation [H.9]

$$\sum_i \frac{\mu_i^q}{l_i{}^\tau(q)} \sim 1, \tag{17.19}$$

the non-uniform grid generalization of 17.18. 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 (17.19) into the form

$$\sum_i \frac{e^{q\gamma n}}{|\Lambda_i|^{q-\tau(q)}} \sim 1, \tag{17.20}$$

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 (17.20) 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

$$P(q-(q-1)D_q) = q\gamma. \tag{17.21}$$

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 (17.21) 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 (17.21) 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

$$D_1 = 1 - \gamma/\overline{\lambda}. \tag{17.22}$$

If the system is bound ($\gamma = 0$) the information dimension and all other dimensions are $D_q = 1$. Also since $D_1 0$ is positive (17.22) proves that the Lyapunov exponent must be larger than the escape rate $\overline{\lambda} > \gamma$ in general.

17.4 page 331
17.5 page 331
17.6 page 331

Commentary

Remark 17.1 Mild phase transition. In non-hyperbolic systems the formulas derived in this chapter should be modified. As we mentioned in 16.1 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 (17.10) while in (17.3)

$e^{\gamma n}/|\Lambda_0|^{\delta\beta}$ 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 (17.10).

Remark 17.2 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 (17.10) and (17.3). 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 17.3 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.

References

[17.1] J. Balatoni and A. Renyi, *Publi. Math. Inst. Hung. Acad.Sci.* **1**, 9 (1956); (english translation **1**, 588 (Akademia Budapest, 1976)).

[17.2] Ya.B. Pesin, *Uspekhi Mat. Nauk* **32,** 55 (1977), [*Russian Math. Surveys* **32,** 55 (1977)].

[17.3] Even though the thermodynamic formalism is of older vintage (we refer the reader to ref. [17.4] for a comprehensive overview), we adhere here to the notational conventions of ref. [17.5] which are more current in the physics literature: we strongly recommend also ref. [17.6], dealing with period doubling universality.

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

[17.5] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, *Phys. Rev.* **A107**, 1141 (1986).

[17.6] E. B. Vul, Ya. G. Sinai, and K. M. Khanin, *Uspekhi Mat. Nauk.* **39**, 3 (1984).

[17.7] C. Shannon, *Bell System Technical Journal,* **27**, 379 (1948).

[17.8] V.I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, (Addison-Wesley, Redwood City 1989)

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

[17.10] A.N. Kolmogorov, *Dokl.Akad.Nauk.* **124**, 754 (1959)

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

[17.12] C.M. Bender and S.A. Orszag S.A, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, Singapore 1978)

[17.13] J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617

[17.14] O. Biham and M. Kvale, *Phys. Rev.* **A 46**, 6334 (1992).

Exercises

Exercise 17.1 Thermodynamics in higher dimensions *Introduce the time averages of the eigenvalues of the Jacobian*

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log |\Lambda_i^t(x_0)|, \tag{17.23}$$

as a generalization of (8.27).

Show that in higher dimensions Pesin's formula is

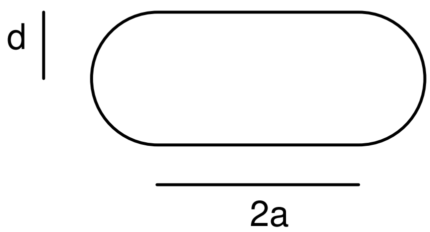
$$K_1 = \sum_i \lambda_i - \gamma, \tag{17.24}$$

where the summation goes for the positive λ_i-s only. (Hint: Use the higher dimensional generalization of (16.10)

$$\mu_i = e^{n\gamma}/|\prod_j \Lambda_{i,j}|,$$

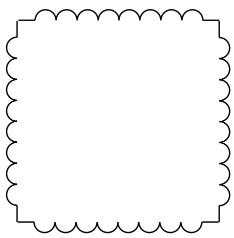
where the product goes for the expanding eigenvalues of the Jacobian of the periodic orbit.

Exercise 17.2 Bunimovich stadium Kolmogorov entropy. *Take for definitiveness $a = 1.6$ and $d = 1$ in the Bunimovich stadium of exercise 5.4,*



estimate the Lyapunov exponent by averaging over a very long trajectory. Biham and Kvale [17.14] estimate the discrete time Lyapunov to $\lambda \approx 1.0 \pm .1$, the continuous time Lyapunov to $\lambda \approx 0.43 \pm .02$, the topological entropy (for their symbolic dynamics) $h \approx 1.15 \pm .03$.

Exercise 17.3 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*

$$K_1 \to \frac{2 \ln L}{\pi L}\,.$$

Exercise 17.4 **Two scale map** *Compute all those quantities - dimensions, escape rate, entropies, etc. - for the repeller of the one dimensional map*

$$f(x) = \begin{cases} 1 + ax & \text{if} \quad x < 0, \\ 1 - bx & \text{if} \quad x > 0. \end{cases} \tag{17.25}$$

where a and b are larger than 2. Compute the fractal dimension, plot the pressure and compute the $f(\alpha)$ spectrum of singularities.

Exercise 17.5 **Four scale map** *Compute the Rényi entropies and $g(\lambda)$ for the four scale map*

$$f(x) = \begin{cases} a_1 x & \text{if} \quad 0 < x < b/a_1, \\ (1-b)((x - b/a_1)/(b - b/a_1)) + b & \text{if} \quad b/a_1 < x < b, \\ a_2(x - b) & \text{if} \quad b < x < b + b/a_2, \\ (1-b)((x - b - b/a_2)/(1 - b - b/a_2)) + b & \text{if} \quad b + b/a_2 < x < 1. \end{cases} \tag{17.26}$$

Hint: Calculate the pressure function and use (17.13).

Exercise 17.6 **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 18

Intermittency

Sometimes They Come Back
Stephen King

(R. Artuso, P. Dahlqvist, G. Tanner and P. Cvitanović)

In the theory of chaotic dynamics developed so far we assumed that the evolution operators have discrete spectra $\{z_0, z_1, z_2, \ldots\}$ given by the zeros of

$$1/\zeta(z) = (\cdots)\prod_k (1 - z/z_k)\,.$$

The assumption was based on the tacit premise that the dynamics is everywhere exponentially unstable. Real life is nothing like that - phase spaces are generically infinitely interwoven patterns of stable and unstable behaviors. The stable (in the case of Hamiltonian flows, integrable) orbits do not communicate with the ergodic components of the phase space, and can be treated by classical methods. In general, one is able to treat the dynamics near stable orbits as well as chaotic components of the phase space dynamics well within a periodic orbit approach. Problems occur at the broderline between chaos and regular dynamics where marginally stable orbits and manifolds present difficulties and still unresolved challenges.

We shall use the simplest example of such behavior - intermittency in 1-dimensional maps - to illustrate effects of marginal stability. The main message will be that spectra of evolution operators are no longer discrete, dynamical zeta functions exhibit branch cuts of the form

$$1/\zeta(z) = (\cdots) + (1 - z)^{\alpha}(\cdots)\,,$$

and correlations decay no longer exponentially, but as power laws.

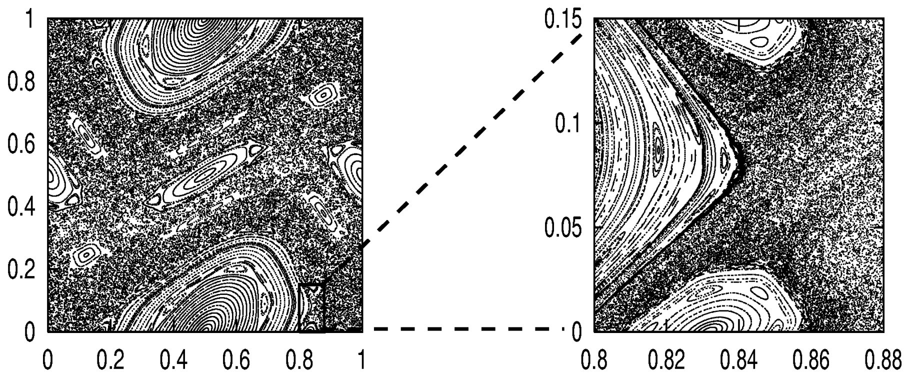


Figure 18.1: Typical phase space for an area-preserving map with mixed phase space dynamics; here the standard map for k=1.2.

18.1 Intermittency everywhere

In many fluid dynamics experiments one observes transitions from regular behaviors to behaviors where long time intervals of regular behavior ("laminar phases") are interrupted by fast irregular bursts. The closer the parameter is to the onset of such bursts, the longer are the intervals of regular behavior. The distributions of laminar phase intervals are well described by power laws.

This phenomenon is called *intermittency*, and it is a very general aspect of dynamics, a shadow cast by non-hyperbolic, marginally stable phase space regions. Complete hyperbolicity assumed in (11.5) 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. 18.1. Wherever stable islands are interspersed with chaotic regions, trajectories which come close to the stable islands can stay 'glued' for arbitrarily long times. These intervals of regular motion are interrupted by irregular bursts as the trajectory is re-injected into the chaotic part of the phase space. How the trajectories are precisely 'glued' to the marginally stable region is often hard to describe. What coarsely looks like a border of an island will under magnification dissolve into infinities of island chains of decreasing sizes, broken tori and bifurcating orbits, as illustrated in fig. 18.1.

Intermittency is due to the existence of fixed points and cycles of marginal stability (4.35), or (in studies of the onset of intermittency) to the proximity of a nearly marginal complex or unstable orbits. In Hamiltonian systems intermittency goes hand in hand with the existence of (marginally stable)

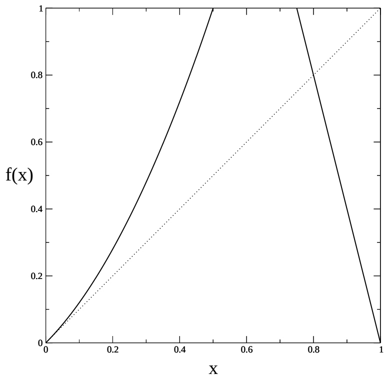


Figure 18.2: A complete binary repeller with a marginal fixed point.

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. 10.2). 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 bifurcations - a stable/unstable pair of periodic orbits is destroyed or created by coalescing into a marginal orbit, so the pruning and the intermittency discussed here are two sides of the same coin.

How to deal with the full complexity of a typical Hamiltonian system with mixed phase space is a very difficult, still open problem. Nevertheless, it is possible to learn quite a bit about intermittency by considering rather simple examples. Here we shall restrict our considerations to 1-dimensional maps which in the neighborhood of a single marginally stable fixed point at x=0 take the form

$$x \mapsto f(x) = x + O(x^{1+s}) \,, \tag{18.1}$$

and are expanding everywhere else. Such a map may allow for escape, like the map shown in fig. 18.2 or the dynamics may be bounded, like the Farey map (15.27) 163,164c153,154

$$x \mapsto f(x) = \left\{ \begin{array}{ll} x/(1-x) & x \in [0, 1/2[\\ (1-x)/x & x \in [1/2, 1] \end{array} \right.$$

introduced in sect. 15.4.

Fig. 18.3 compares a trajectory of the tent map (9.10) side by side with a trajectory of the Farey map. In a stark contrast to the uniformly chaotic trajectory of the tent map, the Farey map trajectory alternates

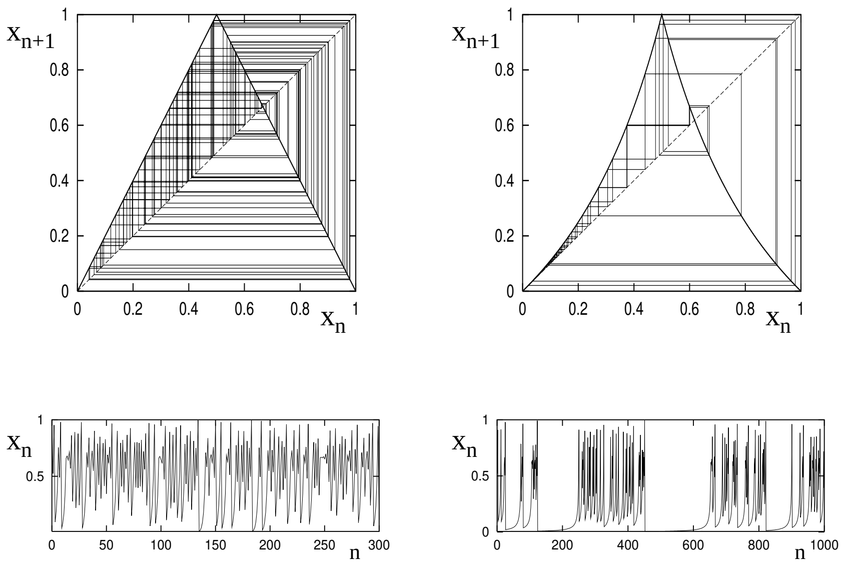


Figure 18.3: (a) A tent map trajectory. (b) A Farey map trajectory.

intermittently between slow regular motion close to the marginally stable fixed point, and chaotic bursts.

sect. 15.4.3

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. 18.2 may be algebraic rather than exponential. In long time explorations of the dynamics intermittency manifests itself by enhancement of natural measure in the proximity of marginally stable cycles.

The questions we shall address here are: how does marginal stability affect zeta functions or spectral determinants? And, can we deduce power law decays of correlations from cycle expansions?

In sect. 13.2.2 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 (11.3) or the spectral determinant (12.3) diverges for marginal orbits with $|\Lambda_p| = 1$ tells us that we have to treat these orbits with care.

In the following we will incorporate marginal stability orbits into cycle-expansions in a systematic manner. To get to know the difficulties lying ahead, we will start in sect. 18.2 with a piecewise linear map, with the asymptotics (18.1). We will construct a dynamical zeta function in the usual way without worrying too much about its justification and show that it has a branch cut singularity. We will calculate the rate of escape from our piecewise linear map and find that it is characterized by decay, rather than exponential decay, a power law. We will show that dynamical zeta functions in the presence of marginal stability can still be written in terms of periodic orbits, exactly as in chapters 8 and 16, with one exception: the marginally stable orbits have to be explicitly excluded. This innocent looking step has far reaching consequences; it forces us to change the symbolic dynamics from a finite to an infinite alphabet, and entails a reorganization of the order of summations in cycle expansions, sect. 18.2.4.

Branch cuts are typical also for smooth intermittent maps with isolated marginally stable fixed points and cycles. In sect. 18.3, we discuss the cycle expansions and curvature combinations for zeta functions of smooth maps tailored to intermittency. The knowledge of the type of singularity one encounters enables us to develop the efficient resummation method presented in sect. 18.3.1.

Finally, in sect. 18.4, we discuss a probabilistic approach to intermittency that yields approximate dynamical zeta functions and provides valuable information about more complicated systems, such as billiards.

18.2 Intermittency for pedestrians

Intermittency does not only present us with a large repertoire of interesting dynamics, it is also at the root of many sorrows such as slow convergence of cycle expansions. In order to get to know the kind of problems which arise when studying dynamical zeta functions in the presence of marginal stability we will consider an artfully concocted piecewise linear model first. From there we will move on to the more general case of smooth intermittant maps, sect. 18.3.

18.2.1 A toy map

The Bernoulli shift map (13.10) is an idealized, but highly instructive, example of a hyperbolic map. To study intermittency we will now construct a likewise piecewise linear model, an intermittent map stripped down to its bare essentials.

Consider a map $x \mapsto f(x)$ on the unit interval $\mathcal{M} = [0,1]$ with two monotone branches

$$f(x) = \begin{cases} f_0(x) & \text{for } x \in \mathcal{M}_0 = [0,a] \\ f_1(x) & \text{for } x \in \mathcal{M}_1 = [b,1] \end{cases} . \tag{18.2}$$

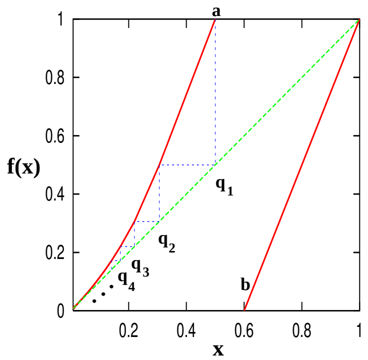


Figure 18.4: A piecewise linear intermittent map of (18.2) type: more specifically, the map piecewise linear over intervals (18.8) of the toy example studied below, $a = .5$, $b = .6$, $s = 1.0$.

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. 18.2 and fig. 18.4). We take the right branch to be expanding and linear:

$$f_1(x) = \frac{1}{1-b}(x-b)\,.$$

Next, we will construct the left branch in a way, which will allow us to model the intermittent behavior (18.1) 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

$$\mathcal{M}_n =]q_n, q_{n-1}] \qquad \text{and} \qquad \mathcal{M}_0 = \bigcup_{n=2}^{\infty} \mathcal{M}_n. \tag{18.3}$$

The map $f_0(x)$ is now specified by the following requirements

- $f_0(x)$ is continuous.
- $f_0(x)$ is linear on the intervals $\mathcal{M}_n$ for $n \geq 2$.
- $f_0(q_n) = q_{n-1}$, that is $\mathcal{M}_n = f_0^{-n+1}([a,1])$.

This fixes the map for any given sequence $\{q_n\}$. The last condition ensures the existence of a simple Markov partition. The slopes of the various linear segments are

$$\begin{aligned}
f_0'(x) &= \tfrac{f_0(q_{n-1})-f_0(q_n)}{q_{n-1}-q_n} = \tfrac{|\mathcal{M}_{n-1}|}{|\mathcal{M}_n|} && \text{for } x \in \mathcal{M}_n,\ n \geq 3 \\
f_0'(x) &= \tfrac{f_0(q_1)-f_0(q_2)}{q_1-q_2} = \tfrac{1-a}{|\mathcal{M}_2|} && \text{for } x \in \mathcal{M}_2 \\
f_0'(x) &= \tfrac{1}{1-b} = \tfrac{|\mathcal{M}|}{|\mathcal{M}_1|} && \text{for } x \in \mathcal{M}_1
\end{aligned} \tag{18.4}$$

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 (18.1). An orbit 10^n enters the intervals $\mathcal{M}_1 \to \mathcal{M}_{n+1} \to \mathcal{M}_n \to \ldots \to \mathcal{M}_2$ successively and the family approaches the marginal stable fixed point at $x = 0$ for $n \to \infty$. The stability of a cycle 10^n for $n \geq 1$ is given by the chain rule (4.30),

$$\Lambda_{10^n} = f_0'(x_{n+1}) f_0'(x_n) \ldots f_0'(x_2) f_1'(x_1) = \frac{1}{|\mathcal{M}_{n+1}|} \frac{1-a}{1-b}\,, \tag{18.5}$$

with $x_i \in \mathcal{M}_i$.

The properties of the map (18.2) are completely determined by the sequence $\{q_n\}$. By choosing $q_n = 2^{-n}$, for example, we recover the uniformly hyperbolic Bernoulli shift map (13.10). An intermittent map of the form (18.3) having the asymptotic behavior (18.1) can be constructed by choosing an algebraically decaying sequence $\{q_n\}$ behaving asymptotically like

$$q_n \sim \frac{1}{n^{1/s}}, \tag{18.6}$$

where s is the intermittency exponent in (18.1). Such a partition leads to intervals whose length decreases asymptotically like a power-law, that is,

$$|\mathcal{M}_n| \sim \frac{1}{n^{1+1/s}}\,. \tag{18.7}$$

As can be seen from (18.5), the stability eigenvalues of periodic orbit families approaching the marginal fixed point, such as the 10^n family increase in turn only algebraically with the cycle length.

It may now seem natural to construct an intermittent toy map in terms of a partition $|\mathcal{M}_n| = 1/n^{1+1/s}$, that is, a partition which follows (18.7) 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 (15.27), and the anomalous diffusion of sect. 20.3. We will, however, not go along this route here; ☞ remark 20.8
instead, we will engage in a bit of reverse engineering and construct a less obvious partition which will simplify the algebra considerably later without loosing any of the key features typical for intermittent systems. We fix the intermittent toy map by specifying the intervals $\mathcal{M}_n$ in terms of Gamma functions according to

$$|\mathcal{M}_n| = \mathcal{C}\,\frac{\Gamma(n+m-1/s-1)}{\Gamma(n+m)} \qquad \text{for} \quad n \geq 2, \tag{18.8}$$

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,

$$\mathcal{C} = a \left[\sum_{n=m+1}^{\infty} \frac{\Gamma(n - 1/s)}{\Gamma(n + 1)} \right]^{-1} . \tag{18.9}$$

Using Stirling's formula for the Gamma function

$$\Gamma(z) \sim e^{-z} z^{z-1/2} \sqrt{2\pi} \, (1 + 1/12z + \ldots) \, ,$$

we verify that the intervals decay asymptotically like $n^{-(1+1/s)}$, as required by the condition (18.7).

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. 15.4 that such cycle expansion converges extremely slowly. The shadowing mechanism between orbits and pseudo-orbits fails for orbits of the form 10^n with stabilities given by (18.5), due to the marginal stability of the fixed point $\overline{0}$. It is therefore advantageous to choose as the fundamental cycles the family of orbits with code 10^n or, equivalently, switch from the finite (binary) alphabet to an infinite alphabet given by

$$10^{n-1} \rightarrow n.$$

Due to the piecewise-linear form of the map which maps intervals $\mathcal{M}_n$ exactly onto $\mathcal{M}_{n-1}$, all periodic orbits entering the left branch at least twice are canceled exactly by pseudo cycles, and the cycle expanded dynamical zeta function depends only on the fundamental series $1, 10, 100, \ldots$:

$$\begin{aligned} 1/\zeta(z) &= \prod_{p \neq 0} \left(1 - \frac{z^{n_p}}{|\Lambda_p|} \right) = 1 - \sum_{n=1}^{\infty} \frac{z^n}{|\Lambda_{10^{n-1}}|} \\ &= 1 - (1-b)z - \mathcal{C} \frac{1-b}{1-a} \sum_{n=2}^{\infty} \frac{\Gamma(n + m - 1/s - 1)}{\Gamma(n + m)} z^n \, . \end{aligned} \tag{18.10}$$

The fundamental term (15.5) consists here of an infinite sum over algebraically decaying cycle weights. The sum is divergent for $|z| \geq 1$. We will see that this behavior is due to a branch cut of $1/\zeta$ starting at $z = 1$. We need to find analytic continuations of sums over algebraically decreasing terms in (18.10). 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. 18.2.4.

18.2.2 Branch cuts

Starting from the dynamical zeta function (18.10), 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 (18.8). The sum over ratios of Gamma functions in (18.10) can be evaluated analytically by using the following identities valid for $1/s = \alpha > 0$ (the famed binomial theorem in disguise),

- α non-integer

$$(1-z)^\alpha = \sum_{n=0}^{\infty} \frac{\Gamma(n-\alpha)}{\Gamma(-\alpha)\Gamma(n+1)} z^n \tag{18.11}$$

- α integer

$$\begin{aligned}(1-z)^\alpha \log(1-z) &= \sum_{n=1}^{\alpha} (-1)^n c_n z^n \\ &+ (-1)^{\alpha+1} \alpha! \sum_{n=\alpha+1}^{\infty} \frac{(n-\alpha-1)!}{n!} z^n\end{aligned} \tag{18.12}$$

with

$$c_n = \binom{\alpha}{n} \sum_{k=0}^{n-1} \frac{1}{\alpha-k}.$$

In order to simplify the notation, we restrict the intermittency parameter to the range $1 \leq 1/s < 2$ with $[1/s] = m = 1$. All what follows can easily be generalized to arbitrary $s > 0$ using equations (18.11) and (18.12). The infinite sum in (18.10) can now be evaluated with the help of (18.11) or (18.12), that is,

$$\sum_{n=2}^{\infty} \frac{\Gamma(n-1/s)}{\Gamma(n+1)} z^n = \begin{cases} \Gamma(-\frac{1}{s}) \left[(1-z)^{1/s} - 1 + \frac{1}{s} z\right] & \text{for} \quad 1 < 1/s < 2; \\ (1-z)\log(1-z) + z & \text{for} \quad s = 1\,. \end{cases}$$

The normalization constant $\mathcal{C}$ in (18.8) can be evaluated explicitly using (18.9) and the dynamical zeta function can be given in closed form. We obtain for $1 < 1/s < 2$

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{1/s - 1}\frac{1-b}{1-a}\left((1-z)^{1/s} - 1 + \frac{1}{s}z\right). \tag{18.13}$$

and for $s = 1$,

$$1/\zeta(z) = 1 - (1-b)z - a\frac{1-b}{1-a}\left((1-z)\log(1-z) + z\right). \tag{18.14}$$

It now becomes clear why the particular choice of intervals $\mathcal{M}_n$ made in the last section is useful; by summing over the infinite family of periodic orbits 0^n1 explicitly, we have found the desired analytical continuation for the dynamical zeta function for $|z| \geq 1$. The function has a branch cut starting at the branch point $z = 1$ and running along the positive real axis. That means, the dynamical zeta function takes on different values when approaching the positive real axis for $\mathrm{Re}\, z > 1$ from above and below. The dynamical zeta function for general $s > 0$ takes on the form

$$1/\zeta(z) = 1 - (1-b)z - \frac{a}{g_s(1)}\frac{1-b}{1-a}\frac{1}{z^{m-1}}\left((1-z)^{1/s} - g_s(z)\right) \tag{18.15}$$

for non-integer s with $m = [1/s]$ and

$$1/\zeta(z) = 1-(1-b)z-\frac{a}{g_m(1)}\frac{1-b}{1-a}\frac{1}{z^{m-1}}\left((1-z)^m\log(1-z) - g_m(z)\right) \tag{18.16}$$

for $1/s = m$ integer and $g_s(z)$ are polynomials of order $m = [1/s]$ which can be deduced from (18.11) or (18.12). 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. 18.3 that branch cuts of that form are generic for 1-dimensional intermittent maps.

Branch cuts are the all important new feature of dynamical zeta functions due to intermittency. So, how do we calculate averages or escape rates of the dynamics of the map from a dynamical zeta function with branch cuts? We take ‘a learning by doing’ approach and calculate the escape from our toy map for $a < b$.

18.2.3 Escape rate

Our starting point for the calculation of the fraction of survivors after n time steps, is the integral representation (12.16)

$$\Gamma_n = \frac{1}{2\pi i}\oint_{\gamma_r^-} z^{-n}\left(\frac{d}{dz}\log\zeta^{-1}(z)\right)dz\,, \tag{18.17}$$

where the contour encircles the origin in the clockwise direction. If the contour lies inside the unit circle $|z| = 1$, we may expand the logarithmic derivative of $\zeta^{-1}(z)$ as a convergent sum over all periodic orbits. Integrals and sums can be interchanged, the integrals can be solved term by term, and the formula (11.22) is recovered. For hyperbolic maps, cycle expansion methods or other techniques may provide an analytic extension of the dynamical zeta function beyond the leading zero; we may therefore deform the original contour into a larger circle with radius R which encircles both poles and zeros of $\zeta^{-1}(z)$, see fig. 18.5(a). Residue calculus turns this into a sum over the zeros z_α and poles z_β of the dynamical zeta function, that is

$$\Gamma_n = \sum_{|z_\alpha|<R}^{\text{zeros}}\frac{1}{z_\alpha^n} - \sum_{|z_\beta|<R}^{\text{poles}}\frac{1}{z_\beta^n} + \frac{1}{2\pi i}\oint_{\gamma_R^-} dz\, z^{-n}\frac{d}{dz}\log\zeta^{-1}, \tag{18.18}$$

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 17.1 for more details.

Things change considerably in the intermittent case. The point $z = 1$ is a branch cut singularity and there exists no Taylor series expansion of ζ^{-1} around $z = 1$. Secondly, the path deformation that led us to (18.18) requires more care, as it must not cross the branch cut. When expanding the contour to large $|z|$ values, we have to deform it along the branch $\mathrm{Re}\,(z) \geq 1$, $\mathrm{Im}\,(z) = 0$ encircling the branch cut in anti-clockwise direction, see fig. 18.5(b). We will denote the detour around the cut as γ_{cut}. We may write symbolically

$$\oint_{\gamma_r} = \sum^{\text{zeros}} - \sum^{\text{poles}} + \oint_{\gamma_R} + \oint_{\gamma_{cut}}$$

where the sums include only the zeros and the poles in the area enclosed by the contours. The asymptotics is controlled by the zero, pole or cut closest to the origin.

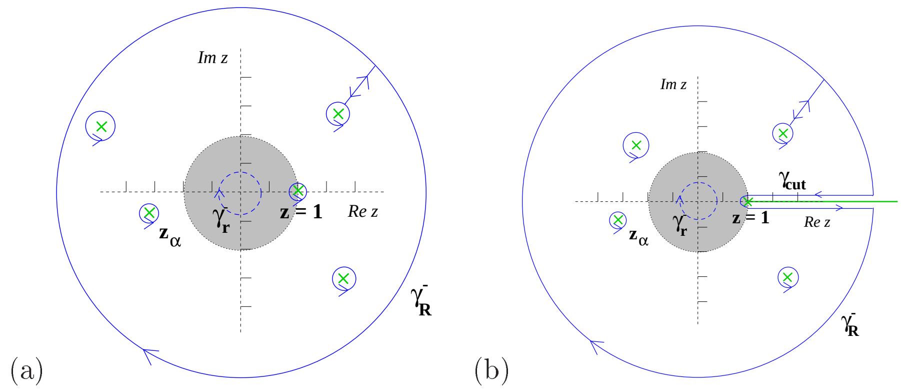


Figure 18.5: The survival probability Γ_n calculated by contour integration; integrating (18.17) inside the domain of convergence $|z| < 1$ (shaded area) of $1/\zeta(z)$ in periodic orbit representation yields (11.22). A deformation of the contour γ_r^- (dashed line) to a larger circle γ_R^- gives contributions from the poles and zeros (x) of $1/\zeta(z)$ between the two circles. These are the only contributions for hyperbolic maps (a), for intermittent systems additional contributions arise, given by the contour γ_{cut} running along the branch cut (b).

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 (18.17) at the branch point $z = 1$. We restrict ourselves again to the case $1 < 1/s < 2$ first and write the dynamical zeta function (18.13) 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

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} (1-u)^{-n} \frac{d}{du} \log G(u) du \qquad (18.19)$$

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

$$\frac{d}{du} \log G(u) = \frac{a_1}{a_0} + \frac{1}{s}\frac{b_0}{a_0} u^{1/s-1} + O(u)\,. \qquad (18.20)$$

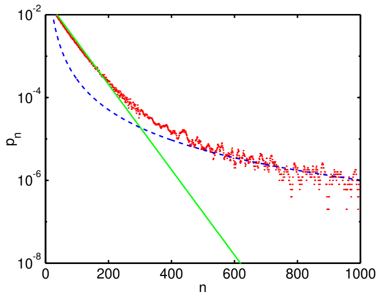


Figure 18.6: The asymptotic escape from an intermittent repeller is a power law. Normally it is preceded by an exponential, which can be related to zeros close to the cut but beyond the branch point $z = 1$, as in fig. 18.5(b).

The integrals along the cut may be evaluated using the general formula

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} u^\alpha (1-u)^{-n} du = \frac{\Gamma(n-\alpha-1)}{\Gamma(n)\Gamma(-\alpha)} \sim \frac{1}{n^{\alpha+1}}(1+O(1/n)) \tag{18.21}$$

which can be obtained by deforming the contour back to a loop around the point $u = 1$, now in positive (anti-clockwise) direction. The contour integral then picks up the $(n-1)$st term in the Taylor expansion of the function u^α at $u = 1$, cf. (18.11). For the continuous time case the corresponding formula is

$$\frac{1}{2\pi i}\oint_{\gamma_{cut}} z^\alpha e^{zt} dz = \frac{1}{\Gamma(-\alpha)}\frac{1}{t^{\alpha+1}}. \tag{18.22}$$

Plugging (18.20) into (18.19) and using (18.21) we get the asymptotic result

$$\Gamma_n \sim \frac{b_0}{a_0}\frac{1}{s}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}} = \frac{a}{s-1}\frac{1-b}{b-a}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}}. \tag{18.23}$$

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. 18.6.

For general non-integer $1/s > 0$, we write

$$1/\zeta(z) = A(u) + (u)^{1/s} B(u) \equiv G(u)$$

with $u = 1 - z$ and $A(u)$, $B(u)$ are functions analytic in a disc of radius 1 around $u = 0$. The leading terms in the Taylor series expansions of $A(u)$ and $B(u)$ are

$$a_0 = \frac{b-a}{1-a}, \quad b_0 = \frac{a}{g_s(1)}\frac{1-b}{1-a},$$

see (18.15). Expanding $\frac{d}{du}\log G(u)$ around $u = 0$, one again obtains leading order contributions according to (18.20) and the general result follows immediately using (18.21), that is,

$$\Gamma_n \sim \frac{a}{sg_s(1)}\frac{1-b}{b-a}\frac{1}{\Gamma(1-1/s)}\frac{1}{n^{1/s}}\,. \tag{18.24}$$

Applying the same arguments for integer intermittency exponents $1/s = m$, one obtains

$$\Gamma_n \sim (-1)^{m+1}\frac{a}{sg_m(1)}\frac{1-b}{b-a}\frac{m!}{n^m}\,. \tag{18.25}$$

So far, we have considered the survival probability for a repeller, that is we assumed $a < b$. The formulas (18.24) and (18.25) do obviously not apply for the case $a = b$, that is, for the bounded map. The coefficient $a_0 = (b-a)/(1-a)$ in the series representation of $G(u)$ is zero, and the expansion of the logarithmic derivative of $G(u)$ (18.20) is no longer valid. We get instead

$$\frac{d}{du}\log G(u) = \left\{ \begin{array}{ll} \frac{1}{u}\left(1 + O(u^{1/s-1})\right) & s < 1 \\ \frac{1}{u}\left(\frac{1}{s} + O(u^{1-1/s})\right) & s > 1 \end{array} \right. ,$$

assuming non-integer $1/s$ for convenience. One obtains for the survival probability.

$$\Gamma_n \sim \left\{ \begin{array}{ll} 1 + O(n^{1-1/s}) & s < 1 \\ 1/s + O(n^{1/s-1}) & s > 1 \end{array} \right. .$$

For $s > 1$, this is what we expect. There is no escape, so the survival probability is equal to 1, which we get as an asymptotic result here. The result for $s > 1$ is somewhat more worrying. It says that Γ_n defined as sum over the instabilities of the periodic orbits as in (16.12) does not tend to unity for large n. However, the case $s > 1$ is in many senses anomalous. For instance, the invariant density cannot be normalized. It is therefore not reasonable to expect that periodic orbit theories will work without complications.

18.2.4 Why does it work (anyway)?

Due to the piecewise linear nature of the map constructed in the previous section, we had the nice property that interval lengths did exactly coincide with the inverse of the stability of periodic orbits of the system, that is

$$|\mathcal{M}_n| = |/\Lambda_{10}^{n-1}\,.$$

There is thus no problem in replacing the survival probability Γ_n given by (1.2), (16.2), that is the fraction of phase space $\mathcal{M}$ surviving n iterations of the map,

$$\Gamma_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i| \,.$$

by a sum over periodic orbits of the form (11.22). The only orbit to worry about is the marginal fixed point $\overline{0}$ itself which we excluded from the zeta function (18.10).

For smooth intermittent maps, things are less clear and the fact that we had to prune the marginal fixed point is a warning sign that interval estimates by periodic orbit stabilities might go horribly wrong. The derivation of the survival probability in terms of cycle stabilities in chapter 16 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. 18.2).

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's are words $i = s_2 s_2 \ldots s_n$ of length n, and the intervals $\mathcal{M}_i$ are constructed recursively

$$\mathcal{M}_{sj} = f_s^{-1}(\mathcal{M}_j) \,, \tag{18.26}$$

where sj is the concatenation of letter s with word j of length $n_j < n$.

In what follows we will concentrate on the survival probability Γ_n, postponing other quantities of interest, such as averages, to later considerations. In establishing the equivalence of the survival probability and the periodic orbit formula for the escape rate for hyperbolic systems we have assumed that the map is expanding, with a minimal expansion rate $|f'(x)| \geq \Lambda_{\min} > 1$. This enabled us to bound the size of every survivor strip $\mathcal{M}_i$ by (16.6), the stability Λ_i of the periodic orbit i within the $\mathcal{M}_i$, and bound the survival probability by the periodic orbit sum (16.7).

The bound (16.6)

$$\mathcal{C}_1 \frac{1}{|\Lambda_i|} < \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \mathcal{C}_2 \frac{1}{|\Lambda_i|}$$

relies on hyperbolicity, and is thus indeed violated for intermittent systems. The problem is that now there is no lower bound on the expansion rate, the minimal expansion rate is $\Lambda_{\min} = 1$. The survivor strip $\mathcal{M}_{0^n}$ which includes the marginal fixed point is thus completely overestimated by $1/|\Lambda_{0^n}| = 1$ which is constant for all n.

12.7
page 233

However, bounding survival probability strip by strip is not what is required for establishing the bound (16.7). For intermittent systems a somewhat weaker bound can be established, saying that the average size of intervals *along a periodic orbit* can be bounded close to the stability of the periodic orbit for all but the interval $\mathcal{M}_{0^n}$. The weaker bound applies to averaging over each prime cycle p separately

$$\mathcal{C}_1 \frac{1}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{|\mathcal{M}_i|}{|\mathcal{M}|} < \mathcal{C}_2 \frac{1}{|\Lambda_p|}, \tag{18.27}$$

where the word i represents a code of the periodic orbit p and all its cyclic permutations. It can be shown that one can find positive constants $\mathcal{C}_1$, $\mathcal{C}_2$ independent of p. Summing over all periodic orbits leads then again to (16.7).

To study averages of multiplicative weights we follow sect. 8.1 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 (8.10)

$$\langle e^{\beta\ A^n(x)} \rangle,$$

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

$$\mathcal{C}_1 \frac{e^{\beta A_p}}{|\Lambda_p|} < \frac{1}{n_p} \sum_{i \in p} \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}_Q} e^{\beta A^n(x)} dx < \mathcal{C}_2 \frac{e^{\beta A_p}}{|\Lambda_p|}, \tag{18.28}$$

is valid for all p. After performing the above average one gets

$$\mathcal{C}_1 \Gamma_n(\beta) < \frac{1}{|\mathcal{M}|} \int_{\mathcal{M}} e^{\beta A(x,n)} dx < \mathcal{C}_2 \Gamma_n(\beta), \tag{18.29}$$

with

$$\Gamma_n(\beta) = \sum_p^n \frac{e^{\beta A_p}}{|\Lambda_p|}. \tag{18.30}$$

and a dynamical zeta function can be derived. In the intermittent case one can expect that the bound (18.28) holds using an averaging argument similar to the one discussed in (18.27). 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 .

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 13

In the intermittent case, the fixed point $\overline{0}$ does not provide any shadowing (cf. sect. J.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 (12.12), 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 expansion with - in the toy model case - vanishing curvature contributions and we can calculate dynamical properties like escape after having understood, how to handle the branch cut. But you might still be worried about leaving out the extra factor $1 - z$ all together. It turns out, that this is not only a matter of convenience, omitting the marginal $\overline{0}$ cycle is a dire necessity. The cycle weight $\Lambda_0^n = 1$ 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 behavior.

Note also, that if we are working with the spectral determinant (12.3), 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 introduces an (unphysical) essential singularity at $z = 1$ due the presence of a factor $(1 - z)^\infty$ stemming from the $\overline{0}$ cycle.

18.3 Intermittency for cyclists

Admittedly, the toy map is what is says - a toy model. The piece wise linearity of the map led to exact cancellations of the curvature contributions leaving only the fundamental terms. There are still infinitely many orbits included in the fundamental term, but the cycle weights were chosen in such a way that the zeta function could be written in closed form. For a smooth intermittent map this all will not be the case in general; still, we will argue that we have already seen almost all the fundamentally new features due to intermittency. What remains are technicalities - not necessarily easy to handle, but nothing very surprise any more.

In the following we will sketch, how to make cycle expansion techniques work for general 1-dimensional maps with a single isolated marginal fixed point. To keep the notation simple, we will consider two-branch maps with a complete binary symbolic dynamics as for example the Farey map, fig. 18.3, or the repeller depicted in fig. 18.2. We again assume that the behavior near the fixed point is given by (18.1). This implies that the stability of a family of periodic orbits approaching the marginally stable orbit, as for example the family 10^n, will increase only algebraically, that is we find again for large n

$$\frac{1}{\Lambda_{10^n}} \sim \frac{1}{n^{1+1/s}} \, ,$$

where s denotes the intermittency exponent.

When considering zeta functions or trace formulas, we again have to take out the marginal orbit $\overline{0}$; periodic orbit contributions of the form $t_{0^n 1}$

∞ – alphabet		binary alphabet				
		$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
1-cycles	n	1	10	100	1000	10000
2-cycles	mn					
	$1n$	11	110	1100	11000	110000
	$2n$	101	0101	10100	101000	1010000
	$3n$	1001	10010	100100	1001000	10010000
	$4n$	10001	100010	1000100	10001000	100010000
3-cycles	kmn					
	$11n$	111	1110	11100	111000	1110000
	$12n$	1101	11010	110100	1101000	11010000
	$13n$	11001	110010	1100100	11001000	110010000
	$21n$	1011	10110	101100	1011000	10110000
	$22n$	10101	101010	1010100	10101000	101010000
	$23n$	101001	1010010	10100100	101001000	1010010000
	$31n$	10011	100110	1001100	10011000	100110000
	$32n$	100101	1001010	10010100	100101000	1001010000
	$33n$	1001001	10010010	100100100	1001001000	10010010000

Table 18.1: Infinite alphabet versus the original binary alphabet for the shortest periodic orbit families. Repetitions of prime cycles ($11 = 1^2, 0101 = 01^2, \ldots$) and their cyclic repeats ($110 = 101, 1110 = 1101, \ldots$) are accounted for by cancellations and combination factors in the cycle expansion (18.31).

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 18.1. 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.

For generic intermittent maps, curvature contributions in the cycle expanded zeta function will not vanish exactly. The most natural way to organize the cycle expansion is to collect orbits and pseudo orbits of the same topological length with respect to the infinite alphabet. Denoting cycle weights in the new alphabet as $t_{nm\ldots} = t_{10^{n-1}10^{m-1}\ldots}$, one obtains

$$
\begin{aligned}
\zeta^{-1} &= \prod_{p\neq 0}(1-t_p) = 1 - \sum_{n=1}^{\infty} ce \\
&= 1 - \sum_{n=1}^{\infty} t_n - \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{1}{2}(t_{mn} - t_m t_n) \\
&- \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(\frac{1}{3}t_{kmn} - \frac{1}{2}t_{km}t_n + \frac{1}{6}t_k t_m t_n) - \sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \ldots .
\end{aligned}
\tag{18.31}
$$

The first sum is the fundamental term, which we have already seen in the toy model, (18.10). The curvature terms c_n in the expansion are now e-fold infinite sums where the prefactors take care of double counting of prime periodic orbits.

Let us consider the fundamental term first. For generic intermittent maps, we can not expect to obtain an analytic expression for the infinite sum of the form

$$
f(z) = \sum_{n=0}^{\infty} h_n z^n . \tag{18.32}
$$

with algebraically decreasing coefficients

$$
h_n \sim \frac{1}{n^\alpha} \quad \text{with} \quad \alpha > 0
$$

To evaluate the sum, we face the same problem as for our toy map: the power series diverges for $z > 1$, that is, exactly in the 'interesting' region where poles, zeros or branch cuts of the zeta function are to be expected. By carefully subtracting the asymptotic behavior with the help of (18.11) or (18.12), one can in general construct an analytic continuation of $f(z)$ around $z = 1$ of the form

$$
\begin{aligned}
f(z) &\sim A(z) + (1-z)^{\alpha-1} B(z) && \alpha \notin \mathbb{N} \\
f(z) &\sim A(z) + (1-z)^{\alpha-1} \ln(1-z) && \alpha \in \mathbb{N} ,
\end{aligned}
\tag{18.33}
$$

where $A(z)$ and $B(z)$ are functions analytic in a disc around $z = 1$. We thus again find that the zeta function (18.31) has a branch cut along the real axis

$\mathrm{Re}\, z \geq 1$. From here on we can switch to auto-pilot and derive algebraic escape, decay of correlation and all the rest. We find in particular that the asymptotic behavior derived in (18.24) and (18.25) is a general result, that is, the survival probability is given asymptotically by

$$\Gamma_n \sim C \frac{1}{n^{1/s}} \tag{18.34}$$

for all 1-dimensional maps of the form (18.1). We have to work a bit harder if we want more detailed information like the prefactor C, exponential precursors given by zeros or poles of the dynamical zeta function or higher order corrections. This information is buried in the functions $A(z)$ and $B(z)$ or more generally in the analytically continued zeta function. To get this analytic continuation, one may follow either of the two different strategies which we will sketch next.

18.3.1 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 (18.33) at $z = 1$. This can be done in principle, if the coefficients h_n in sums like (18.32) are known (as for our toy model). One then considers a resummation of the form

$$\sum_{j=0}^{\infty} h_j z^j = \sum_{j=0}^{\infty} a_j (1-z)^j + (1-z)^{\alpha-1} \sum_{j=0}^{\infty} b_j (1-z)^j , \tag{18.35}$$

and the coefficients a_j and b_j are obtained in terms of the h_j's by expanding $(1-z)^j$ and $(1-z)^{j+\alpha-1}$ on the right hand side around $z = 0$ using (18.11) and equating the coefficients.

In practical calculations one often has only a finite number of coefficients h_j, $0 \leq j \leq N$, which may have been obtained by finding periodic orbits and their stabilities numerically. One can still design a resummation scheme for the computation of the coefficients a_j and b_j in (18.35). We replace the infinite sums in (18.35) by finite sums of increasing degrees n_a and n_b, and require that

$$\sum_{i=0}^{n_a} a_i (1-z)^i + (1-z)^{\alpha-1} \sum_{i=0}^{n_b} b_i (1-z)^i = \sum_{i=0}^{N} h_i z^i + O(z^{N+1}) \ . \tag{18.36}$$

One proceeds again by expanding the right hand side around $z = 0$, skipping all powers z^{N+1} and higher, and then equating coefficients. It is natural to require that $|n_b + \alpha - 1 - n_a| < 1$, so that the maximal powers of the two sums in (18.36) are adjacent. If one chooses $n_a + n_b + 2 = N + 1$, then, for each cutoff length N, the integers n_a and n_b are uniquely determined from a linear system of equations. The price we pay is that the so obtained coefficients depend on the cutoff N. One can now study convergence of the coefficients a_j, and b_j, with respect to increasing values of N, or various quantities derived from a_j and b_j. Note that the leading coefficients a_0 and b_0 determine the prefactor C in (18.34), cf. (18.23). The resummed expression can also be used to compute zeros, inside or outside the radius of convergence of the cycle expansion $\sum h_j z^j$.

The scheme outlined in this section tacitly assumes that a representation of form (18.33) holds in a disc of radius 1 around $z = 1$. Convergence is improved further if additional information about the asymptotics of sums like (18.32) is used to improve the ansatz (18.35).

18.3.2 Analytical continuation by integral transformations

We will now introduce a method which provides an analytic continuation of sums of the form (18.32) without explicitly relying on an ansatz (18.35). The main idea is to rewrite the sum (18.32) as a sum over integrals with the help of the Poisson summation formula and find an analytic continuation of each integral by contour deformation. In order to do so, we need to know the n dependence of the coefficients $h_n \equiv h(n)$ explicitly for all n. If the coefficients are not known analytically, one may proceed by approximating the large n behavior in the form

$$h(n) = n^{-\alpha}(C_1 + C_2 n^{-1} + \ldots)\,, \qquad n \neq 0\,,$$

and determine the constants C_i numerically from periodic orbit data. By using the Poisson resummation identity

$$\sum_{n=-\infty}^{\infty} \delta(x - n) = \sum_{m=-\infty}^{\infty} \exp(2\pi i m x)\,, \tag{18.37}$$

we may write the sum as (18.32)

$$f(z) = \frac{1}{2}h(0) + \sum_{m=-\infty}^{\infty} \int_0^{\infty} dx\, e^{2\pi i m x} h(x) z^x. \tag{18.38}$$

The continuous variable x corresponds to the discrete summation index n and it is convenient to write $z = r\exp(i\sigma)$ from now on. The integrals are still not convergent for $r > 0$, but an analytical continuation can be found by considering the contour integral, where the contour goes out along the real axis, makes a quarter circle to either the positive or negative imaginary axis and goes back to zero. By letting the radius of the circle go to infinity, we essentially rotate the line of integration from the real onto the imaginary axis. For the $m = 0$ term in (18.38), we transform $x \to ix$ and the integral takes on the form

$$\int_0^\infty dx\, h(x)\, r^x\, e^{ix\sigma} = i \int_0^\infty dx\, h(ix)\, r^{ix} e^{-x\sigma}.$$

The integrand is now exponentially decreasing for all $r > 0$ and $\sigma \neq 0$ or 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 (18.38). The real axis is then rotated according to $x \to \mathrm{sign}(m) ix$ where $\mathrm{sign}(m)$ refers to the sign of m.

$$\int_0^\infty dx\, e^{\pm 2\pi i|m|x} h(x)\, r^x e^{ix\sigma} = \pm \mathrm{i} \int_0^\infty dx\, h(\pm ix)\, r^{\pm ix} e^{-x(2\pi|m| \pm \sigma)}.$$

Changing summation and integration, we can carry out the sum over $|m|$ explicitly and one finally obtains the compact expression

$$\begin{aligned} f(z) &= \frac{1}{2} h(0) + i \int_0^\infty dx\, h(ix)\, r^{ix} e^{-x\sigma} \qquad (18.39) \\ &+ i \int_0^\infty dx\, \frac{e^{-2\pi x}}{1 - e^{-2\pi x}} \left[h(ix) r^{ix} e^{-x\sigma} - h(-ix) r^{-ix} e^{x\sigma} \right]. \end{aligned}$$

The transformation from the original sum to the two integrals in (18.39) is exact for $r \leq 1$, and provides an analytic continuation for $r > 0$. The expression (18.39) is especially useful for an efficient numerical calculations of a dynamical zeta function for $|z| > 1$, which is essential when searching for its zeros and poles.

18.3.3 Curvature contributions

So far, we have discussed only the fundamental term $\sum_{n=1}^\infty t_n$ in (18.31), and showed how to deal with such power series with algebraically decreasing coefficients. The fundamental term determines the main structure of the zeta function in terms of the leading order branch cut. Corrections to both the zeros and poles of the dynamical zeta function as well as the leading and subleading order terms in expansions like (18.33) are contained in the curvature terms in (18.31). The first curvature correction is the 2-cycle sum

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{1}{2}(t_{mn}-t_m t_n)\,,$$

with algebraically decaying coefficients which again diverge for $|z|>1$. The analytically continued curvature terms have as usual branch cuts along the positive real z axis. Our ability to calculate the higher order curvature terms depends on how much we know about the cycle weights t_{mn}. The form of the cycle stability (18.5) suggests that t_{mn} decrease asymptotically as

$$t_{mn} \sim \frac{1}{(nm)^{1+1/s}} \tag{18.40}$$

for 2-cycles, and in general for n-cycles as

$$t_{m_1 m_2 \ldots m_n} \sim \frac{1}{(m_1 m_2 \ldots m_n)^{1+1/s}}\,.$$

If we happen to know the cycle weights $t_{m_1 m_2 \ldots m_n}$ analytically, we may proceed as in sect. 18.3.2, transform the multiple sums into multiple integrals and rotate the integration contours.

We have reached the edge of what has been accomplished so far in computing and what is worth the dynamical zeta functions from periodic orbit data. In the next section, we describe a probabilistic method applicable to intermittent maps which does not rely on periodic orbits.

18.4 BER zeta functions

So far we have focused on 1-d models as the simplest setting in which to investigate dynamical implications of marginal fixed points. We now take an altogether different track and describe how probabilistic methods may be employed in order to write down approximate dynamical zeta functions for intermittent systems.

We will discuss the method in a very general setting, for a flow in arbitrary dimension. The key idea is to introduce a surface of section $\mathcal{P}$ such that all trajectories traversing this section will have spent some time both near the marginal stable fixed point and in the chaotic phase. An important quantity in what follows is (3.2), the *first return time* $\tau(x)$, or the time of flight of a trajectory starting in x to the next return to the surface of section $\mathcal{P}$. The period of a periodic orbit p intersecting the $\mathcal{P}$ section n_p times is

$$T_p = \sum_{k=0}^{n_p-1} \tau(f^k(x_p)),$$

where $f(x)$ is the Poincaré map, and $x_p \in \mathcal{P}$ is a cycle point. The dynamical zeta function (12.12)

$$1/\zeta(z,s,\beta) = \prod_p \left(1 - \frac{z^{n_p} e^{\beta A_p - sT_p}}{|\Lambda_p|}\right), \qquad A_p = \sum_{k=0}^{n_p-1} a(f^k(x_p)), \tag{18.41}$$

chapter 8

associated with the observable $a(x)$ captures the dynamics of both the flow *and* the Poincaré map. The dynamical zeta function for the flow is obtained as $1/\zeta(s,\beta) = 1/\zeta(1,s,\beta)$, and the dynamical zeta function for the discrete time Poincaré map is $1/\zeta(z,\beta) = 1/\zeta(z,0,\beta)$.

Our basic assumption will be *probabilistic*. We assume that the chaotic interludes render the consecutive *return* (or *recurrence*) *times* $T(x_i)$, $T(x_{i+1})$ and observables $a(x_i)$, $a(x_{i+1})$ effectively uncorrelated. Consider the quantity $e^{\beta A(x_0,n) - sT(x_0,n)}$ averaged over the surface of section $\mathcal{P}$. With the above probabilistic assumption the large n behavior is

$$\langle e^{\beta A(x_0,n) - sT(x_0,n)} \rangle_{\mathcal{P}} \sim \left(\int_{\mathcal{P}} e^{\beta a(x) - s\tau} \rho(x) dx \right)^n,$$

where $\rho(x)$ is the invariant density of the Poincaré map. This type of behavior is equivalent to there being only one zero $z_0(s,\beta) = \int e^{\beta a(x) - s\tau(x)} \rho(x) dx$ of $1/\zeta(z,s,\beta)$ in the z-β plane. In the language of Ruelle-Pollicott resonances this means that there is an infinite gap to the first resonance. This in turn implies that $1/\zeta(z,s,\beta)$ may be written as

remark 8.1

$$1/\zeta(z,s,\beta) = z - \int_{\mathcal{P}} e^{\beta a(x) - s\tau(x)} \rho(x) dx \quad , \tag{18.42}$$

where we have neglected a possible analytic and non-zero prefactor. The dynamical zeta function of the flow is now

$$1/\zeta(s,\beta) = 1/\zeta(1,s,\beta) = 1 - \int_{\mathcal{P}} e^{\beta a(x)} \rho(x) e^{-s\tau(x)} dx \quad . \tag{18.43}$$

Normally, the best one can hope for is a finite gap to the leading resonance of the Poincaré map. with the above dynamical zeta function only approximatively valid. As it is derived from an approximation due to Baladi, Eckmann, and Ruelle, we shall refer to it as the BER zeta function $1/\zeta_{\mathrm{BER}}(s,\beta)$ in what follows.

A central role is played by the probability distribution of return times

$$\psi(\tau) = \int_{\mathcal{P}} \delta(\tau - \tau(x))\rho(x)dx \tag{18.44}$$

20.7
page 414

The BER zeta function at $\beta = 0$ is then given in terms of the Laplace transform of this distribution

$$1/\zeta_{\mathrm{BER}}(s) = 1 - \int_0^\infty \psi(\tau)e^{-s\tau}d\tau.$$

18.5
page 362

Example 18.1 Return times for the Bernoulli map. *For the Bernoulli shift map (13.10)*

$$x \mapsto f(x) = 2x \textit{ mod } 1,$$

one easily derives the distribution of return times

$$\psi_n = \frac{1}{2^n} \qquad n \geq 1.$$

The BER zeta function becomes (by the discrete Laplace transform (11.8))

$$\begin{aligned} 1/\zeta_{BER}(z) &= 1 - \sum_{n=1}^{\infty} \psi_n z^n = 1 - \sum_{n=1}^{\infty} \frac{z^n}{2^n} \\ &= \frac{1-z}{1-z/2} = \zeta^{-1}(z)/(1-z/\Lambda_0)\,. \end{aligned} \tag{18.45}$$

Thanks to the uniformity of the piecewise linear map measure (7.13) the "approximate" zeta function is in this case the exact *dynamical zeta function, with the cycle point $\overline{0}$ pruned.*

Example 18.2 Return times for the model of sect. 18.2.1. *For the toy model of sect. 18.2.1 one gets $\psi_1 = |\mathcal{M}_1|$, and $\psi_n = |\mathcal{M}_n|(1-b)/(1-a)$, for $n \geq 2$, leading to a BER zeta function*

$$1/\zeta_{BER}(z) = 1 - z|\mathcal{M}_1| - \sum_{n=2}^{\infty} |\mathcal{M}_n| z^n,$$

which again coincides with the exact result, (18.10).

It may seem surprising that the BER approximation produces exact results in the two examples above. The reason for this peculiarity is that both these systems are piecewise linear and have complete Markov partitions. As long as the map is piecewise linear and complete, and the probabilistic approximation is exactly fulfilled, the cycle expansion curvature terms vanish. The BER zeta function and the fundamental part of a cycle expansion discussed in sect. 15.1.1 are indeed intricately related, but not identical in general. In particular, note that the BER zeta function obeys the flow conservation sum rule (16.11) by construction, whereas the fundamental part of a cycle expansion as a rule does not.

Commentary

Remark 18.1 What about the evolution operator formalism? The main virtue of evolution operators was their semigroup property (8.21). This was natural for hyperbolic systems where instabilities grow exponentially, and evolution operators capture this behavior due to their multiplicative nature. Whether the evolution operator formalism is a good way to capture the slow, power law instabilities of intermittent dynamics is less clear. The approach taken here leads us to a formulation in terms of *dynamical zeta functions* rather than spectral determinants, circumventing evolution operators altogether. It is not known if the spectral determinants formulation would yield any benefits when applied to intermittent chaos. Some results on spectral determinants and intermittency can be found in [18.2]. A useful mathematical technique to deal with isolated marginally stable fixed point is that of *inducing*, that is, replacing the intermittent map by a completely hyperbolic map with infinite alphabet and redefining the discrete time; we have used this method implicitly by changing from a finite to an infinite alphabet. We refer to refs. [18.3, 18.19] for detailed discussions of this technique, as well as applications to 1-dimensional maps.

Remark 18.2 Intermittency. Intermittency was discovered by Manneville and Pomeau [18.1] in their study of the Lorentz system. They demonstrated that in neighborhood of parameter value $r_c = 166.07$ the mean duration of the periodic motion scales as $(r - r_c)^{1/2}$. In ref. [18.5] they explained this phenomenon in terms of a 1-dimensional map (such as (18.1)) near tangent bifurcation, and classified possible types of intermittency.

Piecewise linear models like the one considered here have been studied by Gaspard and Wang [18.6]. The escape problem has here been treated following ref. [18.7], resummations following ref. [18.8]. The proof of the bound (18.27) can be found in P. Dahlqvist's notes on `www.nbi.dk/ChaosBook/extras/PDahlqvistEscape.ps.gz`.

Farey map (15.27) has been studied widely in the context of intermittent dynamics, for example in refs. [18.16, 18.17, 15.3, 18.18, L.23, 15.14, 18.2]. The Fredholm determinant and the dynamical zeta functions for the Farey map (15.27) and the related Gauss shift map (21.38) have been studied by Mayer [18.16]. 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 [18.23] for "Axiom A" hyperbolic systems.

Remark 18.3 Tauberian theorems. In this chapter we used Tauberian theorems for power series and Laplace transforms: Feller's monograph [18.9] is a highly recommended introduction to these methods.

Remark 18.4 Probabilistic methods, BER zeta functions. Probabilistic description of intermittent chaos was introduced by Geisal and Thomae [18.10]. The BER approximation studied here is inspired by Baladi, Eckmann and Ruelle [18.14], with further developments in refs. [18.13, 18.15].

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. The marginal orbits have to be omitted, and the cycle expansions now need to include families of infinitely many longer and longer unstable orbits which accumulate toward the marginally stable cycles. Correlations for such non-hyperbolic systems may decay algebraically with the decay rates controlled by the branch cuts of dynamical zeta functions. Compared to pure hyperbolic systems, the physical consequences are drastic: exponential decays are replaced by slow power-law decays, and transport properties, such as the diffusion may become anomalous.

References

[18.1] P. Manneville and Y. Pomeau, *Phys. Lett.* **75A**, 1 (1979).

[18.2] H.H. Rugh, *Inv. Math.* **135**, 1 (1999).

[18.3] 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).

[18.4] T. Prellberg, *Towards a complete determination of the spectrum of a transfer operator associated with intermittency*, *J. Phys. A* **36**, 2455 (2003).

[18.5] Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980).

[18.6] 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).

[18.7] P. Dahlqvist, *Phys. Rev. E* **60**, 6639 (1999).

[18.8] P. Dahlqvist, *J. Phys. A* **30**, L351 (1997).

[18.9] W. Feller, *An introduction to probability theory and applications, Vol. II* (Wiley, New York 1966).

[18.10] T. Geisel and S. Thomae, *Phys. Rev. Lett.* **52**, 1936 (1984).

[18.11] T. Geisel, J. Nierwetberg and A. Zacherl, *Phys. Rev. Lett.* **54**, 616 (1985).

[18.12] R. Artuso, G. Casati and R. Lombardi, *Phys. Rev. Lett.* **71**, 62 (1993).

[18.13] P. Dahlqvist, *Nonlinearity* **8**, 11 (1995).

[18.14] V. Baladi, J.-P. Eckmann and D. Ruelle, *Nonlinearity* **2**, 119 (1989).

[18.15] P. Dahlqvist, *J. Phys. A* **27**, 763 (1994).

[18.16] D.H. Mayer, *Bull. Soc. Math. France* **104**, 195 (1976).

[18.17] D. Mayer and G. Roepstorff, *J. Stat. Phys.* **47**, 149 (1987).

[18.18] D. H. Mayer, *Continued fractions and related transformations*, in ref. [9.11].

[18.19] S. Isola, *J. Stat. Phys.* **97**, 263 (1999).

[18.20] S. Isola, "On the spectrum of Farey and Gauss maps," `mp-arc 01-280`.

[18.21] B. Fornberg and K.S. Kölbig, *Math. of Computation* **29**, 582 (1975).

[18.22] A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi, *Higher trashendental functions, Vol. I* (McGraw-Hill, New York, 1953).

[18.23] D. Ruelle, *Inventiones math.* **34**, 231 (1976)

[18.24] S. Grossmann and H. Fujisaka, *Phys. Rev. A* **26**, 1779 (1982).

[18.25] R. Lombardi, Laurea thesis, Universitá degli studi di Milano (1993).

Exercises

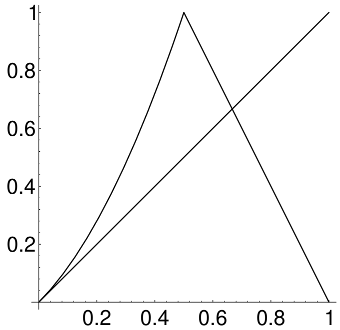
Exercise 18.1 Integral representation of Jonquière functions. *Check the integral representation*

$$J(z,\alpha) = \frac{z}{\Gamma(\alpha)}\int_0^\infty d\xi\,\frac{\xi^{\alpha-1}}{e^\xi - z} \qquad \text{for} \quad \alpha > 0\,. \tag{18.46}$$

Note how the denominator is connected to Bose-Einstein distribution. Compute $J(x+i\epsilon) - J(x-i\epsilon)$ *for a real* $x > 1$.

Exercise 18.2 Power law correction to a power law. *Expand (18.20) further and derive the leading power law correction to (18.23).*

Exercise 18.3 Power-law fall off. *In cycle expansions the stabilities of orbits do not always behave in a geometric fashion. Consider the map* f



This map behaves as $f \to x$ *as* $x \to 0$. *Define a symbolic dynamics for this map by assigning 0 to the points that land on the interval* $[0, 1/2)$ *and 1 to the points that land on* $(1/2, 1]$. *Show that the stability of orbits that spend a long time on the 0 side goes as* n^2. *In particular, show that*

$$\Lambda_{\underbrace{00\cdots0}_{n}1} \sim n^2$$

Exercise 18.4 Power law fall-off of stability eigenvalues in the stadium billiard****.** *From the cycle expansions point of view, the most important consequence of the shear in* $\mathbf{J}^n$ *for long sequences of rotation bounces* n_k *in (5.23) is that the* Λ_n *grows only as a power law in number of bounces:*

$$\Lambda_n \propto n_k^2\,. \tag{18.47}$$

Check.

Exercise 18.5 Probabilistic zeta function for maps. *Derive the probabilistic zeta function for a map with recurrence distribution* ψ_n.

Exercise 18.6 Accelerated diffusion. *Consider a map h, such that $\hat{h} = \hat{f}$, but now running branches are turner into standing branches and vice versa, so that $1, 2, 3, 4$ are standing while 0 leads to both positive and negative jumps. Build the corresponding dynamical zeta function and show that*

$$\sigma^2(t) \sim \begin{cases} t & \text{for } \alpha > 2 \\ t \ln t & \text{for } \alpha = 2 \\ t^{3-\alpha} & \text{for } \alpha \in (1,2) \\ t^2/\ln t & \text{for } \alpha = 1 \\ t^2 & \text{for } \alpha \in (0,1) \end{cases}$$

Exercise 18.7 Anomalous diffusion (hyperbolic maps). *Anomalous diffusive properties are associated to deviations from linearity of the variance of the phase variable we are looking at: this means the the diffusion constant (8.13) either vanishes or diverges. We briefly illustrate in this exercise how the local local properties of a map are crucial to account for anomalous behavior even for hyperbolic systems.*

Consider a class of piecewise linear maps, relevant to the problem of the onset of diffusion, defined by

$$f_\epsilon(x) = \begin{cases} \Lambda x & \text{for } x \in \left[0, x_1^+\right] \\ a - \Lambda_{\epsilon,\gamma}|x - x^+| & \text{for } x \in \left[x_1^+, x_2^+\right] \\ 1 - \Lambda'(x - x_2^+) & \text{for } x \in \left[x_2^+, x_1^-\right] \\ 1 - a + \Lambda_{\epsilon,\gamma}|x - x^-| & \text{for } x \in \left[x_1^-, x_2^-\right] \\ 1 + \Lambda(x-1) & \text{for } x \in \left[x_2^-, 1\right] \end{cases} \qquad (18.48)$$

where $\Lambda = (1/3 - \epsilon^{1/\gamma})^{-1}$, $\Lambda' = (1/3 - 2\epsilon^{1/\gamma})$, $\Lambda_{\epsilon,\gamma} = \epsilon^{1-1/\gamma}$, $a = 1 + \epsilon$, $x^+ = 1/3$, $x_1^+ = x^+ - \epsilon^{1/\gamma}$, $x_2^+ = x^+ + \epsilon^{1/\gamma}$, and the usual symmetry properties (20.11) are satisfied.

Thus this class of maps is characterized by two escaping windows (through which the diffusion process may take place) of size $2\epsilon^{1/\gamma}$: the exponent γ mimicks the order of the maximum for a continuous map, while piecewise linearity, besides making curvatures vanish and leading to finite cycle expansions, prevents the appearance of stable cycles. The symbolic dynamics is easily described once we consider a sequence of parameter values $\{\epsilon_m\}$, where $\epsilon_m = \Lambda^{-(m+1)}$: we then partition the unit interval though the sequence of points $0, x_1^+, x^+, x_2^+, x_1^-, x^-, x_2^-, 1$ and label the corresponding sub–intervals $1, s_a, s_b, 2, d_b, d_a, 3$: symbolic dynamics is described by an unrestricted grammar over the following set of symbols

$$\{1, 2, 3, s_\# \cdot 1^i, d_\# \cdot 3^k\} \qquad \# = a, b \quad i, k = m, m+1, m+2, \ldots$$

This leads to the following dynamical zeta function:

$$\zeta_0^{-1}(z, \alpha) = 1 - \frac{2z}{\Lambda} - \frac{z}{\Lambda'} - 4\cosh(\alpha)\epsilon_m^{1/\gamma - 1}\frac{z^{m+1}}{\Lambda^m}\left(1 - \frac{z}{\Lambda}\right)^{-1}$$

from which, by (20.8) we get

$$D = \frac{2\epsilon_m^{1/\gamma-1}\Lambda^{-m}(1-1/\Lambda)^{-1}}{1-\frac{2}{\Lambda}-\frac{1}{\Lambda'}-4\epsilon_m^{1/\gamma-1}\left(\frac{m+1}{\Lambda^m(1-1/\Lambda)}+\frac{1}{\Lambda^{m+1}(1-1/\Lambda)^2}\right)} \tag{18.49}$$

The main interest in this expression is that it allows exploring how D vanishes in the $\epsilon \mapsto 0$ ($m \mapsto \infty$) limit: as a matter of fact, from (18.49) we get the asymptotic behavior $D \sim \epsilon^{1/\gamma}$, which shows how the onset of diffusion is governed by the order of the map at its maximum.

Remark 18.5 Onset of diffusion for continuous maps. The zoology of behavior for continuous maps at the onset of diffusion is described in refs. [20.11, 20.12, 18.24]: our treatment for piecewise linear maps was introduced in ref. [18.25].

Chapter 19

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,

$$(1-t_p)^{m_p} = \prod_\alpha \det\left(1 - D_\alpha(h_{\tilde{p}})t_{\tilde{p}}\right)^{d_\alpha}, \quad t_p = t_{\tilde{p}}^{g/m_p} ,$$

where $t_{\tilde{p}}$ is the cycle weight evaluated on the 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 I.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 I below.

19.1 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.

19.1.1 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. 9.5. 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 9.1 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. 19.3.

We shall return to the 3-disk game of pinball desymmetrization in sects. 19.2.2 and 19.6, but first we develop a feeling for discrete symmetries by working out a simple 1-d example.

19.1.2 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. 19.1. 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

$$
\begin{aligned}
P_{A_1} &= \frac{1}{2}(\mathbf{e}+\mathbf{C}) \,, \qquad P_{A_2} = \frac{1}{2}(\mathbf{e}-\mathbf{C}) \,, \\
\mathcal{L}_{A_1}(y,x) &= P_{A_1}\mathcal{L}(y,x) = \frac{1}{2}\left(\mathcal{L}(y,x) + \mathcal{L}(-y,x)\right) \,, \\
\mathcal{L}_{A_2}(y,x) &= P_{A_2}\mathcal{L}(y,x) = \frac{1}{2}\left(\mathcal{L}(y,x) - \mathcal{L}(-y,x)\right) \,. \qquad (19.1)
\end{aligned}
$$

To compute the traces of the symmetrization and antisymmetrization projection operators (19.1), we have to distinguish three kinds of cycles:

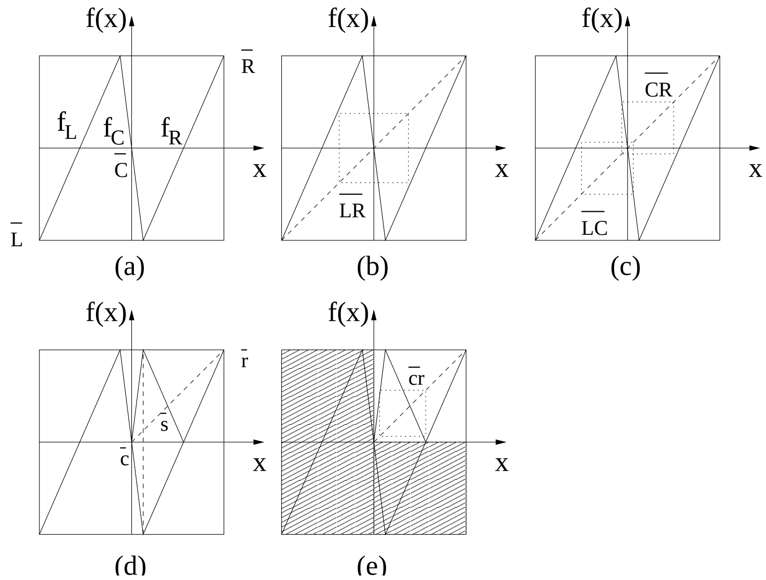


Figure 19.1: The Ulam sawtooth map with the C_2 symmetry $f(-x) = -f(x)$. (a) boundary fixed point $\overline{C}$, (b) symmetric 2-cycle $\overline{LR}$, (c) asymmetric 2-cycles pair $\{\overline{LC},\overline{CR}\}$. The Ulam sawtooth map restricted to the fundamental domain; pieces of the global map (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, symmetric cycles s built by repeats of irreducible segments $\tilde{s}$, and boundary cycles b. Now we show that the spectral determinant can be written as the product over the three kinds of cycles: $\det(1-\mathcal{L}) = \det(1-\mathcal{L})_a \det(1-\mathcal{L})_{\tilde{s}} \det(1-\mathcal{L})_b$.

Asymmetric cycles: A periodic orbits is not symmetric if $\{x_a\} \cap \{\mathbf{C}x_a\} = \emptyset$, 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 (19.1); as they are degenerate, the prefactor 1/2 cancels. Resumming as in the derivation of (12.12) 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 $\mathrm{tr}\,\mathcal{L}_{\pm}^n$:

$$\begin{aligned}
z^n \mathrm{tr}\,\mathcal{L}_{\pm}^n \quad &\to \quad \int_{\mathcal{M}_s} dx\, z^n \, \frac{1}{2} \left(\delta(x - f^n(x)) \pm \delta(x + f^n(x))\right) \\
&= \quad n_{\tilde{s}} \left(\sum_{r=2}^{\mathrm{even}} \delta_{n,rn_{\tilde{s}}} \frac{t_{\tilde{s}}^r}{1 - 1/\Lambda_{\tilde{s}}^r} \pm \sum_{r=1}^{\mathrm{odd}} \delta_{n,rn_{\tilde{s}}} \frac{t_{\tilde{s}}^r}{1 - 1/\Lambda_{\tilde{s}}^r} \right) \\
&= \quad n_{\tilde{s}} \sum_{r=1}^{\infty} \delta_{n,rn_{\tilde{s}}} \frac{(\pm t_{\tilde{s}})^r}{1 - 1/\Lambda_{\tilde{s}}^r}.
\end{aligned}$$

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 $\delta(x + f^n(x))$:

$$\begin{aligned} z^n \mathrm{tr}\, \mathcal{L}_\pm^n &\to \int_{\mathcal{M}_b} dx\, z^n \, \frac{1}{2}\, \left(\delta(x - f^n(x)) \pm \delta(x + f^n(x))\right) \\ &= \sum_{r=1}^{\infty} \delta_{n,r}\, t_b^r \frac{1}{2} \left(\frac{1}{1-1/\Lambda_b^r} \pm \frac{1}{1+1/\Lambda_b^r}\right) \\ z^n\, \mathrm{tr}\, \mathcal{L}_+^n &\to \sum_{r=1}^{\infty} \delta_{n,r} \frac{t_b^r}{1-1/\Lambda_b^{2r}}\,; \qquad z^n\, \mathrm{tr}\, \mathcal{L}_-^n \to \sum_{r=1}^{\infty} \delta_{n,r} \frac{1}{\Lambda_b^r} \frac{t_b^r}{1-1/\Lambda_b^{2r}}\,. \end{aligned}$$

Boundary orbit contributions to the factorized spectral determinants follow by resummation:

$$\det(1-\mathcal{L}_+)_b = \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k}}\right), \qquad \det(1-\mathcal{L}_-)_b = \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k+1}}\right)$$

Only 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:

$$F_+(z) = \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k}\right) \prod_{\tilde{s}} \prod_{k=0}^{\infty} \left(1 - \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k}\right) \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k}}\right)$$

$$F_-(z) = \prod_a \prod_{k=0}^{\infty} \left(1 - \frac{t_a}{\Lambda_a^k}\right) \prod_{\tilde{s}} \prod_{k=0}^{\infty} \left(1 + \frac{t_{\tilde{s}}}{\Lambda_{\tilde{s}}^k}\right) \prod_{k=0}^{\infty} \left(1 - \frac{t_b}{\Lambda_b^{2k+1}}\right) \tag{19.2}$$

We shall work out the symbolic dynamics of such reflection symmetric systems in some detail in sect. 19.5. As reflection symmetry is essentially the only discrete symmetry that a map of the interval can have, this example completes the group-theoretic factorization of determinants and zeta functions for 1-d maps. We now turn to discussion of the general case.

19.1
page 388

19.2 Discrete symmetries

A dynamical system is invariant under a symmetry group $G = \{e, g_2, \dots, g_{|G|}\}$ 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 (8.19) this means

$$
\begin{aligned}
f(x) &= \mathbf{g}^{-1} f(\mathbf{g}x) \\
\mathcal{L}(y, x) &= \mathcal{L}(\mathbf{g}y, \mathbf{g}x) \ .
\end{aligned}
\tag{19.3}
$$

Bold face letters for group elements indicate a suitable representation on 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

$$
\mathbf{C} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} .
\tag{19.4}
$$

$\mathbf{C}$ 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 (19.1). 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 $\chi_\alpha(h) = \mathrm{tr}\, D_\alpha(h)$ are the group characters, and the transfer operator $\mathcal{L}$ splits into a sum of inequivalent irreducible subspace contributions $\sum_\alpha \mathrm{tr}\, \mathcal{L}_\alpha$,

$$
\mathcal{L}_\alpha(y, x) = \frac{d_\alpha}{|G|} \sum_{h \in G} \chi_\alpha(h) \mathcal{L}(\mathbf{h}^{-1} y, x) \ .
\tag{19.5}
$$

The prefactor d_α in the above reflects the fact that a d_α-dimensional representation occurs d_α times.

19.2.1 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 $\mathcal{H}_p \subseteq G$,

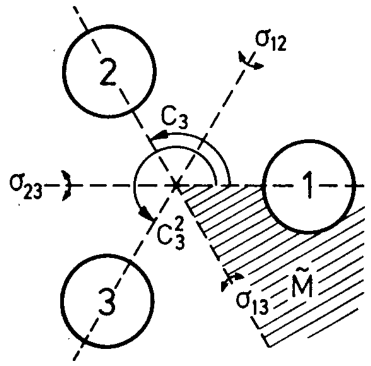


Figure 19.2: The symmetries of three disks on an equilateral triangle. The fundamental domain is indicated by the shaded wedge.

$\mathcal{H}_p = \{e, b_2, b_3, \ldots, b_h\}$ 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 (12.12) takes the form

$$\prod_p (1 - t_p) = \prod_{\hat{p}} (1 - t_{\hat{p}})^{m_{\hat{p}}}. \tag{19.6}$$

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.

19.2.2 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. 19.2. 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. 19.3a.

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.

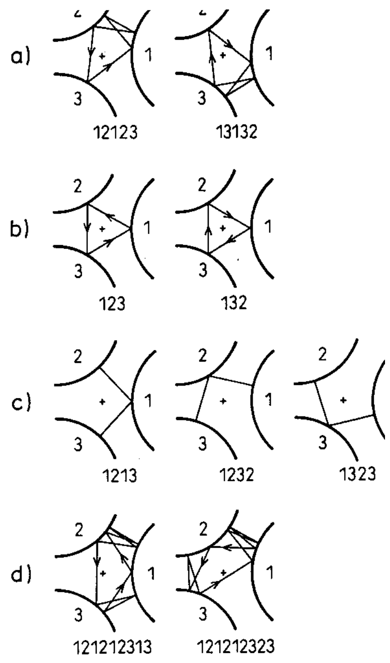


Figure 19.3: Some examples of 3-disk cycles: (a) $\overline{12123}$ and $\overline{13132}$ are mapped into each other by σ_{23}, the flip across 1 axis; this cycle has degeneracy 6 under C_{3v} symmetries. (C_{3v} is the symmetry group of the equilateral triangle.) Similarly (b) $\overline{123}$ and $\overline{132}$ 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 19. (from ref. [1.2])

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. 19.3b; $\mathcal{H}_p = \{e, C_3, C_3^2\}$, 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. 19.3(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. 19.3a.

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. 19.3(d)).

The Euler product (12.12) for the C_{3v} symmetric 3-disk problem is given in (15.32).

19.3 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$\rightarrow$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. 19.4. 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 (19.4) 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 (19.4) $h \in C_2$ and $D(h)$ can be either the identity or the interchange of the two domain labels,

$$D(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \,, \quad D(C) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \,. \tag{19.7}$$

Note that $D(h)$ is a permutation matrix, mapping a tile M_a into a different tile $M_{ha} \neq M_a$ if $h \neq e$. Consequently only $D(e)$ has diagonal elements, and $\mathrm{tr}\, D(h) = g\delta_{h,e}$. However, the 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 (8.19) a $[g \times g]$ matrix evolution operator defined by

$$\mathcal{L}_{ba}(y, x) = D(h)_{ba} \mathcal{L}(y, x) \,,$$

if $x \in M_a$ and $y \in M_b$, and zero otherwise. Now we can use the invariance condition (19.3) to move the starting point x into the fundamental domain $x = \mathbf{a}\tilde{x}$, $\mathcal{L}(y, x) = \mathcal{L}(\mathbf{a}^{-1}y, \tilde{x})$, and then use the relation $a^{-1}b = h^{-1}$ to also relate the endpoint y to its image in the fundamental domain, $\tilde{\mathcal{L}}(\tilde{y}, \tilde{x}) := \mathcal{L}(\mathbf{h}^{-1}\tilde{y}, \tilde{x})$. With this operator which is restricted to the fundamental domain, the global dynamics reduces to

$$\mathcal{L}_{ba}(y, x) = D(h)_{ba} \tilde{\mathcal{L}}(\tilde{y}, \tilde{x}) \,.$$

While the global trajectory runs over the full space M, the restricted trajectory is brought back into the fundamental domain $\tilde{M}$ any time it crosses into adjoining tiles; the two trajectories are related by the symmetry operation h which maps the global endpoint into its fundamental domain image.

Now the traces (12.3) required for the evaluation of the eigenvalues of the transfer operator can be evaluated on the fundamental domain alone

$$\mathrm{tr}\, \mathcal{L} = \int_M dx \mathcal{L}(x, x) = \int_{\tilde{M}} d\tilde{x} \, \sum_h \mathrm{tr}\, D(h) \, \mathcal{L}(\mathbf{h}^{-1}\tilde{x}, \tilde{x}) \tag{19.8}$$

The fundamental domain integral $\int d\tilde{x}\ \mathcal{L}(\mathbf{h}^{-1}\tilde{x}, \tilde{x})$ picks up a contribution from every global cycle (for which $h = e$), but it also picks up contributions from shorter segments of global cycles. The permutation matrix $D(h)$ guarantees by the identity $\mathrm{tr}\, D(h) = 0$, $h \neq e$, that only those repeats of the fundamental domain cycles $\tilde{p}$ that correspond to complete global cycles p contribute. Compare, for example, the contributions of the $\overline{12}$ and $\overline{0}$ cycles of fig. 9.5. $\mathrm{tr}\, D(h)\tilde{\mathcal{L}}$ 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

$$(1 - t_p)^{m_p} = \det\,(1 - D(h_{\tilde{p}})t_{\tilde{p}})\ , \tag{19.9}$$

valid cycle by cycle in the Euler products (12.12) 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

$$t_p = t_{\tilde{p}}^{h_p} \tag{19.10}$$

computed on the irreducible segment that coresponds to a fundamental domain cycle. For example, in fig. 9.5 we see by inspection that $t_{12} = t_0^2$ and $t_{123} = t_1^3$.

19.3.1 Boundary orbits

Before we can turn to a presentation of the factorizations of dynamical zeta functions for the different symmetries we have to discuss an effect that arises for orbits that run on a symmetry line that borders a fundamental domain. In our 3-disk example, no such orbits are possible, but they exist in other systems, such as in the bounded region of the Hénon-Heiles potential and in 1-d maps. For the symmetrical 4-disk billiard, there are in principle two kinds of such orbits, one kind bouncing back and forth between two diagonally opposed disks and the other kind moving along the other axis of reflection symmetry; the latter exists for bounded systems only. While there are typically very few boundary orbits, they tend to be among the shortest orbits, and their neglect can seriously degrade the convergence of cycle expansions, as those are dominated by the shortest cycles.

While such orbits are invariant under some symmetry operations, their neighborhoods are not. This affects the 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. 19.1.2.

Consequences for the dynamical zeta function factorizations are that sometimes a boundary orbit does not contribute. A derivation of a dynamical zeta function (12.12) from a determinant like (12.9) 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 (12.12). 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 (19.9) will be replaced by a projector. If the weights are insensitive to the signs of the eigenvalues, then this projector is

$$g_p = \frac{1}{h}\sum_{i=1}^{h} b_i \,. \tag{19.11}$$

In the cases that we have considered, the change of sign may be taken into account by defining a sign function $\epsilon_p(g) = \pm 1$, with the "-" sign if the symmetry element g flips the neigborhood. Then (19.11) is replaced by

$$g_p = \frac{1}{h}\sum_{i=1}^{h} \epsilon(b_i)\, b_i \,. \tag{19.12}$$

We have illustrated the above in sect. 19.1.2 by working out the full factorization for the 1-dimensional reflection symmetric maps.

19.4 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.

19.4.1 Factorizations of dynamical dynamical zeta functions

According to (19.9) and (19.10), the contribution of a degenerate class of global cycles (cycle p with multiplicity $m_p = g/h_p$) to a dynamical zeta function is given by the corresponding fundamental domain cycle $\tilde{p}$:

$$(1 - t_{\tilde{p}}^{h_p})^{g/h_p} = \det\,(1 - D(h_{\tilde{p}})t_{\tilde{p}}) \tag{19.13}$$

Let $D(h) = \bigoplus_\alpha d_\alpha D_\alpha(h)$ be the decomposition of the matrix representation $D(h)$ into the d_α dimensional irreducible representations α of a finite group G. Such decompositions are block-diagonal, so the corresponding contribution to the Euler product (12.9) factorizes as

$$\det(1 - D(h)t) = \prod_\alpha \det(1 - D_\alpha(h)t)^{d_\alpha}\ , \tag{19.14}$$

where now the product extends over all distinct d_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 (19.13); but from the transfer operator point of view, the key observation is that the symmetry reduces the transfer operator to a block diagonal form; this block diagonalization implies that the dynamical zeta functions (12.12) factorize as

$$\frac{1}{\zeta} = \prod_\alpha \frac{1}{\zeta_\alpha^{d_\alpha}}\ , \qquad \frac{1}{\zeta_\alpha} = \prod_{\tilde{p}} \det\,(1 - D_\alpha(h_{\tilde{p}})t_{\tilde{p}})\ . \tag{19.15}$$

Determinants of d-dimensional irreducible representations can be evaluated using the expansion of determinants in terms of traces,

$$\det(1 + M) \;\;=\;\; 1 + \mathrm{tr}\, M + \frac{1}{2}\left((\mathrm{tr}\, M)^2 - \mathrm{tr}\, M^2\right)$$

$$+\frac{1}{6}\left((\mathrm{tr}\, M)^3 - 3\,(\mathrm{tr}\, M)(\mathrm{tr}\, M^2) + 2\,\mathrm{tr}\, M^3\right)$$
$$+ \cdots + \frac{1}{d!}\left((\mathrm{tr}\, M)^d - \cdots\right) \,, \tag{19.16}$$

(see (K.26), for example) and each factor in (19.14) can be evaluated by looking up the characters $\chi_\alpha(h) = \mathrm{tr}\, D_\alpha(h)$ in standard tables [19.14]. 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.

19.2
page 388

19.4.2 Factorizations of spectral determinants

Factorization of the full spectral determinant (12.3) proceeds in essentially the same manner as the factorization of dynamical zeta functions outlined above. By (19.5) and (19.8) the trace of the transfer operator $\mathcal{L}$ splits into the sum of inequivalent irreducible subspace contributions $\sum_\alpha \mathrm{tr}\, \mathcal{L}_\alpha$, with

$$\mathrm{tr}\, \mathcal{L}_\alpha = d_\alpha \sum_{h \in G} \chi_\alpha(h) \int_{\tilde{M}} d\tilde{x}\, \mathcal{L}(\mathbf{h}^{-1}\tilde{x}, \tilde{x}) \,.$$

This leads by standard manipulations to the factorization of (12.9) into

$$\begin{aligned} F(z) &= \prod_\alpha F_\alpha(z)^{d_\alpha} \\ F_\alpha(z) &= \exp\left(-\sum_{\tilde{p}} \sum_{r=1}^{\infty} \frac{1}{r} \frac{\chi_\alpha(h_{\tilde{p}}^r) z^{n_{\tilde{p}} r}}{|\det\left(\mathbf{1} - \tilde{\mathbf{J}}_{\tilde{p}}^r\right)|}\right) \,, \end{aligned} \tag{19.17}$$

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. 19.3.1, with examples given in the next section as well as in the specific factorizations discussed below.

The factorizations (19.15), (19.17) are the central formulas of this chapter. We now work out the group theory factorizations of cycle expansions of dynamical zeta functions for the cases of C_2 and C_{3v} symmetries. The cases of the C_{2v}, C_{4v} symmetries are worked out in appendix I below.

19.5 C_2 factorization

As the simplest example of implementing the above scheme consider the C_2 symmetry. For our purposes, all that we need to know here is that each orbit or configuration is uniquely 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.

19.5 page 389

The symmetry reduced labelling $\rho_i \in \{0, 1\}$ is related to the standard $s_i \in \{+, -\}$ Ising spin labelling by

$$\begin{array}{llcll} \text{If} & s_i & = & s_{i-1} & \text{then} \quad \rho_i = 1 \\ \text{If} & s_i & \neq & s_{i-1} & \text{then} \quad \rho_i = 0 \end{array} \tag{19.18}$$

For example, $\overline{+} = \cdots + + + + \cdots$ maps into $\cdots 111 \cdots = \overline{1}$ (and so does $\overline{-}$), $\overline{-+} = \cdots - + - + \cdots$ maps into $\cdots 000 \cdots = \overline{0}$, $\overline{- + + -} = \cdots - - + + - - + + \cdots$ maps into $\cdots 0101 \cdots = \overline{01}$, and so forth. A list of such reductions is given in table 19.1.

Depending on the maximal symmetry group $\mathcal{H}_p$ that leaves an orbit p invariant (see sects. 19.2 and 19.3 as well as sect. 19.1.2), the contributions to the dynamical zeta function factor as

$$\begin{array}{rcccl} & & & A_1 \qquad A_2 & \\ \mathcal{H}_p = \{e\}: & (1 - t_{\tilde{p}})^2 & = & (1 - t_{\tilde{p}})(1 - t_{\tilde{p}}) & \\ \mathcal{H}_p = \{e, \sigma\}: & (1 - t_{\tilde{p}}^2) & = & (1 - t_{\tilde{p}})(1 + t_{\tilde{p}}) \,, & \end{array} \tag{19.19}$$

$\tilde{p}$	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 19.1: 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 (15.5). 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:

$$\begin{aligned}
1/\zeta_{A_2} &= (1 + t_0)(1 - t_1)(1 + t_{10})(1 - t_{100})(1 + t_{101})(1 + t_{1000}) \\
&\quad (1 - t_{1001})(1 + t_{1011})(1 - t_{10000})(1 + t_{10001}) \\
&\quad (1 + t_{10010})(1 - t_{10011})(1 - t_{10101})(1 + t_{10111}) \ldots \\
&= 1 + t_0 - t_1 + (t_{10} - t_1 t_0) - (t_{100} - t_{10} t_0) + (t_{101} - t_{10} t_1) \\
&\quad -(t_{1001} - t_1 t_{001} - t_{101} t_0 + t_{10} t_0 t_1) - \ldots\ldots \qquad (19.20)
\end{aligned}$$

Note that the group theory factors do not destroy the curvature corrections (the cycles and pseudo cycles are still arranged into shadowing combinations).

If the system under consideration has a boundary orbit (*cf.* sect. 19.3.1) with group-theoretic factor $\mathbf{h}_p = (\mathbf{e} + \sigma)/2$, the boundary orbit does not contribute to the antisymmetric subspace

$$A_1 \qquad A_2$$

$$\text{boundary:} \quad (1 - t_p) \;=\; (1 - t_{\tilde{p}})(1 - 0t_{\tilde{p}}) \tag{19.21}$$

This is the $1/\zeta$ part of the boundary orbit factorization of sect. 19.1.2.

19.6 C_{3v} factorization: 3-disk game of pinball

The next example, the C_{3v} symmetry, can be worked out by a glance at fig. 9.5a. 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. 9.5b). 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 9.1.

C_{3v} has two 1-dimensional irreducible representations, symmetric and antisymmetric under reflections, denoted A_1 and A_2, and a pair of degenerate 2-dimensional representations of mixed symmetry, denoted E. The contribution of an orbit with symmetry g to the $1/\zeta$ Euler product (19.14) factorizes according to

$$\det(1 - D(h)t) = (1 - \chi_{A_1}(h)t)\,(1 - \chi_{A_2}(h)t)\,\left(1 - \chi_E(h)t + \chi_{A_2}(h)t^2\right)^2 \tag{19.22}$$

with the three factors contributing to the C_{3v} irreducible representations A_1, A_2 and E, respectively, and the 3-disk dynamical zeta function factorizes into $\zeta = \zeta_{A_1}\zeta_{A_2}\zeta_E^2$. Substituting the C_{3v} characters [19.14]

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 (19.22), we obtain for the three classes of possible orbit symmetries (indicated in the first column)

$\mathbf{h}_{\tilde{p}}$ $\qquad$ A_1 $\quad$ A_2 $\quad$ E

$$\begin{aligned}
e: \quad (1-t_{\tilde{p}})^6 &= (1-t_{\tilde{p}})(1-t_{\tilde{p}})(1-2t_{\tilde{p}}+t_{\tilde{p}}^2)^2 \\
C_3, C_3^2: \quad (1-t_{\tilde{p}}^3)^2 &= (1-t_{\tilde{p}})(1-t_{\tilde{p}})(1+\ t_{\tilde{p}}+t_{\tilde{p}}^2)^2 \\
\sigma_v: \quad (1-t_{\tilde{p}}^2)^3 &= (1-t_{\tilde{p}})(1+t_{\tilde{p}})(1+0t_{\tilde{p}}-t_{\tilde{p}}^2)^2 .
\end{aligned} \tag{19.23}$$

where σ_v stands for any one of the three reflections.

The Euler product (12.12) on each irreducible subspace follows from the factorization (19.23). On the symmetric A_1 subspace the ζ_{A_1} is given by the standard binary curvature expansion (15.5). 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 (19.20). For the mixed-symmetry subspace E the curvature expansion is given by

$$\begin{aligned}
1/\zeta_E &= (1+zt_1+z^2t_1^2)(1-z^2t_0^2)(1+z^3t_{100}+z^6t_{100}^2)(1-z^4t_{10}^2) \\
&\quad (1+z^4t_{1001}+z^8t_{1001}^2)(1+z^5t_{10000}+z^{10}t_{10000}^2) \\
&\quad (1+z^5t_{10101}+z^{10}t_{10101}^2)(1-z^5t_{10011})^2 \ldots \\
&= 1+zt_1+z^2(t_1^2-t_0^2)+z^3(t_{001}-t_1t_0^2) \\
&\quad +z^4\left[t_{0011}+(t_{001}-t_1t_0^2)t_1-t_{01}^2\right] \\
&\quad +z^5\left[t_{00001}+t_{01011}-2t_{00111}+(t_{0011}-t_{01}^2)t_1+(t_1^2-t_0^2)t_{100}\right]
\end{aligned} \tag{19.24}$$

We have reinserted the powers of z in order to group together cycles and pseudocycles of the same length. Note that the factorized cycle expansions retain the curvature form; long cycles are still shadowed by (somewhat less obvious) combinations of pseudocycles.

Refering back to the topological polynomial (10.31) obtained by setting $t_p = 1$, we see that its factorization is a consequence of the C_{3v} factorization of the ζ function:

$$1/\zeta_{A_1} = 1-2z\ , \quad 1/\zeta_{A_2} = 1\ , \quad 1/\zeta_E = 1+z\ , \tag{19.25}$$

as obtained from (15.5), (19.20) and (19.24) for $t_p = 1$.

Their symmetry is $K = \{\mathbf{e}, \sigma\}$, so according to (19.11), they pick up the group-theoretic factor $\mathbf{h}_p = (\mathbf{e}+\sigma)/2$. If there is no sign change in t_p, then evaluation of $\det(1-\frac{\mathbf{e}+\sigma}{2}t_{\tilde{p}})$ yields

$$\begin{array}{llcl}
 & & & \quad A_1 \qquad\ \ A_2 \qquad\quad E \\
\text{boundary:} & (1-t_p)^3 & = & (1-t_{\tilde{p}})(1-0t_{\tilde{p}})(1-t_{\tilde{p}})^2\ , \quad t_p = t_{\tilde{p}}
\end{array} \tag{19.26}$$

However, if the cycle weight changes sign under reflection, $t_{\sigma\tilde{p}} = -t_{\tilde{p}}$, the boundary orbit does not contribute to the subspace symmetric under reflection across the orbit;

$$\begin{array}{lcl} & & \quad A_1 \qquad A_2 \qquad E \\ \text{boundary:} \quad (1-t_p)^3 & = & (1-0t_{\tilde{p}})(1-t_{\tilde{p}})(1-t_{\tilde{p}})^2 \,, \quad t_p = t_{\tilde{p}} \end{array} \tag{19.27}$$

Commentary

Remark 19.1 Some examples of systems with discrete symmetries. This chapter is based on ref. [19.1]. One has a C_2 symmetry in the Lorenz system [2.1, 19.15], the Ising model, and in the 3-dimensional anisotropic Kepler potential [27.18, 27.32, 27.33], a C_{3v} symmetry in Hénon-Heiles type potentials [19.2, 19.6, 19.7, 19.5], a C_{4v} symmetry in quartic oscillators [19.9, 19.10], in the pure x^2y^2 potential [19.11, 19.12] and in hydrogen in a magnetic field [19.13], and a $C_{2v} = C_2 \times C_2$ symmetry in the stadium billiard [19.4]. A very nice application of the symmetry factorization is carried out in ref. [19.8].

Remark 19.2 Who did it? This chapter is based on long collaborative effort with B. Eckhardt, ref. [19.1]. The group-theoretic factorizations of dynamical zeta functions that we develop here were first introduced and applied in ref. [3.10]. They are closely related to the symmetrizations introduced by Gutzwiller [27.18] in the context of the semiclassical periodic orbit trace formulas, put into more general group-theoretic context by Robbins [19.4], whose exposition, together with Lauritzen's [19.5] 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. [19.8].

Remark 19.3 Computations The techniques of this chapter have been applied to computations of the 3-disk classical and quantum spectra in refs. [1.2, 28.9], and to a "Zeeman effect" pinball and the x^2y^2 potentials in refs. [19.3, 15.12]. In a larger perspective, the factorizations developed above are special cases of a general approach to exploiting the group-theoretic invariances in spectra computations, such as those used in enumeration of periodic geodesics [19.8, 12.4, 12.14] for hyperbolic billiards [27.16] and Selberg zeta functions [22.2].

Remark 19.4 Other symmetries. In addition to the symmetries exploited here, time reversal symmetry and a variety of other nontrivial 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 19.5 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 19.6 C_2 symmetry The C_2 symmetry arises, for example, in the Lorenz system [19.15], in the 3-dimensional anisotropic Kepler problem [27.18, 27.32, 27.33] or in the cycle expansions treatments of the Ising model [19.16].

Remark 19.7 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 [19.2]

$$V(r, \theta) = \frac{1}{2}r^2 + \frac{1}{3}r^3 \sin(3\theta) \ .$$

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. 19.3.1.

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 (19.23) 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 [19.4]. In conclusion, the manifold advantages of the symmetry reduced dynamics should thus be obvious; full space cycle expansions, such as those of exercise 15.8, are useful only for cross checking purposes.

References

[19.1] P. Cvitanović and B. Eckhardt, *"Symmetry decomposition of chaotic dynamics"*, *Nonlinearity* **6**, 277 (1993).

[19.2] M. Henón and C. Heiles, *J. Astron.* **69**, 73 (1964).

[19.3] G. Russberg, (in preparation)

[19.4] J.M. Robbins, *Phys. Rev.* **A 40**, 2128 (1989).

[19.5] B. Lauritzen, Discrete symmetries and the periodic-orbit expansions, *Phys. Rev.* **A 43** 603, (1991).

[19.6] C. Jung and H.J. Scholz, *J. Phys.* **A 20**, 3607 (1987).

[19.7] C. Jung and P. Richter, *J. Phys.* **A 23**, 2847 (1990).

[19.8] N. Balasz and A. Voros, *Phys. Rep.* **143**, 109 (1986).

[19.9] B. Eckhardt, G. Hose and E. Pollak, *Phys. Rev.* **A 39**, 3776 (1989).

[19.10] C. C. Martens, R. L. Waterland, and W. P. Reinhardt, *J. Chem. Phys.* **90**, 2328 (1989).

[19.11] S.G. Matanyan, G.K. Savvidy, and N.G. Ter-Arutyunyan-Savvidy, *Sov. Phys. JETP 53*, 421 (1981).

[19.12] A. Carnegie and I. C. Percival, *J. Phys.* **A 17**, 801 (1984).

[19.13] B. Eckhardt and D. Wintgen, *J. Phys.* **B 23**, 355 (1990).

[19.14] M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, 1962).

[19.15] G. Ott and G. Eilenberger, private communication.

[19.16] R. Mainieri, Ph. D. thesis, New York University (Aug 1990); *Phys. Rev.* **A 45**, 3580 (1992).

Exercises

Exercise 19.1 **Sawtooth map desymmetrization.** *Work out the some of the shortest global cycles of different symmetries and fundamental domain cycles for the sawtooth map of fig. 19.1. Compute the dynamical zeta function and the spectral determinant of the Perron-Frobenius operator for this map; check explicitely the factorization (19.2).*

Exercise 19.2 **2-d asymmetric representation.** *The above expressions can sometimes be simplified further using standard group-theoretical methods. For example, the* $\frac{1}{2}\left((\mathrm{tr}\, M)^2 - \mathrm{tr}\, M^2\right)$ *term in (19.16) is the trace of the antisymmetric part of the* $M \times M$ *Kronecker product; if* α *is a 2-dimensional representation, this is the* A_2 *antisymmetric representation, so*

$$\text{2-dim:} \quad \det(1 - D_\alpha(h)t) = 1 - \chi_\alpha(h)t + \chi_{A_2}(h)t^2. \tag{19.28}$$

Exercise 19.3 **3-disk desymmetrization.**

a) *Work out the 3-disk symmetry factorization for the* 0 *and* 1 *cycles, i.e. which symmetry do they have, what is the degeneracy in full space and how do they factorize (how do they look in the* A_1, A_2 *and the* E *representations).*

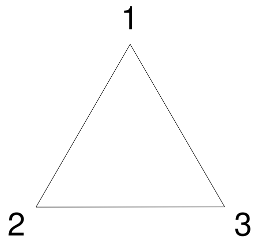
b) *Find the shortest cycle with no symmetries and factorize it like in a)*

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 (19.15) and (19.17).*

(Per Rosenqvist)

Exercise 19.4 **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.*

Exercise 19.5 **C_2 factorizations: the Lorenz and Ising systems.** *In the Lorenz system [2.1, 19.15] 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.*

Exercise 19.6 **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.

Exercise 19.7 **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 20

Deterministic diffusion

> This is a bizzare and discordant situation.
> M.V. Berry

(R. Artuso and P. Cvitanović)

The advances in the theory of dynamical systems have brought a new life to Boltzmann's mechanical formulation of statistical mechanics. Sinai, Ruelle and Bowen (SRB) have generalized Boltzmann's notion of ergodicity for a constant energy surface for a Hamiltonian system in equilibrium to dissipative systems in nonequilibrium stationary states. In this more general setting the attractor plays the role of a constant energy surface, and the SRB measure of sect. 7.1 is a generalization of the Liouville measure. Such measures are purely microscopic and indifferent to whether the system is at equilibrium, close to equilibrium or far from it. "Far for equilibrium" in this context refers to systems with large deviations from Maxwell's equilibrium velocity distribution. Furthermore, the theory of dynamical systems has yielded new sets of microscopic dynamics formulas for macroscopic observables such as diffusion constants and the pressure, to which we turn now.

We shall apply cycle expansions to the analysis of *transport* properties of chaotic systems.

The resulting formulas are exact; no probabilistic assumptions are made, and the all correlations are taken into account by the inclusion of cycles of all periods. The infinite extent systems for which the periodic orbit theory yields formulas for diffusion and other transport coefficients are spatially periodic, the global 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 2-d rotating

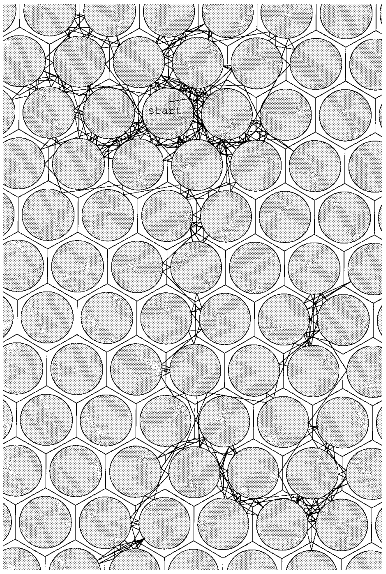


Figure 20.1: Deterministic diffusion in a finite horizon periodic Lorentz gas. (Courtesy of T. Schreiber)

flows, problems which can be described as deterministic diffusion in periodic arrays.

In sect. 20.1 we derive the formulas for diffusion coefficients in a simple physical setting, the 2-d periodic Lorentz gas. This system, however, is not the best one to exemplify the theory, due to its complicated symbolic dynamics. Therefore we apply the theory first to diffusion induced by a 1-d maps in sect. 20.2.

20.1 Diffusion in periodic arrays

The 2-d *Lorentz gas* is an infinite scatterer array in which diffusion of a light molecule in a gas of heavy scatterers is 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 is built up from only defocusing concave surfaces, it is a pure hyperbolic system, and one of the simplest nontrivial dynamical systems that exhibits deterministic diffusion, fig. 20.1. We shall now show that the *periodic* Lorentz gas is amenable to a purely deterministic treatment. In this class of open dynamical systems quantities characterizing global dynamics, such as the Lyapunov exponent, pressure and diffusion constant, can be computed from the dynamics restricted to the elementary cell. The method applies to any hyperbolic dynamical system that is a periodic tiling $\hat{\mathcal{M}} = \bigcup_{\hat{n}\in T} \mathcal{M}_{\hat{n}}$ of the dynamical 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

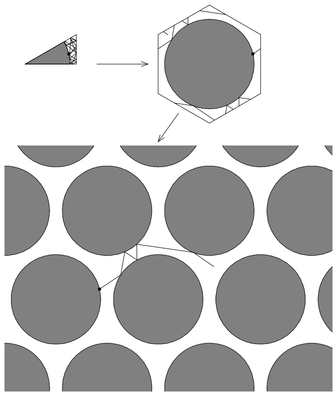


Figure 20.2: Tiling of $\hat{\mathcal{M}}$, a periodic lattice of reflecting disks, by the fundamental domain $\widetilde{\mathcal{M}}$. Indicated is an example of a global trajectory $\hat{x}(t)$ together with the corresponding elementary cell trajectory $x(t)$ and the fundamental domain trajectory $\tilde{x}(t)$. (Courtesy of J.-P. Eckmann)

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 now relate the dynamics in $\mathcal{M}$ to diffusive properties of the Lorentz gas in $\hat{\mathcal{M}}$.

These concepts are best illustrated by a specific example, a Lorentz gas based on the hexagonal lattice Sinai billiard of fig. 20.2. We distinguish two types of diffusive behavior; the *infinite horizon* case, which allows for infinite length flights, and the *finite horizon* case, where any free particle trajectory must hit a disk in finite time. 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. 20.3.

As we will work with three kinds of phase spaces, good manners require that we repeat what hats, tildas and nothings atop symbols signify:

$\tilde{}$	fundamental domain, triangle in fig. 20.2
	elementary cell, hexagon in fig. 20.2
$\hat{}$	full phase space, lattice in fig. 20.2 (20.1)

It is convenient to define an evolution operator for each of the 3 cases of fig. 20.2. $\hat{x}(t) = \hat{f}^t(\hat{x})$ 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

$$\hat{n}_t(x_0) = \hat{f}^t(x_0) - f^t(x_0) \in T\,, \tag{20.2}$$

the translation of the endpoint of the global path into the elementary cell $\mathcal{M}$. The quantity $\tilde{x}(t) = \tilde{f}^t(\tilde{x})$ denotes the flow in the fundamental domain $\widetilde{\mathcal{M}}$; $\tilde{f}^t(\tilde{x})$ is related to $f^t(\tilde{x})$ by a discrete symmetry $g \in G$ which maps $\tilde{x}(t) \in \widetilde{\mathcal{M}}$ to $x(t) \in \mathcal{M}$.

☞ chapter 19

Fix a vector $\beta \in \mathbb{R}^d$, where d is the dimension of the phase space. We will compute the diffusive properties of the Lorentz gas from the leading eigenvalue of the evolution operator (8.11)

$$s(\beta) = \lim_{t\to\infty} \frac{1}{t} \log\langle e^{\beta\cdot(\hat{x}(t)-x)}\rangle_{\mathcal{M}} \ , \tag{20.3}$$

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

$$\left.\frac{\partial}{\partial\beta_i}\frac{\partial}{\partial\beta_j}s(\beta)\right|_{\beta=0} = \lim_{t\to\infty}\frac{1}{t}\langle(\hat{x}(t)-x)_i(\hat{x}(t)-x)_j\rangle_{\mathcal{M}} \ ,$$

yield a (generally anisotropic) diffusion matrix. The spatial diffusion constant is then given by the Einstein relation (8.13)

$$D = \frac{1}{2d}\sum_i \left.\frac{\partial^2}{\partial\beta_i^2}s(\beta)\right|_{\beta=0} = \lim_{t\to\infty}\frac{1}{2dt}\langle(\hat{q}(t)-q)^2\rangle_{\mathcal{M}} \ ,$$

where the i sum is restricted to the spatial components q_i of the phase space vectors $x = (q,p)$, that is if the dynamics is Hamiltonian to the number of the degrees of freedom.

We now turn to the connection between (20.3) and periodic orbits in the elementary cell. As the full $\hat{\mathcal{M}} \to \widetilde{\mathcal{M}}$ reduction is complicated by the nonabelian nature of G, we shall introduce the main ideas in the abelian $\hat{\mathcal{M}} \to \mathcal{M}$ context.

☞ remark 20.6

20.1.1 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. 8.2, we have used the identity $1 = \int_{\mathcal{M}} dy\, \delta(y - \hat{x}(t))$ to motivate the introduction of the evolution operator $\mathcal{L}^t(\hat{y}, x)$. There is a unique lattice translation $\hat{n}$ such that $\hat{y} = y - \hat{n}$, with $y \in \mathcal{M}$, and $f^t(x)$ given by (20.2). The difference is a translation by a constant, and the Jacobian for changing integration from $d\hat{y}$ to dy equals unity. Therefore, and this is the main point, translation invariance can be used to reduce this average to the elementary cell:

$$\langle e^{\beta\cdot(\hat{x}(t)-x)}\rangle_{\mathcal{M}} = \frac{1}{|\mathcal{M}|}\int_{x,y\in\mathcal{M}} dx dy\, e^{\beta\cdot(\hat{f}^t(x)-x)}\delta(y - f^t(x)) \,. \tag{20.4}$$

As this is a translation, the Jacobian is $\delta\hat{y}/\delta y = 1$. In this way the global $\hat{f}^t(x)$ flow averages can be computed by following the flow $f^t(x_0)$ restricted to the elementary cell $\mathcal{M}$. The equation (20.4) suggests that we study the evolution operator

$$\mathcal{L}^t(y, x) = e^{\beta\cdot(\hat{x}(t)-x)}\delta(y - f^t(x)) \,, \tag{20.5}$$

where $\hat{x}(t) = \hat{f}^t(x) \in \hat{\mathcal{M}}$, but x, $f^t(x)$, $y \in \mathcal{M}$. It is straightforward to check that this operator satisfies the semigroup property (8.21), $\int_{\mathcal{M}} dz\, \mathcal{L}^{t_2}(y, z)\mathcal{L}^{t_1}(z, x) = \mathcal{L}^{t_2+t_1}(y, x)$. For $\beta = 0$, the operator (20.5) is the Perron-Frobenius operator (7.10), with the leading eigenvalue $e^{s_0} = 1$ because there is no escape from this system (this will lead to the flow conservation sum rule (16.11) later on).

The rest is old hat. The spectrum of $\mathcal{L}$ is evaluated by taking the trace ☞ sect. 11.3

$$\mathrm{tr}\, \mathcal{L}^t = \int_{\mathcal{M}} dx\, e^{\beta\cdot\hat{n}_t(x)}\delta(x - x(t)) \,.$$

Here $\hat{n}_t(x)$ is the discrete lattice translation defined in (20.2). 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 travelled 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 (12.6)

$$\det\left(s(\beta) - \mathcal{A}\right) = \prod_p \exp\left(-\sum_{r=1}^{\infty}\frac{1}{r}\frac{e^{(\beta\cdot\hat{n}_p - sT_p)r}}{\left|\det\left(\mathbf{1} - \mathbf{J}_p^r\right)\right|}\right) \,, \tag{20.6}$$

or the corresponding dynamical zeta function (12.12)

$$1/\zeta(\beta,s) = \prod_p \left(1 - \frac{e^{(\beta\cdot\hat{n}_p - sT_p)}}{|\Lambda_p|}\right) . \tag{20.7}$$

The dynamical zeta function cycle averaging formula (15.17) for the diffusion constant (8.13), zero mean drift $\langle \hat{x}_i \rangle = 0$, is given by

$$D = \frac{1}{2d}\frac{\left\langle \hat{x}^2 \right\rangle_\zeta}{\langle T \rangle_\zeta} = \frac{1}{2d}\frac{1}{\langle T \rangle_\zeta}\sum{}' \frac{(-1)^{k+1}(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1}\cdots\Lambda_{p_k}|} . \tag{20.8}$$

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 $\left\langle \hat{x}^2 \right\rangle_\zeta$ gets contributions only from the periodic runaway trajectories; they are closed in the elementary cell, but on the periodic lattice each one grows like $\hat{x}(t)^2 = (\hat{n}_p/T_p)^2 = v_p^2 t^2$. So the orbits that contribute to the trace formulas and spectral determinants exhibit either ballistic transport or no transport at all: diffusion arises as a balance between the two kinds of motion, weighted by the $1/|\Lambda_p|$ measure. If the system is not hyperbolic such weights may be abnormally large, with $1/|\Lambda_p| \approx 1/{T_p}^\alpha$ rather than $1/|\Lambda_p| \approx e^{-T_p\lambda}$, where λ is the Lyapunov exponent, and they may lead to anomalous diffusion - accelerated or slowed down depending on whether the probabilities of the running or the standing orbits are enhanced.

☞ sect. 20.3

We illustrate the main idea, tracking of a globally diffusing orbit by the associated confined orbit restricted to the elementary cell, with a class of simple 1-d dynamical systems where all transport coefficients can be evaluated analytically.

20.2 Diffusion induced by chains of 1-d maps

In a typical deterministic diffusive process, trajectories originating from a given scatterer reach a finite set of neighboring scatterers in one bounce, and then the process is repeated. As was shown in chapter 9, 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-d expanding maps. This observation motivates introduction of a class of particularly simple 1-d systems, chains of piecewise linear maps.

We start by defining the map $\hat{f}$ on the unit interval as

$$\hat{f}(\hat{x}) = \begin{cases} \Lambda\hat{x} & \hat{x} \in [0, 1/2) \\ \Lambda\hat{x} + 1 - \Lambda & \hat{x} \in (1/2, 1] \end{cases} , \qquad \Lambda > 2 , \tag{20.9}$$

and then extending the dynamics to the entire real line, by imposing the translation property

$$\hat{f}(\hat{x} + \hat{n}) = \hat{f}(\hat{x}) + \hat{n} \qquad \hat{n} \in \mathbb{Z} . \tag{20.10}$$

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. The map is antisymmetric under the $\hat{x}$-coordinate flip

$$\hat{f}(\hat{x}) = -\hat{f}(-\hat{x}) , \tag{20.11}$$

so the dynamics will exhibit no mean drift; all odd derivatives of the generating function (8.11) with respect to β, evaluated at $\beta = 0$, will vanish.

The map (20.9) is sketched in fig. 20.3(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 random walk for almost every initial condition.

The translational symmetry (20.10) 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

$$f(x) = \hat{f}(\hat{x}) - \left[\hat{f}(\hat{x})\right] , \qquad x = \hat{x} - [\hat{x}] \in [0, 1] \tag{20.12}$$

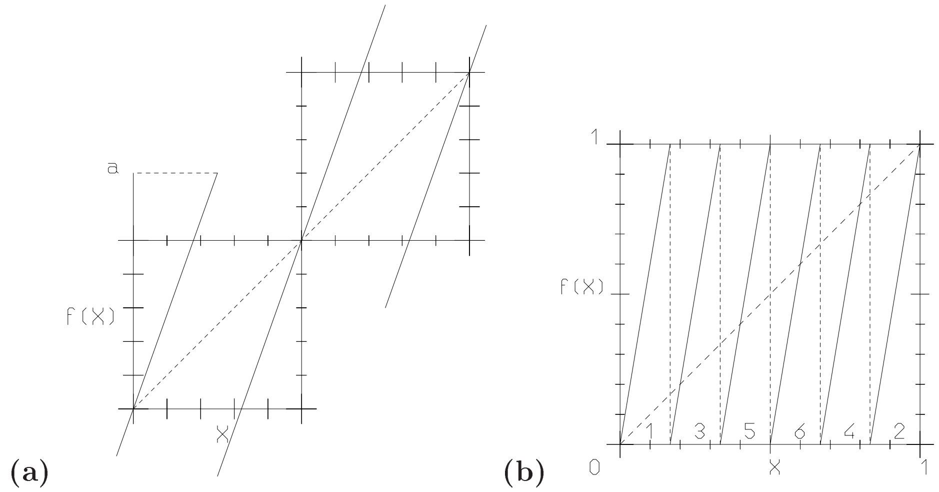


Figure 20.3: (a) $\hat{f}(\hat{x})$, the full space sawtooth map (20.9), $\Lambda > 2$. (b) $f(x)$, the sawtooth map restricted to the unit circle (20.12), $\Lambda = 6$.

fig. 20.3(b), where $[\cdots]$ stands for the integer part (20.2). As noted above, the elementary cell cycles correspond to either standing or running orbits for the map on the full line: we shall refer to $\hat{n}_p \in \mathbb{Z}$ as the *jumping number* of the p cycle, and take as the cycle weight

$$t_p = z^{n_p} e^{\beta \hat{n}_p} / |\Lambda_p| \,. \tag{20.13}$$

For the piecewise linear map of fig. 20.3 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 by a single parameter, for example its critical value $a = \hat{f}(1/2)$, the absolute maximum on the interval $[0, 1]$ related to the slope of the map by $a = \Lambda/2$. The larger Λ is, the stronger is the stretching action of the map.

The diffusion constant formula (20.8) for 1-d maps is

$$D = \frac{1}{2} \frac{\langle \hat{n}^2 \rangle_\zeta}{\langle n \rangle_\zeta} \tag{20.14}$$

where the "mean cycle time" is given by (15.18)

$$\langle n \rangle_\zeta = z \frac{\partial}{\partial z} \frac{1}{\zeta(0, z)} \bigg|_{z=1} = -\sum{}' (-1)^k \frac{n_{p_1} + \cdots + n_{p_k}}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} \,, \tag{20.15}$$

and the "mean cycle displacement squared" by (17.1)

$$\langle \hat{n}^2 \rangle_\zeta = \frac{\partial^2}{\partial \beta^2} \frac{1}{\zeta(\beta, 1)} \bigg|_{\beta=0} = -\sum{}' (-1)^k \frac{(\hat{n}_{p_1} + \cdots + \hat{n}_{p_k})^2}{|\Lambda_{p_1} \cdots \Lambda_{p_k}|} \,, \tag{20.16}$$

the primed sum indicating all distinct non-repeating combinations of prime cycles. The evaluation of these formulas in this simple system will require nothing more than pencil and paper.

20.2.1 Case of unrestricted symbolic dynamics

Whenever Λ is an integer number, the symbolic dynamics is exceedingly simple. For example, for the case $\Lambda = 6$ illustrated in fig. 20.3(b), the elementary cell map consists of 6 full branches, with uniform stretching factor $\Lambda = 6$. The branches have different jumping numbers: for branches 1 and 2 we have $\hat{n} = 0$, for branch 3 we have $\hat{n} = +1$, for branch 4 $\hat{n} = -1$, and finally for branches 5 and 6 we have respectively $\hat{n} = +2$ and $\hat{n} = -2$. The same structure reappears whenever Λ is an even integer $\Lambda = 2a$: all branches are mapped onto the whole unit interval and we have two $\hat{n} = 0$ branches, one branch for which $\hat{n} = +1$ and one for which $\hat{n} = -1$, and so on, up to the maximal jump $|\hat{n}| = a-1$. The symbolic dynamics is thus full, unrestricted shift in $2a$ symbols $\{0_+, 1_+, \ldots, (a-1)_+, (a-1)_-, \ldots, 1_-, 0_-\}$, where the symbol indicates both the length and the direction of the corresponding jump.

For the piecewise linear maps with uniform stretching the weight associated with a given symbol sequence is a product of weights for individual steps, $t_{sq} = t_s t_q$. For the map of fig. 20.3 there are 6 distinct weigths (20.13):

$$
\begin{aligned}
t_1 &= t_2 = z/\Lambda \\
t_3 &= e^{\beta} z/\Lambda\,, \quad t_4 = e^{-\beta} z/\Lambda\,, \quad t_5 = e^{2\beta} z/\Lambda\,, \quad t_6 = e^{-2\beta} z/\Lambda\,.
\end{aligned}
$$

The piecewise linearity and the simple symbolic dynamics lead to the full cancellation of all curvature corrections in (15.5). The *exact* dynamical zeta function (10.13) is given by the fixed point contributions:

$$
\begin{aligned}
1/\zeta(\beta, z) &= 1 - t_{0_+} - t_{0_-} - \cdots - t_{(a-1)_+} - t_{(a-1)_-} \\
&= 1 - \frac{z}{a}\left(1 + \sum_{j=1}^{a-1} \cosh(\beta j)\right)\,. \qquad (20.17)
\end{aligned}
$$

The leading (and only) eigenvalue of the evolution operator (20.5) is

$$
s(\beta) = \log\left\{\frac{1}{a}\left(1 + \sum_{j=1}^{a-1}\cosh(\beta j)\right)\right\}\,, \qquad \Lambda = 2a,\ a \text{ integer}\,. \qquad (20.18)
$$

The flow conservation (16.11) sum rule is manifestly satisfied, so $s(0) = 0$. The first derivative $s(0)'$ vanishes as well by the left/right symmetry of the dynamics, implying vanishing mean drift $\langle \hat{x} \rangle = 0$. The second derivative $s(\beta)''$ yields the diffusion constant (20.14):

$$\langle n \rangle_\zeta = 2a\frac{1}{\Lambda} = 1\,, \qquad \left\langle \hat{x}^2 \right\rangle_\zeta = 2\frac{0^2}{\Lambda}+2\frac{1^2}{\Lambda}+2\frac{2^2}{\Lambda}+\cdots+2\frac{(a-1)^2}{\Lambda} \tag{20.19}$$

Using the identity $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$ we obtain

$$D = \frac{1}{24}(\Lambda-1)(\Lambda-2)\,, \qquad \Lambda \text{ even integer}\,. \tag{20.20}$$

20.1 page 413

Similar calculation for odd integer $\Lambda = 2k-1$ yields

$$D = \frac{1}{24}(\Lambda^2-1)\,, \qquad \Lambda \text{ odd integer}\,. \tag{20.21}$$

20.2.2 Higher order transport coefficients

The same approach yields higher order transport coefficients

$$\mathcal{B}_k = \frac{1}{k!}\frac{d^k}{d\beta^k}s(\beta)\bigg|_{\beta=0}\,, \qquad \mathcal{B}_2 = D\,, \tag{20.22}$$

known for $k > 2$ as the Burnett coefficients. The behavior of the higher order coefficients yields information on the relaxation to the asymptotic distribution function generated by the diffusive process. Here $\hat{x}_t$ is the relevant dynamical variable and $\mathcal{B}_k$'s are related to moments $\left\langle \hat{x}_t^k \right\rangle$ of arbitrary order.

Were the diffusive process purely gaussian

$$e^{ts(\beta)} = \frac{1}{\sqrt{4\pi Dt}}\int_{-\infty}^{+\infty} d\hat{x}\, e^{\beta\hat{x}} e^{-\hat{x}^2/(4Dt)} = e^{\beta^2 Dt} \tag{20.23}$$

the only $\mathcal{B}_k$ coefficient different from zero would be $\mathcal{B}_2 = D$. Hence, nonvanishing higher order coefficients signal deviations of deterministic diffusion from a gaussian stochastic process.

For the map under consideration the first Burnett coefficient coefficient $\mathcal{B}_4$ is easily evaluated. For example, using (20.18) in the case of even integer slope $\Lambda = 2a$ we obtain

20.2 page 413

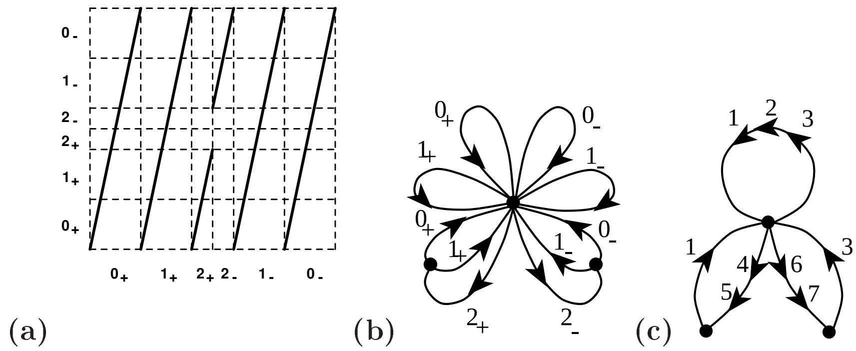


Figure 20.4: (a) A partition of the unit interval into six intervals, labeled by the jumping number $\hat{n}(x)$ $I = \{0_+, 1_+, 2_+, 2_-, 1_-, 0_-\}$. The partition is Markov, as the critical point is mapped onto the right border of $\mathcal{M}_{1_+}$. (b) The Markov graph for this partition. (c) The Markov graph in the compact notation of (20.26) (introduced by Vadim Moroz).

$$\mathcal{B}_4 = -\frac{1}{4! \cdot 60}(a-1)(2a-1)(4a^2 - 9a + 7)\,. \tag{20.24}$$

We see that deterministic diffusion is <u>not</u> a gaussian stochastic process. Higher order even coefficients may be calculated along the same lines.

20.2.3 Case of finite Markov partitions

For piecewise-linear maps exact results may be obtained whenever the critical points are mapped in finite numbers of iterations onto partition boundary points, or onto unstable periodic orbits. We will work out here an example for which this occurs in two iterations, leaving other cases as exercises.

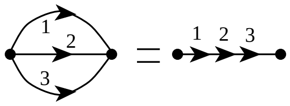
The key idea is to construct a *Markov partition* (9.4), with intervals mapped *onto* unions of intervals. As an example we determine a value of the parameter $4 \le \Lambda \le 6$ for which $f\,(f(1/2)) = 0$. As in the integer Λ case, we partition the unit interval into six intervals, labeled by the 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. 20.4(a).

In general the critical value $a = \hat{f}(1/2)$ will not correspond to an interval border, but now we choose a such that the critical point is mapped onto the right border of $\mathcal{M}_{1_+}$. Equating $f(1/2)$ with the right border of $\mathcal{M}_{1_+}$, $x = 1/\Lambda$, we obtain a quadratic equation with the expanding solution $\Lambda = 2(\sqrt{2}+1)$. For this parameter value $f(\mathcal{M}_{1_+}) = \mathcal{M}_{0_+} \bigcup \mathcal{M}_{1_+}$,

$f(\mathcal{M}_{2_-}) = \mathcal{M}_{0_-} \bigcup \mathcal{M}_{1_-}$, while the remaining intervals map onto the whole unit interval $\mathcal{M}$. The transition matrix (9.2) is given by

$$\phi' = T\phi \; = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{0_+} \\ \phi_{1_+} \\ \phi_{2_+} \\ \phi_{2_-} \\ \phi_{1_-} \\ \phi_{0_-} \end{pmatrix} . \qquad (20.25)$$

One could diagonalize (20.25) on a computer, but, as we saw in sect. 9.6, the Markov graph fig. 20.4(b) corresponding to fig. 20.4(a) offers more insight into the dynamics. The graph fig. 20.4(b) can be redrawn more compactly as Markov graph fig. 20.4(c) by replacing parallel lines in a graph by their sum



$$= t_1 + t_2 + t_3 \, . \qquad (20.26)$$

The dynamics is unrestricted in the alphabet

$$\mathcal{A} = \{0_+, \, 1_+, \, 2_+0_+, \, 2_+1_+, \, 2_-1_-, \, 2_-0_-, \, 1_-, \, 0_-\} \, .$$

Applying the loop expansion (10.13) of sect. 10.3, we are led to the dynamical zeta function

$$\begin{aligned} 1/\zeta(\beta, z) &= 1 - t_{0_+} - t_{1_+} - t_{2_+0_+} - t_{2_+1_+} - t_{2_-1_-} - t_{2_-0_-} - t_{1_-} - t_{0_-} \\ &= 1 - \frac{2z}{\Lambda}\,(1 + \cosh(\beta)) - \frac{2z^2}{\Lambda^2}\,(\cosh(2\beta) + \cosh(3\beta)) \end{aligned} \qquad (20.27)$$

For grammar as simple as this one, the dynamical zeta function is the sum over fixed points of the unrestricted alphabet. As the first check of this expression for the dynamical zeta function we verify that

$$1/\zeta(0,1) \; = 1 - \frac{4}{\Lambda} - \frac{4}{\Lambda^2} \; = \; 0 \, ,$$

as required by the flow conservation (16.11). 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 (20.35), is helpful in reducing the order of the polynomial condition that fixes Λ.

20.3
page 413

The diffusion constant follows from (20.14)

$$\begin{aligned}\langle n\rangle_\zeta &= 4\frac{1}{\Lambda}+4\frac{2}{\Lambda^2}\,, \quad \langle \hat{n}^2\rangle_\zeta = 2\frac{1^2}{\Lambda}+2\frac{2^2}{\Lambda^2}+2\frac{3^2}{\Lambda^2} \\ D &= \frac{15+2\sqrt{2}}{16+8\sqrt{2}}\,. \end{aligned} \tag{20.28}$$

It is by now clear how to build an infinite hierarchy of finite Markov partitions: tune the slope in such a way that the critical value $f(1/2)$ is mapped into the fixed point at the origin in a finite number of iterations p $f^P(1/2)=0$. By taking higher and higher values of p one constructs a dense set of Markov parameter values, organized into a hierarchy that resembles the way in which rationals are densely embedded in the unit interval. For example, each of the 6 primary intervals can be subdivided into 6 intervals obtained by the 2-nd iterate of the map, and for the critical point mapping into any of those in 2 steps the grammar (and the corresponding cycle expansion) is finite. So, if we can prove continuity of $D=D(\Lambda)$, we can apply the periodic orbit theory to the sawtooth map (20.9) for a random "generic" value of the parameter Λ, for example $\Lambda=4.5$. The idea is to bracket this value of Λ by a sequence of nearby Markov values, compute the exact diffusion constant for each such Markov partition, and study their convergence toward the value of D for $\Lambda=4.5$. Judging how difficult such problem is already for a tent map (see sect. 10.6 and appendix E.1), this is not likely to take only a week of work.

Expressions like (20.20) 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 might expect that the diffusion coefficient increases smoothly and monotonically as the slope Λ of the map (20.9) 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 illustrated in fig. 20.5. 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, even of purely hyperbolic systems such as the Lozi map and the 1-d diffusion map (20.9). 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 parameters. If you want to master the material, working through the project P.1 and/or project P.2 is strongly recommended.

20.3 Marginal stability and anomalous diffusion

What effect does the intermittency of chapter 18 have on transport properties of 1-d maps? Consider a $1-d$ map of the real line on itself with the same properties as in sect. 20.2, except for a marginal fixed point at $x = 0$.

A marginal fixed point unbalances the role of running and standing orbits, thus generating a mechanism that may result in anomalous diffusion. Our model example is the map shown in fig. 20.6(a), with the corresponding circle map shown in fig. 20.6(b). As in sect. 18.2.1, a branch with support in $\mathcal{M}_i$, $i = 1, 2, 3, 4$ has constant slope Λ_i, while $f|_{\mathcal{M}_0}$ is of intermittent form. To keep you nimble, this time we take a slightly different choice of slopes. The toy example of sect. 18.2.1 was cooked up so that the $1/s$ branch cut in dynamical zeta function was the whole answer. Here we shall take a slightly different route, and pick piecewise constant slopes such that the dynamical zeta function for intermittent system can be expressed in terms of the Jonquière function

☞ remark 20.8

$$J(z, s) = \sum_{k=1}^{\infty} z^k / k^s \tag{20.29}$$

Once the $\overline{0}$ fixed point is pruned away, the symbolic dynamics is given by the infinite alphabet $\{1, 2, 3, 4, 0^i 1, 0^j 2, 0^k 3, 0^l 4\}$, $i, j, k, l = 1, 2, \ldots$ (compare with table 18.1). The partitioning of the subinterval $\mathcal{M}_0$ is induced by $\mathcal{M}_{0^k(right)} = \phi^k_{(right)}(\mathcal{M}_3 \bigcup \mathcal{M}_4)$ (where $\phi_{(right)}$ denotes the inverse of the right branch of $\hat{f}|_{\mathcal{M}_0}$) and the same reasoning applies to the leftmost branch. These are regions over which the slope of $\hat{f}|_{\mathcal{M}_0}$ is constant. Thus we have the following stabilities and jumping numbers associated to letters:

$$\begin{array}{rll} 0^k 3,\, 0^k 4 & \Lambda_p = \frac{k^{1+\alpha}}{q/2} & \hat{n}_p = 1 \\ 0^l 1,\, 0^l 2 & \Lambda_p = \frac{l^{1+\alpha}}{q/2} & \hat{n}_p = -1 \\ 3,\, 4 & \Lambda_p = \pm\Lambda & \hat{n}_p = 1 \\ 2,\, 1 & \Lambda_p = \pm\Lambda & \hat{n}_p = -1\,, \end{array} \tag{20.30}$$

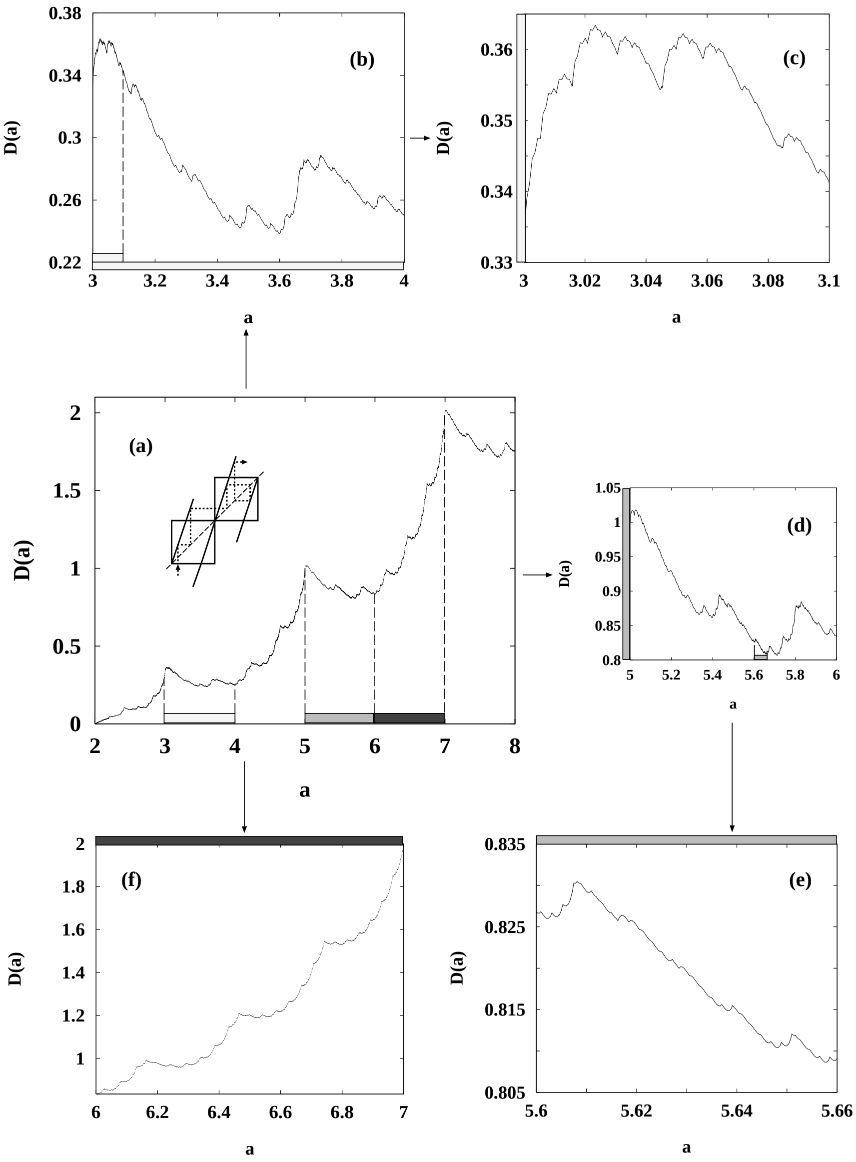


Figure 20.5: The dependence of D on the map parameter a is continuous, but not monotone. (From ref. [20.7]). Here a stands for the slope Λ in (20.9).

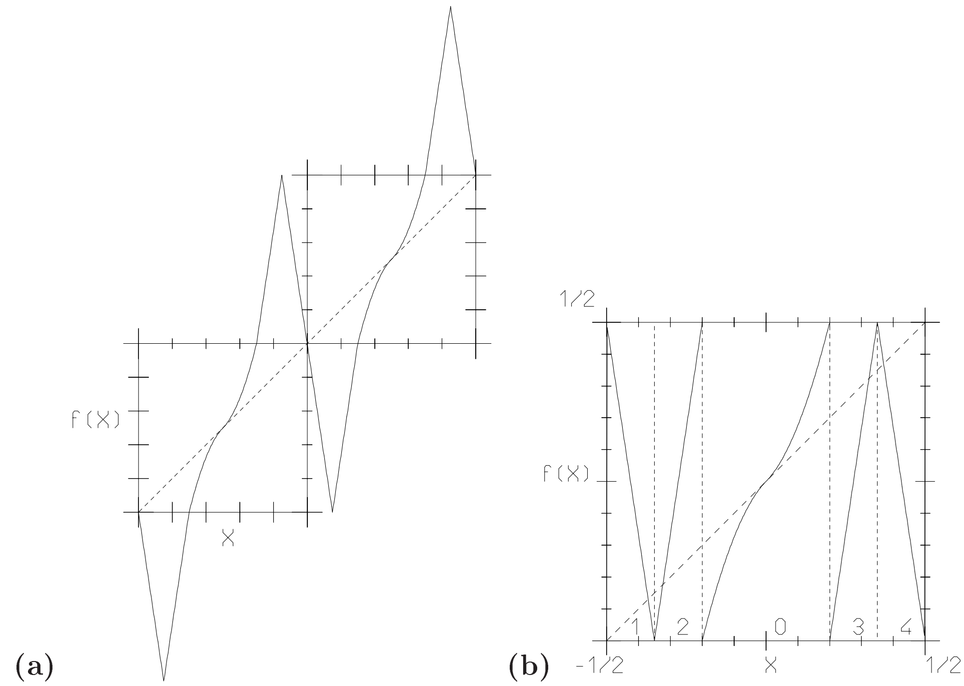


Figure 20.6: (a) A map with marginal fixed point. (b) The map restricted to the unit circle.

where $\alpha = 1/s$ is determined by the intermittency exponent (18.1), while q is to be determined by the flow conservation (16.11) for $\hat{f}$: —PCdefine R

$$\frac{4}{\Lambda} + 2q\zeta(\alpha+1) = 1$$

so that $q = (\Lambda-4)/2\Lambda\zeta(\alpha+1)$. The dynamical zeta function picks up contributions just by the alphabet's letters, as we have imposed piecewise linearity, and can be expressed in terms of a Jonguiere function (20.29):

$$1/\zeta_0(z,\beta) \;=\; 1 - \frac{4}{\Lambda}z\cosh\beta - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}z\cosh\beta\cdot J(z,\alpha+1)\,. \quad (20.31)$$

Its first zero $z(\beta)$ is determined by

$$\frac{4}{\Lambda}z + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}z\cdot J(z,\alpha+1) \;=\; \frac{1}{\cosh\beta}\,.$$

By using implicit function derivation we see that D vanishes (that is $z''(\beta)|_{\beta=1} = 0$) when $\alpha \le 1$. The physical interpretion is that a typical orbit will stick for long times near the $\overline{0}$ marginal fixed point, and the 'trapping time' will be larger for higher values of the intermittency parameter s (recall $\alpha = s^{-1}$). Hence, we need to look more closely at the behavior of traces of high powers of the transfer operator.

The evaluation of transport coefficient requires one more derivative with respect to expectation values of phase functions (see sect. 20.1): if we use the diffusion dynamical zeta function (20.7), we may write the diffusion coefficient as an inverse Laplace transform,in such a way that any distinction between maps and flows has vanished. In the case of 1-d diffusion we thus have

$$D = \lim_{t\to\infty} \frac{d^2}{d\beta^2} \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} ds\, e^{st} \frac{\zeta'(\beta,s)}{\zeta(\beta,s)}\bigg|_{\beta=0} \tag{20.32}$$

where the ζ' refers to the derivative with respect to s.

The evaluation of inverse Laplace transforms for high values of the argument is most conveniently performed using Tauberian theorems. We shall take

$$\omega(\lambda) = \int_0^\infty dx\, e^{-\lambda x} u(x)\,,$$

with $u(x)$ monotone as $x \to \infty$; then, as $\lambda \mapsto 0$ and $x \mapsto \infty$ respectively (and $\rho \in (0,\infty)$,

$$\omega(\lambda) \sim \frac{1}{\lambda^\rho} L\left(\frac{1}{\lambda}\right)$$

if and only if

$$u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x)\,,$$

where L denotes any slowly varying function with $\lim_{t\to\infty} L(ty)/L(t) = 1$. Now

$$\frac{1/{\zeta_0}'(e^{-s},\beta)}{1/\zeta_0(e^{-s},\beta)} = \frac{\left(\frac{4}{\Lambda} + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}\left(J(e^{-s},\alpha+1) + J(e^{-s},\alpha)\right)\right)\cosh\beta}{1 - \frac{4}{\Lambda}e^{-s}\cosh\beta - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}e^{-s}(e^{-s},\alpha+1)\cosh\beta J}\,.$$

We then take the double derivative with respect to β and obtain

$$\frac{d^2}{d\beta^2}\left(1/{\zeta_0}'(e^{-s},\beta)/\zeta^{-1}(e^{-s},\beta)\right)_{\beta=0}$$

$$= \frac{\frac{4}{\Lambda} + \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}\left(J(e^{-s},\alpha+1) + J(e^{-s},\alpha)\right)}{\left(1 - \frac{4}{\Lambda}e^{-s} - \frac{\Lambda-4}{\Lambda\zeta(1+\alpha)}e^{-s}J(e^{-s},\alpha+1)\right)^2} = g_\alpha(s) \qquad (20.33)$$

The asymptotic behavior of the inverse Laplace transform (20.32) may then be evaluated via Tauberian theorems, once we use our estimate for the behavior of Jonquière functions near $z = 1$. The deviations from normal behavior correspond to an explicit dependence of D on time. Omitting prefactors (which can be calculated by the same procedure) we have

$$g_\alpha(s) \sim \begin{cases} s^{-2} & \text{for } \alpha > 1 \\ s^{-(\alpha+1)} & \text{for } \alpha \in (0,1) \\ 1/(s^2 \ln s) & \text{for } \alpha = 1\,. \end{cases}$$

20.6 page 413

The anomalous diffusion exponents follow:

$$\langle (x - x_0)^2 \rangle_t \sim \begin{cases} t & \text{for } \alpha > 1 \\ t^\alpha & \text{for } \alpha \in (0,1) \\ t/\ln t & \text{for } \alpha = 1\,. \end{cases} \qquad (20.34)$$

Commentary

Remark 20.1 Lorentz gas. The original pinball model proposed by Lorentz [20.3] consisted of randomly, rather than regularly placed scatterers.

Remark 20.2 Who's dun it? Cycle expansions for the diffusion constant of a particle moving in a periodic array have been introduced independently by R. Artuso [20.4] (exact dynamical zeta function for 1-d chains of maps (20.8)), by W.N. Vance [20.5], and by P. Cvitanović, J.-P. Eckmann, and P. Gaspard [20.6] (the dynamical zeta function cycle expansion (20.8) applied to the Lorentz gas).

Remark 20.3 Lack of structural stability for D. Expressions like (20.20) 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 shown in fig. 20.5. 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 refs. [20.7, 20.8, 20.9]. 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 [20.10].

Remark 20.4 Diffusion induced by 1-d maps. We refer the reader to refs. [20.11, 20.12] for early work on the deterministic diffusion induced by 1-dimenional maps. The sawtooth map (20.9) was introduced by Grossmann and Fujisaka [20.13] who derived the integer slope formulas (20.20) for the diffusion constant. The sawtooth map is also discussed in refs. [20.14].

Remark 20.5 Symmetry factorization in one dimension. In the $\beta = 0$ limit the dynamics (20.11) is symmetric under $x \to -x$, and the zeta functions factorize into products of zeta functions for the symmetric and antisymmetric subspaces, as described in sect. 19.1.2:

$$\frac{1}{\zeta(0,z)} = \frac{1}{\zeta_s(0,z)}\,\frac{1}{\zeta_a(0,z)}\,, \qquad \frac{\partial}{\partial z}\frac{1}{\zeta} = \frac{1}{\zeta_s}\frac{\partial}{\partial z}\frac{1}{\zeta_a} + \frac{1}{\zeta_a}\frac{\partial}{\partial z}\frac{1}{\zeta_s}\,. \tag{20.35}$$

The leading (material flow conserving) eigenvalue $z = 1$ belongs to the symmetric subspace $1/\zeta_s(0,1) = 0$, so the derivatives (20.15) also depend only on the symmetric subspace:

$$\langle n\rangle_\zeta = z\frac{\partial}{\partial z}\frac{1}{\zeta(0,z)}\bigg|_{z=1} = \frac{1}{\zeta_a(0,z)}\,z\frac{\partial}{\partial z}\,\frac{1}{\zeta_s(0,z)}\bigg|_{z=1}\,. \tag{20.36}$$

Implementing the symmetry factorization is convenient, but not essential, at this level of computation.

Remark 20.6 Lorentz gas in the fundamental domain. The vector valued nature of the generating function (20.3) 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. [20.15], numerical estimates for scalar quantities are reported in table 20.1, taken from ref. [20.16].

In order to perform the full reduction for diffusion one should express the dynamical zeta function (20.7) in terms of the prime cycles of the fundamental domain $\tilde{\mathcal{M}}$ of the lattice (see fig. 20.2) 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 non-commutativity of translations and rotations: see ref. [20.6].

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 20.1: Fundamental domain, w=0.3 .

Remark 20.7 Anomalous diffusion. Anomalous diffusion for 1-d intermittent maps was studied in the continuous time random walk approach in refs. [18.10, 18.11]. The first approach within the framework of cycle expansions (based on truncated dynamical zeta functions) was proposed in ref. [18.12]. Our treatment follows methods introduced in ref. [18.13], applied there to investigate the behavior of the Lorentz gas with unbounded horizon.

Remark 20.8 Jonquière functions. In statistical mechanics Jonquière functions

$$J(z,s) = \sum_{k=1}^{\infty} z^k/k^s \tag{20.37}$$

appear in the theory of free Bose-Einstein gas, see refs. [18.21, 18.22].

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

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

Such formulas are *exact*; the issue in their applications is what are the most effective schemes of estimating the infinite cycle sums required for their evaluation. Unlike most statistical mechanics, here there are no phenomenological macroscopic parameters; quantities such as transport coefficients are calculable to any desired accuracy from the microscopic dynamics.

Though superficially indistinguishable from the probabilistic random walk diffusion, deterministic diffusion is quite recognizable, at least in low dimensional settings, through fractal dependence of the diffusion constant on the system parameters, and through non-gaussion relaxation to equilibrium (non-vanishing Burnett coefficients).

For systems of a few degrees of freedom these results are on rigorous footing, but there are indications that they capture the essential dynamics of systems of many degrees of freedom as well.

Actual evaluation of transport coefficients is a test of the techniques developped above in physical settings. In cases of severe pruning the trace formulas and ergodic sampling of dominant cycles might be more effective strategy than the cycle expansions of dynamical zeta functions and systematic enumeration of all cycles.

References

[20.1] J. Machta and R. Zwanzig, *Phys. Rev. Lett.* **50**, 1959 (1983).

[20.2] G.P. Morriss and L. Rondoni, *J. Stat. Phys.* **75**, 553 (1994).

[20.3] H.A. Lorentz, *Proc. Amst. Acad.* **7**, 438 (1905).

[20.4] R. Artuso, *Phys. Lett.* **A 160**, 528 (1991).

[20.5] W.N. Vance, *Phys. Rev. Lett.* **96**, 1356 (1992).

[20.6] P. Cvitanović, J.-P. Eckmann, and P. Gaspard, *Chaos, Solitons and Fractals* **6**, 113 (1995).

[20.7] R. Klages, *Deterministic diffusion in one-dimensional chaotic dynamical systems* (Wissenschaft & Technik-Verlag, Berlin, 1996); `www.mpipks-dresden.mpg.de/~rklages/publ/phd.html`.

[20.8] R. Klages and J.R. Dorfman, *Phys. Rev. Lett.* **74**, 387 (1995); *Phys. Rev. E* **59**, 5361 (1999).

[20.9] R. Klages and J.R. Dorfman, "Dynamical crossover in deterministic diffusion", *Phys. Rev.* **E 55**, R1247 (1997).

[20.10] J. Lloyd, M. Niemeyer, L. Rondoni and G.P. Morriss, *CHAOS* **5**, 536 (1995).

[20.11] T. Geisel and J. Nierwetberg, *Phys. Rev. Lett.* **48**, 7 (1982).

[20.12] M. Schell, S. Fraser and R. Kapral, *Phys. Rev.* **A 26**, 504 (1982).

[20.13] S. Grossmann, H. Fujisaka, Phys. Rev. **A 26**, 1179 (1982); H. Fujisaka and S. Grossmann, Z. Phys. **B 48**, 261 (1982).

[20.14] P. Gaspard and F. Baras, in M. Mareschal and B.L. Holian, eds., *Microscopic simulations of Complex Hydrodynamic Phenomena* (Plenum, NY 1992).

[20.15] F. Christiansen, Master's Thesis, Univ. of Copenhagen (June 1989)

[20.16] P. Cvitanović, P. Gaspard, and T. Schreiber, "Investigation of the Lorentz Gas in terms of periodic orbits", *CHAOS* **2**, 85 (1992).

[20.17] S. Grossmann and S. Thomae, *Phys. Lett.***A 97**, 263 (1983).

[20.18] R. Artuso, G. Casati and R. Lombardi, *Physica* **A 205**, 412 (1994).

Exercises

Exercise 20.1 **Diffusion for odd integer Λ.** *Show that when the slope $\Lambda = 2k - 1$ in (20.9) is an odd integer, the diffusion constant is given by $D = (\Lambda^2 - 1)/24$, as stated in (20.21).*

Exercise 20.2 **Fourth-order transport coefficient.** *Verify (20.24). You will need the identity*

$$\sum_{k=1}^{n} k^4 \;=\; \frac{1}{30} n(n+1)(2n+1)(3n^2+3n-1)\,.$$

Exercise 20.3 **Finite Markov partitions.** *Verify (20.28).*

Exercise 20.4 **Maps with variable peak shape:**
Consider the following piecewise linear map

$$f_\delta(x) \;=\; \begin{cases} \frac{3x}{1-\delta} & \text{for } x \in \left[0, \frac{1}{3}(1-\delta)\right] \\ \frac{3}{2} - \left(\frac{2}{\delta}\left|\frac{4-\delta}{12} - x\right|\right) & \text{for } x \in \left[\frac{1}{3}(1-\delta), \frac{1}{6}(2+\delta)\right] \\ 1 - \frac{3}{1-\delta}\left(x - \frac{1}{6}(2+\delta)\right) & \text{for } x \in \left[\frac{1}{6}(2+\delta), \frac{1}{2}\right] \end{cases} \tag{20.38}$$

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. [20.17, 20.18] for further details.

Exercise 20.5 **Two-symbol cycles for the Lorentz gas.** *Write down all cycles labelled by two symbols, such as $(0\ 6)$, $(1\ 7)$, $(1\ 5)$ and $(0\ 5)$.*

Appendix P contains several project-length deterministic diffusion exercises.

Exercise 20.6 **Accelerated diffusion.** *Consider a map h, such that $\hat{h} = \hat{f}$ of fig. 20.6(b), but now running branches are turner into standing branches and vice versa, so that $1, 2, 3, 4$ are standing while 0 leads to both positive and negative jumps. Build the corresponding dynamical zeta function and show that*

$$\sigma^2(t) \sim \begin{cases} t & \textit{for } \alpha > 2 \\ t\ln t & \textit{for } \alpha = 2 \\ t^{3-\alpha} & \textit{for } \alpha \in (1,2) \\ t^2/\ln t & \textit{for } \alpha = 1 \\ t^2 & \textit{for } \alpha \in (0,1) \end{cases}$$

Exercise 20.7 **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*

$$\left| \frac{d_{m,n}}{R} \sin\left(\phi - \alpha - \theta_{m,n}\right) + \sin\alpha \right| \leq 1 \tag{20.39}$$

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 (20.39).

(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.*

Chapter 21

Irrationally winding

I don't care for islands, especially very small ones.
D.H. Lawrence

(R. Artuso and P. Cvitanović)

This chapter is concerned with the mode locking problems for circle maps: besides its physical relevance it nicely illustrates the use of cycle expansions away from the dynamical setting, in the realm of renormalization theory at the transition to chaos.

The physical significance of circle maps is connected with their ability to model the two–frequencies mode–locking route to chaos for dissipative systems. In the context of *dissipative* dynamical systems one of the most common and experimentally well explored routes to chaos is the two-frequency mode-locking route. Interaction of pairs of frequencies is of deep theoretical interest due to the generality of this phenomenon; as the energy input into a dissipative dynamical system (for example, a Couette flow) is increased, typically first one and then two of intrinsic modes of the system are excited. After two Hopf bifurcations (a fixed point with inward spiralling stability has become unstable and outward spirals to a limit cycle) a system lives on a two-torus. Such systems tend to mode-lock: the system adjusts its internal frequencies slightly so that they fall in step and minimize the internal dissipation. In such case the ratio of the two frequencies is a rational number. An irrational frequency ratio corresponds to a quasiperiodic motion - a curve that never quite repeats itself. If the mode-locked states overlap, chaos sets in. The likelihood that a mode-locking occurs depends on the strength of the coupling of the two frequencies.

Our main concern in this chapter is to illustrate the "global" theory of circle maps, connected with universality properties of the whole irrational winding set. We shall see that critical global properties may be expressed via cycle expansions involving "local" renormalization critical exponents. The renormalization theory of critical circle maps demands rather tedious numerical computations, and our intuition is much facilitated by approximating circle maps by number-theoretic 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.

21.1 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*

$$x_{n+1} = f(x_n) = x_n + \Omega - \frac{k}{2\pi}\sin(2\pi x_n) \qquad \mathrm{mod}\ 1 . \tag{21.1}$$

$f(x)$ 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

$$\hat{x}_{n+1} = \hat{f}(\hat{x}_n) = \hat{x}_n + \Omega - \frac{k}{2\pi}\sin(2\pi\hat{x}_n) . \tag{21.2}$$

where $\hat{}$ is used in the same sense as in chapter 20.

The winding number is defined as

$$W(k,\Omega) = \lim_{n\to\infty}(\hat{x}_n - \hat{x}_0)/n. \tag{21.3}$$

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. 21.1

$$x_{n+1} = x_n + \Omega \qquad \mathrm{mod}\ 1 , \tag{21.4}$$

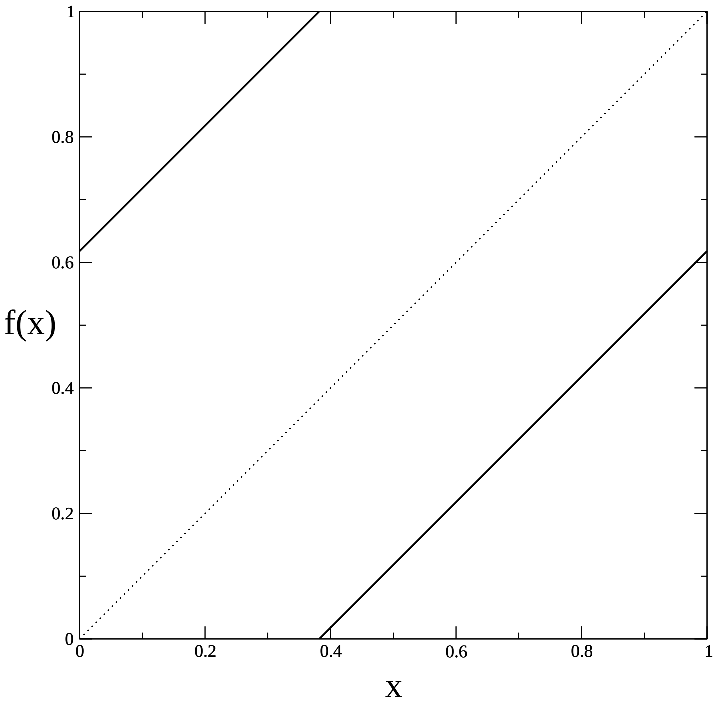


Figure 21.1: Unperturbed circle map ($k = 0$ in (21.1)) with golden mean rotation number.

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

21.1 page 437

$$\Delta_{P/Q} = Q^{-2\mu_{P/Q}} = \Omega^{right}_{P/Q} - \Omega^{left}_{P/Q} \ . \tag{21.5}$$

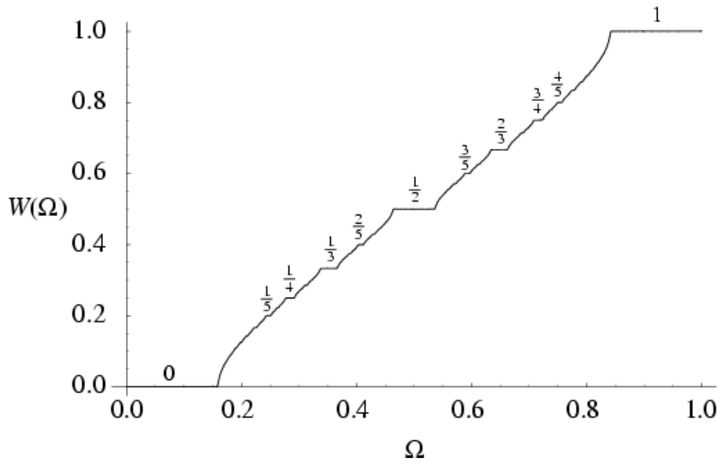


Figure 21.2: The critical circle map ($k = 1$ in (21.1)) devil's staircase [21.3]; the winding number W as function of the parameter Ω.

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 (21.5)). For the shift map (21.4), 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. 21.2 whose self-similar structure we are to unravel. Circle maps with zero slope at the inflection point x_c (see fig. 21.3)

$$f'(x_c) = 0\,, \qquad f''(x_c) = 0$$

($k = 1$, $x_c = 0$ in (21.1)) 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

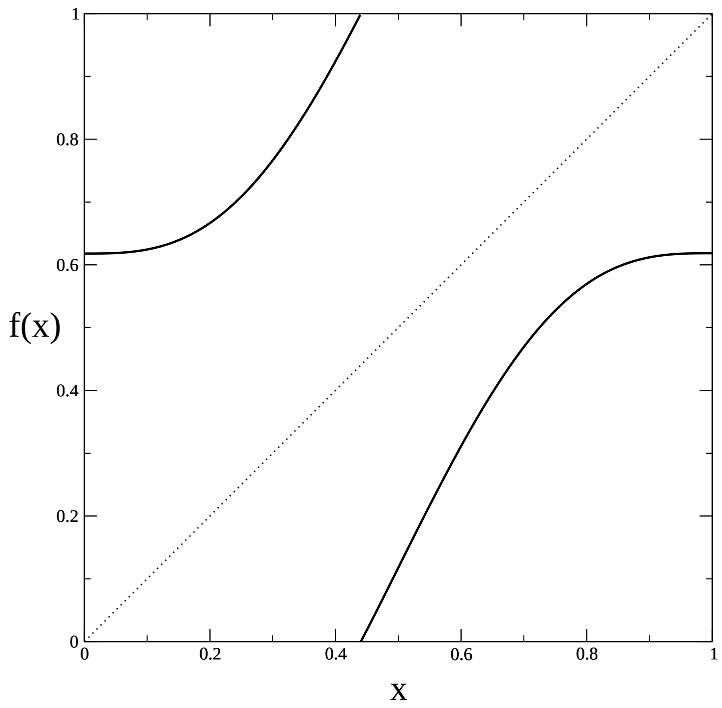


Figure 21.3: Critical circle map ($k = 1$ in (21.1)) with golden mean bare rotation number.

whole interval. A critical map has a superstable P/Q cycle for any rational P/Q, as the stability of any cycle that includes the inflection point equals zero. If the map is non-invertible ($k > 1$), it is called supercritical; the bifurcation structure of this regime is extremely rich and beyond the scope of this exposition.

The physically relevant transition to chaos is connected with the critical case, however the apparently simple "free" shift map limit is quite instructive: in essence it involves the problem of ordering rationals embedded in the unit interval on a hierarchical structure. From a physical point of view, the main problem is to identify a (number-theoretically) consistent hierarchy susceptible of experimental verification. We will now describe a few ways of organizing rationals along the unit interval: each has its own advantages as well as its drawbacks, when analyzed from both mathematical and physical perspective.

21.1.1 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

$$\Phi(Q) = \sum_{n=1}^{Q} \phi(Q) = \frac{3Q^2}{\pi^2} + O(Q \ln Q) . \tag{21.6}$$

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

$$\begin{aligned} T(x) &= \frac{1}{x} - \left[\frac{1}{x}\right] \quad x \neq 0 \\ & 0 \ , \qquad\quad x = 0 \end{aligned} \tag{21.7}$$

($[\cdots]$ denotes the integer part) acts as a shift on the continued fraction representation of numbers on the unit interval

$$x = [a_1, a_2, a_3, \ldots] \to T(x) = [a_2, a_3, \ldots] \ . \tag{21.8}$$

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. 21.2 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* $a_i \geq 1$, $i = 1, 2, \ldots, k-1$, $\quad a_k \geq 2$, *add up to* $\sum_{i=1}^{k} a_i = n+2$. For example

$$T_2 = \{[4], [2,2], [1,1,2], [1,3]\} = \left(\frac{1}{4}, \frac{1}{5}, \frac{3}{5}, \frac{3}{4}\right). \tag{21.9}$$

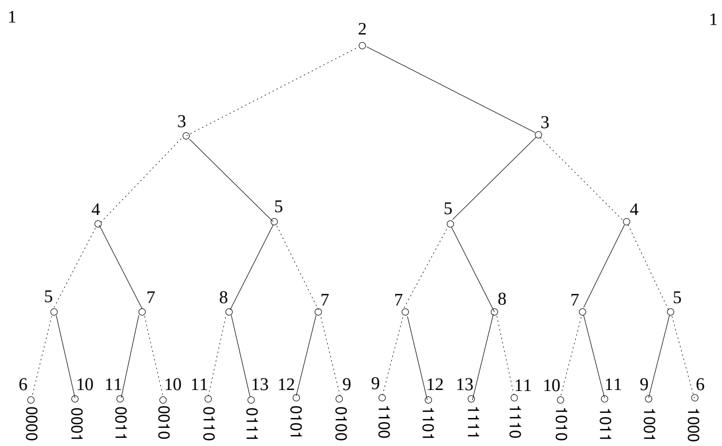


Figure 21.4: Farey tree: alternating binary ordered labelling of all Farey denominators on the nth Farey tree level.

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. 21.4.

The smallest and the largest denominator in T_n are respectively given by

$$[n-2] = \frac{1}{n-2}, \qquad [1,1,\ldots,1,2] = \frac{F_{n+1}}{F_{n+2}} \propto \rho^n , \tag{21.10}$$

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

$$\rho = \frac{1+\sqrt{5}}{2} = 1.61803\ldots \tag{21.11}$$

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.

21.2 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 (21.10). 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 (21.38); 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 $[b, c, \ldots]$:

$$R_a \begin{pmatrix} g \\ f \end{pmatrix} = \begin{pmatrix} \alpha g^{a-1} \circ f \circ \alpha^{-1} \\ \alpha g^{a-1} \circ f \circ g \circ \alpha^{-1} \end{pmatrix}, \tag{21.12}$$

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 (21.1) 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

$$\Omega_r - \Omega_{r+1} \propto \delta_p^{-r}. \tag{21.13}$$

What this means is that the "devil's staircase" of fig. 21.2 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 (21.13) 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

$$\Delta_{1/Q} \propto Q^{-3} \tag{21.14}$$

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

$$\Delta_{P/Q} \propto Q^{-2\mu_1} \tag{21.15}$$

where $P = F_{n-1}$, $Q = F_n$ and μ_1 is related to the universal Shenker number δ_1 (21.13) and the golden mean (21.11) by

$$\mu_1 = \frac{\ln|\delta_1|}{2\ln\rho} = 1.08218\ldots \tag{21.16}$$

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:

$$3/2 > \mu_{m/n} \geq 1.08218 \cdots . \tag{21.17}$$

21.3 Global theory: Thermodynamic averaging

Consider the following average over mode-locking intervals (21.5):

$$\Omega(\tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} \Delta_{P/Q}^{-\tau}. \tag{21.18}$$

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 (21.18) 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 [21.11]. 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 (21.18) 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:

$$\bar{Z}_n(\tau) = \sum_{i \in \mathcal{S}_n} \Delta_i^{-\tau}. \tag{21.19}$$

The original sum (21.18) 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 (21.18):

$$\Omega(q, \tau) = \sum_{Q=1}^{\infty} \sum_{(P|Q)=1} e^{-q\nu_{P/Q}} Q^{2\tau\mu_{P/Q}} . \tag{21.20}$$

The sum (21.18) 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

$$\Omega(q,\tau) = \sum_{Q=1}^{\infty} \Phi(Q)^{-q} \sum_{(P|Q)=1} Q^{2\tau\mu_{P/Q}} \ , \tag{21.21}$$

the continued fraction partitioning

$$\Omega(q,\tau) = \sum_{n=1}^{\infty} e^{-qn} \sum_{[a_1,\ldots,a_n]} Q^{2\tau\mu_{[a_1,\ldots,a_n]}} \ , \tag{21.22}$$

and the Farey tree partitioning

$$\Omega(q,\tau) = \sum_{k=n}^{\infty} 2^{-qn} \sum_{i=1}^{2^n} Q_i^{2\tau\mu_i} \ , \quad Q_i/P_i \in T_n \ . \tag{21.23}$$

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.

21.4 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

$$\ell_i = \Omega^{min}_{P_r/Q_r} - \Omega^{max}_{P_l/Q_l} \ . \tag{21.24}$$

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

$$Z_n(\tau) = \sum_{i \in \mathcal{S}_n} \ell_i^{-\tau} \tag{21.25}$$

The value of $-\tau$ for which the $n \to \infty$ limit of the sum (21.25) 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.

21.2
page 437

21.4.1 The Hausdorff dimension in terms of cycles

Estimating the $n \to \infty$ limit of (21.25) 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 (21.4), the covering intervals are generated by iterations of the Farey map (21.37) or the Gauss shift (21.38). The nth level sum (21.25) 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

$$\begin{aligned} \det(1 - z\mathcal{L}_\tau) &= \exp\left(-\sum_p \sum_{r=1}^\infty \frac{z^{rn_p}}{r} \frac{|\Lambda_p^r|^\tau}{1 - 1/\Lambda_p^r}\right) \\ &= \prod_p \prod_{k=0}^\infty \left(1 - z^{n_p}|\Lambda_p|^\tau/\Lambda_p^k\right) . \end{aligned} \tag{21.26}$$

The sum (21.25) 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 (21.26); putting all these pieces together, we obtain a pretty formula relating the Hausdorff dimension to the prime cycles of the map $f(x)$:

$$0 = \prod_p \left(1 - 1/|\Lambda_p|^{D_H}\right) \ . \tag{21.27}$$

For the Gauss shift (21.38) 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

$$x_a = \frac{-a + \sqrt{a^2+4}}{2}, \qquad \Lambda_a = -\left(\frac{a + \sqrt{a^2+4}}{2}\right)^2 \tag{21.28}$$

and the $x_{ab} = [a, b, a, b, a, b, \ldots]$ 2–cycles are given by

$$\begin{aligned} x_{ab} &= \frac{-ab + \sqrt{(ab)^2 + 4ab}}{2b} \\ \Lambda_{ab} &= (x_{ab} x_{ba})^{-2} = \left(\frac{ab + 2 + \sqrt{ab(ab+4)}}{2}\right)^2 \end{aligned} \tag{21.29}$$

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 (21.27) merely a very convoluted way to compute the number 1? Possibly so, but once the meaning of (21.27) has been grasped, the corresponding formula for the *critical* circle maps follows immediately:

$$0 = \prod_p \left(1 - 1/|\delta_p|^{D_H}\right) \ . \tag{21.30}$$

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 (21.30) 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 21.1 we give a small selection of computed Shenker's scalings.

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 21.1: Shenker's δ_p for a few periodic continued fractions, from ref. [21.1].

21.5 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 17, by looking at the free energy.

Consider the Farey tree partition sum (21.23): the narrowest mode-locked interval (21.15) at the nth level of the Farey tree partition sum (21.23) is the golden mean interval

$$\Delta_{F_{n-1}/F_n} \propto |\delta_1|^{-n}. \tag{21.31}$$

It shrinks exponentially, and for τ positive and large it dominates $q(\tau)$ and bounds $dq(\tau)/d\tau$:

$$q'_{max} = \frac{\ln|\delta_1|}{\ln 2} = 1.502642\ldots \tag{21.32}$$

However, for τ large and negative, $q(\tau)$ is dominated by the interval (21.14) which shrinks only harmonically, and $q(\tau)$ approaches 0 as

$$\frac{q(\tau)}{\tau} = \frac{3\ln n}{n\ln 2} \to 0. \tag{21.33}$$

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 $-D_H \le \tau$. 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}:

$$Z_n(\tau) = \sum_{i=1}^{2^n} Q_i^{2\tau}. \tag{21.34}$$

Here Q_i is the denominator of the ith Farey rational P_i/Q_i. For example (see fig. 21.4),

$$Z_2(1/2) = 4 + 5 + 5 + 4.$$

By the annihilation property (21.38) 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

$$Z_n(\tau) = \int dx\mathcal{L}_\tau^n(0,x)\,,$$

with the sum restricted to the Farey level $a_1 + \ldots + a_k = n+2$. It is easily checked that $f'_{a_1\ldots a_k}(0) = (-1)^k Q^2_{[a_1,\ldots,a_k]}$, so the Farey model sum is a partition generated by the Gauss map preimages of $x = 0$, that is by rationals, rather than by the quadratic irrationals as in (21.26). 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 21.2, they are consequence of the fact that the denominators on a given level are Farey sums of denominators on preceding levels.

21.3
page 437

A bound on D_H can be obtained by approximating (21.34) by

$$Z_n(\tau) = n^{2\tau} + 2^n\rho^{2n\tau}. \tag{21.35}$$

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 (21.15)). The crossover from

$\tau/2$	$Z_n(\tau/2)/Z_{n-1}(\tau/2)$
0	2
1	3
2	$(5+\sqrt{17})/2$
3	7
4	$(5+\sqrt{17})/2$
5	$7+4\sqrt{6}$
6	$26.20249\ldots$

Table 21.2: Partition function sum rules for the Farey model.

the harmonic dominated to the golden mean dominated behavior occurs at the τ value for which the two terms in (21.35) contribute equally:

$$D_n = \hat{D} + O\left(\frac{\ln n}{n}\right), \quad \hat{D} = \frac{\ln 2}{2\ln\rho} = .72\ldots \tag{21.36}$$

For negative τ the sum (21.35) is the lower bound on the sum (21.25) , so $\hat{D}$ is a lower bound on D_H.

From a general perspective the analysis of circle maps thermodynamics has revealed the fact that physically interesting dynamical systems often exhibit mixtures of hyperbolic and marginal stabilities. In such systems there are orbits that stay 'glued' arbitrarily close to stable regions for arbitrarily long times. This is a generic phenomenon for Hamiltonian systems, where elliptic islands of stability coexist with hyperbolic homoclinic webs. Thus the considerations of chapter 18 are important also in the analysis of renomarmalization at the onset of chaos.

Commentary

Remark 21.1 The physics of circle maps. Mode–locking phenomenology is reviewed in ref. [21.5], a more theoretically oriented discussion is contained in ref. [21.3]. 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 [21.8, 21.9]. By losing all of the "island-within-island" structure of real systems, circle map models skirt the problems of determining the symbolic dynamics for a realistic Hamiltonian system, but they do retain some of the essential features of such systems, such as the golden mean renormalization [14.7, 21.8] and non-hyperbolicity in form of sequences of cycles accumulating toward the borders of stability. In particular, in such systems there are orbits that stay "glued" arbitrarily close to stable regions for arbitrarily long times. As this is a generic phenomenon in physically interesting dynamical systems, such as the Hamiltonian systems with coexisting elliptic islands of stability and hyperbolic homoclinic webs, development of good computational techniques is here of utmost practical importance.

Remark 21.2 Critical mode–locking set The fact that modelockings completely fill the unit interval at the critical point has been proposed in refs. [21.3, 21.10]. The proof that the set of irrational windings is of zero Lebesgue measure in given in ref. [21.11].

Remark 21.3 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 (21.25) of any methods used so far; for example the computation in ref. [21.12] for critical sine map (21.1), 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 21.4 Farey tree presentation function. The Farey tree rationals can be generated by backward iterates of 1/2 by the Farey presentation function [21.13]:

$$\begin{array}{rcll} f_0(x) & = & x/(1-x) & 0 \leq x < 1/2 \\ f_1(x) & = & (1-x)/x & 1/2 < x \leq 1 \,. \end{array} \tag{21.37}$$

The Gauss shift (21.7) corresponds to replacing the binary Farey presentation function branch f_0 in (21.37) by an infinity of branches

$$\begin{array}{rcll} f_a(x) & = & f_1 \circ f_0^{(a-1)}(x) = \dfrac{1}{x} - a, & \dfrac{1}{a-1} < x \leq \dfrac{1}{a}, \\ f_{ab\cdots c}(x) & = & f_c \circ \cdot \circ f_b \circ f_a(x) \,. & \end{array} \tag{21.38}$$

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 (21.37) and by its nontrivial 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. [21.4].

An alternative labelling of Farey denominators has been introduced by Knauf [21.6] in context of number-theoretical modeling of ferromagnetic spin chains: it allows for a number of elegant manipulations in thermodynamic averages connected to the Farey tree hierarchy.

Remark 21.5 Circle map renormalization The idea underlying golden mean renormalization goes back to Shenker [21.9]. A renormalization group procedure was formulated in refs. [21.7, 21.14], where moreover the uniqueness of the relevant eigenvalue is claimed. This statement has been confirmed by a computer–assisted proof [21.15], and in the following we will always assume it. There are a number of experimental evidences for local universality, see refs. [21.16, 21.17].

On the other side of the scaling tale, the power law scaling for harmonic fractions (discussed in refs. [21.2, 21.3, 21.4]) is derived by methods akin to those used in describing intermittency [21.21]: $1/Q$ cycles accumulate toward the edge of $0/1$ mode-locked interval, and as the successive mode-locked intervals $1/Q$, $1/(Q-1)$ lie on a parabola, their differences are of order Q^{-3}.

Remark 21.6 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 [21.22, 21.23] . The distribution of the Farey series rationals across the unit interval is surprisingly uniform - indeed, so uniform that in the pre-computer days it has motivated a compilation of an entire handbook of Farey series [21.24]. 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 [21.25, 21.22, 21.23], the Riemann hypothesis is equivalent to the statement that

$$\sum_{Q_i \le Q} |\delta_i| = o(Q^{\frac{1}{2}+\epsilon})$$

for all 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 21.7 Farey tree partitioning. The Farey tree partitioning was introduced in refs. [21.26, 21.27, 21.4] and its thermodynamics is discussed in detail in refs. [21.12, 21.13]. The Farey tree hierarchy of rationals is rather new, and, as far as we are aware, not previously studied by number theorists. It is appealing both from the experimental and from the the golden-mean renormalization point of view, but it has a serious drawback of lumping together mode-locking intervals of wildly different sizes on the same level of the Farey tree.

Remark 21.8 Local and global universality. Numerical evidences for global universal behavior have been presented in ref. [21.3]. The question was reexamined in ref. [21.12], where it was pointed out how a high-precision numerical estimate is in practice very hard to obtain. It is not at all clear whether this is the optimal global quantity to test but at least the Hausdorff dimension has the virtue of being independent of how one partitions mode-lockings and should thus be the same for the variety of thermodynamic averages in the literature.

The formula (21.30), linking local to global behavior, was proposed in ref. [21.1].

The derivation of (21.30) relies only on the following aspects of the "hyperbolicity conjecture" of refs. [21.4, 21.18, 21.19, 21.20]:

1. *limits* for Shenker δ's *exist* and are universal. This should follow from the renormalization theory developed in refs. [21.7, 21.14, 21.15], 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 $p = a_1 a_2 \ldots n$ with a large continued fraction entry n grows as a *power* of n. According to (21.14), $\lim_{n\to\infty} \delta_p \propto n^3$. In the calculation of ref. [21.1] 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 21.9 Farey model. The Farey model (21.33) has been proposed in ref. [21.12]; though it might seem to have been pulled out of a hat, the Farey model is as sensible description of the distribution of rationals as the periodic orbit expansion (21.26).

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

[21.1] P. Cvitanović, G.H. Gunaratne and M. Vinson, *Nonlinearity* **3** (1990)

[21.2] K. Kaneko, *Prog. Theor. Phys.* **68**, 669 (1982); **69**, 403 (1983); **69**, 1427 (1983)

[21.3] 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)

[21.4] P. Cvitanović, B. Shraiman and B. Söderberg, *Physica Scripta* **32**, 263 (1985).

[21.5] J.A. Glazier and A. Libchaber, *IEEE Trans. Circ. Syst.*, **35**, 790 (1988)

[21.6] A. Knauf, "On a ferromagnetic spin chain", *Commun. Math. Phys.* **153**, 77 (1993).

[21.7] M.J. Feigenbaum, L.P. Kadanoff, S.J. Shenker, *Physica* **5D**, 370 (1982)

[21.8] S.J. Shenker and L.P. Kadanoff, *J. Stat. Phys.* **27**, 631 (1982)

[21.9] S.J. Shenker, *Physica* **5D**, 405 (1982)

[21.10] O.E. Lanford, *Physica* **14D**, 403 (1985)

[21.11] G. Swiatek, *Commun. Math. Phys.* **119**, 109 (1988)

[21.12] R. Artuso, P. Cvitanović and B.G. Kenny, *Phys. Rev.* **A39**, 268 (1989); P. Cvitanović, in R. Gilmore (ed), *Proceedings of the XV International Colloquium on Group Theoretical Methods in Physics*, (World Scientific, Singapore, 1987)

[21.13] M.J. Feigenbaum, *J.Stat.Phys.* **52**, 527 (1988)

[21.14] S. Ostlund, D.A. Rand, J. Sethna and E. Siggia, *Physica* **D 8**, 303 (1983)

[21.15] B.D. Mestel, Ph.D. Thesis (U. of Warwick 1985).

[21.16] J. Stavans, F. Heslot and A. Libchaber, *Phys. Rev. Lett.* **55**, 569 (1985)

[21.17] E.G. Gwinn and R.M. Westervelt, *Phys. Rev. Lett.* **59**, 157 (1987)

[21.18] O.E. Lanford, in M. Mebkhout and R. Sénéor, eds., *Proc. 1986 IAMP Conference in Mathematical Physics* (World Scientific, Singapore 1987); D.A. Rand, *Proc. R. Soc. London* **A 413**, 45 (1987); *Nonlinearity* **1**, 78 (1988)

[21.19] S.-H. Kim and S. Ostlund, *Physica* **D 39**, 365, (1989)

[21.20] M.J. Feigenbaum, *Nonlinearity* **1**, 577 (1988)

[21.21] Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**, 189 (1980); P. Manneville, *J. Phys.* (Paris) **41**, 1235 (1980)

[21.22] H.M. Edwards, *Riemann's Zeta Function* (Academic, New York 1974)

[21.23] E.C. Titchmarsh, *The Theory of Riemann Zeta Function* (Oxford Univ. Press, Oxford 1951); chapter XIV.

[21.24] E.H. Neville, *Roy. Soc. Mathematical Tables* (Cambridge U. Press, Cambridge 1950)

[21.25] J. Franel and E. Landau, *Göttinger Nachr. 198* (1924)

[21.26] G. T. Williams and D. H. Browne, *Amer. Math. Monthly* **54**, 534 (1947)

[21.27] P. Cvitanović and J. Myrheim, *Phys. Lett.* **A94**, 329 (1983); *Commun. Math. Phys.* **121**, 225 (1989)

[21.28] P. Contucci and A. Knauf, *Forum Math.* **9**, 547 (1997)

[21.29] G.H. Hardy and E.M. Wright, *Theory of Numbers* (Oxford Univ. Press, Oxford 1938)

[21.30] A. Csordás and P. Szépfalusy, *Phys. Rev. A* **40**, 2221 (1989) and references therein.

[21.31] P. Dahlqvist, unpublished notes.

[21.32] D. Levin, *Inter. J. Computer Math.* **B3**, 371 (1973).

[21.33] N. Osada, *SIAM J. Numer. Anal.* **27**, 178 (1990).

Exercises

Exercise 21.1 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.*

Exercise 21.2 Bounds on Hausdorff dimension. *By making use of the bounds (21.17) show that the Hausdorff dimension for critical mode lockings may be bounded by*

$$2/3 \leq D_H \leq .9240\ldots$$

Exercise 21.3 Farey model sum rules. *Verify the sum rules reported in table 21.2. 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. [21.28].*

Exercise 21.4 Metric entropy of the Gauss shift. *Check that the Lyapunov exponent of the Gauss map (21.7) is given by $\pi^2/6\ln 2$. This result has been claimed to be relevant in the discussion of "mixmaster" cosmologies, see ref. [21.30].*

Exercise 21.5 Refined expansions. *Show that the above estimates can be refined as follows:*

$$F(z,2) \sim \zeta(2) + (1-z)\log(1-z) - (1-z)$$

and

$$F(z,s) \sim \zeta(s) + \Gamma(1-s)(1-z)^{s-1} - S(s)(1-z)$$

for $s \in (1,2)$ ($S(s)$ being expressed by a converging sum). You may use either more detailed estimate for $\zeta(s,a)$ (via Euler summation formula) or keep on subtracting leading contributions [21.31].

Exercise 21.6 Hitting condition. *Prove (20.39). Hint: together with the real trajectory consider the line passing through the starting point, with polar angle $\theta_{m,n}$: then draw the perpendiculars to the actual trajectory, passing through the center of the $(0,0)$ and (m,n) disks.*

Exercise 21.7 j_n and α_{cr}. *Look at the integration region and how it scales by plotting it for increasing values of n.*

Exercise 21.8 Estimates of the Riemann zeta function. *Try to approximate numerically the Riemann zeta function for $s = 2, 4, 6$ using different acceleration algorithms: check your results with refs. [21.32, 21.33].*

Exercise 21.9 **Farey tree and continued fractions I.** *Consider the Farey tree presentation function* $f : [0,1] \mapsto [0,1]$*, such that if* $I = [0,1/2)$ *and* $J = [1/2,1]$*,* $f|_I = x/(1-x)$ *and* $f|_J = (1-x)/x$*. Show that the corresponding induced map is the Gauss map* $g(x) = 1/x - [1/x]$*.*

Exercise 21.10 **Farey tree and continued fraction II. (Lethal weapon II).** *Build the simplest piecewise linear approximation to the Farey tree presentation function (hint: substitute first the righmost, hyperbolic branch with a linear one): consider then the spectral determinant of the induced map* $\hat{g}$*, and calculate the first two eigenvalues besides the probability conservation one. Compare the results with the rigorous bound deduced in ref. [18.17].*