Part I

Geometry of chaos

WE START OUT with a recapitulation of the basic notions of dynamics. Our aim is narrow; we keep the exposition focused on prerequisites to the applications to be developed in this text. We assume that the reader is familiar with dynamics on the level of the introductory texts mentioned in remark 1.1, and concentrate here on developing intuition about what a dynamical system can do. It will be a coarse brush sketch–a full description of all possible behaviors of dynamical systems is beyond human ken. While for a novice there is no shortcut through this lengthy detour, a sophisticated traveler might bravely skip this well-trodden territory and embark upon the journey at chapter 15.

The fate has handed you a flow. What are you to do about it?

1. Define your *dynamical system* $(\mathcal{M}, f)$: the space of its possible states $\mathcal{M}$, and the law f^t of their evolution in time.

2. Pin it down locally–is there anything about it that is stationary? Try to determine its *equilibria / fixed points* (Chapter 2).

3. Slice it, represent as a map from a section to a section (Chapter 3).

4. Explore the neighborhood by *linearizing* the flow–check the *linear stability* of its equilibria / fixed points, their stability eigen-directions (Chapter 4).

5. Go global: *partition the state space* of 1-dimensional maps. Label the regions by *symbolic dynamics* (Chapter 11).

6. Now venture global distances across the system by continuing eigenvectors into *stable / unstable manifolds*. Their intersections *partition the state space* in a dynamically invariant way (Chapter 12).

7. Guided by this topological partition, compute a set of *periodic orbits* up to a given topological length (Chapter 13).

Along the way you might want to learn about dynamical invariants (chapter 5), nonlinear transformations (chapter 6), classical mechanics (chapter 7), billiards (chapter 8), and discrete (chapter 9) and continuous (chapter 10) symmetries of dynamics.

Chapter 1

Overture

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

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

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

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

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

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

 fast track points you where to skip to

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

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

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

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

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

1.1 Why ChaosBook?

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

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

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

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

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

1.2 Chaos ahead

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

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

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

In the traditional approach the integrable motions are used as zeroth-order approximations to physical systems, and weak nonlinearities are then accounted for perturbatively. For strongly nonlinear, non-integrable systems such expansions fail completely; at asymptotic times the dynamics exhibits amazingly rich structure which is not at all apparent in the integrable approximations. However, hidden in this apparent chaos is a rigid skeleton, a self-similar tree of *cycles* (periodic orbits) of increasing lengths. The insight of the modern dynamical systems theory is that the zeroth-order approximations to the harshly chaotic dynamics should be very different from those for the nearly integrable systems: a good starting approximation here is the stretching and folding of baker's dough, rather than the periodic motion of a harmonic oscillator.

So, what is chaos, and what is to be done about it? To get some feeling for how and why unstable cycles come about, we start by playing a game of pinball. The reminder of the chapter is a quick tour through the material covered in ChaosBook. Do not worry if you do not understand every detail at the first reading–the intention

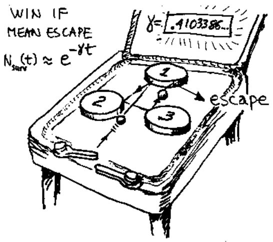


Figure 1.1: A physicist's bare bones game of pinball.

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, [section 1.4] on the margin points at the appropriate section. section 1.4

1.3 The future as in a mirror

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

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

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

At the beginning of the 18th century Baron Gottfried Wilhelm Leibniz was confident that given the initial conditions one knew everything a deterministic system would do far into the future. He wrote [1.2], anticipating by a century and a half the oft-quoted Laplace's "Given for one instant an intelligence which could comprehend all the forces by which nature is animated...":

> That everything is brought forth through an established destiny is just

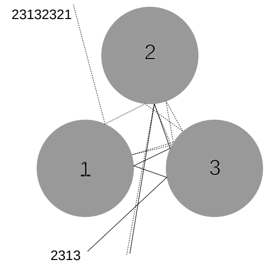


Figure 1.2: Sensitivity to initial conditions: two pinballs that start out very close to each other separate exponentially with time.

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

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

1.3.1 What is 'chaos'?

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

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

For a stochastic system the initial conditions determine the future only partially, due to noise, or other external circumstances beyond our control: the present state reflects the past initial conditions plus the particular realization of the noise encountered along the way.

A deterministic system with sufficiently complicated dynamics can fool us into regarding it as a stochastic one; disentangling the deterministic from the stochastic is the main challenge in many real-life settings, from stock markets to palpitations of chicken hearts. So, what is 'chaos'?

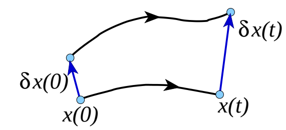


Figure 1.3: Unstable trajectories separate with time.

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

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

where λ, the mean rate of separation of trajectories of the system, is called the *Lyapunov exponent*. For any finite accuracy $\delta x = |\delta\mathbf{x}(0)|$ of the initial data, the dynamics is predictable only up to a finite *Lyapunov time* section 17.3

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

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

A positive Lyapunov exponent does not in itself lead to chaos. One could try to play 1- or 2-disk pinball game, but it would not be much of a game; trajectories would only separate, never to meet again. What is also needed is *mixing*, the coming together again and again of trajectories. While locally the nearby trajectories separate, the interesting dynamics is confined to a globally finite region of the state space and thus the separated trajectories are necessarily folded back and can re-approach each other arbitrarily closely, infinitely many times. For the case at hand there are 2^n topologically distinct n bounce trajectories that originate from a given disk. More generally, the number of distinct trajectories with n bounces can be quantified as section 15.1

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

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

The appellation 'chaos' is a confusing misnomer, as in deterministic dynamics there is no chaos in the everyday sense of the word; everything proceeds mathematically–that is, as Baron Leibniz would have it, infallibly. When a physicist says that a certain system exhibits 'chaos,' he means that the system obeys deterministic laws of evolution, but that the outcome is highly sensitive to small uncertainties in the specification of the initial state. The word 'chaos' has in this

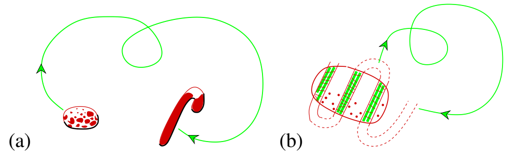


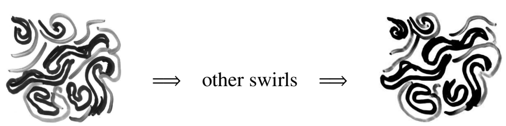
Figure 1.4: Dynamics of a *chaotic* dynamical system is (a) everywhere locally unstable (positive Lyapunov exponent) and (b) globally mixing (positive entropy). (A. Johansen)

context taken on a narrow technical meaning. If a deterministic system is locally unstable (positive Lyapunov exponent) and globally mixing (positive entropy)–figure 1.4–it is said to be *chaotic*.

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

A definition of 'turbulence' is even harder to come by. Intuitively, the word refers to irregular behavior of an infinite-dimensional dynamical system described by deterministic equations of motion–say, a bucket of sloshing water described by the Navier-Stokes equations. But in practice the word 'turbulence' tends to refer to messy dynamics which we understand poorly. As soon as a phenomenon is understood better, it is reclaimed and renamed: 'a route to chaos', 'spatiotemporal chaos', and so on.

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



For any finite spatial resolution, a deterministic flow follows approximately for a finite time an unstable pattern belonging to a finite alphabet of admissible patterns,

and the long term dynamics can be thought of as a walk through the space of such patterns. In ChaosBook we recast this image into mathematics.

1.3.2 When does 'chaos' matter?

> In dismissing Pollock's fractals because of their limited magnification range, Jones-Smith and Mathur would also dismiss half the published investigations of physical fractals.
>
> — Richard P. Taylor [1.4, 1.5]

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

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

At present the theory that will be developed here is in practice applicable only to systems of a low intrinsic *dimension* – the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent (a description of its long time dynamics requires a space of high intrinsic dimension) we are out of luck. Hence insights that the theory offers in elucidating problems of fully developed turbulence, quantum field theory of strong interactions and early cosmology have been modest at best. Even that is a caveat with qualifications. There are applications–such as spatially extended (non-equilibrium) systems, plumber's turbulent pipes, etc.,–where the few important degrees of freedom can be isolated and studied profitably by methods to be described here.

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

In 1980's something happened that might be without parallel; this is an area of science where the advent of cheap computation had actually subtracted from our collective understanding. The computer pictures and numerical plots of fractal science of the 1980's have overshadowed the deep insights of the 1970's, and these pictures have since migrated into textbooks. By a regrettable oversight, ChaosBook has none, so 'Untitled 5' of figure 1.5 will have to do as the illustration of the power of fractal analysis. Fractal science posits that certain quantities remark 1.6

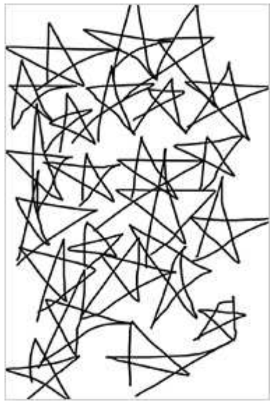


Figure 1.5: Katherine Jones-Smith, 'Untitled 5,' the drawing used by K. Jones-Smith and R.P. Taylor to test the fractal analysis of Pollock's drip paintings [1.6].

(Lyapunov exponents, generalized dimensions, ...) can be estimated on a computer. While some of the numbers so obtained are indeed mathematically sensible characterizations of fractals, they are in no sense observable and measurable on the length-scales and time-scales dominated by chaotic dynamics.

Even though the experimental evidence for the fractal geometry of nature is circumstantial [1.7], in studies of probabilistically assembled fractal aggregates we know of nothing better than contemplating such quantities. In deterministic systems we can do *much* better.

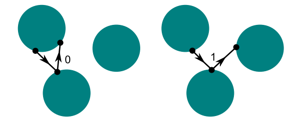
1.4 A game of pinball

Formulas hamper the understanding.
—S. Smale

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

Confronted with a potentially chaotic dynamical system, our analysis proceeds in three stages; I. diagnose, II. count, III. measure. First, we determine the intrinsic *dimension* of the system–the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent we are, at present, out of luck. We know only how to deal with the transitional regime

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



between regular motions and chaotic dynamics in a few dimensions. That is still something; even an infinite-dimensional system such as a burning flame front can turn out to have a very few chaotic degrees of freedom. In this regime the chaotic dynamics is restricted to a space of low dimension, the number of relevant parameters is small, and we can proceed to step II; we *count* and *classify* all possible chapter 11
topologically distinct trajectories of the system into a hierarchy whose successive chapter 15
layers require increased precision and patience on the part of the observer. This we shall do in sect. 1.4.2. If successful, we can proceed with step III: investigate the *weights* of the different pieces of the system.

We commence our analysis of the pinball game with steps I, II: diagnose, count. We shall return to step III–measure–in sect. 1.5. The three sections that chapter 20
follow are *highly* technical, they go into the guts of what the book is about. Is today is not your thinking day, skip them, jump straight to sect. 1.7.

1.4.1 Symbolic dynamics

With the game of pinball we are in luck–it is a low dimensional system, free motion in a plane. The motion of a point particle is such that after a collision with one disk it either continues to another disk or it escapes. If we label the three disks by 1, 2 and 3, we can associate every trajectory with an *itinerary*, a sequence of labels indicating the order in which the disks are visited; for example, the two trajectories in figure 1.2 have itineraries 2313, 23132321 respectively. exercise 1.1
Such labeling goes by the name *symbolic dynamics*. As the particle cannot collide section 2.1
two times in succession with the same disk, any two consecutive symbols must differ. This is an example of *pruning*, a rule that forbids certain subsequences of symbols. Deriving pruning rules is in general a difficult problem, but with the game of pinball we are lucky–for well-separated disks there are no further pruning rules. chapter 12

The choice of symbols is in no sense unique. For example, as at each bounce we can either proceed to the next disk or return to the previous disk, the above 3-letter alphabet can be replaced by a binary $\{0,1\}$ alphabet, figure 1.6. A clever choice of an alphabet will incorporate important features of the dynamics, such as its symmetries. section 11.6

Suppose you wanted to play a good game of pinball, that is, get the pinball to bounce as many times as you possibly can–what would be a winning strategy? The simplest thing would be to try to aim the pinball so it bounces many times between a pair of disks–if you managed to shoot it so it starts out in the periodic orbit bouncing along the line connecting two disk centers, it would stay there forever. Your game would be just as good if you managed to get it to keep bouncing

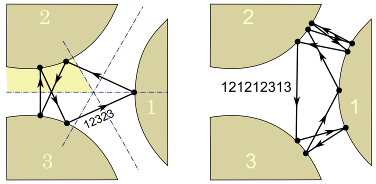


Figure 1.7: The 3-disk pinball cycles $\overline{1232}$ and $\overline{121212313}$.

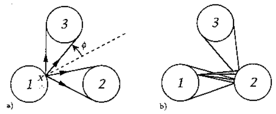


Figure 1.8: (a) A trajectory starting out from disk 1 can either hit another disk or escape. (b) Hitting two disks in a sequence requires a much sharper aim, with initial conditions that hit further consecutive disks nested within each other, as in Fig. 1.9.

between the three disks forever, or place it on any periodic orbit. The only rub is that any such orbit is *unstable*, so you have to aim very accurately in order to stay close to it for a while. So it is pretty clear that if one is interested in playing well, unstable periodic orbits are important–they form the *skeleton* onto which all trajectories trapped for long times cling.

1.4.2 Partitioning with periodic orbits

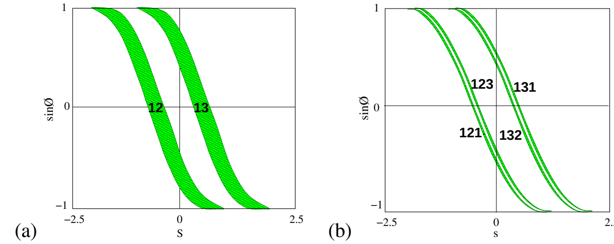
A trajectory is periodic if it returns to its starting position and momentum. We shall sometimes refer to the set of periodic points that belong to a given periodic orbit as a *cycle*.

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

Suppose that the pinball has just bounced off disk 1. Depending on its position and outgoing angle, it could proceed to either disk 2 or 3. Not much happens in between the bounces–the ball just travels at constant velocity along a straight line–so we can reduce the 4-dimensional flow to a 2-dimensional map P that takes the coordinates of the pinball from one disk edge to another disk edge. The trajectory just after the moment of impact is defined by s_n, the arc-length position of the 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, see figure 1.9. Such section of a flow is called a *Poincaré section*. In terms of Poincaré sections, the dynamics is reduced to the set of six *maps* $P_{s_k \leftarrow s_j} : (s_n, p_n) \mapsto (s_{n+1}, p_{n+1})$, with $s \in \{1, 2, 3\}$, from the boundary of the disk j to the boundary of the next disk k.

example 3.2
section 8

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



Next, we mark in the Poincaré section those initial conditions which do not escape in one bounce. There are two strips of survivors, as the trajectories originating from one disk can hit either of the other two disks, or escape without further ado. We label the two strips $\mathcal{M}_{12}$, $\mathcal{M}_{13}$. Embedded within them there are four strips $\mathcal{M}_{121}$, $\mathcal{M}_{123}$, $\mathcal{M}_{131}$, $\mathcal{M}_{132}$ of initial conditions that survive for two bounces, and so forth, see figures 1.8 and 1.9. Provided that the disks are sufficiently separated, after n bounces the survivors are divided into 2^n distinct strips: the $\mathcal{M}_i$th strip consists of all points with itinerary $i = s_1 s_2 s_3 \ldots s_n$, $s = \{1, 2, 3\}$. The unstable cycles as a skeleton of chaos are almost visible here: each such patch contains a periodic point $\overline{s_1 s_2 s_3 \ldots s_n}$ with the basic block infinitely repeated. Periodic points are skeletal in the sense that as we look further and further, the strips shrink but the periodic points stay put forever.

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

1.4.3 Escape rate

example 17.4

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

exercise 1.2

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

Consider figure 1.9 again. In each bounce the initial conditions get thinned out, yielding twice as many thin strips as at the previous bounce. The total area that remains at a given time is the sum of the areas of the strips, so that the fraction of survivors after n bounces, or the *survival probability* is given by

$$\hat{\Gamma}_1 = \frac{|\mathcal{M}_0|}{|\mathcal{M}|} + \frac{|\mathcal{M}_1|}{|\mathcal{M}|}, \qquad \hat{\Gamma}_2 = \frac{|\mathcal{M}_{00}|}{|\mathcal{M}|} + \frac{|\mathcal{M}_{10}|}{|\mathcal{M}|} + \frac{|\mathcal{M}_{01}|}{|\mathcal{M}|} + \frac{|\mathcal{M}_{11}|}{|\mathcal{M}|},$$

$$\hat{\Gamma}_n = \frac{1}{|\mathcal{M}|} \sum_i^{(n)} |\mathcal{M}_i|, \tag{1.2}$$

where i is a label of the ith strip, $|\mathcal{M}|$ is the initial area, and $|\mathcal{M}_i|$ is the area of the ith strip of survivors. $i = 01, 10, 11, \ldots$ is a label, not a binary number. Since at each bounce one routinely loses about the same fraction of trajectories, one expects the sum (1.2) to fall off exponentially with n and tend to the limit chapter 22

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

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

1.5 Chaos for cyclists

> Étant données des équations ... et une solution particuliére quelconque de ces équations, on peut toujours trouver une solution périodique (dont la période peut, il est vrai, étre trés longue), telle que la différence entre les deux solutions soit aussi petite qu'on le veut, pendant un temps aussi long qu'on le veut. D'ailleurs, ce qui nous rend ces solutions périodiques si précieuses, c'est qu'elles sont, pour ansi dire, la seule bréche par où nous puissions esseyer de pénétrer dans une place jusqu'ici réputée inabordable.
>
> —H. Poincaré, *Les méthodes nouvelles de la méchanique céleste*

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

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

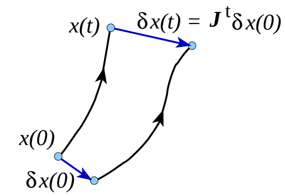


Figure 1.10: The Jacobian matrix J^t maps an infinitesimal displacement δx at x_0 into a displacement $J^t(x_0)\delta x$ finite time t later.

1.5.1 How big is my neighborhood?

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

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

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

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

A trajectory of a pinball moving on a flat surface is specified by two position coordinates and the direction of motion, so in this case $d = 3$. Evaluation of a cycle Jacobian matrix is a long exercise - here we just state the result. The Jacobian section 8.2 matrix describes the deformation of an infinitesimal neighborhood of $x(t)$ along the flow; its eigenvectors and eigenvalues give the directions and the corresponding rates of expansion or contraction, figure 1.10. The trajectories that start out in an infinitesimal neighborhood separate along the unstable directions (those whose eigenvalues are greater than unity in magnitude), approach each other along the stable directions (those whose eigenvalues are less than unity in magnitude), and maintain their distance along the marginal directions (those whose eigenvalues equal unity in magnitude).

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

As the heights of the strips in figure 1.9 are effectively constant, we can concentrate on their thickness. If the height is $\approx L$, then the area of the ith strip is $\mathcal{M}_i \approx L l_i$ for a strip of width l_i.

Each strip i in figure 1.9 contains a periodic point x_i. The finer the intervals, the smaller the variation in flow across them, so the contribution from the strip of width l_i is well-approximated by the contraction around the periodic point x_i within the interval,

$$l_i = a_i/|\Lambda_i| \,, \tag{1.5}$$

where Λ_i is the unstable eigenvalue of the Jacobian matrix $J^t(x_i)$ 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). Only the magnitude of this eigenvalue matters, we can disregard its sign. The prefactors a_i reflect the overall size of the system and the particular distribution of starting values of x. As the asymptotic trajectories are strongly mixed by bouncing chaotically around the repeller, we expect their distribution to be insensitive to smooth variations in the distribution of initial points. section 16.4

To proceed with the derivation we need the *hyperbolicity* assumption: for large n the prefactors $a_i \approx O(1)$ are overwhelmed by the exponential growth of Λ_i, so we neglect them. If the hyperbolicity assumption is justified, we can replace $|\mathcal{M}_i| \approx Ll_i$ in (1.2) by $1/|\Lambda_i|$ and consider the sum section 18.1.1

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

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:

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

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 (1.6) diverges:

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

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

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

For sufficiently small z this sum is convergent. The escape rate γ is now given by the leading pole of (1.7), rather than by a numerical extrapolation of a sequence of γ_n extracted from (1.3). As any finite truncation $n < n_{\mathrm{trunc}}$ of (1.8) is a polynomial in z, convergent for any z, finding this pole requires that we know something about Γ_n for any n, and that might be a tall order. section 18.3

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

1.5.2 Dynamical zeta function

If a trajectory retraces a *prime* cycle r times, its expanding eigenvalue is Λ_p^r. A prime cycle p is a single traversal of the orbit; its label is a non-repeating symbol string of n_p symbols. There is only one prime cycle for each cyclic permutation class. For example, $p = \overline{0011} = \overline{1001} = \overline{1100} = \overline{0110}$ 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 (1.8). Hence (1.8) can be rewritten as exercise 15.2 section 4.5

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

where the index p runs through all distinct *prime* cycles. Note that we have resummed the contribution of the cycle p to all times, so truncating the summation up to given p is *not* a finite time $n \leq n_p$ approximation, but an asymptotic, *infinite* time estimate based by approximating stabilities of all cycles by a finite number of the shortest cycles and their repeats. The $n_p z^{n_p}$ factors in (1.9) suggest rewriting the sum as a derivative

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

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

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

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

(1.10). The escape rate is related by (1.7) to a divergence of $\Gamma(z)$, and $\Gamma(z)$ diverges whenever $1/\zeta(z)$ has a zero. section 22.1 section 19.4

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

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

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

1.5.3 Cycle expansions

How are formulas such as (1.10) used? We start by computing the lengths and eigenvalues of the shortest cycles. This usually requires some numerical work, such as the Newton method searches for periodic solutions; we shall assume that the numerics are under control, and that *all* short cycles up to given length have been found. In our pinball example this can be done by elementary geometrical optics. It is very important not to miss any short cycles, as the calculation is as accurate as the shortest cycle dropped–including cycles longer than the shortest omitted does not improve the accuracy (unless exponentially many more cycles are included). The result of such numerics is a table of the shortest cycles, their periods and their stabilities. chapter 13 section 29.3

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

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

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 20 section 20.1

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

1.5.4 Shadowing

When you actually start computing this escape rate, you will find out that the convergence is very impressive: only three input numbers (the two fixed points $\overline{0}$,

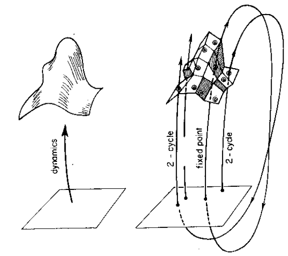


Figure 1.11: Approximation to a smooth dynamics (left frame) by the skeleton of periodic points, together with their linearized neighborhoods, (right frame). Indicated are segments of two 1-cycles and a 2-cycle that alternates between the neighborhoods of the two 1-cycles, shadowing first one of the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

$\overline{1}$ and the 2-cycle $\overline{10}$) already yield the pinball escape rate to 3-4 significant digits! We have omitted an infinity of unstable cycles; so why does approximating the dynamics by a finite number of the shortest cycle eigenvalues work so well? section 20.2.2

The convergence of cycle expansions of dynamical zeta functions is a consequence of the smoothness and analyticity of the underlying flow. Intuitively, one can understand the convergence in terms of the geometrical picture sketched in figure 1.11; the key observation is that the long orbits are *shadowed* by sequences of shorter orbits.

A typical term in (1.11) is a difference of a long cycle $\{ab\}$ minus its shadowing approximation by shorter cycles $\{a\}$ and $\{b\}$

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

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

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

To understand this on a more general level, try to visualize the partition of a chaotic dynamical system’s state space in terms of cycle neighborhoods as a tessellation (a tiling) of the dynamical system, with smooth flow approximated by its periodic orbit skeleton, each ‘tile’ centered on a periodic point, and the scale of the ‘tile’ determined by the linearization of the flow around the periodic point, as illustrated by figure 1.11.

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

chapter 23

1.6 Change in time

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

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

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

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

is the Dirac delta function, as for a deterministic flow the initial point x maps into a unique point y at time t. For discrete time, $f^n(x)$ is the 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 $\mathcal{A}$, the generator of infinitesimal time translations

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

section 16.6

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

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

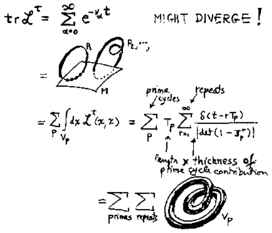


Figure 1.12: The trace of an evolution operator is concentrated in tubes around prime cycles, of length T_p and thickness $1/|\Lambda_p|^r$ for the rth repetition of the prime cycle p.

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

1.6.1 Trace formula

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

—Matthias Neubert

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

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

section 18.2

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

where T_p is the period of prime cycle p, and the monodromy matrix M_p is the flow-transverse part of Jacobian matrix J (1.4). This formula has a simple geometrical interpretation sketched in figure 1.12. After the 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 $1/\left|\det\left(\mathbf{1} - M_p^r\right)\right| \to 1/|\Lambda_p|$, the same weight we obtained heuristically in sect. 1.5.1.

The 'spiky' sum (1.14) is disquieting in the way reminiscent of the Poisson resummation formulas of Fourier analysis; the left-hand side is the smooth eigenvalue sum $\mathrm{tr}\, e^{\mathcal{A}t} = \sum e^{s_\alpha t}$, while the right-hand side equals zero everywhere except for the set $t = rT_p$. A Laplace transform smooths the sum over Dirac delta functions in cycle periods and yields the *trace formula* for the eigenspectrum $s_0, s_1, \cdots$ of the classical evolution operator: chapter 18

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

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

1.6.2 Spectral determinant

The eigenvalues of a linear operator are given by the zeros of the appropriate determinant. One way to evaluate determinants is to expand them in terms of traces, using the identities exercise 4.1

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

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 19

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

The $1/r$ factor is due to the s integration, leading to the replacement $T_p \to T_p/rT_p$ in the periodic orbit expansion (1.15). section 19.5

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

1.7 From chaos to statistical mechanics

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

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

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

A century ago it seemed reasonable to assume that statistical mechanics applies only to systems with very many degrees of freedom. More recent is the realization that much of statistical mechanics follows from chaotic dynamics, and already at the level of a few degrees of freedom the evolution of densities is irreversible. Furthermore, the theory that we shall develop here generalizes notions of 'measure' and 'averaging' to systems far from equilibrium, and transports us into regions hitherto inaccessible with the tools of equilibrium statistical mechanics.

By going to a description in terms of the asymptotic time evolution operators we give up tracking individual trajectories for long times, but trade in the uncontrollable trajectories for a powerful description of the asymptotic trajectory densities. This will enable us, for example, to give exact formulas for transport coefficients such as the diffusion constants without *any* probabilistic assumptions. chapter 25
The classical Boltzmann equation for evolution of 1-particle density is based on *stosszahlansatz*, neglect of particle correlations prior to, or after a 2-particle collision. It is a very good approximate description of dilute gas dynamics, but a difficult starting point for inclusion of systematic corrections. In the theory developed here, no correlations are neglected - they are all included in the cycle averaging formulas such as the cycle expansion for the diffusion constant $2dD = \lim_{T\to\infty} \left\langle x(T)^2 \right\rangle / T$ of a particle diffusing chaotically across a spatially-periodic array, section 25.1

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

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

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

chapter 24

1.8 Chaos: what is it good for?

> Happy families are all alike; every unhappy family is unhappy in its own way.
>
> — *Anna Karenina*, by Leo Tolstoy

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

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

Several tabletop experiments that could measure transport on macroscopic scales are sketched in figure 1.13 (each a tabletop, but an expensive tabletop). Figure 1.13 (a) depicts a 'slanted washboard;' a particle in a gravity field bouncing down the washboard, losing some energy at each bounce, or a charged particle in a constant electric field trickling across a periodic condensed-matter device. The interplay between chaotic dynamics and energy loss results in a terminal mean velocity/conductance, a function of the washboard slant or external electric field that the periodic theory can predict accurately. Figure 1.13 (b) depicts a 'cold atom lattice' of very accurate spatial periodicity, with a dilute cloud of atoms placed onto a standing wave established by strong laser fields. Interaction of gravity with gentle time-periodic jiggling of the EM fields induces a diffusion of the atomic cloud, with a diffusion constant predicted by the periodic orbit theory. Figure 1.13 (c) depicts a tip of an atomic force microscope (AFM) bouncing against a periodic atomic surface moving at a constant velocity. The frictional drag experienced is the interplay of the chaotic bouncing of the tip and the energy loss at each

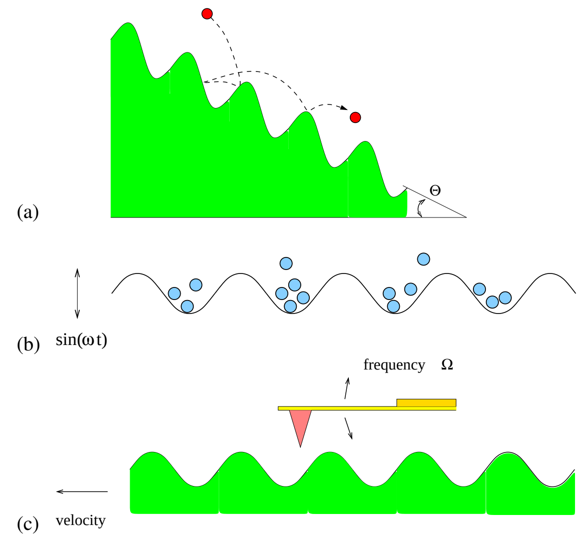


Figure 1.13: (a) Washboard mean velocity, (b) cold atom lattice diffusion, and (c) AFM tip drag force. (Y. Lan)

tip/surface collision, accurately predicted by the periodic orbit theory. None of these experiments have actually been carried out, (save for some numerical experimentation), but are within reach of what can be measured today.

ChaosBook.org/projects

Given microscopic dynamics, periodic orbit theory predicts observable macroscopic transport quantities such as the washboard mean velocity, cold atom lattice diffusion constant, and AFM tip drag force. But the experimental proposal is sexier than that, and goes into the heart of dynamical systems theory.

remark A.1

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

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

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

But dynamical systems families are ‘families’ only in a name. That the structural stability conjecture turned out to be badly wrong is, however, not a blow for chaotic dynamics. Quite to the contrary, it is actually a virtue, perhaps the most section 12.2 dramatic experimentally measurable prediction of chaotic dynamics.

As long as microscopic periodicity is exact, the prediction is counterintuitive for a physicist - transport coefficients are *not* smooth functions of system parame- section 25.2 ters, rather they are non-monotonic, *nowhere differentiable* functions. Conversely, if the macroscopic measurement yields a smooth dependence of the transport on system parameters, the periodicity of the microscopic lattice is degraded by impurities, and probabilistic assumptions of traditional statistical mechanics apply. So the proposal is to –by measuring *macroscopic transport*– conductance, diffusion, drag –observe determinism on *nanoscales*, and –for example– determine whether an atomic surface is clean.

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

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

1.9 What is not in ChaosBook

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

This book offers everyman a breach into a domain hitherto reputed unreachable, a domain traditionally traversed only by mathematical physicists and mathematicians. What distinguishes it from mathematics is the insistence on computability and numerical convergence of methods offered. A rigorous proof, the end of the story as far as a mathematician is concerned, might state that in a given setting, for times in excess of 10^{32} years, turbulent dynamics settles onto an attractor of dimension less than 600. Such a theorem is of a little use to an honest, hard-working plumber, especially if her hands-on experience is that within the span of a few typical ‘turnaround’ times the dynamics seems to settle on a (transient?) attractor of dimension less than 3. If rigor, magic, fractals or brains is your thing, read remark 1.4 and beyond.

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

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

Résumé

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

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

The theory of evaluation of the spectra of evolution operators presented here is based on the observation that the motion in dynamical systems of few degrees of freedom is often organized around a few *fundamental* cycles. These short cycles capture the skeletal topology of the motion on a strange attractor/repeller in the sense that any long orbit can approximately be pieced together from the nearby periodic orbits of finite length. This notion is made precise by approximating orbits by prime cycles, and evaluating the associated curvatures. A curvature measures the deviation of a longer cycle from its approximation by shorter cycles; smoothness and the local instability of the flow implies exponential (or faster) fall-off for (almost) all curvatures. Cycle expansions offer an efficient method for evaluating classical and quantum observables.

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

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

1. the weight associated with the observable under consideration is multiplicative along the trajectory,

2. the set is organized in such a way that the nearby points in the symbolic dynamics have nearby weights.

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

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

—Gibbon

Commentary

Remark 1.1 Nonlinear dynamics texts. This text aims to bridge the gap between the physics and mathematics dynamical systems literature. The intended audience is Henri Roux, the perfect physics graduate student with a theoretical bent who does not believe anything he is told. As a complementary presentation we recommend Gaspard's monograph [1.8] which covers much of the same ground in a highly readable and scholarly manner.

As far as the prerequisites are concerned–ChaosBook is not an introduction to nonlinear dynamics. Nonlinear science requires a one semester basic course (advanced undergraduate or first year graduate). A good start is the textbook by Strogatz [1.9], an introduction to the applied mathematician's visualization of flows, fixed points, manifolds, bifurcations. It is the most accessible introduction to nonlinear dynamics–a book on differential equations in nonlinear disguise, and its broadly chosen examples and many exercises make it a favorite with students. It is not strong on chaos. There the textbook of Alligood, Sauer and Yorke [1.10] is preferable: an elegant introduction to maps, chaos, period doubling, symbolic dynamics, fractals, dimensions–a good companion to ChaosBook. Introduction more comfortable to physicists is the textbook by Ott [1.11], with the baker's map used to illustrate many key techniques in analysis of chaotic systems. Ott is perhaps harder than the above two as first books on nonlinear dynamics. Sprott [1.12] and Jackson [1.13] textbooks are very useful compendia of the '70s and onward 'chaos' literature which we, in the spirit of promises made in sect. 1.1, tend to pass over in silence.

An introductory course should give students skills in qualitative and numerical analysis of dynamical systems for short times (trajectories, fixed points, bifurcations) and familiarize them with Cantor sets and symbolic dynamics for chaotic systems. For the dynamical systems material covered here in chapters 2 to 4, as well as for the in-depth study of bifurcation theory we warmly recommend Kuznetsov [1.14]. A good introduction to numerical experimentation with physically realistic systems is Tufillaro, Abbott, and Reilly [1.15]. Korsch and Jodl [1.16] and Nusse and Yorke [1.17] also emphasize hands-on approach to dynamics. With this, and a graduate level-exposure to statistical mechanics, partial differential equations and quantum mechanics, the stage is set for any of the one-semester advanced courses based on ChaosBook.

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

Deterministic chaos. Chaotic averaging, evolution operators, trace formulas, zeta functions, cycle expansions, Lyapunov exponents, billiards, transport coefficients, thermodynamic formalism, period doubling, renormalization operators. A graduate level introduction to statistical mechanics from the dynamical point view is given by Dorfman [1.18]; the Gaspard monograph [1.8] covers the same ground in more depth. Driebe monograph [1.19] 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?"* [1.20].

Spatiotemporal dynamical systems. Partial differential equations for dissipative systems, weak amplitude expansions, normal forms, symmetries and bifurcations, pseudospectral methods, spatiotemporal chaos, turbulence. Holmes, Lumley and Berkooz [1.21] offer a delightful discussion of why the Kuramoto-Sivashinsky equation deserves study as a staging ground for a dynamical approach to study of turbulence in full-fledged Navier-Stokes boundary shear flows.

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

Remark 1.3 Periodic orbit theory. This book puts more emphasis on periodic orbit theory than any other current nonlinear dynamics textbook. The role of unstable periodic orbits was already fully appreciated by Poincaré [1.22, 1.23], 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 [1.24] for an overview of some of that literature.

Remark 1.4 If you seek rigor? If you find ChaosBook not rigorous enough, you should turn to the mathematics literature. We recommend Robinson's advanced graduate level exposition of dynamical systems theory [1.25] from Smale perspective. The most extensive reference is the treatise by Katok and Hasselblatt [1.26], an impressive compendium of modern dynamical systems theory. The fundamental papers in this field, all still valuable reading, are Smale [1.27], Bowen [1.28] and Sinai [1.29]. Sinai's paper is prescient and offers a vision and a program that ties together dynamical systems and statistical mechanics. It is written for readers versed in statistical mechanics. For a dynamical systems exposition, consult Anosov and Sinai [1.30]. Markov partitions were introduced by Sinai in ref. [1.31]. The classical text (though certainly not an easy read) on the subject of dynamical zeta functions is Ruelle's *Statistical Mechanics, Thermodynamic Formalism* [1.32]. 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 [1.33]. Further excellent mathematical references on thermodynamic formalism are Parry and Pollicott's monograph [1.34] with emphasis on the symbolic dynamics aspects of the formalism, and Baladi's clear and compact reviews of the theory of dynamical zeta functions [1.35, 1.36].

Remark 1.5 If you seek magic? ChaosBook resolutely skirts number-theoretical magic such as spaces of constant negative curvature, Poincaré tilings, modular domains, Selberg Zeta functions, Riemann hypothesis, . . . Why? While this beautiful mathematics has been very inspirational, especially in studies of quantum chaos, almost no powerful method in its repertoire survives a transplant to a physical system that you are likely to care about.

Remark 1.6 Sorry, no schmactals! ChaosBook skirts mathematics and empirical practice of fractal analysis, such as Hausdorff and fractal dimensions. Addison's introduction to fractal dimensions [1.37] offers a well-motivated entry into this field. While in studies of probabilistically assembled fractals such as diffusion limited aggregates (DLA) better measures of 'complexity' are lacking, for deterministic systems there are much better, physically motivated and experimentally measurable quantities (escape rates, diffusion coefficients, spectrum of helium, ...) that we focus on here.

Remark 1.7 Rat brains? If you were wondering while reading this introduction 'what's up with rat brains?', the answer is yes indeed, there is a line of research in neuronal dynamics that focuses on possible unstable periodic states, described for example in refs. [1.38, 1.39, 1.40, 1.41].

A guide to exercises

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

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

exercise 20.2

- **Dynamics**
  1. count prime cycles, exercise 1.1, exercise 9.6, exercise 11.1
  2. pinball simulator, exercise 8.1, exercise 13.4
  3. pinball stability, exercise 13.7, exercise 13.4
  4. pinball periodic orbits, exercise 13.5, exercise 13.6
  5. helium integrator, exercise 2.10, exercise 13.11
  6. helium periodic orbits, exercise 13.12
- **Averaging, numerical**
  1. pinball escape rate, exercise 17.3
- **Averaging, periodic orbits**
  1. cycle expansions, exercise 20.1, exercise 20.2
  2. pinball escape rate, exercise 20.4, exercise 20.5
  3. cycle expansions for averages, exercise 20.1, exercise 22.3
  4. cycle expansions for diffusion, exercise 25.1
  5. pruning, transition graphs, exercise 15.6
  6. desymmetrization exercise 21.1
  7. intermittency, phase transitions, exercise 24.6

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

Exercises

1.1. **3-disk symbolic dynamics.** As periodic trajectories will turn out to be our main tool to breach deep into the realm of chaos, it pays to start familiarizing oneself with them now by sketching and counting the few shortest prime cycles (we return to this in sect. 15.4). Show that the 3-disk pinball has $3 \cdot 2^{n-1}$ itineraries of length n. List periodic orbits of lengths 2, 3, 4, 5, $\cdots$. Verify that the shortest 3-disk prime cycles are 12, 13, 23, 123, 132, 1213, 1232, 1323, 12123, $\cdots$. Try to sketch them. (continued in exercise 12.6)

1.2. **Sensitivity to initial conditions.** Assume that two pinball trajectories start out parallel, but separated by 1 Ångström, and the disks are of radius $a = 1$ cm and center-to-center separation $R = 6$ cm. Try to estimate in how many bounces the separation will grow to the size of system (assuming that the trajectories have been picked so they remain trapped for at least that long). Estimate the Who's *Pinball Wizard*'s typical score (number of bounces) in a game without cheating, by hook or crook (by the end of chapter 20 you should be in position to make very accurate estimates).

References

[1.1] J. Guckenheimer, "Piecewise-Smooth Dynamical Systems (*Book Review*)" *SIAM Rev.* **50**, 606 (2008).

[1.2] G.W. Leibniz, *Von dem Verhängnisse*, in *Hauptschriften zur Grundlegung der Philosophie*, Vol II, , pp. 129-134 (Ernst Cassirer, Leibzig 1906).

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

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

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

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

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

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

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

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