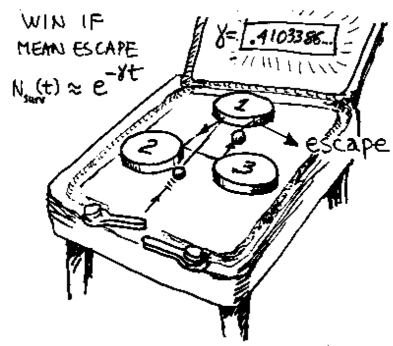

Figure 1.1: A physicist's bare bones game of pinball.

1.3 The future as in a mirror

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

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

We all have an intuitive feeling for what a ball does as it bounces among the pinball machine's disks, and only high-school level Euclidean geometry is needed to describe its trajectory. A physicist's pinball game is the game of pinball stripped to its bare essentials: three equidistantly placed reflecting disks in a plane, 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 [1.1], anticipating by a century and a half the oft-quoted Laplace's "Given for one instant an intelligence which could comprehend all the forces by which nature is animated...":

> That everything is brought forth through an established destiny is just as certain as that three times three is nine. [. . .] If, for example, one sphere meets another sphere in free space and if their sizes and their paths and directions before collision are known, we can then foretell and calculate how they will rebound and what course they will take after the impact. Very simple laws are followed which also apply,

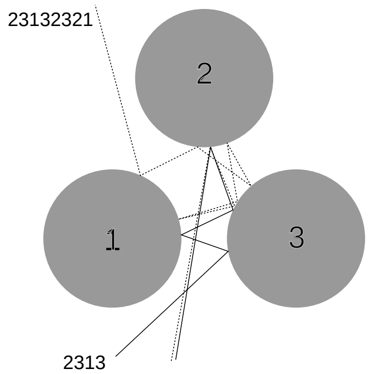

Figure 1.2: Sensitivity to initial conditions: two pinballs that start out very close to each other separate exponentially with time.

> no matter how many spheres are taken or whether objects are taken other than spheres. From this one sees then that everything proceeds mathematically – that is, infallibly – in the whole wide world, so that if someone could have a sufficient insight into the inner parts of things, and in addition had remembrance and intelligence enough to consider all the circumstances and to take them into account, he would be a prophet and would see the future in the present as in a mirror.

Leibniz chose to illustrate his faith in determinism precisely with the type of physical system that we shall use here as a paradigm of "chaos". His claim is wrong in a deep and subtle way: a state of a physical system can *never* be specified to infinite precision, there is no way to take all the circumstances into account, and a single trajectory cannot be tracked, only a ball of nearby initial points makes physical sense.

1.3.1 What is "chaos"?

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

A deterministic system is a system whose present state is *in principle* fully determined by its initial conditions, in contrast to a stochastic system. For a stochastic system the initial conditions determine the future only partially, due to noise, or other external circumstances beyond our control: the present state reflects the past initial conditions plus the particular realization of the noise encountered along the way.

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

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

This property of *sensitivity to initial conditions* can be quantified as

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

☞ sect. 10.3

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*

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

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 phase 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

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

☞ sect. 13.1

☞ sect. 20.1

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

The appellation "chaos" is a confusing misnomer, as in deterministic dynamics there is no chaos in the everyday sense of the word; everything proceeds mathematically – that is, as Baron Leibniz would have it, infallibly. When a physicist says that a certain system exhibits "chaos," he means that the system obeys deterministic laws of evolution, but that the outcome is highly sensitive to small uncertainties in the specification of the initial state. The word "chaos" has in this context taken on a narrow technical meaning. If a deterministic system is locally unstable (positive Lyapunov exponent) and globally mixing (positive entropy) - figure 1.3 - 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

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)

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 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.

☞ appendix B

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:

 $\Longrightarrow$ 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. In ChaosBook we recast this image into mathematics.

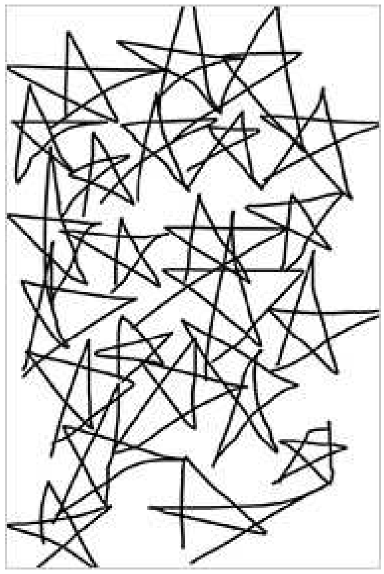

Figure 1.4: Katherine Jones-Smith, "Untitled 5," the drawing used by K. Jones-Smith and R.P. Taylor to test the fractal analysis of Pollock's drip paintings [1.3].

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 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.

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 is in practice applicable only to systems with a low intrinsic *dimension* – the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent (a description of its long time dynamics requires a space of 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 (nonequilibrium) 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.

☞ chapter 23

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 interburst 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 1.4 will have to do as the illustration of the power of fractal analysis. Fractal science posits that certain quantities (Lyapunov exponents, generalized dimensions, . . .) can be estimated on a computer. While some of the numbers so obtained are indeed mathematically sensible characterizations of fractals, they are in no sense observable and measurable on the length-scales and time-scales dominated by chaotic dynamics.

Even though the experimental evidence for the fractal geometry of nature is circumstantial [1.2], 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 brasstacks. Time to fasten your seatbelts and turn off all electronic devices. But first, a disclaimer: If you understand the rest of this chapter on the first reading, you either do not need this book, or you are delusional. If you do not understand it, it is not because the people who wrote it are smarter than you: the most you can hope for at this stage is to get a flavor of what lies ahead. If a statement in this chapter mystifies/intrigues, fast forward to a section indicated by ☞ on the margin, read only the parts that you feel you need. Of course, we think that you need to learn ALL of it, or otherwise we would not have selected it in the first place.

Confronted with a potentially chaotic dynamical system, we analyze it through a sequence of three distinct stages; I. diagnose, II. count, III. measure. First we determine the intrinsic *dimension* of the system – the minimum number of coordinates necessary to capture its essential dynamics. If the system is very turbulent we are, at present, out of luck. We know only how to deal with the transitional regime between regular motions and chaotic dynamics in a few dimensions. That is still something; even an

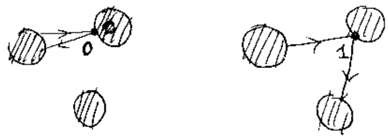

Figure 1.5: Binary labeling of the 3-disk pinball trajectories; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

infinite-dimensional system such as a burning flame front can turn out to have a very few chaotic degrees of freedom. In this regime the chaotic dynamics is restricted to a space of low dimension, the number of relevant parameters is small, and we can proceed to step II; we *count* and *classify* all possible topologically distinct trajectories of the system into a hierarchy whose successive layers require increased precision and patience on the part of the observer. This we shall do in sect. 1.4.1. If successful, we can proceed with step III: investigate the *weights* of the different pieces of the system.

☞ chapter 11
☞ chapter 13

We commence our analysis of the pinball game with steps I, II: diagnose, count. We shall return to step III – measure – in sect. 1.5.

☞ chapter 18

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. The itinerary is finite for a scattering trajectory, coming in from infinity and escaping after a finite number of collisions, infinite for a trapped trajectory, and infinitely repeating for a periodic orbit. 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.

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

☞ sect. 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 between the three disks

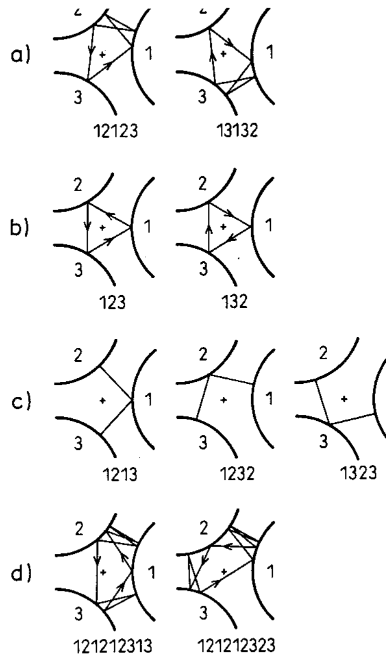

Figure 1.6: Some examples of 3-disk cycles: (a) $\overline{12123}$ and $\overline{13132}$ are mapped into each other by the flip across 1 axis. Similarly (b) $\overline{123}$ and $\overline{132}$ are related by flips, and (c) $\overline{1213}$, $\overline{1232}$ and $\overline{1323}$ by rotations. (d) The cycles $\overline{121212313}$ and $\overline{121212323}$ are related by rotaion *and* time reversal. These symmetries are discussed in more detail in chapter 22. (from ref. [1.6])

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. 35.2

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 figure 1.6 - 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 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 phase space trajectories onto a 2-dimensional subspace, the configuration space. A clearer picture of the dynamics is obtained by constructing a phase space Poincaré section.

Suppose that the pinball has just bounced off disk 1. Depending on its position and outgoing angle, it could proceed to either disk 2 or 3. Not much happens in between the bounces – the ball just travels at constant velocity along a straight line – so we can reduce the four-dimensional flow to a two-dimensional map f that takes the coordinates of the pinball from one disk edge to another disk edge. Let us state this more precisely: the trajectory

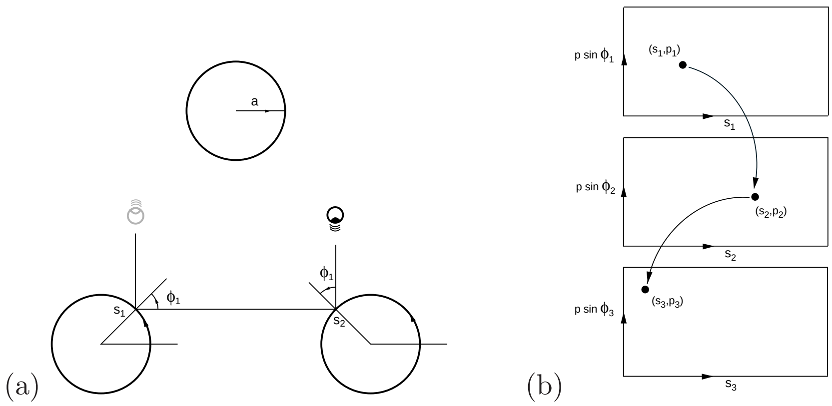

Figure 1.7: (a) The Poincaré section coordinates for the 3-disk game of pinball. (b) Collision sequence $(s_1, p_1) \mapsto (s_2, p_2) \mapsto (s_3, p_3)$ from the boundary of a disk to the boundary of the next disk presented in the Poincaré section coordinates.

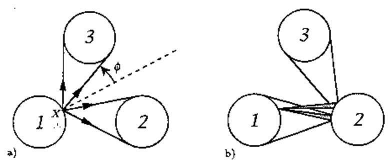

Figure 1.8: (a) A trajectory starting out from disk 1 can either hit another disk or escape. (b) Hitting two disks in a sequence requires a much sharper aim. The cones of initial conditions that hit more and more consecutive disks are nested within each other, as in figure 1.9.

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, figure 1.7. Such a 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. 6

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 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 $i = s_1 s_2 s_3 \ldots s_n$, $s = \{0, 1\}$. 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 utilize a symbolic dynamics; it provides a navigation chart through chaotic phase space. There exists a unique trajectory for every admissible infinite length itinerary, and a unique itinerary

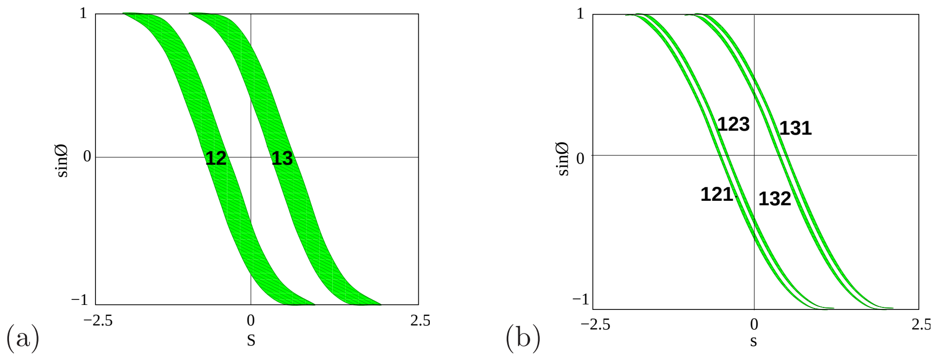

Figure 1.9: The 3-disk game of pinball Poincaré section, trajectories emanating from the disk 1 with $x_0 =$ (arclength, parallel momentum) $= (s_0, p_0)$, disk radius : center separation ratio a:R = 1:2.5. (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in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

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

☞ example 10.1

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 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 such a non-confining 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.

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 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*

Chapter 11

Qualitative dynamics, for pedestrians

> The classification of the constituents of a chaos, nothing less is here essayed.
> Herman Melville, *Moby Dick*, chapter 32

In this chapter we begin to learn how to use qualitative properties of a flow in order to *partition* the phase space in a topologically invariant way, and *name* topologically distinct orbits. This will enable us – in chapter 13 – to *count* the distinct orbits, and in the process touch upon all the main themes of this book, going the whole distance from diagnosing chaotic dynamics to computing zeta functions.

We start by a simple physical example, symbolic dynamics of a 3-disk game of pinball, and then show that also for smooth flows the qualitative dynamics of stretching and folding flows enables us to partition the phase space and assign symbolic dynamics itineraries to trajectories. Here we illustrate the method on a $1-d$ approximation to Rössler flow. In chapter 13 we turn this topological dynamics into a multiplicative operation on the phase space partitions by means of transition matrices/Markov graphs, the simplest examples of evolution operators. Deceptively simple, this subject can get very difficult very quickly, so in this chapter we do the first pass, at a pedestrian level, postponing the discussion of higher-dimensional, cyclist level issues to chapter 12.

Even though by inclination you might only care about the serious stuff, like Rydberg atoms or mesoscopic devices, and resent wasting time on things formal, this chapter and chapter 13 are good for you. Read them.

11.1 Qualitative dynamics

(R. Mainieri and P. Cvitanović)

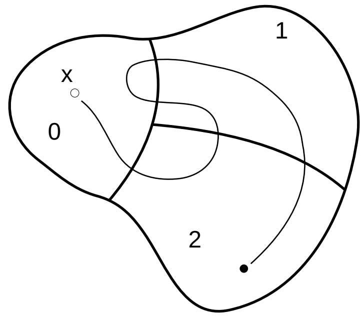

Figure 11.1: A trajectory with itinerary 021012.

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, and call it the neighborhood $\mathcal{M}_0$, as long as it encloses x_0. For chaotic dynamical systems, the evolution might bring the point back to the starting neighborhood infinitely often. That is, the set

$$\left\{y \in \mathcal{M}_0 : \quad y = f^t(x_0), \quad t > t_0\right\} \tag{11.1}$$

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. 11.6; 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 figure 11.1, the phase space is divided into three regions $\mathcal{M}_0$,

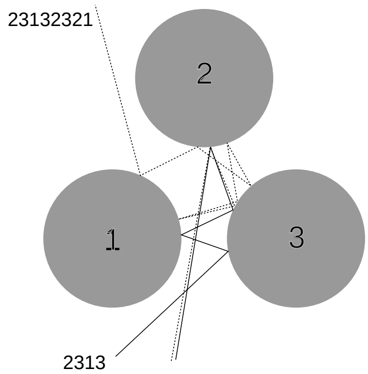

Figure 11.2: Two pinballs that start out very close to each other exhibit the same qualitative dynamics _2313_ for the first three bounces, but due to the exponentially growing separation of trajectories with time, follow different itineraries thereafter: one escapes after _2313_, the other one escapes after _23132321_.

$\mathcal{M}_1$, 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, i.e., one should be able to go from any region $\mathcal{M}_i$ to any other region $\mathcal{M}_j$ in a finite number of steps. A dynamical system with such partition is said to be *metrically indecomposable*.

In general one also encounters transient regions - regions to which the dynamics does not return to once they are exited. Hence we have to distinguish between (for us uninteresting) wandering trajectories that never return to the initial neighborhood, and the non–wandering set (2.2) of the *recurrent* trajectories.

The allowed transitions between the regions of a partition are encoded in the $[N \times N]$-dimensional *transition matrix* whose elements take values

$$T_{ij} \quad = \quad \begin{cases} 1 & \text{if a transition } \mathcal{M}_j \to \mathcal{M}_i \text{ is possible} \\ 0 & \text{otherwise}\,. \end{cases} \tag{11.2}$$

The transition matrix encodes the topological dynamics as an invariant law of motion, with the allowed transitions at any instant independent of the trajectory history, requiring no memory.

Example 11.1 Complete N-ary dynamics: *All transition matrix entries equal unity (one can reach any region from any other region in one step):*

$$T_c = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} . \tag{11.3}$$

Further examples of transition matrices, such as the 3-disk transition matrix (11.5) and the 1-step memory sparse matrix (11.15), are peppered throughout the text.

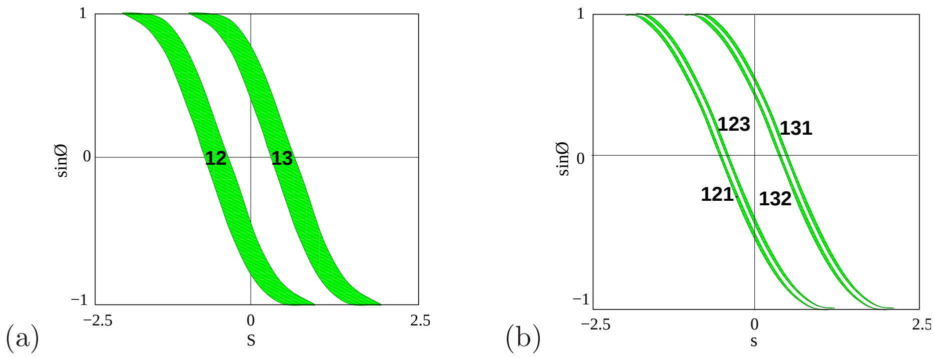

Figure 11.3: The 3-disk game of pinball Poincaré section, trajectories emanating from the disk 1 with $x_0 =$ (arclength, parallel momentum) $= (s_0, p_0)$, disk radius : center separation ratio a:R = 1:2.5. (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in 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)

However, knowing that a point from $\mathcal{M}_i$ reaches $\mathcal{M}_j$ in one step is not quite good enough. We would be happier if we knew that *any* point in $\mathcal{M}_i$ reaches $\mathcal{M}_j$; otherwise we have to subpartition $\mathcal{M}_i$ into the points which land in $\mathcal{M}_j$, and those which do not, and often we will find ourselves partitioning *ad infinitum*.

Such considerations motivate the notion of a *Markov partition*, a partition for which no memory of preceding steps is required to fix the transitions allowed in the next step. Dynamically, *finite Markov partitions* can be generated by *expanding* d-dimensional iterated mappings $f : \mathcal{M} \to \mathcal{M}$, if $\mathcal{M}$ can be divided into N regions $\{\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_{N-1}\}$ such that in one step points from an initial region $\mathcal{M}_i$ either fully cover a region $\mathcal{M}_j$, or miss it altogether,

$$\text{either} \quad \mathcal{M}_j \cap f(\mathcal{M}_i) = \emptyset \quad \text{or} \quad \mathcal{M}_j \subset f(\mathcal{M}_i) \,. \tag{11.4}$$

Let us illustrate what this means by our favorite example, the game of pinball.

1.1
page 32

Example 11.2 3-disk symbolic dynamics: *Consider the motion of a free point particle in a plane with 3 elastically reflecting convex disks. After a collision with a disk a particle either continues to another disk or escapes, and any trajectory can be labeled by the disk sequence. For example, if we label the three disks by 1, 2 and 3, the two trajectories in figure 11.2 have itineraries* $\_2313\_$, $\_23132321\_$ *respectively. The 3-disk prime cycles given in figures 1.6 and 11.6 are further examples of such itineraries.*

At each bounce a cone of initially nearby trajectories defocuses (see figure 1.8), and in order to attain a desired longer and longer itinerary of bounces the initial point $x_0 = (s_0, p_0)$ *has to be specified with a larger and larger precision, and lie within initial phase space strips drawn in figure 11.3. 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*

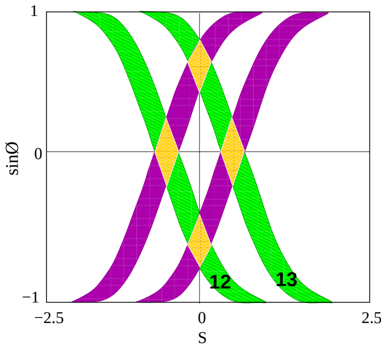

Figure 11.4: The Poincaré section of the phase space for the binary labeled pinball. For definitiveness, this set is generated by starting from disk 1, preceded by disk 2. Indicated are the fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle periodic points $\overline{01}$, $\overline{10}$, together with strips which survive 1, 2, . . . bounces. Iteration corresponds to the decimal point shift; for example, all points in the rectangle $[01.01]$ map into the rectangle $[010.1]$ in one iteration. See also figure 11.6 (b).

changes sign $\sin\phi \;\to\; -\sin\phi$*).* $\mathcal{M}_{s_1.s_2}$*, figure 11.4, 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 is: its itinerary is an infinitely repeating block of symbols. More generally, for hyperbolic flows the intersection of the future and past itineraries, the bi-infinite itinerary $S^-.S^+ = \cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots$ specifies a unique trajectory. This is intuitively clear for our 3-disk game of pinball, and is stated more formally in the definition (11.4) 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.

Example 11.3 Pruning rules for a 3-disk alphabet: *As the disks are convex, there can be no two consecutive reflections off the same disk, hence the covering symbolic dynamics consists of all sequences which include no symbol repetitions* $\_11\_$*,* $\_22\_$*,* $\_33\_$*. This is a finite set of finite length* pruning rules*, hence, the dynamics is a subshift of finite type (see (11.24) for definition), with the transition matrix (11.2) given by*

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} . \tag{11.5}$$

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 (5.23), the other transverse eigenvalue is contracting. This example demonstrates that finite Markov partitions can be constructed for hyperbolic *dynamical systems which are expanding in some directions, contracting in others. Further examples are the* 1*-dimensional expanding mapping sketched in figure 11.8, and more examples are worked out in sect. 23.2.*

Determining whether the symbolic dynamics is complete (as is the case for sufficiently separated disks), pruned (for example, for touching or over-

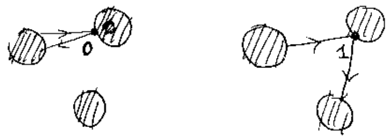

Figure 11.5: Binary labeling of the 3-disk pinball trajectories; a bounce in which the trajectory returns to the preceding disk is labeled 0, and a bounce which results in continuation to the third disk is labeled 1.

lapping disks), or only a first coarse graining of the topology (as, for example, for smooth potentials with islands of stability) requires case-by-case investigation, a discussion we postpone to sect. 11.4 and chapter 12. 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. 11.3, p. 166

11.2 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. In chapter 22 we shall be handsomely rewarded for our labors.

As the three disks are equidistantly spaced, our game of pinball has a sixfold symmetry. For instance, the cycles $\overline{12}$, $\overline{23}$, and $\overline{13}$ are related to each other by rotation by $\pm 2\pi/3$ or, equivalently, by a relabeling of the disks. Further examples of such symmetries are shown in figure 1.6. The disk labels are arbitrary; what is important is how a trajectory evolves as it hits subsequent disks, not what label the starting disk had. We exploit this symmetry by *recoding*, in this case replacing the absolute disk labels by relative symbols, indicating the type of the collision. For the 3-disk game of pinball there are two topologically distinct kinds of collisions, figure 11.5:

11.1 page 182
11.2 page 182

$$s_i = \begin{cases} 0 & : \quad \text{pinball returns to the disk it came from} \\ 1 & : \quad \text{pinball continues to the third disk}\,. \end{cases} \tag{11.6}$$

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,

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
7	0001001
	0000111
	0001011
	0001101
	0010011
	0010101
	0001111
	0010111
	0011011
	0011101
	0101011
	0011111
	0101111
	0110111
	0111111
8	00000001
	00000011
	00000101
	00001001
	00000111
	00001011
	00001101
	00010011
	00010101
	00011001
	00100101
8	00001111
	00010111
	00011011
	00011101
	00100111
	00101011
	00101101
	00110101
	00011111
	00101111
	00110111
	00111011
	00111101
	01010111
	01011011
	00111111
	01011111
	01101111
	01111111
9	000000001
	000000011
	000000101
	000001001
	000010001
	000000111
	000001011
9	000001101
	000010011
	000010101
	000011001
	000100011
	000100101
	000101001
	000001111
	000010111
	000011011
	000011101
	000100111
	000101011
	000101101
	000110011
	000110101
	000111001
	001001011
	001001101
	001010011
	001010101
	000011111
	000101111
	000110111
	000111011
	000111101
9	001001111
	001010111
	001011011
	001011101
	001100111
	001101011
	001101101
	001110101
	010101011
	000111111
	001011111
	001101111
	001110111
	001111011
	001111101
	010101111
	010110111
	010111011
	001111111
	010111111
	011011111
	011101111
	011111111

Table 11.1: Prime cycles for the binary symbolic dynamics up to length 9.

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 11.1.

11.3
page 182

Example 11.4 Recoding ternary symbolic dynamics in binary: *Given a ternary sequence and labels of 2 preceding disks, rule (11.6) fixes the subsequent binary symbols. Here we list an arbitrary ternanry itinerary, and the corresponding binary sequence:*

$$\begin{array}{lcl} \textit{ternary} & : & 3\ 1\ 2\ 1\ 3\ 1\ 2\ 3\ 2\ 1\ 2\ 3\ 1\ 3\ 2\ 3 \\ \textit{binary} & : & \cdot\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0 \end{array} \tag{11.7}$$

The first 2 disks initialize the trajectory and its direction; $3 \mapsto 1 \mapsto 2 \mapsto \cdots$. *Due to the 3-disk symmetry the six 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 11.2, figure 11.6 and figure 22.3.*

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 figure 11.6 (b). Every global 3-disk trajectory has a corresponding fundamental domain mirror trajectory obtained by replacing every crossing of a symmetry axis by a reflection. Depending on the symmetry of the global trajectory, a repeating binary symbols block corresponds either to the full periodic orbit or to an irreducible segment

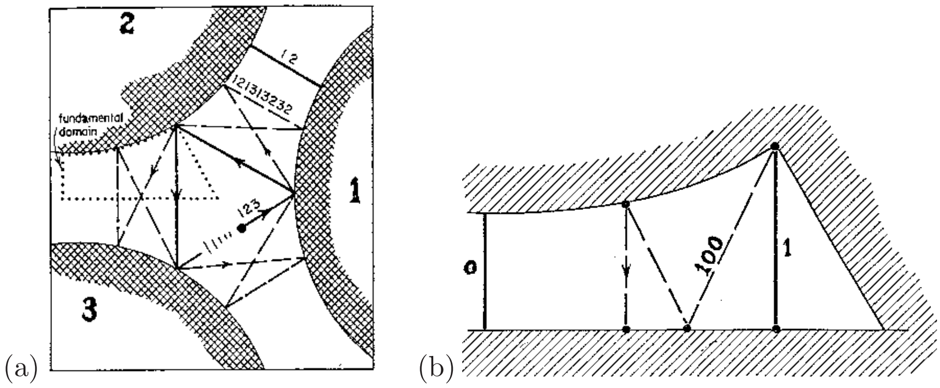

Figure 11.6: The 3-disk game of pinball with the disk radius : center separation ratio a:R = 1:2.5. (a) The three disks, with $\overline{12}$, $\overline{123}$ and $\overline{121232313}$ cycles indicated. (b) The fundamental domain, i.e., the small 1/6th wedge indicated in (a), consisting of a section of a disk, two segments of symmetry axes acting as straight mirror walls, and an escape gap. The above cycles restricted to the fundamental domain are now the two fixed points $\overline{0}$, $\overline{1}$, and the $\overline{100}$ cycle.

11.5
page 182

(examples are shown in figure 11.6 and table 11.2). An irreducible segment corresponds to a periodic orbit in the fundamental domain. Table 11.2 lists some of the shortest binary periodic orbits, together with the corresponding full 3-disk symbol sequences and orbit symmetries. For a number of reasons that will be elucidated in chapter 22, 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.

Inspecting the figure 11.3 we see that the relative ordering of regions with differing finite itineraries is a qualitative, topological property of the flow, so it makes sense to define a simple “canonical” representative partition which in a simple manner exhibits spatial ordering common to an entire class of topologically similar nonlinear flows.


in depth:
chapter 22, p. 387

11.3 Stretch and fold

Symbolic dynamics for N-disk game of pinball is so straightforward that one may altogether fail to see the connection between the topology of hyperbolic flows and their symbolic dynamics. This is brought out more clearly by the 1-dimensional visualization of “stretch & fold” flows to which we turn now.

Suppose concentrations of certain chemical reactants worry you, or the variations in the Chicago temperature, humidity, pressure and winds affect your mood. All such properties vary within some fixed range, and so do their rates of change. Even if we are studying an open system such as the 3-disk pinball game, we tend to be interested in a finite region around the