Chapter 11

Charting the state space

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

In this chapter and the next we learn to *partition* state space in a topologically invariant way, and *identify* topologically distinct orbits.

We start in sect. 11.1 with a simple and intuitive example, a 3-disk game of pinball. The qualitative dynamics of stretching/shrinking strips of surviving state space regions enables us to partition the state space and assign *symbolic dynamics* itineraries to trajectories. For the 3-disk game of pinball all possible symbol sequences enumerate all possible orbits.

In sect. 11.2 we use Rössler and Lorenz flows to motivate modeling of higher-dimensional flows by iteration of 1-dimensional maps. For these two flows the 1-dimensional maps capture essentially all of the higher-dimensional flow dynamics, both qualitatively and quantitatively. 1-dimensional maps suffice to explain the two key aspects of qualitative dynamics; *temporal ordering*, or *itinerary* with which a trajectory visits state space regions (sect. 11.3), and the *spatial ordering* between trajectory points (sect. 11.4), which is the key to determining the admissibility of an orbit with a prescribed itinerary. In a generic dynamical system not every symbol sequence is realized as a dynamical trajectory; as one looks further and further, one discovers more and more 'pruning' rules which prohibit families of itineraries. For 1-dimensional 'stretch & fold' maps the *kneading theory* (sect. 11.5) provides the definitive answer as to which temporal itineraries are *admissible* as trajectories of the dynamical system. Finally, sect. 11.6 is meant serve as a guide to the basic concepts of symbolic dynamics.

Deceptively simple, this subject can get very difficult very quickly, so in this chapter we do the first, 1-dimensional 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

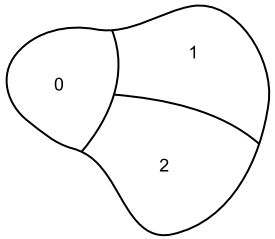


Figure 11.1: A coarse partition of $\mathcal{M}$ into regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$, labeled by ternary alphabet $\mathcal{A} = \{1, 2, 3\}$.

Rydberg atoms or mesoscopic devices, and resent wasting time on formal things, this chapter and chapters 14 and 15 are good for you. Study them.

11.1 Qualitative dynamics

(R. Mainieri and P. Cvitanović)

What can a flow do to points in state space? 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. The first answer was inspired by the motion of the planets: they appear to repeat their motion through the firmament, so the ancients' 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 state space, the important first step on the way to a theory of dynamical systems: qualitative, topological dynamics, or *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 and references cited in remark 11.1 whenever you run into baffling jargon.

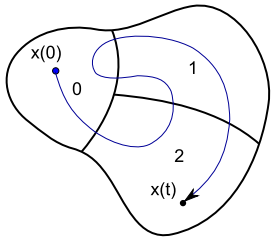


Figure 11.2: A trajectory with itinerary 021012.

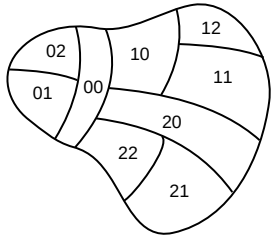


Figure 11.3: A 1-step memory refinement of the partition of figure 11.1, with each region $\mathcal{M}_i$ subdivided into $\mathcal{M}_{i0}$, $\mathcal{M}_{i1}$, and $\mathcal{M}_{i2}$, labeled by nine 'words' $\{00, 01, 02, \cdots, 21, 22\}$.

We start by dividing the state space up into regions $\mathcal{M}_A, \mathcal{M}_B, \ldots, \mathcal{M}_Z$, as in figure 11.1. This can be done in many ways, not all equally clever. Any such division of state space into distinct regions constitutes 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 state evolves, 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 11.2, the state space is divided into three regions $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$, the 'letters' are the integers $\{0, 1, 2\}$, and the itinerary for the trajectory sketched in the figure is $0 \mapsto 2 \mapsto 1 \mapsto 0 \mapsto 1 \mapsto 2 \mapsto \cdots$.

Example 11.1 ***3-disk symbolic dynamics:*** *Consider the motion of a free point particle in a plane with* 3 *elastically reflecting convex disks, figure 11.4. After a collision with a disk a particle either continues to another disk or escapes, so a trajectory can be labeled by the disk sequence. Sets of configuration space pinball trajectories of figure 11.4 become quickly hard to disentangle. As we shall see in what follows, their state space visualization in terms of Poincaré sections $\mathcal{P} = [s, p]$ (figure 11.5, see also figure 3.9 (b)) is much more powerful. (continued in example 11.2)* example 11.1

In general only a subset of points in $\mathcal{M}_B$ reaches $\mathcal{M}_A$. This observation offers a systematic way to refine a partition by introducing *m-step memory*: the region $\mathcal{M}_{s_m \cdots s_1 s_0}$ consists of the subset of points of $\mathcal{M}_{s_0}$ whose trajectory for the next m time steps will be $s_0 \mapsto s_1 \mapsto \cdots \mapsto s_m$, see figure 11.3.

Example 11.2 ***3-disk state space partition:*** *(continued from example 11.1) Embedded within $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ are four strips $\mathcal{M}_{121}$, $\mathcal{M}_{123}$, $\mathcal{M}_{131}$, $\mathcal{M}_{132}$ of initial conditions that survive two bounces, and so forth. At each bounce a cone of initially nearby trajectories disperses (see figures 1.8 and 11.4). Also in order to attain a desired longer and*

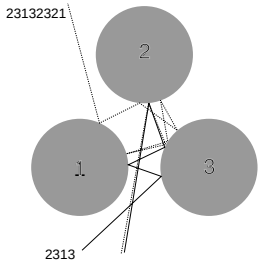


Figure 11.4: Two pinballs that start out very close to each other exhibit the same qualitative dynamics _2313_ for the first three bounces, but due to the exponentially growing separation of trajectories with time, follow different itineraries thereafter: one escapes after _2313_, the other one escapes after _23132321_. (Notation _2313_ is explained in sect. 11.6.)

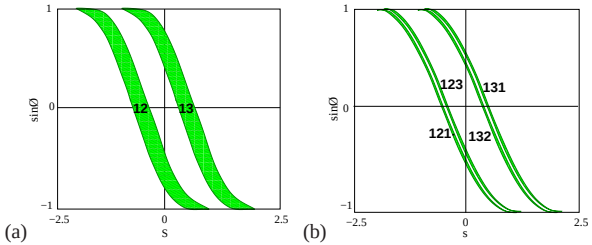


Figure 11.5: The 3-disk game of pinball Poincaré section, trajectories emanating from the disk 1 with x = (arclength, parallel momentum) = (s, p), where $p = \sin\theta$. (a) Strips of initial points $\mathcal{M}_{12}$, $\mathcal{M}_{13}$ which reach disks 2, 3 in one bounce, respectively. (b) 1-step memory refinement of partition (see figure 11.3): 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. Disk radius : center separation ratio a:R = 1:2.5. (Y. Lan)

longer itinerary of bounces, the strip of initial points $x_0 = (s_0, p_0)$ requires exponentially finer precision, nested within the initial state space strips drawn in figure 11.5. Provided that the disks are sufficiently separated, after n bounces the survivors are labeled by 2^n distinct itineraries $s_1 s_2 s_3 \ldots s_n$. (continued in example 12.2)

If there is no way to reach partition $\mathcal{M}_i$ from partition $\mathcal{M}_j$, and conversely, partition $\mathcal{M}_j$ from partition $\mathcal{M}_i$, the state space consists of at least two disconnected pieces, and we can analyze it piece by piece. An interesting partition should be dynamically connected, i.e., one should be able to go from any region $\mathcal{M}_i$ to any other region $\mathcal{M}_j$ in a finite number of steps. A dynamical system with such a partition is said to be *metrically indecomposable*.

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

However, knowing that a point from $\mathcal{M}_i$ reaches $\{\mathcal{M}_j, \cdots, \mathcal{M}_k\}$ in one step is not quite good enough. We would be happier if we knew that the map of the entire initial region $f(\mathcal{M}_i)$ overlaps nicely with the entire $\mathcal{M}_j$; otherwise we have to subpartition $\mathcal{M}_j$ into the subset $f(\mathcal{M}_i)$ and the reminder, and often we will find ourselves partitioning *ad infinitum*, a difficult topic that we shall return to sect. 12.4.

Such considerations motivate the notion of a *Markov partition*, a partition for

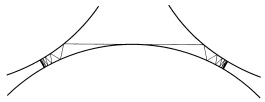


Figure 11.6: For the 3-disk game of pinball no itineraries are pruned as long as the inter-disk spacing exceeds $R : a > 2.04821419\ldots$. (from K.T. Hansen [12.22])

which no memory of preceding steps is required to fix the transitions allowed in the next step. *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,

$$\text{either} \quad \mathcal{M}_j \cap f(\mathcal{M}_i) = \emptyset \quad \text{or} \quad \mathcal{M}_j \subset f(\mathcal{M}_i)\,. \tag{11.2}$$

Whether such partitions can be found is not clear at all - the borders need to be lower-dimensional sets invariant under dynamics, and there is no guarantee that these are topologically simple objects. However, the game of pinball (and many other non-wandering repeller sets) is especially nice: the issue of determining the partition borders does not arise, as the survivors live on disconnected pieces of the state space, separated by a chasm of escaping trajectories.

The itinerary of a billiard trajectory is finite for a scattering trajectory, coming in from infinity and escaping after a finite number of collisions, infinite for a trapped trajectory, and infinitely repeating for a periodic orbit. A finite length trajectory is not uniquely specified by its finite itinerary, but an isolated unstable cycle is: its itinerary is an infinitely repeating block of symbols. 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 orbit. Almost all infinite length trajectories (orbits) are aperiodic. Still, the longer the trajectory is, the closer to it is a periodic orbit whose itinerary shadows the trajectory for its whole length: think of the state space as the unit interval, aperiodic orbits as normal numbers, and periodic ones as fractions whose denominators correspond to cycle periods, as is literally the case for the Farey map (24.42), to be discussed in sect. 24.3.4.

Determining whether the symbolic dynamics is complete (as is the case for sufficiently separated disks, see figure 11.6), pruned (for example, for touching or overlapping disks), or only a first coarse-graining of the topology (as, for example, for smooth potentials with islands of stability) requires a case-by-case investigation, a discussion we postpone until sect. 11.5 and chapter 12. For now, we assume that the disks are sufficiently separated that there is no additional pruning beyond the prohibition of self-bounces.

Inspecting figure 11.5 we see that the relative ordering of regions with differing finite itineraries is a qualitative, topological property of the flow. This observation motivates searches for simple, 'canonical' partitions which exhibit in a simple manner the spatial ordering common to entire classes of topologically similar nonlinear flows.

11.2 From d-dimensional flows to 1-dimensional maps

Symbolic dynamics for the 3-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.

We construct here the return maps (3.4) for two iconic flows, the Rössler and the Lorenz, in order to show how ODEs in higher dimensions can be modeled by low-dimensional maps. In the examples at hand the strong dissipation happens to render the dynamics essentially 1-dimensional, both qualitatively and quantitatively. However, as we shall show in chapter 12, strong dissipation is not essential -the hyperbolicity is- so the method applies to Hamiltonian (symplectic areas preserving) flows as well. The key idea is to replace the original, arbitrarily concocted coordinates by intrinsic, dynamically invariant curvilinear coordinates erected on neighborhoods of unstable manifolds.

fast track:
sect. 11.3, p. 227

Suppose concentrations of certain chemical reactants worry you, or the variations in the Chicago temperature, humidity, pressure and winds affect your mood. Such quantities vary within some fixed range, and so do their rates of change. Even if we are studying an open system such as the 3-disk pinball game, we tend to be interested in a finite region around the disks and ignore the escapees. So a typical dynamical system that we care about is *bounded*. If the price to keep going is high - for example, we try to stir up some tar, and observe it come to a dead stop the moment we cease our labors - the dynamics tends to settle into a simple state. However, as the resistance to change decreases - the tar is heated up and we are more vigorous in our stirring - the dynamics becomes unstable.

Example 11.3 ***Rössler attractor return map: Stretch & fold.*** *(continued from example 4.6) In the Rössler flow (2.17) of example 3.3 we sketched the attractor by running a long chaotic trajectory, and noted that the attractor of figure 3.2 is very thin. For Rössler flow an interval transverse to the attractor is stretched, folded and fiercely pressed back. The attractor is 'fractal,' but for all practical purposes the return map is 1-dimensional; your printer will need a resolution better than* 10^{13} *dots per inch to start resolving its structure. We had attempted to describe this 'stretch & fold' flow by a 1-dimensional return map, but the maps that we plotted in figure 3.3 were disquieting; they did not appear to be a 1-to-1 maps. This apparent non-invertibility is an artifact of projection of a* 2*-dimensional return map* $(R_n, z_n) \to (R_{n+1}, z_{n+1})$ *onto the 1-dimensional subspace* $R_n \to R_{n+1}$*. Now that we understand equilibria and their linear stability, let's do this right.*

The key idea is to measure arclength distances along the unstable manifold of the x_{-} equilibrium point, as in figure 11.7 (a). Luck is with us; figure 11.7 (b) return map $s_{n+1} = P(s_n)$ looks much like a parabola of example 3.8, so we shall take the unimodal map symbolic dynamics, sect. 11.3, as our guess for the covering symbolic dynamics.
(continued in example 11.11)

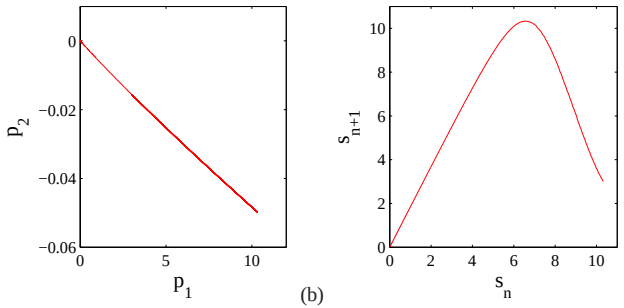


Figure 11.7: (a) $x = 0$, $y > 0$ Poincaré section of the x_- unstable manifold, Rössler flow figure 2.6. (p_1, p_2) are measured with the origin placed at x_-. (b) $s \to P(s)$ return map, where s is the arc-length distance measured along the unstable manifold of equilibrium point x_-. (A. Basu and J. Newman)

You get the idea - Rössler flow winds around the stable manifold of the 'central' equilibrium, stretches and folds, and the dynamics on the Poincaré section of the flow can be reduced to a 1-dimensional map. The next example is similar, but the folding mechanism is very different: the unstable manifold of one of the equilibria collides with the stable manifold of the other one, forcing a robust *heteroclinic connection* between the two.

fast track:
sect. 11.3, p. 227

11.2.1 Heteroclinic connections

In general, two manifolds can intersect in a stable way if the sum of their dimensions is greater than or equal to the dimension of the state space, hence an unstable manifold of dimension k is likely to intersect a stable manifold whose codimension in state space is less than or equal to k (i.e., robustly with respect to small changes of system parameters). Trajectories that leave a fixed point along its unstable manifold and reach another fixed point along its stable manifold are called *heteroclinic* if the two fixed points are distinct or *homoclinic* if the initial and the final point are the same point. Whether the two manifolds actually intersect is a subtle question that is central to the issue of "structural stability" of ergodic dynamical systems. remark 11.3

Example 11.4 ***Lorenz flow: Stretch & crease.*** *We now deploy the symmetry of Lorenz flow to streamline and complete analysis of the Lorenz strange attractor commenced in example 9.14. There we showed that the dihedral $D_1 = \{e, R\}$ symmetry identifies the two equilibria EQ_1 and EQ_2, and the traditional 'two-eared' Lorenz flow figure 2.5 is replaced by the 'single-eared' flow of figure 9.6 (a). Furthermore, symmetry identifies two sides of any plane through the z axis, replacing a full-space Poincaré section plane by a half-plane, and the two directions of a full-space eigenvector of EQ_0 by a one-sided eigenvector, see figure 9.6 (a).*

Example 4.8 explained the genesis of the x_{EQ_1} equilibrium unstable manifold, its orientation and thickness, its collision with the z-axis, and its heteroclinic connection to the $x_{EQ_0} = (0, 0, 0)$ equilibrium. All that remains is to describe how the EQ_0 neighborhood connects back to the EQ_1 unstable manifold.

Figure 9.6 and figure 11.8 (a) show clearly how the Lorenz dynamics is pieced together from the 2 equilibria and their unstable manifolds: Having completed the descent to EQ_0, the infinitesimal neighborhood of the heteroclinic $EQ_1 \to EQ_0$ trajectory is ejected along the unstable manifold of EQ_0 and is re-injected into the unstable manifold of EQ_1. Both sides of the narrow strip enclosing the EQ_0 unstable manifold lie above it, and they get folded onto each other with a knife-edge crease (contracted exponentially for infinite time to the EQ_0 heteroclinic point), with the heteroclinic outtrajectory defining the outer edge of the strange attractor. This leads to the folding of the outer branch of the Lorenz strange attractor, illustrated in figure 11.8 (b), with the outermost edge following the unstable manifold of EQ_0.

Now the stage is set for construction of Poincaré sections and associated Poincaré return maps. There are two natural choices; the section at EQ_0, lower part of figure 11.8 (b), and the section (blue) above EQ_1. The first section, together with the blowup of the EQ_0 neighborhood, figure 4.7 (b), illustrates clearly the scarcity of trajectories (vanishing natural measure) in the neighborhood of EQ_0. The flat section above EQ_1 (which is, believe it or not, a smooth conjugacy by the flow of the knife-sharp section at EQ_0) is more convenient for our purposes. Its return map (3.4) is given by figure 11.9.

The rest is straight sailing: to accuracy 10^{-4} the return map is unimodal, its critical point's forward trajectory yields the kneading sequence (11.13), and the admissible binary sequences, so any number of periodic points can be accurately determined from

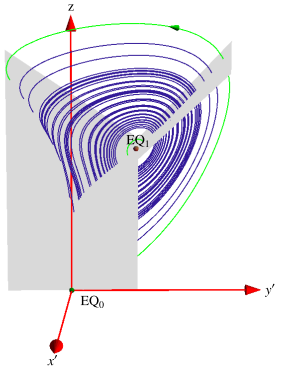
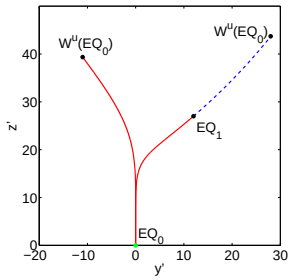


Figure 11.8: (a) A Poincaré section of the Lorenz flow in the doubled-polar angle representation, figure 9.6, given by the $[y', z]$ plane that contains the z-axis and the equilibrium EQ_1. x' axis points toward the viewer. (b) The Poincaré section of the Lorenz flow by the section Crossings *into* the section are marked red (solid) and crossings *out* of the section are marked blue (dashed). Outermost points of both in- and out-sections are given by the EQ_0 unstable manifold $W^u(EQ_0)$ intersection. (E. Siminos)



(b)

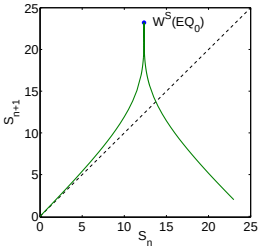


Figure 11.9: The Poincaré return map $s_{n+1} = P(s_n)$ parameterized by Euclidean arclength s measured along the EQ_1 unstable manifold, from x_{EQ_1} to $W^u(EQ_0)$ section point, uppermost right point of the blue (dashed) segment in figure 11.8 (b). The critical point (the 'crease') of the map is given by the section of the heteroclinic orbit $W^s(EQ_0)$ that descends all the way to EQ_0, in infinite time and with infinite slope. (E. Siminos)

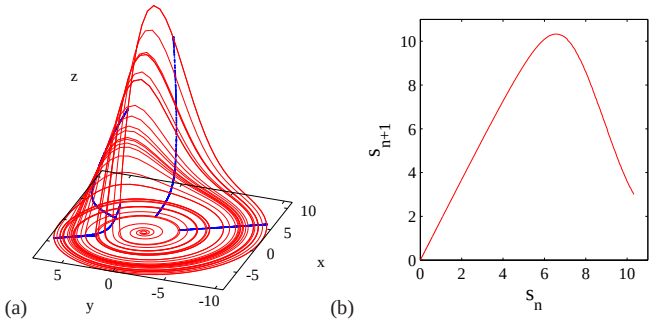


Figure 11.10: (a) The Rössler flow, figure 3.2, is an example of a recurrent flow that stretches and folds. (b) The Rössler 'stretch & fold' return map, figure 11.7 (b). (R. Paškauskas and A. Basu)

this 1-dimensional return map, and the 3-dimensional cycles then verified by integrating the Lorenz differential equations (2.12). As already observed by Lorenz, such a map is everywhere expanding on the strange attractor, so it is no wonder mathematicians can here make the ergodicity rigorous. section 20.6

(E. Siminos and J. Halcrow)

What have we learned from the above two exemplary 3-dimensional flows? If a flow is locally unstable but globally bounded, any open ball of initial points will be stretched out and then folded back. If the equilibria are hyperbolic, the trajectories will be attracted along some eigen-directions and ejected along others. The unstable manifold of one equilibrium can avoid stable manifolds of other equilibria, as is the case for Rössler, or slice them head on, as is the case for Lorenz. A typical trajectory wanders through state space, alternatively attracted into equilibria neighborhoods, and then ejected again. What is important is the motion along the unstable manifolds – that is where 1*d* maps come from.

At this juncture we proceed to show how this works on the simplest example: unimodal mappings of the interval. The erudite reader may skim through this chapter and then take a more demanding path, via the Smale horseshoes of chapter 12. Unimodal maps are easier, but less physically compelling. Smale horseshoes offer the high road, more complicated, but the right tool to generalize what we learned from the 3-disk dynamics, and begin analysis of general dynamical systems. It is up to you - unimodal maps suffice to get quickly to the heart of this treatise.

11.3 Temporal ordering: Itineraries

In this section we learn to *name* topologically distinct trajectories for the simple, but instructive case; 1-dimensional maps of an interval.

The simplest mapping of this type is *unimodal;* interval is stretched and folded only once, with at most two points mapping into a point in the refolded inter-

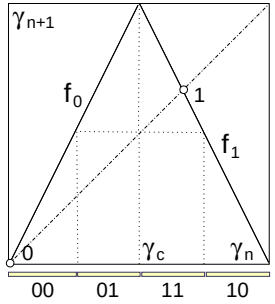


Figure 11.11: The full tent map (11.4) partition $\{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{11}, \mathcal{M}_{10}\}$ together with the fixed points x_0, x_1.

val, as in the Rössler return map figure 11.10 (b). A unimodal map $f(x)$ is a 1-dimensional function $\mathbb{R} \to \mathbb{R}$ defined on an interval $\mathcal{M} \in \mathbb{R}$ with a monotonically increasing (or decreasing) branch, a *critical point* (or interval) x_c for which $f(x_c)$ attains the maximum (minimum) value, followed by a monotonically decreasing (increasing) branch. *Uni*-modal means that the map is a 1-humped map with one critical point within interval $\mathcal{M}$. *Multi*-modal maps, with several critical points within interval $\mathcal{M}$, can be described with a straight-forward generalization of the methods we describe next.

Example 11.5 ***Unimodal maps:*** *(continued from example 3.8) The simplest examples of unimodal maps are the quadratic map*

$$f(x) = Ax(1-x), \qquad x \in \mathcal{M} = [0,1] \tag{11.3}$$

and numerically computed return maps such as figure 11.10 (b). Such dynamical systems are irreversible (the inverse of f is double-valued), but, as we shall show in sect. 12.2, they may nevertheless serve as effective descriptions of invertible 2-dimensional hyperbolic flows. For the unimodal map such as figure 11.12 a Markov partition of the unit interval $\mathcal{M}$ is given by the two intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$. (continued in example 11.6)

Example 11.6 ***Full tent map, Ulam map:*** *(continued from example 11.5) The simplest examples of unimodal maps with complete binary symbolic dynamics are the* full tent map, *figure 11.11,*

$$f(\gamma) = 1 - 2|\gamma - 1/2|, \qquad \gamma \in \mathcal{M} = [0,1], \tag{11.4}$$

the Ulam map (quadratic map (11.3) with $A = 4$) exercise 6.4

$$f(x) = 4x(1-x), \qquad x \in \mathcal{M} = [0,1], \tag{11.5}$$

and the repelling unimodal maps such as figure 11.12. For unimodal maps the Markov partition of the unit interval $\mathcal{M}$ is given by intervals $\{\mathcal{M}_0, \mathcal{M}_1\}$. We refer to (11.4) as the complete *tent map because its symbolic dynamics is completely binary: as both $f(\mathcal{M}_0)$ and $f(\mathcal{M}_1)$ fully cover $\mathcal{M} = \{\mathcal{M}_0, \mathcal{M}_1\}$, all binary sequences are realized as admissible itineraries.*

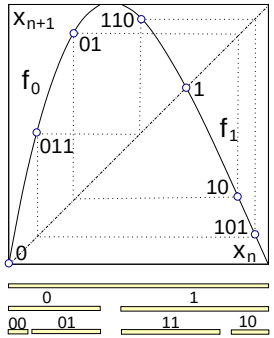


Figure 11.12: A unimodal repeller with the survivor intervals after 1 and 2 iterations. Intervals marked $s_1 s_2 \cdots s_n$ consist of points that do not escape in n iterations, and follow the itinerary $S^+ = s_1 s_2 \cdots s_n$. Note that the spatial ordering does not respect the binary ordering; for example $x_{00} < x_{01} < x_{11} < x_{10}$. Also indicated are the fixed points $\overline{0}$, $\overline{1}$, the 2-cycle $\overline{01}$, and the 3-cycle $\overline{011}$.

For 1d maps the *critical value* denotes either the maximum or the minimum value of $f(x)$ on the defining interval; we assume here that it is a maximum, $f(x_c) \geq f(x)$ for all $x \in \mathcal{M}$. The critical point x_c that yields the critical value $f(x_c)$ belongs to neither the left nor the right partition $\mathcal{M}_i$ and is instead denoted by its own symbol $s = C$. As we shall see, its images and preimages serve as partition boundary points.

The trajectory $x_1, x_2, x_3, \ldots$ of the initial point x_0 is given by the iteration $x_{n+1} = f(x_n)$. Iterating f and checking whether the point lands to the left or to the right of x_c generates a *temporally* ordered topological itinerary (11.17) for a given trajectory,

$$s_n = \begin{cases} 1 & \text{if } x_n > x_c \\ C & \text{if } x_n = x_c \\ 0 & \text{if } x_n < x_c \end{cases} . \tag{11.6}$$

We refer to $S^+(x_0) = .s_1 s_2 s_3 \cdots$ as the *future itinerary*. Our next task is to answer the reverse problem: given an itinerary, what is the *spatial* ordering of points that belong to the corresponding state space trajectory?

11.4 Spatial ordering

> A well-known theorem states that combinatorial factors are impossible to explain. [11.22]
>
> —G. 't Hooft and M. Veltman, DIAGRAMMAR

Suppose you have succeeded in constructing a covering symbolic dynamics, such as the one we constructed for a well-separated 3-disk system. Now start moving the disks toward each other. At some critical separation (see figure 11.6) a disk will start blocking families of trajectories traversing the other two disks. The order in which trajectories disappear is determined by their relative ordering in

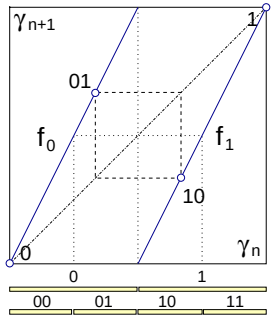


Figure 11.13: The $n = 2$, 4-intervals state space partition for the Bernoulli shift map (11.7), together with the fixed points $\overline{0}$, $\overline{1}$ and the 2-cycle $\overline{01}$.

space; the ones closest to the intervening disk will be pruned first. Determining inadmissible itineraries requires that we relate the spatial ordering of trajectories to their time ordered itineraries. exercise 12.7

The easiest point of departure is to start by working out this relation for the symbolic dynamics of 1-dimensional mappings. As it appears impossible to present this material without getting bogged down in a sea of 0's, 1's and subscripted subscripts, we announce the main result before embarking upon its derivation: section 11.5

The admissibility criterion (sect. 11.5) eliminates *all* itineraries that cannot occur for a given unimodal map.

Example 11.7 ***Bernoulli shift map state space partition.*** *First, an easy example: the* Bernoulli shift map*, figure 11.13,*

$$b(\gamma) = \begin{cases} b_0(\gamma) = 2\gamma\,, & \gamma \in \mathcal{M}_0 = [0, 1/2) \\ b_1(\gamma) = 2\gamma - 1\,, & \gamma \in \mathcal{M}_1 = (1/2, 1] \end{cases} , \tag{11.7}$$

models the 50-50% probability of a coin toss. It maps the unit interval onto itself, with fixed points $\gamma_0 = 0$, $\gamma_1 = 1$. The closely related doubling map *acts on the circle*

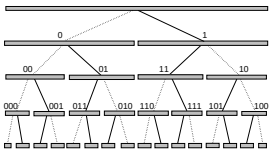
$$x \mapsto 2x \pmod 1\,, \qquad x \in [0, 1] \tag{11.8}$$

and consequently has only one fixed point, $x_0 = 0 = 1 \pmod 1$. The Bernoulli map is called a 'shift' map, as a multiplication by 2 acts on the binary representation of $\gamma = .s_1 s_2 s_3 \ldots$ by shifting its digits, $b(\gamma) = .s_2 s_3 \ldots$. The nth preimages $b^{-n}(\gamma)$ of the critical point $\gamma_c = 1/2$ partition the state space into 2^n subintervals, each labeled by the first n binary digits of points $\gamma = .s_1 s_2 s_3 \ldots$ within the subinterval: figure 11.13 illustrates such 4-intervals state space partition $\{\mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{11}, \mathcal{M}_{10}\}$ for $n = 2$.

Consider a map $f(x)$ topologically conjugate (two monotonically increasing branches) to the Bernoulli shift, with the forward orbit of x generating the itinerary $s_1 s_2 s_3 \ldots$. Convert this itinerary into Bernoulli map point $\gamma = .s_1 s_2 s_3 \ldots$. These values can now be used to spatially order *points with different temporal itineraries: if $\gamma < \gamma'$, then $x < x'$.*

Suppose we have already computed all $(n-1)$-cycles of $f(x)$, and would now like to compute the cycle $p = s_1 s_2 s_3 \ldots s_n$ of period n. Mark γ values on the unit interval

Figure 11.14: An alternating binary tree relates the itinerary labeling of the unimodal map intervals, figure 11.12, to their spatial ordering. The dotted line stands for 0, the full line for 1; the binary sub-tree whose root is a full line with symbol 1 reverses the orientation, due to the orientation-reversing fold in figures 11.10 and 11.12. See also figure 14.4.



for all known periodic points of the Bernoulli shift map, and then insert in between them $\gamma_{\sigma^k p}, k = 0, 1, \cdots, n_p - 1$ corresponding to periodic points of cycle p. In the dynamical state space they will be bracketed by corresponding cycle points x_j from cycles already computed, and thus the knowledge of the topological ordering of all *cycle points provides us with robust initial guesses for periodic-orbit searches for any map with 2 monotonically increasing branches. (continued in example 23.5)*

For the Bernoulli shift converting itineraries into a topological ordering is easy; the binary expansion of coordinate γ is also its temporary itinerary. The tent map (11.4), figure 11.11 is a bit harder. It consists of two straight segments joined at $x = 1/2$. The symbol s_n defined in (11.6) equals 0 if the function increases, and 1 if it decreases. Iteration forward in time generates the time itinerary. More importantly, the piecewise linearity of the map makes the converse possible: determine analytically an initial point given its itinerary, a property that we now use to define a topological coordinatization common to all unimodal maps.

Here we have to face the fundamental problem of pedagogy: combinatorics cannot be taught. The best one can do is to state the answer and hope that you will figure it out by yourself. Then you can also complain that the way the rule is stated here is incomprehensible.

The tent map point $\gamma(S^+)$ with future itinerary S^+ is given by converting the itinerary of s_n's into a binary number γ by the following algorithm:

$$\begin{aligned} w_{n+1} &= \begin{cases} w_n & \text{if} \quad s_n = 0 \\ 1 - w_n & \text{if} \quad s_{n+1} = 1 \end{cases}, \qquad w_1 = s_1 \\ \gamma(S^+) &= 0.w_1 w_2 w_3 \ldots = \sum_{n=1}^{\infty} w_n / 2^n. \end{aligned} \tag{11.9}$$

This follows by inspection from the binary tree of figure 11.14. Once you figure this out, feel free to complain that the way the rule is stated here is incomprehensible, and show us how you did it better. exercise 11.4

Example 11.8 ***Converting γ to S^+:*** *γ whose itinerary is $S^+ = 0110000\cdots$ is given by the binary number $\gamma = .010000\cdots$. Conversely, the itinerary of $\gamma = .01$ is $s_1 = 0$, $f(\gamma) = .1 \to s_2 = 1$, $f^2(\gamma) = f(.1) = 1 \to s_3 = 1$, etc.. Orbit that starts out as a finite block followed by infinite repeats of another block $p = S_p = (s_1 s_2 s_3 \ldots s_n)^\infty$ is said to be* heteroclinic *to the cycle p. An orbit that starts out as p^∞ followed by a finite block followed by infinite repeats of another block p' is said to be* heteroclinic connection *from cycle p to cycle p'.*

We refer to $\gamma(S^+)$ as the *(future) topological coordinate*. The w_t's are the digits in the binary expansion of the starting point γ for the full tent map (11.4). In the left half-interval the map $f(x)$ acts by multiplication by 2, while in the right half-interval the map acts as a flip as well as multiplication by 2, reversing the ordering, and generating in the process the sequence of s_n's from the binary digits w_n.

The mapping $x_0 \to S^+(x_0) \to \gamma_0 = \gamma(S^+)$ is a *topological conjugacy* that maps the trajectory of an initial point x_0 under the iteration of a given unimodal map to that initial point γ_0 for which the trajectory of the 'canonical' unimodal map, the full tent map (11.4), has the same itinerary. The virtue of this conjugacy is that $\gamma(S^+)$ *preserves the ordering* for any unimodal map in the sense that if $\overline{x} > x$, then $\overline{\gamma} > \gamma$.

Example 11.9 ***Periodic orbits of unimodal maps.*** *Let*

$$f(x) = \begin{cases} f_0(x) \text{ if} & x < x_c \\ f_1(x) \text{ if} & x > x_c \end{cases}, \tag{11.10}$$

and assume that all periodic orbits are unstable, i.e., the stability $\Lambda_p = f_a^{k\prime}$ (see (4.46)) satisfies $|\Lambda_p| > 1$. Then the periodic point $x_{s_1 s_2 s_3 \ldots s_n}$ is the only fixed point of the unique composition (3.15) of n maps

$$f_{s_n} \circ \cdots \circ f_{s_2} \circ f_{s_1}(x_{s_1 s_2 s_3 \ldots s_n}) = x_{s_1 s_2 s_3 \ldots s_n} \tag{11.11}$$

(note that successive maps, applied from the left, correspond to later times, i.e., later symbols in the itinerary).

The nth iterate of a unimodal map has at most 2^n monotone segments, and therefore there will be 2^n or fewer periodic points of length n. For the full tent map (11.4) it has exactly 2^n periodic points. A periodic orbit p of length n corresponds to an section 12.2 *infinite repetition of a length $n = n_p$ symbol string block, customarily indicated by a line over the string: $p = S_p = (s_1 s_2 s_3 \ldots s_n)^\infty = \overline{s_1 s_2 s_3 \ldots s_n}$. As all itineraries are infinite, we shall adopt convention that a finite string itinerary $p = s_1 s_2 s_3 \ldots s_n$ stands for infinite repetition of a finite block, and routinely omit the overline. A cycle p is called* prime *if its itinerary S cannot be written as a repetition of a shorter block S'. If the itinerary of x_0 is $p = s_1 s_2 s_3 \ldots s_n$, its cyclic permutation $\sigma^k p = \overline{s_k s_{k+1} \ldots s_n s_1 \ldots s_{k-1}}$ corresponds to the point x_{k-1} in the same cycle.*

Example 11.10 ***Periodic points of the full tent map.*** *Each cycle p is a set of n_p rational-valued full tent map periodic points γ. It follows from (11.9) that if the repeating string $s_1 s_2 \ldots s_n$ contains an odd number of '1's, the string of well ordered symbols $w_1 w_2 \ldots w_{2n}$ has to be of the double length before it repeats itself. The cycle-point γ is a geometrical sum which we can rewrite as the odd-denominator fraction*

$$\begin{aligned} \gamma(\overline{s_1 s_2 \ldots s_n}) &= \sum_{t=1}^{2n} \frac{w_t}{2^t} + \frac{1}{2^{-2n}} \sum_{t=1}^{2n} \frac{w_t}{2^t} + \cdots \\ &= \frac{2^{2n}}{2^{2n} - 1} \sum_{t=1}^{2n} \frac{w_t}{2^t} \end{aligned} \tag{11.12}$$

Using this we can calculate the $\hat{\gamma}_p = \hat{\gamma}(S_p)$ for all short cycles. For orbits up to length 5 this is done in table 11.1.

S	$\hat{\gamma}(S)$			S	$\hat{\gamma}(S)$		
$\overline{0}$	$.\overline{0}$	=	0	$\overline{10111}$	$.\overline{11010}$	=	26/31
$\overline{1}$	$.\overline{10}$	=	2/3	$\overline{10110}$	$.\overline{1101100100}$	=	28/33
$\overline{10}$	$.\overline{1100}$	=	4/5	$\overline{10010}$	$.\overline{11100}$	=	28/31
$\overline{101}$	$.\overline{110}$	=	6/7	$\overline{10011}$	$.\overline{1110100010}$	=	10/11
$\overline{100}$	$.\overline{111000}$	=	8/9	$\overline{10001}$	$.\overline{11110}$	=	30/31
$\overline{1011}$	$.\overline{11010010}$	=	14/17	$\overline{10000}$	$.\overline{1111100000}$	=	32/33
$\overline{1001}$	$.\overline{1110}$	=	14/15				
$\overline{1000}$	$.\overline{11110000}$	=	16/17				

Table 11.1: The maximal values of unimodal map cycles up to length 5. (K.T. Hansen)

Critical points are special - they define the boundary between intervals, i.e., interval is split into 0 [left part], x_c [critical point] and 1 [right part]. For the dike map and the repeller figure 11.12 x_c is the whole interval of points along the flat top of the map, but usually it is a point. As illustrated by figures 11.11 and 11.13, for a unimodal map the preimages $f^{-n}(x_c)$ of the critical point x_c serve as partition boundary points. But not all preimages–one has to ensure that they are within the set of all admissible orbits by checking them against the kneading sequence of the map.

11.5 Kneading theory

> Tired of being harassed by your professors? Finish, get a job, do combinatorics your own way, while you still know everything.
>
> —Professor Gatto Nero

(K.T. Hansen and P. Cvitanović)

The reason we need to be mindful of spatial ordering of temporal itineraries is that this spatial ordering provides us with criteria that separate inadmissible orbits from those realizable by the dynamics. For 1-dimensional mappings the *kneading theory* provides a precise and definitive criterion of admissibility.

If the parameter in the quadratic map (11.3) is $A > 4$, then the iterates of the critical point x_c diverge for $n \to \infty$, and any sequence S^+ composed of letters $s_i = \{0, 1\}$ is admissible, and any value of $0 \le \gamma < 1$ corresponds to an admissible orbit in the non–wandering set of the map. The corresponding repeller is a complete binary labeled Cantor set, the $n \to \infty$ limit of the nth level covering intervals sketched in figure 11.12.

For $A < 4$ only a subset of the points in the interval $\gamma \in [0, 1]$ corresponds to admissible orbits. The forbidden symbolic values are determined by observing that the largest x_n value in an orbit $x_1 \to x_2 \to x_3 \to \ldots$ has to be smaller than or equal to the image of the critical point, *the critical value* $f(x_c)$. Let $K = S^+(x_c)$

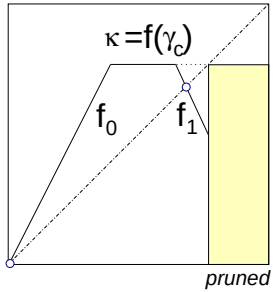


Figure 11.15: The 'dike' map obtained by slicing of the top portion of the tent map in figure 11.11. Any orbit that visits the primary pruning interval $(\kappa, 1]$ is inadmissible. The admissible orbits form the Cantor set obtained by removing from the unit interval the primary pruning interval and all its iterates. Any admissible orbit has the same topological coordinate and itinerary as the corresponding tent map figure 11.11 orbit.

be the itinerary of the critical point x_c, denoted the *kneading sequence* of the map. The corresponding topological coordinate is called the *kneading value*

$$\kappa = \gamma(K) = \gamma(S^+(x_c)). \tag{11.13}$$

The 'canonical' map that has the same kneading sequence K (11.13) as $f(x)$ is the dike map, figure 11.15,

$$f(\gamma) = \begin{cases} f_0(\gamma) = 2\gamma & \gamma \in \mathcal{M}_0 = [0, \kappa/2) \\ f_c(\gamma) = \kappa & \gamma \in \mathcal{M}_c = [\kappa/2, 1 - \kappa/2] \\ f_1(\gamma) = 2(1 - \gamma) & \gamma \in \mathcal{M}_1 = (1 - \kappa/2, 1] \end{cases}, \tag{11.14}$$

obtained by slicing off all $\gamma(S^+(x_0)) > \kappa$. The dike map is the full tent map figure 11.11 with the top sliced off. It is convenient for coding the symbolic dynamics, as those γ values that survive the pruning are the same as for the full tent map figure 11.11, and are easily converted into admissible itineraries by (11.9).

If $\gamma(S^+) > \gamma(K)$, the point x whose itinerary is S^+ would exceed the critical value, $x > f(x_c)$, and hence cannot be an admissible orbit. Let

$$\hat{\gamma}(S^+) = \sup_m \gamma(\sigma^m(S^+)) \tag{11.15}$$

be the *maximal value*, the highest topological coordinate reached by the orbit $x_1 \to x_2 \to x_3 \to \ldots$, where σ is the shift (11.20), $\sigma(\cdots s_{-2}s_{-1}s_0.s_1s_2s_3\cdots) = \cdots s_{-2}s_{-1}s_0s_1.s_2s_3\cdots$. We shall call the interval $(\kappa, 1]$ the *primary pruned interval*. The orbit S^+ is inadmissible if γ of any shifted sequence of S^+ falls into this interval.

Criterion of admissibility: *Let κ be the kneading value of the critical point, and $\hat{\gamma}(S^+)$ be the maximal value of the orbit S^+. Then the orbit S^+ is admissible if and only if $\hat{\gamma}(S^+) \le \kappa$.*

While a unimodal map may depend on many arbitrarily chosen parameters, its dynamics determines the unique kneading value κ. We shall call κ the *topological*

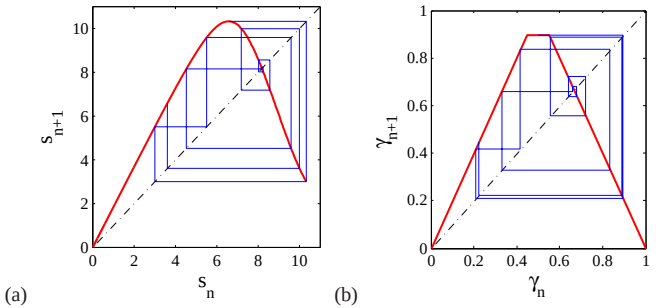


Figure 11.16: (a) Web diagram generated by the trajectory of the critical point the unimodal Rössler return map of figure 11.7 (b). (b) The web diagram for the corresponding 'canonical' dike map (11.14) with the same kneading sequence. (A. Basu and J. Newman)

parameter of the map. Unlike the parameters of the original dynamical system, the topological parameter has no reason to be either smooth or continuous. The jumps in κ as a function of the map parameter such as A in (11.3) correspond to inadmissible values of the topological parameter. Each jump in κ corresponds to a stability window associated with a stable cycle of a smooth unimodal map. For the quadratic map (11.3) κ increases monotonically with the parameter A, but for a general unimodal map such monotonicity need not hold.

Example 11.11 ***Rössler return map web diagram:*** *(continuation of example 11.2) The arclength distance along the unstable manifold of the x_- equilibrium point return map, figure 11.7 (b), generates the kneading sequence (11.13) as the itinerary of the critical point plotted in figure 11.16 (a).*

For further details of unimodal dynamics, the reader is referred to appendix D.1. As we shall see in sect. 12.4, for higher dimensional maps and flows there is no single parameter that orders dynamics monotonically; as a matter of fact, there is an infinity of parameters that need adjustment for a given symbolic dynamics. This difficult subject is beyond our current ambition horizon.

fast track:
chapter 12, p. 244

11.6 Symbolic dynamics, basic notions

> (Mathematics) is considered a specialized dialect of the natural language and its functioning as a special case of speech.
>
> — Yuri I. Manin [11.1]

In this section we collect the basic notions and definitions of symbolic dynamics. The reader might prefer to skim through this material on a first reading and return to it later, as the need arises.

Shifts. We associate with every initial point $x_0 \in \mathcal{M}$ the *future itinerary*, a sequence of symbols $S^+(x_0) = s_1 s_2 s_3 \cdots$ which indicates the order in which the regions are visited. If the trajectory $x_1, x_2, x_3, \ldots$ of the initial point x_0 is generated by

$$x_{n+1} = f(x_n)\,, \tag{11.16}$$

then the itinerary is given by the symbol sequence

$$s_n = s \qquad \text{if} \qquad x_n \in \mathcal{M}_s\,. \tag{11.17}$$

Similarly, the *past itinerary* $S^-(x_0) = \cdots s_{-2} s_{-1} s_0$ describes the history of x_0, the order in which the regions were visited before arriving to the point x_0. To each point x_0 in the dynamical space we thus associate a bi-infinite itinerary

$$S(x_0) = (s_k)_{k\in\mathbb{Z}} = S^-.S^+ = \cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots\,. \tag{11.18}$$

The itinerary will be finite for a scattering trajectory, entering and then escaping $\mathcal{M}$ after a finite time, infinite for a trapped trajectory, and infinitely repeating for a periodic trajectory.

The set of all bi-infinite itineraries that can be formed from the letters of the alphabet $\mathcal{A}$ is called the *full shift* (or *topological Markov chain*)

$$\mathcal{A}^{\mathbb{Z}} = \{(s_k)_{k\in\mathbb{Z}} : s_k \in \mathcal{A} \quad \text{for all} \quad k \in \mathbb{Z}\}\,. \tag{11.19}$$

The jargon is not thrilling, but this is how professional dynamicists talk to each other. We will stick to plain English to the extent possible.

We refer to this set of all conceivable itineraries as the *covering* symbolic dynamics. The name *shift* is descriptive of the way the dynamics acts on these sequences. As is clear from the definition (11.17), a forward iteration $x \to x' = f(x)$ shifts the entire itinerary to the left through the 'decimal point.' This operation, denoted by the shift operator σ,

$$\sigma(\cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots) = \cdots s_{-2} s_{-1} s_0 s_1 . s_2 s_3 \cdots\,, \tag{11.20}$$

demoting the current partition label s_1 from the future S^+ to the 'has been' itinerary S^-. The inverse shift σ^{-1} shifts the entire itinerary one step to the right.

A finite sequence $b = s_k s_{k+1} \cdots s_{k+n_b-1}$ of symbols from $\mathcal{A}$ is called a *block* of length n_b. If the symbols outside of the block remain unspecified, we denote the totality of orbits that share this block by $\_s_k s_{k+1} \cdots s_{k+n_b-1}\_$.

A state space orbit is *periodic* if it returns to its initial point after a finite time; in shift space the orbit is periodic if its itinerary is an infinitely repeating block p^∞. We shall refer to the set of periodic points $\mathcal{M}_p$ that belong to a given periodic orbit as a *cycle*

$$p = \overline{s_1 s_2 \cdots s_{n_p}} = \{x_{s_1 s_2 \cdots s_{n_p}}, x_{s_2 \cdots s_{n_p} s_1}, \cdots, x_{s_{n_p} s_1 \cdots s_{n_p - 1}}\} . \tag{11.21}$$

By its definition, a cycle is invariant under cyclic permutations of the symbols in the repeating block. A bar over a finite block of symbols denotes a periodic itinerary with infinitely repeating basic block; we shall omit the bar whenever it is clear from the context that the orbit is periodic. Each *periodic point* is labeled by the first n_p steps of its future itinerary. For example, the 2nd periodic point is labeled by

$$x_{s_2 \cdots s_{n_p} s_1} = x_{\overline{s_2 \cdots s_{n_p} s_1}.\overline{s_2 \cdots s_{n_p} s_1}} .$$

This - a bit strained - notation is meant to indicate that the symbol block repeats both in the past and in the future. It is helpful for determining spatial ordering of cycles of $2D$-hyperbolic maps, to be undertaken in sect. 12.3.1.

A *prime* cycle p of length n_p is a single traversal of the orbit; its label is a block of n_p symbols that cannot be written as a repeat of a shorter block (in the literature, such cycles are sometimes called *primitive*; we shall refer to it as 'prime' throughout this text).

Partitions. A partition is called *generating* if every infinite symbol sequence corresponds to a distinct point in state space. The finite Markov partition (11.2) is an example. Constructing a generating partition for a given system is a difficult problem. In the examples to follow, we shall concentrate on cases which that permit finite partitions, but in practice almost any generating partition of interest is infinite.

A partition too coarse, coarser than, for example, a Markov partition, would assign the same symbol sequence to distinct dynamical trajectories. To avoid that, we often find it convenient to work with partitions finer than strictly necessary. Ideally the dynamics in the refined partition assigns a unique infinite itinerary $\cdots s_{-2} s_{-1} s_0 . s_1 s_2 s_3 \cdots$ to each distinct orbit, but there might exist full shift symbol sequences (11.19) which are not realized as orbits; such sequences are called *inadmissible*, and we say that the symbolic dynamics is *pruned*. The word is suggested by the 'pruning' of branches corresponding to forbidden sequences for symbolic dynamics organized hierarchically into a tree structure, as explained in chapter 14.

A mapping $f : \mathcal{M} \to \mathcal{M}$ together with a partition $\mathcal{A}$ induces *topological dynamics* (Σ, σ), where the *subshift*

$$\Sigma = \{(s_k)_{k \in \mathbb{Z}}\} , \tag{11.22}$$

is the set of all *admissible* (i.e., 'pruned') infinite itineraries, and $\sigma : \Sigma \to \Sigma$ is the shift operator (11.20). The designation 'subshift' comes form the fact that $\Sigma \subset \mathcal{A}^{\mathbb{Z}}$ is the subset of the full shift (11.19). One of our principal tasks in developing the symbolic dynamics of dynamical systems that occur in nature will be to determine Σ, the set of all bi-infinite itineraries S that are actually realized by the given dynamical system.

Pruning. If the dynamics is pruned, the alphabet must be supplemented by a *grammar*, a set of pruning rules. After the inadmissible sequences have been pruned, it is often convenient to parse the symbolic strings into words of variable length - this is called *coding*. Suppose that the grammar can be stated as a finite number of pruning rules, each forbidding a block of finite length,

$$\mathcal{G} = \{b_1, b_2, \cdots b_k\} , \tag{11.23}$$

where a *pruning block* b is a sequence of symbols $b = s_1 s_2 \cdots s_{n_b}$, $s \in \mathcal{A}$, of finite length n_b. In this case we can always construct a finite Markov partition (11.2) by replacing finite length words of the original partition by letters of a new alphabet. In particular, if the longest forbidden block is of length $M + 1$, we say that the symbolic dynamics is a shift of finite type with M-step memory. In that case we can *recode* the symbolic dynamics in terms of a new alphabet, with each new letter given by an admissible block of at most length M.

A topological dynamical system (Σ, σ) for which all admissible itineraries are generated by a finite transition matrix (14.1)

$$\Sigma = \{(s_k)_{k \in \mathbb{Z}} : T_{s_k s_{k+1}} = 1 \quad \text{for all } k\} \tag{11.24}$$

is called a subshift of *finite type*.

in depth:
chapter 12, p. 244

Résumé

From our initial chapters 2 to 4 fixation on things local: a representative point, a short-time trajectory, a neighborhood, in this chapter we have made a courageous leap and gone global.

The main lesson is that - if one intends to go thoughtfully about globalization - one should trust the dynamics itself, and let it partition the state space, by means of its (topologically invariant) unstable manifolds. This works if every equilibrium and periodic orbit is unstable, so one exits it local neighborhood via its unstable manifold. We delineate the segment of the unstable manifold between the fixed

point and the point where the nonlinearity of the dynamics folds back on itself as the primary segment, and measure location of nearby state space points by arclengths measured along this (curvilinear) segment. For 1-dimensional maps the folding point is the critical point, and easy to determine. In higher dimensions, the situation is not so clear - we shall discuss that in chapter 12.

Trajectories exit a neighborhood of an equilibrium or periodic point along unstable directions, and fall along stable manifolds towards other fixed points, until they again are repelled along their unstable manifolds. Such sequences of visitations can be described by *symbolic dynamics*. As we shall show in chapter 14, they are encoded by transition matrices / transition graphs, and approximated dynamically by sequences of unstable manifold $\to$ unstable manifold maps, or, in case of a return to the initial neighborhood, by return maps $s \to f(s)$.

As 'kneading theory' of sect. 11.5 illustrates, not all conceivable symbol sequences are actually realized (*admissible*). The identification of all inadmissible or *pruned* sequences is in general not possible. However, the theory to be developed here relies on exhaustive enumeration of all admissible itineraries up to a given topological length; chapters 12 and 15 describe several strategies for accomplishing this for physically realistic goals.

Commentary

Remark 11.1 Symbolic dynamics. For a brief history of symbolic dynamics, from Hadamard in 1898, Morse and Hedlund in 1938 and onward, see notes to chapter 1 of Kitchens monograph [11.2], a very clear and enjoyable mathematical introduction to topics discussed here. Diacu and Holmes [11.3] provide an excellent survey of symbolic dynamics applied to celestial mechanics. For a compact survey of symbolic dynamics techniques, consult sects. 3.2 and 8.3 of Robinson [11.4]. The binary labeling of the once-folding map periodic points was introduced by Myrberg [11.5] for 1-dimensional maps, and its utility to 2-dimensional maps has been emphasized in refs. [11.6, 11.7]. For 1-dimensional maps it is now customary to use the *R-L* notation of Metropolis, Stein and Stein [11.8, 11.9], indicating that the point x_n lies either to the left or to the right of the critical point in figure 11.12. The symbolic dynamics of such mappings has been extensively studied by means of the Smale horseshoes, see for example ref. [11.10]. Using letters rather than numerals in symbol dynamics alphabets probably reflects good taste. We prefer numerals for their computational convenience, as they speed up conversions of itineraries into the topological coordinates (δ, γ) introduced in sect. 12.3.1. The alternating binary ordering of figure 11.14 is related to the Gray codes of computer science [11.11]. Kitchens [11.2] convention is $\cdots s_{-2}s_{-1}.s_0s_1s_2s_3\cdots$, with '.' placed differently from our convention (11.18).

Remark 11.2 Kneading theory. The admissible itineraries are studied, for example, in refs. [11.12, 11.8, 11.10, 11.13]. We follow here the Milnor-Thurston exposition [11.14]. They study the topological zeta function for piecewise monotone maps of the interval,

and show that for the finite subshift case it can be expressed in terms of a finite dimensional *kneading determinant*. As the kneading determinant is essentially the topological zeta function of sect. 15.4, we do not discuss it here. Baladi and Ruelle have reworked this theory in a series of papers [11.15, 11.16, 11.17, 11.18]. See also P. Dahlqvist's appendix D.1. Knight and Klages [11.19] in their study of deterministic diffusion (for deterministic diffusion, see chapter 25) refer to the set of iterates of the critical point as 'generating orbit.' They say: "The structure of the Markov partitions varies wildly under parameter variation. The method we employ to understand the Markov partitions involves iterating the critical point. The set of iterates of this point form a set of Markov partition points for the map. Hence we call the orbit of the critical point a 'generating orbit.' If the generating orbit is finite for a particular value of parameters, we obtain a finite Markov partition. We can then use the finite Markov partition to tell us about the diffusive properties of the map and hence the structure of the diffusion coefficient."

Remark 11.3 Heteroclinic connections. For sketches of heteroclinic connections in the nonlinear setting, see Abraham and Shaw illustrated classic [11.20]. Section 5 of ref. [11.21] makes elegant use of stable manifold co-dimension counts and of invariant subspaces implied by discrete symmetries of the underlying PDE to deduce the existence of a heteroclinic connection.

Exercises

11.1. **Binary symbolic dynamics.** Verify that the shortest prime binary cycles of the unimodal repeller of figure 11.12 are $\overline{0}$, $\overline{1}$, $\overline{01}$, $\overline{001}$, $\overline{011}$, $\cdots$. Compare with table 15.1. Sketch them in the graph of the unimodal function $f(x)$; compare the ordering of the periodic points with that in figure 11.14. The point is that while overlayed on each other the longer cycles look like a hopeless jumble, the periodic points are clearly and logically ordered by the alternating binary tree.

11.2. **Generating prime cycles.** Write a program that generates all binary prime cycles up to a given finite length.

11.3. **A contracting baker's map.** Consider the contracting (or "dissipative") baker's map defined in exercise 4.6.

The symbolic dynamics encoding of trajectories is realized via symbols 0 ($y \leq 1/2$) and 1 ($y > 1/2$). Consider the observable $a(x, y) = x$. Verify that for any periodic orbit p ($\epsilon_1 \ldots \epsilon_{n_p}$), $\epsilon_i \in \{0, 1\}$

$$A_p = \frac{3}{4} \sum_{j=1}^{n_p} \delta_{j,1} .$$

11.4. **Unimodal map symbolic dynamics.** Show that the tent map point $\gamma(S^+)$ with future itinerary S^+ is given by converting the sequence of s_n's into a binary number by the algorithm (11.9). This follows by inspection from the binary tree of figure 11.14.

11.5. **Unimodal map kneading value.** Consider the 1-dimensional quadratic map

$$f(x) = Ax(1 - x), \qquad A = 3.8 . \tag{11.25}$$

(a) (easy) Plot (11.25), and the first 4-8 (whatever looks better) iterates of the critical point $x_c = 1/2$.

(b) (hard) Draw corresponding intervals of the partition of the unit interval as levels of a Cantor set, as in the symbolic dynamics partition of figure 11.12. Note, however, that some of the intervals of figure 11.12 do not appear in this case - they are *pruned*.

(c) (easy) Check numerically that $K = S^+(x_c)$, kneading sequence (the itinerary of the critical point (11.13)) is

$$K = 10110111101101111010111101111110 \ldots$$

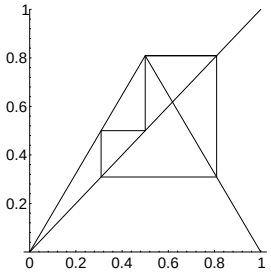
As the orbits of a chaotic map are exponentially unstable, so many digits seem too good to be true - recheck this sequence using arbitrary precision arithmetics.

(d) (medium) The tent map point $\gamma(S^+)$ with future itinerary S^+ is given by converting the sequence of s_n's into a binary number by the algorithm (11.9). List the corresponding kneading value (11.13) sequence $\kappa = \gamma(K)$ to the same number of digits as K.

(e) (hard) Plot the dike map, figure 11.15, with the same kneading sequence K as $f(x)$. The dike map is obtained by slicing off all $\gamma(S^+(x_0)) > \kappa$, from the full tent map figure 11.11, see (11.14).

How this kneading sequence is converted into a series of pruning rules is a dark art, relegated to sect. 15.5.

11.6. **"Golden mean" pruned map.** Consider a symmetric tent map on the unit interval such that its highest point belongs to a 3-cycle:



(a) Find the value $|\Lambda|$ for the slope (the two different slopes $\pm\Lambda$ just differ by a sign) where the maximum at $1/2$ is part of a 3-cycle, as depicted in the figure.

(b) Show that no orbit of this map can visit the region $x > (1 + \sqrt{5})/4$ more than once. Verify also that once an orbit exceeds $x > (\sqrt{5} - 1)/4$, it does not reenter the region $x < (\sqrt{5} - 1)/4$.

(c) If an orbit is in the interval $(\sqrt{5} - 1)/4 < x < 1/2$, where will it be on the next iteration?

(d) If the symbolic dynamics is such that for $x < 1/2$ we use the symbol 0 and for $x > 1/2$ we use the symbol 1, show that no periodic orbit will have the substring _00_ in it.

(e) On a second thought, is there a periodic orbit that violates the above _00_ pruning rule?

To continue with this line of thinking, see exercise 15.7 and exercise 19.2. See also exercise 15.6 and exercise 15.8.

11.7. **Binary 3-step transition matrix.** Construct an [8×8] binary 3-step transition matrix analogous to the 2-step transition matrix (14.10). Convince yourself that the number of terms of contributing to $\mathrm{tr}\, T^n$ is independent of the memory length, and that this $[2^m \times 2^m]$ trace is well defined in the infinite memory limit $m \to \infty$.

11.8. **Full tent map periodic points.** This exercise is easy: just making sure you know how to go back and forth between spatial and temporal ordering of trajectory points.

(a) derive (11.12)

(b) compute the five periodic points of cycle $\overline{10011}$

(c) compute the five periodic points of cycle $\overline{10000}$

(d) (optional) plot the above two cycles on the graph of the full tent map.

(continued in exercise 13.15)

References

[11.1] Yuri I. Manin, *Mathematics as Metaphor: Selected Essays*, (Amer. Math. Soc., Providence RI 2007).

[11.2] B.P. Kitchens, Symbolic dynamics: one-sided, two-sided, and countable state Markov shifts (Springer, Berlin 1998).

[11.3] F. Diacu and P. Holmes, *Celestial Encounters, The Origins of Chaos and Stability* (Princeton Univ. Press, Princeton NJ 1996).

[11.4] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (C. R. C. Press, Boca Raton 1999)

[11.5] P.J. Myrberg, *Ann. Acad. Sc. Fenn., Ser. A,* **256**, 1 (1958); **259**, 1 (1958).

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

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

[11.8] N. Metropolis, M.L. Stein and P.R. Stein, "On Finite Limit Sets for Transformations on the Unit Interval," *J. Comb. Theo.* **15**, 25 (1973).

[11.9] P. Collet and J.P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhauser, Boston 1980).

[11.10] J. Guckenheimer and P. Holmes, *Non-linear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, New York, 1986).

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

[11.12] A.N. Sarkovskii, "Coexistence of cycles of a continuous map of a line into itself," *Ukrainian Math. J.* **16**, 61 (1964).

[11.13] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems* (Addison-Wesley, Reading MA, 1989).

[11.14] J. Milnor and W. Thurston, "On iterated maps of the interval," in A. Dold and B. Eckmann, eds., *Dynamical Systems, Proceedings, U. of Maryland 1986-87, Lec. Notes in Math.* **1342,** 465 (Springer, Berlin 1988).

[11.15] V. Baladi and D. Ruelle, "An extension of the theorem of Milnor and Thurston on the zeta functions of interval maps," *Ergodic Theory Dynamical Systems* **14**, 621 (1994).

[11.16] V. Baladi, "Infinite kneading matrices and weighted zeta functions of interval maps," *J. Functional Analysis* **128**, 226 (1995).

[11.17] D. Ruelle, "Sharp determinants for smooth interval maps," in F. Ledrappier, J. Lewowicz, and S. Newhouse, eds., *Proceedings of Montevideo Conference 1995* (Addison-Wesley, Harlow 1996).

[11.18] V. Baladi and D. Ruelle, "Sharp determinants," *Invent. Math.* **123**, 553 (1996).

[11.19] G. Knight and R. Klages, "Linear and fractal diffusion coefficients in a family of one dimensional chaotic maps," `arXiv:1007.3393`.

[11.20] Abraham, R. H. and Shaw, C. D., *Dynamics - The geometry of behavior* (Addison-Wesley, Reading, MA 1992).

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

[11.22] G. 't Hooft and M. Veltman, *DIAGRAMMAR, CERN report 73/9*, (1973); reprinted in *Particle interactions at very high energies, NATO Adv. Study Inst. Series, Sect. B*, vol. **4B**, 177; www.phys.uu.nl/~thooft/gthpub.html.