2
Discrete Dynamical Systems: Maps

2.1
Introduction

Many physical systems displaying chaotic behavior are accurately described by mathematical models derived from well-understood physical principles. For example, the fundamental equations of fluid dynamics, namely the Navier–Stokes equations, are obtained from elementary mechanical and thermodynamical considerations. The simplest laser models are built from Maxwell's laws of electromagnetism and from the quantum mechanics of a two-level atom.

Except for stationary regimes, it is in general not possible to find closed-form solutions to systems of nonlinear partial or ordinary differential equations (PDEs and ODEs). However, numerical integration of these equations often reproduces surprisingly well the irregular behaviors observed experimentally. Thus, these models must have some mathematical properties that are linked to the occurrence of chaotic behavior. To understand what these properties are, it is clearly desirable to study chaotic dynamical systems whose mathematical structure is as simple as possible.

Because of the difficulties associated with the analytical study of differential systems, a large amount of work has been devoted to dynamical systems whose state is known only at a discrete set of times. These are usually defined by a relation

$$X_{n+1} = f(X_n) \tag{2.1}$$

where $f : M \rightarrow M$ is a map of a state space into itself and X_n denotes the state at the discrete time n. The sequence $\{X_n\}$ obtained by iterating (2.1) starting from an initial condition X_0 is called the *orbit* of X_0, with $n \in \mathbb{N}$ or $n \in \mathbb{Z}$, depending on whether or not map f is invertible. If X_0 is chosen at random, one observes generally that its orbit displays two distinct phases: a *transient* that is limited in time and specific to the orbit, and an *asymptotic regime* that persists for arbitrarily long times and is qualitatively the same for different typical orbits. Rather than study individual orbits, we are interested in classifying all the asymptotic behaviors that can be observed in a dynamical system.

To understand the generic properties of chaotic behavior, there is no loss of generality in restricting ourselves to discrete-time systems; it is often easy to extract such a system from a continuous-time system. A common example is provided by the technique of Poincaré sections, where the study is narrowed to the dynamics of the map relating the successive intersections of a trajectory with a given surface of phase space, which is usually called the *first return map*. Another example is the *time-one map* $X(t+1) = \phi^1(X(t))$ of a differential system, which relates two states located one time unit apart along the same trajectory.

It might seem natural to restrict our study to discrete-time dynamical systems sharing some key properties with differential systems. For example, the solution of a system of ODEs depends continuously on initial conditions, so that continuous maps are naturally singled out. Invertibility is also a crucial property. Given an initial condition, the state of an ODE system can in principle be determined at any time in the future but also in the past; thus, we must be able to go backward in time. Maps satisfying these two requirements (i.e., continuous maps with a continuous inverse) are called *homeomorphisms*. Another important class of maps is made of *diffeomorphisms*: These are homeomorphisms that are differentiable as well as their inverse.

The most important aspects of chaotic behavior should appear in systems of lowest dimension. Thus, we would like in a first step to reduce as much as possible the dimension of state space. However, this quickly conflicts with the requirement of invertibility. On the one hand, it can be shown that maps based on a one-dimensional homeomorphism can only display stationary or periodic regimes, and hence cannot be chaotic. On the other hand, if we sacrifice invertibility temporarily, thereby introducing *singularities*, one-dimensional chaotic systems can easily be found, as illustrated by the celebrated logistic map. Indeed, this simple system will be seen to display many of the essential features of deterministic chaos.

It is, in fact, no coincidence that chaotic behavior appears in its simplest form in a noninvertible system. As emphasized in this book, singularities and noninvertibility are intimately linked to the mixing processes (stretching and squeezing) associated with chaos.

Because of the latter, a dissipative invertible chaotic map becomes formally noninvertible when infinitely iterated (i.e., when the phase space has been infinitely squeezed). Thus the dynamics is, in fact, organized by an underlying singular map of lower dimension, as can be shown easily in model systems such as the horseshoe map. A classical example of this is the Hénon map, a diffeomorphism of the plane into itself that is known to have the logistic map as a backbone.

2.2
Logistic Map

A noninvertible one-dimensional map has at least one point where its derivative vanishes. The simplest such maps are quadratic polynomials, which can always be brought to the form $f(x) = a - x^2$ under a suitable change of variables. The
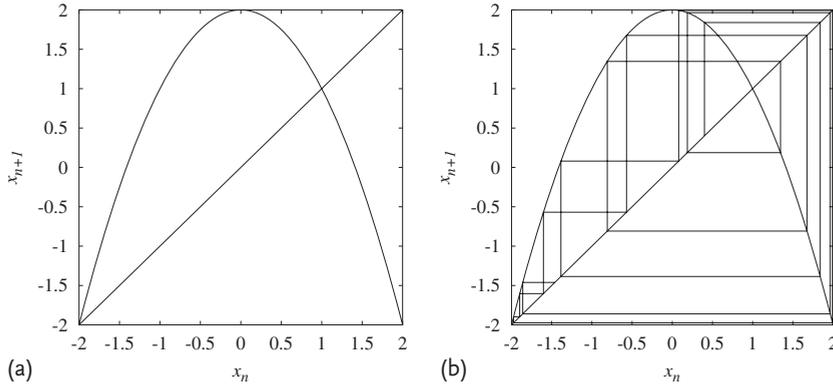
Figure 2.1 (a) Graph of the logistic map for $a = 2$; (b) graphical representation of the iteration of (2.2).

logistic map[1]

$$x_{n+1} = a - x_n^2 \tag{2.2}$$

which depends on a single parameter a, is thus the simplest one-dimensional map displaying a singularity. As can be seen from its graph (Figure 2.1a), the most important consequence of the singularity located at the critical point $x = 0$ is that each value in the range of the map f has exactly two preimages, which will prove to be a key ingredient to generate chaos. Maps with a single critical point are called *unimodal*. It will be seen later that all unimodal maps display very similar dynamical behavior.

As is often the case in dynamical systems theory, the action of the logistic map can be represented not only algebraically, as in (2.2), but also geometrically. Given a point x_n, the graph of the logistic map provides $y = f(x_n)$. To use y as the starting point of the next iteration, we must find the corresponding location in the x space, which is done simply by drawing the line from the point $[x_n, f(x_n)]$ to the diagonal $y = x$. This simple construction is then repeated ad libitum, as illustrated in Figure 2.1b.

The various behaviors displayed by the logistic map are easily explored, as this map depends on a single parameter a. As illustrated in Figure 2.2, one finds quickly that two main types of dynamical regimes can be observed: stationary or periodic regimes on the one hand, and "chaotic" regimes on the other hand. In the first case, iterations eventually visit only a finite set of different values that are forever repeated in a fixed order. In the latter case, the state of the system never repeats itself exactly and seemingly evolves in a disordered way, as in Figure 2.1b. Both types of behaviors have been observed in the experiment discussed in Chapter 1.

What makes the study of the logistic map so important is not only that the organization in parameter space of these periodic and chaotic regimes can be completely

---

1) A popular variant is $x_{n+1} = \lambda x_n(1 - x_n)$, with parameter λ.
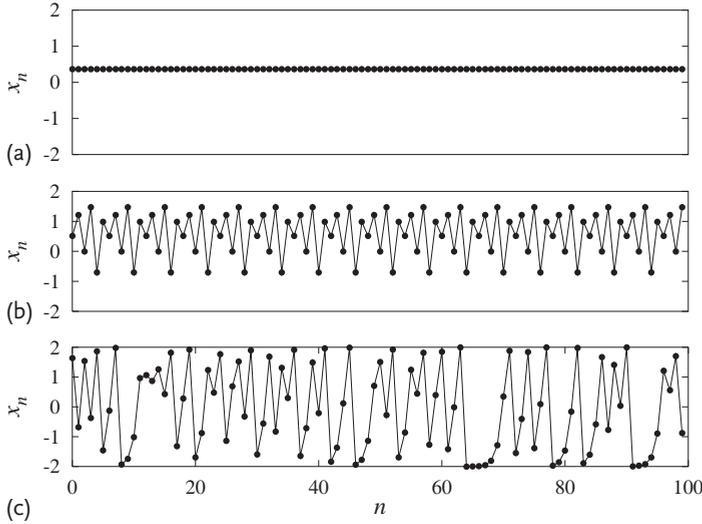
Figure 2.2 Different dynamical behaviors observed in the logistic map system are represented by plotting successive iterates: (a) stationary regime, $a = 0.5$; (b) periodic regime of period 5, $a = 1.476$; (c) chaotic regime, $a = 2.0$.

understood with simple tools, but that, despite its simplicity, it displays the most important features of low-dimensional chaotic behavior. By studying how periodic and chaotic behavior are interlaced, we will learn much about the mechanisms responsible for the appearance of chaotic behavior. Moreover, the logistic map is not only a paradigmatic system: One-dimensional maps will later prove also to be a fundamental tool for understanding the topological structure of flows.

2.3
Bifurcation Diagrams

A first step in classifying the dynamical regimes of the logistic map is to obtain a global representation of the various regimes that are encountered as control parameter a is varied. This can be done with the help of *bifurcation diagrams,* which are tools commonly used in nonlinear dynamics. Bifurcation diagrams display some characteristic property of the asymptotic solution of a dynamical system as a function of a control parameter, allowing one to see at a glance where qualitative changes in the asymptotic solution occur. Such changes are termed *bifurcations*.

In the case of the logistic map that has a single dynamical variable, the bifurcation diagram is readily obtained by plotting a sample set of values of the sequence (x_n) as a function of parameter a, as shown in Figure 2.3.

For $a < a_0 = -1/4$, iterations of the logistic map escape to infinity from all initial conditions. For $a > a_R = 2$ almost all initial conditions escape to infinity.
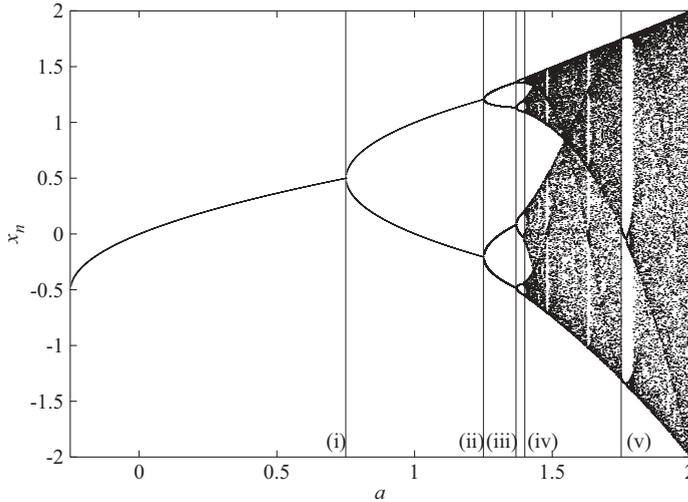
Figure 2.3 Bifurcation diagram of the logistic map. For a number of parameter values between $a = -0.25$ and $a = 2.0$, 50 successive iterates of the logistic map are plotted after transients have died out. From left to right, the vertical lines mark the creations of (i) a period-2 orbit; (ii) a period-4 orbit; (iii) a period-8 orbit, and (iv) the accumulation point of the period-doubling cascade; (v) the starting point of a period-3 window.

The bifurcation diagram is thus limited to the range $a_0 < a < a_R$, where bounded solutions can be observed.

Between $a_0 = -1/4$ and $a_1 = 3/4$, the limit set consists of a single value. This corresponds to a stationary regime, but one that should be considered in this context as a period-1 *periodic orbit*. At $a = a_1$, a bifurcation occurs, giving birth to a period-2 *periodic orbit*: Iterations oscillate between two values. As detailed in Section 2.4.2, this is an example of a *period-doubling bifurcation*. At $a = a_2 = 5/4$, there is another period-doubling bifurcation where the period-2 orbit gives birth to a period-4 orbit.

The period-doubling bifurcations occurring at $a = a_1$ and $a = a_2$ are the first two members of an infinite series, known as the *period-doubling cascade*, in which an orbit of period 2^n is created for every integer n. The bifurcation at $a = a_3$ leading to a period-8 orbit is easily seen in the bifurcation diagram of Figure 2.3, the one at $a = a_4$ is hardly visible, and the following ones are completely indiscernible to the naked eye. This is because the parameter values a_n at which the period-2^n orbit is created converge geometrically to the accumulation point $a_\infty = 1.401\,155\,189\ldots$ with a convergence ratio substantially larger than 1:

$$\lim_{n \to \infty} \frac{a_n - a_{n-1}}{a_{n+1} - a_n} = \delta \sim 4.669\,201\,609\,1 \ldots \tag{2.3}$$

The constant δ appearing in (2.3) was discovered by Feigenbaum [22, 37] and is named after him. This distinction is justified by a remarkable property: Period-doubling cascades observed in an extremely large class of systems (experimental or

Figure 2.4 Enlarged view of the chaotic zone of the bifurcation diagram of Figure 2.3. Inside periodic windows of period up to 8, vertical lines indicate the parameter values where the corresponding orbits are most stable, with the period indicated above the line.

theoretical, defined by maps or differential equations ...) have a convergence rate given by δ.

At the accumulation point a_∞, the period of the solution has become infinite. Right of this point, the system can be found in chaotic regimes, as can be guessed from the abundance of dark regions in this part of the bifurcation diagram, which indicate that the system visits many different states. The period-doubling cascade is one of the best-known *routes to chaos* and can be observed in many low-dimensional systems [10]. It has many universal properties that are in no way restricted to the case of the logistic map.

However, the structure of the bifurcation diagram is more complex than a simple division between periodic and chaotic regions on both sides of the accumulation point of the period-doubling cascade. For example, a relatively large *periodic window*, which corresponds to the domain of stability of a period-3 orbit, is clearly seen to begin at $a = 7/4$, well inside the chaotic zone. In fact, periodic windows and chaotic regions are arbitrarily finely interlaced, as illustrated by Figure 2.4. As will be shown later, there are infinitely many periodic windows between any two periodic windows. To interpret Figure 2.4, it should be noted that periodic windows are visible to the naked eye only for very low periods. For higher periods, (i) the periodic window is too narrow compared to the scale of the plot and (ii) the number of samples is sufficiently large that the window cannot be distinguished from the chaotic regimes.

Ideally, we would like to determine for each periodic solution the range of parameter values over which it is stable. In Section 2.4 we will perform this analysis for the simple cases of the period-1 and period-2 orbits, so that we get a better understanding of the two types of bifurcation that are encountered in a logistic map. This is motivated by the fact that these are the two bifurcations that are generically observed in low-dimensional dynamical systems (omitting the Hopf bifurcation, which we discuss later).

However, we will not attempt to go much further in this direction. First, the complexity of Figures 2.3 and 2.4 shows that this task is out of reach. Moreover, we are only interested in properties of the logistic map that are shared by many other dynamical systems. In this respect, computing exact stability ranges for a large number of regimes would be pointless.

This does not imply that a deep understanding of the structure of the bifurcation diagram of Figure 2.3 cannot be achieved. Quite to the contrary, we will see later that simple topological methods allow us to answer precisely the following questions: How can we classify the different periodic regimes? Does the succession of different dynamical regimes encountered as parameter *a* is increased follow a logical scheme? In particular, a powerful approach to chaotic behavior, *symbolic dynamics*, which we present in Section 2.7, will prove to be perfectly suited for unfolding the complexity of chaos.

2.4
Elementary Bifurcations in the Logistic Map

2.4.1
Saddle–Node Bifurcation

The simplest regime that can be observed in the logistic map is the period-1 orbit. It is stably observed on the left of the bifurcation diagram of Figure 2.3 for $a_0 < a < a_1$. It corresponds to a *fixed point* of the logistic map (i.e., it is mapped onto itself) and is thus a solution of the equation $x = f(x)$. For the logistic map, finding the fixed points merely amounts to solving the quadratic equation

$$x = a - x^2 \tag{2.4}$$

which has two solutions:

$$x_-(a) = \frac{-1 - \sqrt{1 + 4a}}{2} \quad x_+(a) = \frac{-1 + \sqrt{1 + 4a}}{2} \tag{2.5}$$

The fixed points of a one-dimensional map can also be located geometrically, since they correspond to the intersections of its graph with the diagonal (Figure 2.1).

Although a single period-1 regime is observed in the bifurcation diagram, there are actually two period-1 orbits. Later we will see why. Expressions (2.5) are real-valued only for $a > a_0 = -1/4$. Below this value, all orbits escape to infinity. Thus, the point at infinity, which we denote x_∞ in what follows, can formally be considered as another fixed point of the system, albeit unphysical.

The important qualitative change that occurs at $a = a_0$ is our first example of a ubiquitous phenomenon of low-dimensional nonlinear dynamics, a *tangent*, or *saddle–node*, bifurcation: The two fixed points (2.5) become simultaneously real and are degenerate: $x_-(a_0) = x_+(a_0) = -1/2$. The two designations point to two different (but related) properties of this bifurcation.
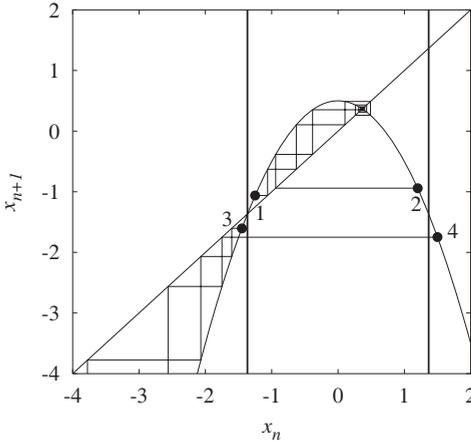
Figure 2.5 The basin of attraction of the x_+ fixed point is located between the left fixed point x_- and its preimage, indicated by two vertical lines. The orbits labeled 1 and 2 are inside the basin and converge toward x_+. The orbits labeled 3 and 4 are outside the basin and escape to infinity (i.e., converge to the point at infinity x_∞).

The saddle–node qualifier is related to the fact that the two bifurcating fixed points have different stability properties. For a slightly above a_0, it is found that orbits located near x_+ converge to it, whereas those starting in the neighborhood of x_- leave it to either converge to x_+ or escape to infinity, depending on whether they are located right or left of x_-. Thus, the fixed point x_+ (and obviously also x_∞) is said to be *stable* while x_- is *unstable*. They are called the *node* and the *saddle*, respectively.

Since trajectories in their respective neighborhoods converge to them, x_+ and x_∞ are *attracting sets*, or *attractors*. The sets of points whose orbits converge to an attractor of a system is called the *basin of attraction* of this point. From Figure 2.5 we see that the unstable fixed point x_- is on the boundary between the *basins of attraction* of the two stable fixed points x_+ and x_∞. The other boundary point is the preimage $f^{-1}(x_-)$ of x_- (Figure 2.5).

It is easily seen that the stability of a fixed point depends on the derivative of the map at the fixed point. Indeed, if we perturb a fixed point $x_* = f(x_*)$ by a small quantity δx_n, the perturbation δx_{n+1} at the next iteration is given by

$$\delta x_{n+1} = f(x_* + \delta x_n) - x_* = \left.\frac{d f(x)}{dx}\right|_{x_*} \delta x_n + O\left(\delta x_n^2\right) \tag{2.6}$$

If we start with an infinitesimally small δx_0, the perturbation after n iterations is thus $\delta x_n \approx (\mu_*)^n \delta x_0$, where μ_*, the *multiplier* of the fixed point, is given by the map derivative at $x = x_*$.

A fixed point is thus stable (resp. unstable) when the absolute value of its multiplier is smaller (resp. greater) than unity. Here the multipliers $\mu_\pm$ of the two fixed
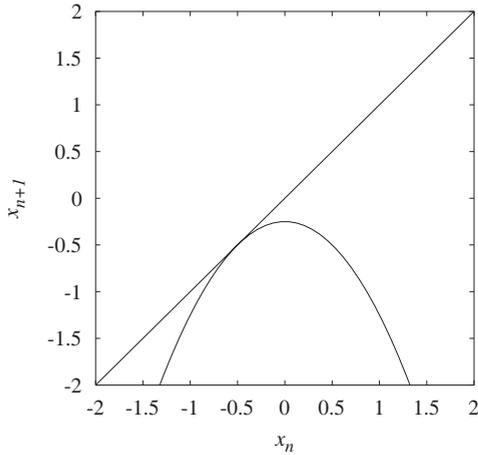
Figure 2.6 Graph of the logistic map at the initial saddle–node bifurcation.

points of the logistic map are given by

$$\mu_- = \left.\frac{d\,f(x)}{d\,x}\right|_{x_-} = -2x_- = 1 + \sqrt{1+4a} \tag{2.7a}$$

$$\mu_+ = \left.\frac{d\,f(x)}{d\,x}\right|_{x_+} = -2x_+ = 1 - \sqrt{1+4a} \tag{2.7b}$$

Equation 2.7a shows that x_- is unconditionally unstable on its entire domain of existence, and hence is generically not observed as a stationary regime, whereas x_+ is stable for parameters a just above $a_0 = -1/4$, as mentioned above. This is why only x_+ can be observed on the bifurcation diagram shown in Figure 2.3.

More precisely, x_+ is stable for $a \in [a_0, a_1]$, where $a_1 = 3/4$ is such that $\mu_+ = -1$. This is consistent with the bifurcation diagram of Figure 2.3. Note that at $a = 0 \in [a_0, a_1]$, the multiplier $\mu_+ = 0$ and thus perturbations are damped out faster than exponentially: The fixed point is then said to be *superstable*.

At the saddle–node bifurcation, both fixed points are degenerate and their multiplier is $+1$. This fundamental property is linked to the fact that at the bifurcation point, the graph of the logistic map is tangent to the diagonal (Figure 2.6), which is why this bifurcation is also known as the tangent bifurcation. Tangency of two smooth curves (here, the graph of f and the diagonal) is generic at a multiple intersection point. This is an example of a *structurally unstable situation*: An arbitrarily small perturbation of f leads to two distinct intersections or no intersection at all (alternatively, to two real or to two complex roots).

It is instructive to formulate the intersection problem in algebraic terms. The fixed points of the logistic equations are zeros of the equation $G(x, a) = f(x; a) - x = 0$. This equation defines implicit functions $x_+(a)$ and $x_-(a)$ of parameter a. In structurally stable situations, these functions can be extended to neighboring parameter values by use of the implicit function theorem.

Assume that $x_*(a)$ satisfies $G(x_*(a), a) = 0$ and that we shift parameter a by an infinitesimal quantity δa. Provided that $\partial G(x_*(a), a)/\partial x \neq 0$, the corresponding variation δx_* in x_* is given by

$$G(x, a) = G(x_* + \delta x_*, a + \delta a) = G(x, a) + \frac{\partial G}{\partial x}\delta x_* + \frac{\partial G}{\partial a}\delta a = 0 \quad (2.8)$$

which yields

$$\delta x_* = -\frac{\dfrac{\partial G}{\partial a}}{\dfrac{\partial G}{\partial x}}\delta a \quad (2.9)$$

showing that $x_*(a)$ is well defined on both sides of a if and only if $\partial G/\partial x \neq 0$. The condition

$$\frac{\partial G(x_*(a), a)}{\partial x} = 0 \quad (2.10)$$

is thus the signature of a bifurcation point. In this case, the Taylor series (2.8) has to be extended to higher orders of δx_*. If $\partial^2 G(x_*(a), a)/\partial x^2 \neq 0$, the variation δx_* in the neighborhood of the bifurcation is given by

$$(\delta x_*)^2 = -2\frac{\dfrac{\partial G}{\partial a}}{\dfrac{\partial^2 G}{\partial x^2}}\delta a \quad (2.11)$$

From (2.11) we recover the fact that there is a twofold degeneracy at the bifurcation point, two solutions on one side of the bifurcation and none on the other side. The stability of the two bifurcating fixed points can also be analyzed: Since $G(x, a) = f(x; a) - x$, their multipliers are given by $\mu_* = 1 + \partial G(x_*, a)/\partial x$ and are thus equal to 1 at the bifurcation.

Just above the bifurcation point, it is easy to show that the multipliers of the two fixed points x_+ and x_- are given to leading order by $\mu_\pm = 1 \mp \alpha\sqrt{|\delta a|}$, where the factor α depends on the derivatives of G at the bifurcation point. It is thus generic that one bifurcating fixed point is stable while the other one is unstable. In fact, this is a trivial consequence of the fact that the two nondegenerate zeros of $G(x, a)$ must have derivatives $\partial G/\partial x$ with opposite signs.

This is linked to a fundamental theorem, which we state below in the one-dimensional case but which can be generalized to arbitrary dimensions by replacing derivatives with Jacobian determinants. Define the degree of a map f as

$$\deg f = \sum_{f(x_i)=y} \mathrm{sgn}\frac{df}{dx}(x_i) \quad (2.12)$$

where the sum extends over all the preimages of the arbitrary point y, and $\mathrm{sgn}z = +1$ (resp. -1) if $z > 0$ (resp. $z < 0$). It can be shown that $\deg f$ does not depend

on the choice of γ provided that it is a regular value (the derivatives at its preimages x_i are not zero) and that it is invariant by homotopy. Let us apply this to $G(x, a)$ for $\gamma = 0$. Obviously, deg $G = 0$ when there are no fixed points, but also for any a since the effect of varying a is a homotopy. We thus see that fixed points must appear in pairs having opposite contributions to deg G. As discussed above, these opposite contributions correspond to different stability properties at the bifurcation.

The discussion above shows that although we have introduced the tangent bifurcation in the context of the logistic map, much of the analysis can be carried to higher dimensions. In an n-dimensional state space, the fixed points are determined by an n-dimensional vector function $\boldsymbol{G}$. In a structurally stable situation, the Jacobian $\partial \boldsymbol{G} / \partial \boldsymbol{X}$ has rank n. As one control parameter is varied, bifurcations will be encountered at parameter values where $\partial \boldsymbol{G} / \partial \boldsymbol{X}$ is of lower rank. If the Jacobian has rank $n - 1$, it has a single null eigenvector, which defines the direction along which the bifurcation takes place. This explains why the essential features of tangent bifurcations can be understood from a one-dimensional analysis.

The theory of bifurcations is in fact a subset of a larger field of mathematics, the *theory of singularities* [50, 51], which includes *catastrophe theory* [35, 49] as a special important case. The tangent bifurcation is an example of the simplest type of singularity: the *fold singularity*, which typically corresponds to twofold degeneracies.

In the next section we see an example of a threefold degeneracy, the *cusp singularity*, in the form of the period-doubling bifurcation.

2.4.2
Period-Doubling Bifurcation

As shown in Section 2.4.1, the fixed point x_+ is stable only for $a \in [a_0, a_1]$, with $\mu_+ = 1$ at $a = a_1 = -1/4$ and $\mu_+ = -1$ at $a = a_1 = 3/4$. For $a > a_1$, both fixed points (2.5) are unstable, which precludes a period-1 regime. Just above the bifurcation, what is observed instead is that successive iterates oscillate between two distinct values (Figure 2.3), which comprise a period-2 orbit. This could have been expected from the fact that at $a = a_1$, $\mu_+ = -1$ indicates that perturbations are reproduced every other period. The qualitative change that occurs at $a = a_1$ (a fixed point becomes unstable and gives birth to an orbit of twice the period) is another important example of bifurcation: the *period-doubling bifurcation*, which is represented schematically in Figure 2.7. Saddle–node and period-doubling bifurcations are the only two types of local bifurcation that are observed for the logistic map. With the Hopf bifurcation, they are also the only bifurcations that occur generically in one-parameter paths in parameter space and, consequently, in low-dimensional systems.

Before we carry out the stability analysis for the period-2 orbit created at $a = a_1$, an important remark has to be made. Expression (2.5) shows that the period-1 orbit x_+ exists for every $a > a_0$; hence it does not disappear at the period-doubling bifurcation but merely becomes unstable. It is thus present in all the dynamical regimes observed after its loss of stability, including in the chaotic regimes of the right part of the bifurcation diagram of Figure 2.3. In fact, this holds for all the
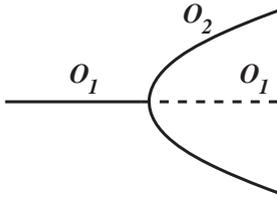
Figure 2.7 Period-doubling bifurcation. Orbit O_1 becomes unstable in giving birth to orbit O_2, whose period is twice that of O_1.
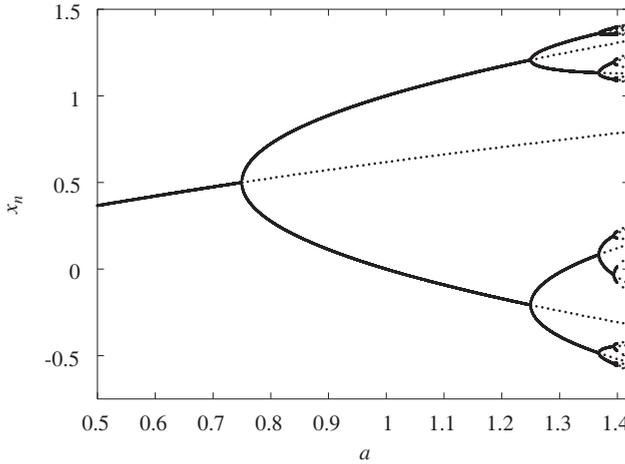
Figure 2.8 Orbits of period up to 16 of the period-doubling cascade. Stable (resp. unstable) periodic orbits are drawn with solid (resp. dashed) lines.

periodic solutions of the logistic map. As an example, the logistic map at the transition to chaos ($a = a_\infty$) has an infinity of (unstable) periodic orbits of periods 2^n for any n, as Figure 2.8 shows.

We thus expect periodic orbits to play an important role in the dynamics even after they have become unstable. We will see later that this is indeed the case and that much can be learned about a chaotic system from its set of periodic orbits, both stable and unstable.

Since the period-2 orbit can be viewed as a fixed point of the second iterate of the logistic map, we can proceed as above to determine its range of stability. The two periodic points $\{x_1, x_2\}$ are solutions of the quartic equation

$$x = f(f(x)) = a - \left(a - x^2\right)^2 \tag{2.13}$$

To solve for x_1 and x_2, we take advantage of the fact that the fixed points x_+ and x_- are obviously solutions of (2.13). Hence, we just have to solve the quadratic equation

$$p(x) = \frac{f(f(x)) - x}{f(x) - x} = 1 - a - x + x^2 = 0 \tag{2.14}$$

whose solutions are

$$x_1 = \frac{1 - \sqrt{-3 + 4a}}{2} \quad x_2 = \frac{1 + \sqrt{-3 + 4a}}{2} \tag{2.15}$$

We recover the fact that the period-2 orbit (x_1, x_2) appears at $a = a_1 = 3/4$ and exists for every $a > a_1$. By using the chain rule for derivatives, we obtain the multiplier of the fixed point x_1 of f^2 as

$$\mu_{1,2} = \left.\frac{d\, f^2(x)}{dx}\right|_{x_1} = \left.\frac{d\, f(x)}{dx}\right|_{x_2} \times \left.\frac{d\, f(x)}{dx}\right|_{x_1} = 4x_1 x_2 = 4(1 - a) \tag{2.16}$$

Note that x_1 and x_2 viewed as fixed points of f^2 have the same multiplier, which is defined to be the multiplier of the orbit (x_1, x_2). At the bifurcation point $a = a_1$, we have $\mu_{1,2} = 1$, a signature of the two periodic points x_1 and x_2 being degenerate at the period-doubling bifurcation.

However, the structure of the bifurcation is not completely similar to that of the tangent bifurcation discussed earlier. Indeed, the two periodic points x_1 and x_2 are also degenerate with the fixed point x_+. The period-doubling bifurcation of the fixed point x_+ is thus a situation where the second iterate f^2 has *three* degenerate fixed points. If we define $G_2(x, a) = f^2(x; a) - x$, the signature of this threefold degeneracy is $G_2 = \partial G_2/\partial x = \partial^2 G_2/\partial x^2 = 0$, which corresponds to a higher-order singularity than the fold singularity encountered in our discussion of the tangent bifurcation. This is, in fact, our first example of the *cusp singularity*. Note that x_+ has a multiplier of -1 as a fixed point of f at the bifurcation and hence exists on both sides of the bifurcation; it merely becomes unstable at $a = a_1$. On the contrary, x_1 and x_2 have multiplier 1 for the lowest iterate of f, of which they are fixed points, and thus exist only on one side of the bifurcation.

We also may want to verify that deg $f^2 = 0$ on both sides of the bifurcation. Let us denote $d(x_*)$ as the contribution of the fixed point x_* to deg f^2. We do not consider x_-, which is not invoved in the bifurcation. Before the bifurcation, we have $d(x_+) = -1$ $(d\, f^2/dx(x_+) < 1)$. After the bifurcation, $d(x_+) = 1$ but $d(x_1) = d(x_2) = -1$, so that the sum is conserved.

The period-2 orbit is stable only on a finite parameter range. The other end of the stability domain is at $a = a_2 = 5/4$, where $\mu_{1,2} = -1$. At this parameter value, a new period-doubling bifurcation takes place, where the period-2 orbit loses its stability and gives birth to a period-4 orbit. As shown in Figures 2.3 and 2.8, period doubling occurs repeatedly until an orbit of infinite period is created.

Although one might in principle repeat the analysis above for the successive bifurcations of the period-doubling cascade, the algebra involved quickly becomes intractable. Anyhow, the sequence of parameters a_n at which a solution of period 2^n emerges converges so quickly to the accumulation point a_∞ that this would be of little use, except perhaps to determine the exact value of a_∞, after which the first chaotic regimes are encountered.

A fascinating property of the period-doubling cascade is that we do not need to analyze directly the orbit of period 2^∞ to determine very accurately a_∞. Indeed,

it can be remarked that the orbit of period 2^∞ is formally its own period-doubled orbit. This indicates some kind of scale invariance. Accordingly, it was recognized by Feigenbaum that the transition to chaos in the period-doubling cascade can be analyzed by means of renormalization group techniques [22, 37].

In this section we have analyzed how the periodic solutions of the logistic map are created. After discussing changes of coordinate systems in the next section, we shall take a closer look at the chaotic regimes appearing in the bifurcation diagram of Figure 2.3. We will then be in a position to introduce more sophisticated techniques to analyze the logistic map, namely *symbolic dynamics*, and to gain a complete understanding of the bifurcation diagram of a large class of maps of the interval.

2.5
Map Conjugacy

2.5.1
Changes of Coordinates

The behavior of a physical system does not depend on how we describe it. Equations defining an abstract dynamical system are meaningful only with respect to a given parameterization of its states (i.e., in a given coordinate system). If we change the parameterization, the dynamical equations should be modified accordingly so that the same physical states are connected by the evolution laws.

Assume that we have a system whose physical states are parameterized by coordinates $x \in X$, with an evolution law given by $f \colon X \to X$ (i.e., $x_{n+1} = f(x_n)$). If we switch to a new coordinate system specified by $y = h(x)$, with $y \in Y$, the dynamical equations become $y_{n+1} = g(y_n)$, where the map $g \colon Y \to Y$ satisfies

$$h(f(x)) = g(h(x)) \tag{2.17}$$

Relation (2.17) simply expresses that, on the one hand, $y_{n+1} = h(x_{n+1}) = h(f(x_n))$, and on the other hand, $y_{n+1} = g(y_n) = g(h(x_n))$. This is summarized by what follows *commutative diagram*:

$$
\begin{array}{ccc}
x_n & \xrightarrow{\;f\;} & x_{n+1} \\
\Big\downarrow{\scriptstyle h} & & \Big\downarrow{\scriptstyle h} \\
y_n & \xrightarrow{\;g\;} & y_{n+1}
\end{array}
\tag{2.18}
$$

where relation (2.17) is recovered by comparing the two paths from x_n to y_{n+1}.

Different types of conjugacy may be defined depending on the class of functions the transformation h belongs to (e.g., see [52]). *Conjugacy*, or *smooth conjugacy*, corresponds to the case where h is a diffeomorphism. If h is a homeomorphism, one has *topological conjugacy*. Note that in some cases, transformation h can be $2 \to 1$. This is referred to as *semiconjugacy*.

2.5.2
Invariants of Conjugacy

Often, the problem is not to compute the evolution equations in a new coordinate system, but to determine whether two maps f and g correspond to the same physical system (i.e., whether or not they are conjugate). A common strategy to address this type of problem is to search for quantities that are invariant under the class of transformations considered. If two objects have different invariants, they cannot be transformed into each other. The knot invariants discussed later provide an important example of this. The ideal case is when there exists a complete set of invariants; equality of the invariants then implies identity of the objects. In this section we present briefly two important invariants of conjugacy.

Spectrum of periodic orbits. An important observation is that there is a one-to-one correspondence between periodic orbits of two conjugate maps. Assume that x_* is a period-p orbit of f: $f^p(x_*) = x_*$. If $f = h^{-1} \circ g \circ h$, we have

$$f^p = \left(h^{-1} \circ g \circ h\right)^p = h^{-1} \circ g^p \circ h \tag{2.19}$$

Thus, $y_* = h(x_*)$ satisfies

$$y_* = h\left(f^p(x_*)\right) = g^p\left(h(x_*)\right) = g^p\left(y_*\right) \tag{2.20}$$

This shows that y_* is itself a period-p orbit of g. If h is a one-to-one transformation, it follows immediately that f and g have the same number of period-p orbits.

Of course, this should have been expected: The existence of a periodic solution does not depend on the coordinate system. Yet this provides a useful criterion to test whether two maps are conjugate.

Multipliers of periodic orbits. Similarly, the stability and the asymptotic evolution of a system are coordinate independent. In algebraic terms, this translates into the invariance of the multipliers of a periodic orbit when transformation h is a diffeomorphism.

To show this, let us compute the tangent map Dg^p of $g^p = h \circ f^p \circ h^{-1}$ at a point y_0, using the chain rule for derivatives:

$$Dg^p(y_0) = Dh\left(\left(f^p \circ h^{-1}\right)(y_0)\right) \times Df^p\left(h^{-1}(y_0)\right) \times Dh^{-1}(y_0) \tag{2.21}$$

To simplify notations, we set $x_0 = h^{-1}(y_0)$ and rewrite (2.21) as

$$Dg^p(y_0) = Dh\left(f^p(x_0)\right) \times Df^p(x_0) \times Dh^{-1}(y_0) \tag{2.22}$$

Relation (2.22) yields no special relation between $Df^p(x_0)$ and $Dg^p(y_0)$ unless x_0 is a period-p orbit and satisfies $f^p(x_0) = x_0$. In this case, indeed, we note that since $Dh(f^p(x_0)) = Dh(x_0)$, and because

$$Dh\left(h^{-1}(y_0)\right) \times Dh^{-1}(y_0) = 1 \tag{2.23}$$

we have

$$Dg^p(y_0) = P \times Df^p(x_0) \times P^{-1} \tag{2.24}$$

where $P = Dh(x_0)$. Equation 2.24 indicates that the matrices $Df^p(x_0)$ and $Dg^p(y_0)$ are *similar*: They can be viewed as two representations of the same linear operator in two bases, with matrix P (obviously nonsingular since h is a diffeomorphism) specifying the change of basis. Two matrices that are similar have, accordingly, the same eigenvalue spectrum.

This shows that the multipliers of a periodic orbit do not depend on the coordinate system chosen to parameterize the states of a system, and hence that they are invariants of (smooth) conjugacy. Note, however, that they need not be preserved under a topological conjugacy.

2.6
Fully Developed Chaos in the Logistic Map

The first chaotic regime that we study in the logistic map is the one observed at the right end of the bifurcation diagram, namely at $a = 2$. At this point, the logistic map is surjective on the interval $I = [-2, 2]$: Every point $y \in I$ is the image of two different points, $x_1, x_2 \in I$. I is then an invariant set since $f(I) = I$.

It turns out that the dynamical behavior of a surjective logistic map can be analyzed in a particularly simple way by using a suitable change of coordinates, namely $x = 2\sin(\pi x'/4)$. This is a one-to-one transformation between I and itself, which is a diffeomorphism everywhere except at the endpoints $x = \pm 2$, where the inverse function $x'(x)$ is not differentiable. With the help of a few trigonometric identities,
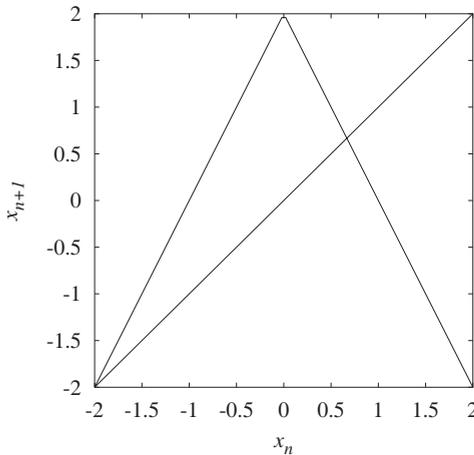
Figure 2.9 Graph of tent map (2.25).

the action of the logistic map in the x' space can be written as

$$x'_{n+1} = g(x'_n) = 2 - 2\left|x'_n\right| \tag{2.25}$$

a piecewise linear map known as the *tent map*.

Figure 2.9 shows that the graph of a tent map is extremely similar to that of a logistic map (Figure 2.1). In both cases, interval I is decomposed into two subintervals: $I = I_0 \cup I_1$, such that each restriction $f_k: I_k \to f(I_k)$ of f is a homeomorphism, with $f(I_0) = f(I_1) = I$. Moreover, f_0 (resp. f_1) is orientation-preserving (resp. orientation-reversing).

In fact, these topological properties suffice to determine the dynamics completely and are characteristic features of what is often called a *topological horseshoe*. In the remainder of Section 2.6, we review a few fundamental properties of chaotic behavior that can be shown to be direct consequences of these properties.

2.6.1
Iterates of the Tent Map

The advantage of the tent map over the logistic map is that calculations are simplified dramatically. In particular, higher-order iterates of the tent map, which are involved in the study of the asymptotic dynamics, are themselves piecewise-linear maps and are easy to compute. For illustration, the graphs of the second iterate g^2 and of the fourth iterate g^4 are shown in Figure 2.10. Their structure is seen to be directly related to that of the tent map.

Much of the structure of the g^n iterates can be understood from the fact that g maps linearly each of the two subintervals I_0 and I_1 to the whole interval I. Thus, the graph of the restriction of g^2 to each of the two components I_k reproduces the graph of g on I. This explains the two-"hump" structure of g^2. Similarly, the trivial
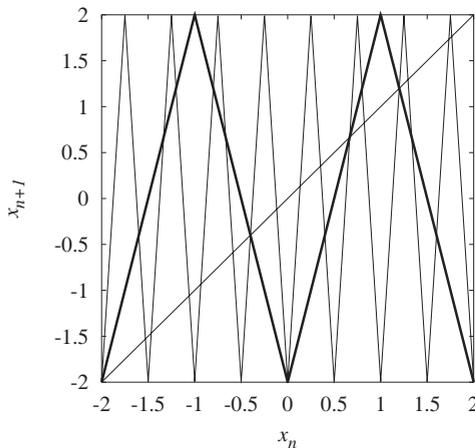
Figure 2.10 Graphs of the second (heavy line) and fourth (light line) iterates of the tent map (2.25).

relation

$$\forall x \in I_k \quad g^n(x) = g\left(g^{n-1}(x)\right) = g^{n-1}(g(x)) \tag{2.26}$$

shows that the graph of g^n consists of two copies of that of g^{n-1}. Indeed, (2.26) can be viewed as a semiconjugacy between g^n and g^{n-1} via the two-to-one transformation $x' = g(x)$.

By recursion, the graph of g^n shows 2^{n-1} scaled copies of the graph of g, each contained in a subinterval $I_k^n = [X_k - \epsilon_n, X_k + \epsilon_n]$ $(0 \leq k < 2^{n-1})$, where $\epsilon_n = 1/2^{n-2}$ and $X_k = -2 + (2k+1)\epsilon_n$. The expression of g^n can thus be obtained from that of g by

$$\forall x \in I_k^n = [X_k - \epsilon_n, X_k + \epsilon_n] \quad g^n(x) = g\left(\epsilon_n(x - X_k)\right) \tag{2.27}$$

An important consequence of (2.27) is that each subinterval I_k^n is mapped to the whole interval I in no more than n iterations of g:

$$\forall k = 0 \dots 2^{n-1} \quad g^n\left(I_k^n\right) = g(I) = I \tag{2.28}$$

More precisely, one has $g(I_k^n) = I_{k'}^{n-1}$, where $k' = k$ (resp. $k' = 2^{n-1} - k$) if $k < 2^{n-2}$ (resp. $k \geq 2^{n-2}$). Note also that each I_k^n can itself be split into two intervals $I_{k,i}^n$ on which g^n is monotonic and such that $g^n(I_{k,i}^n) = I$.

Because the diameter of I_k^n is $|I_k^n| = 2^{3-n}$ and can be made arbitrarily small if n is chosen sufficiently large, this implies that an arbitrary subinterval $J \subset I$, however small, contains at least one interval I_k^n:

$$\forall J \subset I \quad \exists N_0 \quad n > N_0 \Rightarrow \exists k \quad I_k^n \subset J \tag{2.29}$$

Thus, how the g^n iterates act on the I_k^n intervals can help us to understand how they act on an arbitrary interval, as we will see later. In general, chaotic dynamics is better characterized by studying how sets of points are globally mapped rather than by focusing on individual orbits.

2.6.2
Lyapunov Exponents

An important feature of tent map (2.25) is that the slope $|dg(x)/dx| = 2$ is constant on the whole interval $I = [-2, 2]$. This simplifies significantly the study of the stability of solutions of (2.25). From (2.6) an infinitesimal perturbation δx_0 from a reference state will grow after n iterations to $|\delta x_n| = 2^n |\delta x_0|$. Thus, any two distinct states, however close they may be, will eventually be separated by a macroscopic distance. This shows clearly that no periodic orbit can be stable (Section 2.4.1) and thus that the asymptotic motion of (2.25) is aperiodic.

This exponential divergence of neighboring trajectories, or *sensitivity to initial conditions*, can be characterized quantitatively by *Lyapunov exponents*, which correspond to the average separation rate. For a one-dimensional map, there is only one

Lyapunov exponent, defined by

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \frac{|\delta x_{n+1}|}{|\delta x_n|} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \left| \frac{d f}{d x} \left(f^i(x_0) \right) \right| \qquad (2.30)$$

which is a geometric average of the stretching rates experienced at each iteration. It can be shown that Lyapunov exponents are independent of the initial condition x_0, except perhaps for a set of measure zero [53].

Since the distance between infinitesimally close states grows exponentially as $\delta x_n \sim e^{n\lambda} \delta x_0$, sensitivity to initial conditions is associated with a strictly positive Lyapunov exponent. It is easy to see that the Lyapunov exponent of surjective tent map (2.25) is $\lambda = \ln 2$.

2.6.3
Sensitivity to Initial Conditions and Mixing

Sensitivity to initial conditions can also be expressed in a way that is more topological, without using distances. The key property we use here is that any subinterval $J \subset I$ is eventually mapped to the whole I:

$$\forall J \subset I \quad \exists N_0 \quad n > N_0 \Rightarrow g^n(J) = I \qquad (2.31)$$

This follows directly from the fact that J contains one of the basis intervals I_k^n and that these expand to I under the action of g; see (2.28) and (2.29).

We say that a map is *expansive* if it satisfies property (2.31). In plain words, the iterates of points in any subinterval can take every possible value in I after a sufficient number of iterations. Assume that J represents the uncertainty in the location of an initial condition x_0: We merely know that $x_0 \in J$, but not its precise position. Then (2.31) shows that chaotic dynamics is, though deterministic, asymptotically unpredictable: After a certain amount of time, the system can be anywhere in the state space. Note that the time after which all the information about the initial condition has been lost depends only logarithmically on the diameter $|J|$ of J. Roughly, (2.28) indicates that $N_0 \simeq -\ln|J|/\ln 2 \sim -\ln|J|/\lambda$.

In what follows, we use property (2.31) as a topological definition of chaos in one-dimensional noninvertible maps. To illustrate it, we recall the definitions of various properties that have been associated with chaotic behavior [12] and which can be shown to follow from (2.31).

A map $f: I \to I$:

- Has *sensitivity to initial conditions* if $\exists \delta > 0$ such that for all $x \in I$ and any interval $J \ni x$, there is a $y \in J$ and an $n > 0$ such that $|f^n(x) - f^n(y)| > \delta$.
- Is *topologically transitive* if for each pair of open sets $A, B \subset I$ there exists n such that $f^n(A) \cap B \neq \emptyset$.
- Is *mixing* if for each pair of open sets $A, B \subset I$ there exists $N_0 > 0$ such that $n > N_0 \Rightarrow f^n(A) \cap B \neq \emptyset$. A mixing map is obviously topologically transitive.

Sensitivity to initial conditions trivially follows from (2.31) since any neighborhood of $x \in I$ is eventually mapped to I. The mixing property, and hence transitivity, is also a consequence of expansiveness because the N_0 in the definition can be chosen so that $f^{N_0}(A) = I$ intersects any $B \subset I$. It can be shown that a topologically transitive map has at least a dense orbit (i.e., an orbit that passes arbitrarily close to any point of the invariant set).

Note that (2.31) precludes the existence of an invariant subinterval $J \subset I$ other than I itself: We would have simultaneously $f(J) = J$ and $f^{N_0}(J) = I$ for some N_0. Thus, invariant sets contained in I necessarily consist of isolated points; these are the periodic orbits discussed in the next section.

2.6.4
Chaos and Density of (Unstable) Periodic Orbits

It has been proposed by Devaney [12] to say that a map f is chaotic if it:

- Displays sensitivity to initial conditions
- Is topologically transitive
- Has a set of periodic orbits that is dense in the invariant set.

The first two properties were established in Section 2.6.3. It remains to be proved that (2.31) implies the third. When studying the bifurcation diagram of the logistic map (Section 2.4.2), we have noted that chaotic regimes contain many (unstable) periodic orbits. We are now in a position to make this observation more precise. We begin by showing that the tent map $x' = g(x)$ has infinitely many periodic orbits.

2.6.4.1 Number of Periodic Orbits of the Tent Map

A periodic orbit of g of period p is a fixed point of the pth iterate g^p. Thus, it satisfies $g^p(x) = x$ and is associated with an intersection of the graph of g^p with the diagonal. Since g itself has exactly two such intersections (corresponding to period-1 orbits), (2.27) shows that g^p has

$$N_f(p) = 2^p \tag{2.32}$$

fixed points (see Figure 2.10 for an illustration).

Some of these intersections might actually be orbits of lower period. For example, the four fixed points of g^2 consist of two period-1 orbits and of two points constituting a period-2 orbit. As another example, note in Figure 2.10 that fixed points of g^2 are also fixed points of g^4. The number of periodic orbits of lowest period p is thus

$$N(p) = \frac{N_f(p) - \sum_q q N(q)}{p} \tag{2.33}$$

where the q are the divisors of p. Note that this is a recursive definition of $N(p)$. As an example, $N(6) = [N_f(6) - 3N(3) - 2N(2) - N(1)]/6 = (2^6 - 3 \times 2 - 2 \times 1 - 2)/6 = 9,$

with the computation of $N(3)$, $N(2)$, and $N(1)$ being left to the reader. As detailed in Section 2.7.5.3, one of these nine orbits appears in a period doubling and the eight others are created by pairs in saddle–node bifurcations. Because $N_f(p)$ increases exponentially with p, $N(p)$ is well approximated for large p by $N(p) \simeq N_f(p)/p$.

We thus have the important property that there are an infinite number of periodic points and that the number $N(p)$ of periodic orbits of period p increases exponentially with the period. The corresponding growth rate,

$$h_P = \lim_{p \to \infty} \frac{1}{p} \ln N(p) = \lim_{p \to \infty} \frac{1}{p} \ln \frac{N_f(p)}{p} = \ln 2 \tag{2.34}$$

provides an accurate estimate of a central measure of chaos, the *topological entropy* h_T. In many cases it can be proven rigorously that $h_P = h_T$. Topological entropy itself can be defined in several different but equivalent ways.

2.6.4.2 Expansiveness Implies Infinitely Many Periodic Orbits

We now prove that if a continuous map $f : I \to I$ is expansive, *its unstable periodic orbits are dense in I*: Any point $x \in I$ has periodic points arbitrarily close to it. Equivalently, any subinterval $J \subset I$ contains periodic points.

We first note that if $J \subset f(J)$ (this is a particular case of a topological covering), then J contains a fixed point of f as a direct consequence of the intermediate value theorem.[2] Similarly, J contains at least one periodic point of period p if $J \subset f^p(J)$.

Now, if (2.31) is satisfied, every interval $J \subset I$ is eventually mapped to I: $f^n(J) = I$ (and thus $f^n(J) \subset J$) for $n > N_0(J)$. Using the remark above, we deduce that J contains periodic points of period p for any $p > N_0(J)$, but also possibly for smaller p. Therefore, *any interval contains an infinity of periodic points with arbitrarily high periods*. A graphical illustration is provided by Figure 2.10: Each intersection of a graph with the diagonal corresponds to a periodic point.

Thus, the expansiveness property (2.31) implies that unstable periodic points are dense. We showed earlier that it also implies topological transitivity and sensitivity to initial conditions. Therefore, any map satisfying (2.31) is chaotic according to the definition given at the beginning of this section.

It is quite fascinating that sensitivity to initial conditions, which makes the dynamics unpredictable, and unstable periodic orbits, which correspond to perfectly ordered motion, are so deeply linked: In a chaotic regime, order and disorder are intimately entangled.

Unstable periodic orbits will prove to be a powerful tool to analyze chaos. They form a skeleton around which the dynamics is organized. Although they can be characterized in a finite time, they provide invaluable information on the asymptotic dynamics because of the density property: The dynamics in the neighborhood of an unstable periodic orbit is governed largely by that orbit.

---

2) Denote by $x_a, x_b \in J$ the points such that $f(J) = [f(x_a), f(x_b)]$. If $J \subset f(J)$, one has $f(x_a) \le x_a$ and $x_b \le f(x_b)$. Thus, the function $F(x) = f(x) - x$ has opposite signs in x_a and x_b. If f is continuous, F must take all the values between $F(x_a)$ and $F(x_b)$. Thus, there exists $x_* \in J$ such that $F(x_*) = f(x_*) - x_* = 0$: x_* is a fixed point of f.

2.6.5
Symbolic Coding of Trajectories: First Approach

We showed above that because of sensitivity to initial conditions, the dynamics of the surjective tent map is asymptotically unpredictable (Section 2.6.3). However, we would like to have a better understanding of how irregular, or random, typical orbits can be. We also learned that there is a dense set of unstable periodic orbits embedded in the invariant set I and that this set has a well-defined structure. What about the other orbits, which are aperiodic?

In this section we introduce a powerful approach to chaotic dynamics that answers these questions: *symbolic dynamics*. To do so as simply as possible, let us consider a dynamical system extremely similar to the surjective tent map, defined by the map

$$x_{n+1} = 2x_n \quad (\text{mod } 1) \tag{2.35}$$

It only differs from the tent map in that the two branches of its graph are both orientation-preserving (Figure 2.11). As with the tent map, the interval $[0, 1]$ is decomposed in two subintervals I_k such that the restrictions $f_k \colon I_k \rightarrow f_k(I_k)$ are homeomorphisms.

The key step is to recognize that because the slope of the graph is 2 everywhere, the action of (2.35) is trivial if the coordinates $x \in [0, 1]$ are represented in base 2. Let x_n have the binary expansion $x_n = 0.d_0 d_1 \ldots d_k \ldots$, with $d_k \in \{0, 1\}$. It is easy to see that the next iterate will be

$$x_{n+1} = (d_0.d_1 d_2 \ldots d_k \ldots) \quad (\text{mod } 1) = 0.d_1 d_2 \ldots d_k \ldots \tag{2.36}$$

Thus, the base-2 expansion of x_{n+1} is obtained by dropping the leading digit in the expansion of x_n. This leading digit indicates whether x is greater than or equal to
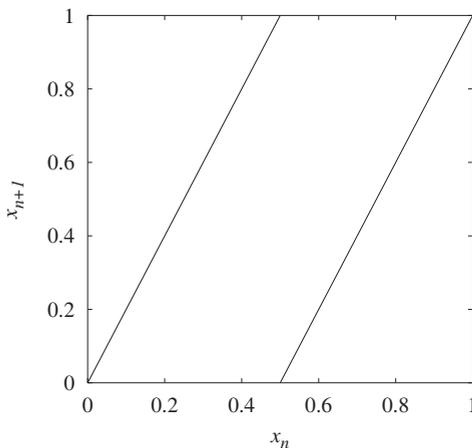
Figure 2.11 Graph of map (2.35).

$1/2 = 0.1\overline{0}$ ($\overline{s}$ represents an infinite repetition of the string s) and, hence, which interval $I_0 = [0, 0.5)$ or $I_1 = [0.5, 1]$ the point belongs to. Note that in the present case, $0.1\overline{0}$ and $0.0\overline{1}$, which usually represent the same number $1/2$, correspond here to different trajectories because of the discontinuity. The former is located at $(1/2)^+$ and remains on the fixed point $x = 1$ forever, while the latter is associated with $(1/2)^-$ and converges to the fixed point $x = 0$.

Thus, there is a $1 : 1$ correspondence between orbits of dynamical system (2.35) (parameterized by their initial condition x) and infinite digit sequences $(d_k) \in \{0, 1\}^\mathbb{N}$. Moreover, the action of the map in the latter space has a particularly simple form. This correspondence allows one to establish extremely easily all the properties derived for the tent map in previous sections.

- *Sensitivity to initial conditions:* Whether the nth iterate of x falls in I_0 or I_1 is determined by the nth digit of the binary expansion of x. A small error in the initial condition (e.g., the nth digit is false) becomes macroscopic after a sufficient amount of time (i.e., after n iterations).
- *Existence of a dense orbit:* Construct an infinite binary sequence such that it contains all possible finite sequences. For example, concatenate all sequences of length $1, 2, \dots, n$ for arbitrarily large n. The iterates of the associated point $x = 0.0|1|00|01|10|11|000|001 \dots$ will pass arbitrarily close to any point of the interval. The existence of a dense orbit implies topological transitivity.
- *Density of periodic orbits:* Each periodic point of (2.35) obviously corresponds to a periodic binary sequence. It is known that a periodic or eventually periodic digit expansion is a characteristic property of rational numbers. Since it is a classical result that rational numbers are dense in $[0, 1]$, we deduce immediately that periodic or eventually periodic points are dense in the interval $[0, 1]$. Alternatively, each point x can be approximated arbitrarily well by a sequence of periodic points $x_*(n)$ whose sequences consist of the infinitely repeated n first digits of x, with $n \to \infty$.

This analysis can easily be transposed to the case of the surjective tent map. Since its right branch is orientation-reversing, the action of this map on the binary expansion of a point x located in this branch differs slightly from that of (2.36). Assuming that the tent map is defined on $[0, 1]$, its expression at the right (resp. left) of the critical point is $x' = 2(1-x)$ (resp. $x' = 2x$). Consequently, we have the additional rule that if the leading digit is $d_0 = 1$, all the digits d_i, $i \in \mathbb{N}$, should be replaced by $\tilde{d}_i = 1 - d_i$ before dropping the leading digit d_0 as with the left branch (in fact, the two operations can be carried out in any order). The operation $d_i \to \tilde{d}_i$ is known as *complement to one*.

Example 2.1

Under the tent map, $0.010\,010\,11 \to 0.100\,101\,1 \to 0.110\,100$. For the first transition, since $0.010\,010\,0 < 1/2$, we shift the separator one digit to the right. In the

second transition, since $x = 0.100\,101\,1 > 1/2$, we first complement x and obtain $x' = (1 - x) = 0.011\,010\,0$, then multiply by 2: $2x' = 0.110\,110\,0$.

Except for this minor difference in the coding of trajectories, the arguments used above to show the existence of chaos in the map $x' = 2x$ (mod 1) can be followed without modification. The binary coding we have used is thus a powerful method to prove that the tent map displays chaotic behavior.

The results of this section naturally highlight two important properties of chaotic dynamics:

- A series of coarse-grained measurements of the state of a system can suffice to estimate it with arbitrary accuracy if carried out over a sufficiently long time. By merely noting which branch is visited (one-bit digitizer) at each iteration of the map (2.35), all the digits of an initial condition can be extracted.
- Although a system such as (2.35) is perfectly deterministic, its asymptotic dynamics is as random as coin flipping (all sequences of 0 and 1 can be observed).

However, the coding used in these two examples (*n*-ary expansion) is too naive to be extended to maps that do not have a constant slope equal to an integer. In the next section we discuss the general theory of symbolic dynamics for one-dimensional maps. This topological approach will prove to be an extremely powerful tool to characterize the dynamics of the logistic map, not only in the surjective case but for any value of parameter a.

2.7
One-Dimensional Symbolic Dynamics

2.7.1
Partitions

Consider a continuous map $f : I \rightarrow I$ that is singular. We would like to extend the symbolic dynamical approach introduced in Section 2.6.5 in order to analyze its dynamics. To this end, we have to construct a coding associating each orbit of the map with a symbol sequence.

We note that in the previous examples, each digit of the binary expansion of a point x indicates whether x belongs to the left or right branch of the map. Accordingly, we decompose interval I in N disjoint intervals I_α, $\alpha = 0 \ldots N-1$ (numbered from left to right), such that

- $I = I_0 \cup I_1 \cup \cdots I_{N-1}$
- In each interval I_α, the restriction $f|_{I_\alpha} : I_\alpha \rightarrow f(I_\alpha)$, which we denote f_α, is a homeomorphism.
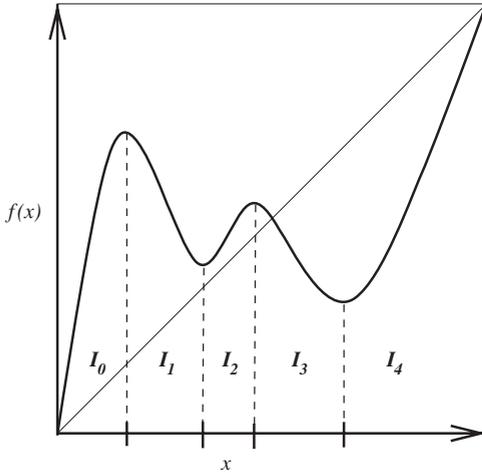
Figure 2.12 Decomposition of the domain of a map f into intervals I_α such that the restrictions $f : I_\alpha \to f(I_\alpha)$ to the intervals I_α are homeomorphisms.

For one-dimensional maps, such a *partition* can easily be constructed by choosing the critical points of the map as endpoints of the intervals I_α, as Figure 2.12 illustrates. At each iteration, we record the symbol $\alpha \in \mathcal{A} = \{0, \ldots, N-1\}$ that identifies the interval to which the current point belongs. The alphabet $\mathcal{A}$ consists of the N values that the symbol can assume.

We denote by $s(x)$ the corresponding coding function:

$$s(x) = \alpha \iff x \in I_\alpha \tag{2.37}$$

Any orbit $\{x, f(x), f^2(x), \ldots, f^i(x), \ldots\}$ of initial condition x can then be associated with the infinite sequence of symbols indicating the intervals visited successively by the orbit:

$$\Sigma(x) = \left\{ s(x), s(f(x)), s\left(f^2(x)\right), \ldots, s\left(f^i(x)\right), \ldots \right\} \tag{2.38}$$

The sequence $\Sigma(x)$ is called the *itinerary* of x. We will also use the compact notation $\Sigma = s_0 s_1 s_2 \ldots s_i \ldots$, with the s_i being the successive symbols of the sequence (e.g., $\Sigma = 01101001 \ldots$). The set of all possible sequences in the alphabet $\mathcal{A}$ is denoted $\mathcal{A}^{\mathbb{N}}$, and $\Sigma(I) \subset \mathcal{A}^{\mathbb{N}}$ represents the set of sequences actually associated with a point of I:

$$\Sigma(I) = \{\Sigma(x); x \in I\} \tag{2.39}$$

The finite sequence made of the n leading digits of $\Sigma(x)$ will later be useful. We denote it $\Sigma_n(x)$. For example, if $\Sigma(x) = 10110 \ldots$, then $\Sigma_3(x) = 101$. Accordingly, the set of finite sequences of length n involved in the dynamics is $\Sigma_n(I)$.

An important property of the symbolic representation (2.38) is that the expression of the time-one map becomes particularly simple. Indeed, if we compare the

symbolic sequence of $f(x)$

$$\Sigma(f(x)) = \left\{ s(f(x)), s(f^2(x)), s(f^3(x)), \ldots, s(f^{i+1}(x)), \ldots \right\} \quad (2.40)$$

with that of x given in (2.38), we observe that the former can be obtained from the latter by dropping the leading symbol and shifting the remaining symbols to the left. Accordingly, we define the *shift operator σ* by

$$\Sigma = \{s_0, s_1, s_2, \ldots, s_i, \ldots\} \xrightarrow{\sigma} \{s_1, s_2, \ldots, s_i, \ldots\} = \sigma\Sigma \quad (2.41)$$

Applying f to a point $x \in I$ is equivalent to applying the shift operator σ on its symbolic sequence $\Sigma(x) \in \Sigma(I)$:

$$\Sigma(f(x)) = \sigma\Sigma(x) \quad (2.42)$$

which corresponds to the commutative diagram

$$
\begin{array}{ccc}
x & \xrightarrow{\quad f \quad} & f(x) \\[4pt]
\Big\downarrow{\scriptstyle\Sigma} & & \Big\downarrow{\scriptstyle\Sigma} \\[4pt]
\{s_i\}_{i \in \mathbb{N}} & \xrightarrow{\quad \sigma \quad} & \{s_{i+1}\}_{i \in \mathbb{N}}
\end{array}
\quad (2.43)
$$

Note that because only forward orbits $\{f^n(x)\}_{n \geq 0}$ can be computed with a noninvertible map, the associated symbolic sequences are one-sided and extend to infinity only in the direction of forward time. This makes the operator σ noninvertible, as f is itself. Formally, we can define several "inverse" operators σ_α^{-1} acting on a sequence by inserting the symbol α at its head:

$$\Sigma = \{s_0, s_1, s_2, \ldots, s_i, \ldots\} \xrightarrow{\sigma_\alpha^{-1}} \{\alpha, s_0, s_1, \ldots, s_{i-1}, \ldots\} = \sigma_\alpha^{-1}\Sigma \quad (2.44)$$

However, note that $\sigma \circ \sigma_\alpha^{-1} = \mathrm{Id} \neq \sigma_\alpha^{-1} \circ \sigma$.

Periodic sequences $\Sigma = \{s_i\}$ with $s_i = s_{i+p}$ for all $i \in \mathbb{N}$ will be of particular importance in what follows. Indeed, they satisfy $\sigma^p\Sigma = \Sigma$, which translates into $f^p(x) = x$ for the associated point that is thus periodic. Infinite periodic sequences will be represented by overlining the base pattern (e.g., $\overline{01011} = 010110101101011\ldots$). When there is no ambiguity, the base pattern will be used as the name of the corresponding periodic orbit (e.g., the orbit 01011 has sequence $\overline{01011}$).

2.7.2
Symbolic Dynamics of Expansive Maps

To justify the relevance of symbolic coding, we now show that it is a faithful representation. That is to say, the correspondence $x \in I \leftrightarrow \Sigma(x) \in \Sigma(I)$ defined by (2.37) and (2.38) can under appropriate conditions be made a bijection, that is,

$$x_1 \neq x_2 \iff \Sigma(x_1) \neq \Sigma(x_2) \quad (2.45)$$

We might additionally require some form of continuity so that sequences that are close according to some metric are associated with points that are close in space.

In plain words, the symbolic sequence associated to a given point is sufficient to distinguish it from any other point in interval I. The two dynamical systems (I, f) and $(\Sigma(I), \sigma)$ can then be considered as equivalent, with $\Sigma(x)$ playing the role of a change of coordinate. Partitions of state space that satisfy (2.45) are said to be *generating*.

In Section 2.6.5, we saw two particular examples of one-to-one correspondence between orbits and symbolic sequences. Here we show that such a bijection holds if the following two conditions are true: (i) The restriction of the map to each member of the partition is a homeomorphism (Section 2.7.1) and (ii) the map satisfies the expansiveness property (2.31). This will illustrate the intimate connection between symbolic dynamics and chaotic behavior.

In the tent map example, it is obvious how the successive digits of the binary expansion of a point x specify the position of x with increasing accuracy. As we show below, this is also true for general symbolic sequences under appropriate conditions.

As a simple example, assume that a point x has a symbol sequence $\Sigma(x) = 101\dots$ From the leading symbol we extract the top-level information about the position of x, namely that $x \in I_1$. Since the second symbol is 0, we deduce that $f(x) \in I_0$ (i.e., $x \in f^{-1}(I_0)$). This second-level information combined with the first-level information indicates that $x \in I_1 \cap f^{-1}(I_0) \equiv I_{10}$. Using the first three symbols, we obtain $x \in I_{101} = I_1 \cap f^{-1}(I_0) \cap f^{-2}(I_1) = I_1 \cap f^{-1}(I_0 \cap f^{-1}(I_1))$. We note that longer symbol sequences localize the point with higher accuracy: $I_1 \supset I_{10} \supset I_{101}$.

More generally, define the interval $I_\Lambda = I_{s_0 s_1 \dots s_{n-1}}$ as the set of points whose symbolic sequence begins by the finite sequence $\Lambda = s_0 s_1 \dots s_{n-1}$, the remaining part of the sequence being arbitrary:

$$
\begin{aligned}
I_\Lambda = I_{s_0 s_1 \dots s_{n-1}} &= \{x; \Sigma_n(x) = s_0 s_1 \dots s_{n-1}\} \\
&= \{x; s(f^i(x)) = s_i, i < n\} \\
&= \{x; f^i(x) \in I_{s_i}, i < n\}
\end{aligned}
\tag{2.46}
$$

Such sets are usually termed *cylinders*, with an n-cylinder being defined by a sequence of length n. We now show that cylinders can be expressed simply using inverse branches of function f. We first define

$$
\forall J \subset I \quad f_\alpha^{-1}(J) = I_\alpha \cap f^{-1}(J)
\tag{2.47}
$$

This is a slight abuse of notation since we have only that $f_\alpha(f_\alpha^{-1}(J)) \subset J$ without the equality being always satisfied, but it makes the notation more compact. With this convention, the base intervals I_α can be written as

$$
I_\alpha = \{x; s(x) = \alpha\} = f_\alpha^{-1}(I) \quad \forall \alpha \in \mathcal{A}
\tag{2.48}
$$

To generate the whole set of cylinders, this expression can be generalized to longer sequences by noting that

$$I_{a\Lambda} = f_a^{-1}(I_\Lambda) \tag{2.49}$$

which follows directly from definitions (2.46) and (2.47). Alternatively, (2.49) can be seen merely to express that $a\Lambda = \sigma_a^{-1}\Lambda$. By applying (2.49) recursively, one obtains

$$I_\Lambda = f_\Lambda^{-n}(I) \tag{2.50}$$

where f_Λ^{-n} is defined by

$$f_\Lambda^{-n} = f_{s_0 s_1 \ldots s_{n-1}}^{-n} = f_{s_0}^{-1} \circ f_{s_1}^{-1} \circ \cdots \circ f_{s_{n-1}}^{-1} \tag{2.51}$$

Just as the restriction of f to any interval I_a is a homeomorphism $f_a : I_a \to f(I_a)$, the restriction of f^n to any set I_Λ with Λ of length n is a homeomorphism[3] $f_\Lambda^n : I_\Lambda \to f^n(I_\Lambda)$. The function f_Λ^{-n} defined by (2.51) is the inverse of this homeomorphism, which explains the notation. For a graphical illustration, see Figure 2.10: Each interval of monotonic behavior of the graph of g^2 (resp. g^4) corresponds to a different interval I_Λ, with Λ of length 2 (resp. 4).

Note that because I is connected and the f_a^{-1} are homeomorphisms, all the I_Λ are connected sets and, hence, are intervals in the one-dimensional context. This follows directly from (2.49) and the fact that the image of a connected set by a homeomorphism is a connected set. This property will be important in what follows.

To illustrate relation (2.50), we apply it to the case $\Lambda = 101$ considered in the example above:

$$\begin{aligned} I_{101} = (f_1^{-1} \circ f_0^{-1} \circ f_1^{-1})(I) &= (f_1^{-1} \circ f_0^{-1})(I_1) \\ &= f_1^{-1}(I_0 \cap f^{-1}(I_1)) \\ &= I_1 \cap f^{-1}(I_0 \cap f^{-1}(I_1)) \end{aligned}$$

and verify that it reproduces the expression obtained previously.

The discussion above shows that the set of n cylinders $\mathcal{C}_n = \{I_\Lambda; \Lambda \in \Sigma_n(I)\}$ is a partition of I:

$$I = \bigcup_{\Lambda \in \Sigma_n(I)} I_\Lambda \qquad I_\Lambda \cap I_{\Lambda'} = \emptyset \tag{2.52}$$

with $\mathcal{C}_n$ being a refinement of $\mathcal{C}_{n-1}$ (i.e., each member of $\mathcal{C}_n$ is a subset of a member of $\mathcal{C}_{n-1}$). As $n \to \infty$, the partition $\mathcal{C}_n$ becomes finer and finer (again, see Figure 2.10). What we want to show is that the partition is arbitrarily fine in this limit, with the size of *each* interval of the partition converging to zero.

---

3) Note that f^n is a homeomorphism only on set I_Λ defined by symbolic strings Λ of length $p \geq n$ (e.g., f^2 has a singularity in the middle of I_0).
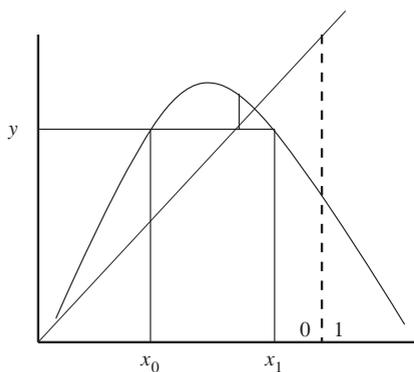
Figure 2.13 The dashed line indicates the border of a partition such that the preimages x_0 and x_1 of the same point y are coded with the same symbol (0). As a consequence, the symbolic sequences associated to x_0 and x_1 are identical.

Consider an arbitrary symbolic sequence $\Sigma(x)$, with $\Sigma_n(x)$ listing its n leading symbols. Since by definition $I_{\Sigma_{n+1}(x)} \subset I_{\Sigma_n(x)}$, the sequence $(I_{\Sigma_n(x)})_{n \in \mathbb{N}}$ is decreasing, hence it converges to a limit $I_{\Sigma(x)}$. All the points in $I_{\Sigma(x)}$ share the same infinite symbolic sequence.

As the limit of a sequence of connected sets, $I_{\Sigma(x)}$ is itself a connected set and, hence, an interval or an isolated point. Assume that $I_{\Sigma(x)}$ is an interval. Then, because of the expansiveness property (2.31), there is N_0 such that $f^{N_0}(I_{\Sigma(x)}) = I$. This implies that for points $x \in I_{\Sigma(x)}$, the symbol $s_{N_0} = s(f^{N_0})(x)$ can take any value $\alpha \in \mathcal{A}$, in direct contradiction of $I_{\Sigma(x)}$ corresponding to a unique sequence. Thus the only possible solution is that the limit $I_{\Sigma(x)}$ is an isolated point, showing that the correspondence between points and symbolic sequences is one to one. Therefore, the symbolic dynamical representation of the dynamics is faithful.

This demonstration assumes a partition of I constructed so that each interval of monotonicity corresponds to a different symbol (Figure 2.12). This guarantees that all the preimages of a given point will be associated with different symbols since they belong to different intervals.

It is easy to see that partitions not respecting this rule cannot be generating. Assume that two points x_0 and x_1 have the same image $f(x_0) = f(x_1) = y$ and that they are coded with the same symbol $s(x_0) = s(x_1) = \alpha_k$ (Figure 2.13). They are then necessarily associated with the same symbolic sequence, consisting of the common symbol α_k concatenated with the symbolic sequence of their common image: $\Sigma(x_0) = \Sigma(x_1) = \alpha_k \Sigma(y)$. In other words, associating x_0 and x_1 with different symbols is the only chance to distinguish them because they have exactly the same future.

In one-dimensional maps, two preimages of a given point are always separated by a critical point. Hence, the simplest generating partition is obtained by merely dividing the base interval I into intervals connecting two adjacent critical points and associating each with a different symbol.

- *Remark 1:* This is no longer true for higher-dimensional noninvertible maps, which introduces some ambiguity in the symbolic coding of trajectories.
- *Remark 2:* Invertible chaotic maps do not have singularities, hence the construction of generating partitions is more involved. The forthcoming examples of the horseshoe and of the Hénon map will help us to understand how the rules established in the present section can be generalized.

2.7.3
Grammar of Chaos: First Approach

Symbolic dynamics provides a simple but faithful representation of a chaotic dynamical system (Section 2.7.2). It has allowed us to understand the structure of the chaotic and periodic orbits of the surjective tent map, and hence of the surjective logistic map (Section 2.6.5). But there is more.

As a control parameter of a one-dimensional map varies, the structure of its invariant set and of its orbits changes (perestroika). Symbolic dynamics is a powerful tool to analyze these modifications: As orbits are created or destroyed, symbolic sequences appear or disappear from the associated symbolic dynamics. Thus, a regime can be characterized by a description of its set of forbidden sequences. We refer to such a description as the *grammar of chaos*. Changes in the structure of a map are characterized by changes in this grammar.

As we illustrate below with simple examples, which sequences are allowed and which are not can be determined entirely geometrically. In particular, the orbit of the critical point plays a crucial role. The complete theory, namely *kneading theory*, is detailed in Section 2.7.4.

2.7.3.1 Interval Arithmetics and Invariant Interval

We begin by determining the smallest invariant interval I (i.e., such that $f(I) = I$). This is where the asymptotic dynamics will take place. Let us first show how to compute the image of an arbitrary interval $J = [x_l, x_h]$. If J is located entirely to the left or right of the critical point x_c, one merely needs to take into account that the logistic map is orientation-preserving (resp. orientation-reversing) at the left (resp. right) of x_c. Conversely, if $x_c \in J$, then J can be decomposed as $[x_l, x_h] = [x_l, x_c) \cup [x_c, x_h]$. This gives

$$f([x_l, x_h]) = \begin{cases} [f(x_l), f(x_h)] & \text{if} \quad x_l, x_h \leq x_c \\ [f(x_h), f(x_l)] & \text{if} \quad x_c \leq x_l, x_h \\ [\min\{f(x_l), f(x_h)\}, f(x_c)] & \text{if} \quad x_l \leq x_c \leq x_h \end{cases} \tag{2.53}$$

As was noted by Poincaré, the apparent complexity of chaotic dynamics is such that it makes little sense to follow individual orbits; what is relevant is how regions of the state space are mapped between each other. One-dimensional maps are no exception, and in fact many properties of the logistic map can be extracted from

the interval arithmetics defined by (2.53). Here we use them to show in a simple way that some symbolic sequences are forbidden.

Let us now determine $I = [x_{\min}, x_{\max}]$ such that $f(I) = I$. We are interested only in situations where this interval contains the critical point, so that the dynamics is nontrivial. Note that this implies that $x_c \leq f(x_c)$ because we must have $x_c = f(\gamma) \leq f(x_c)$ (hence the top of the parabola must be above the diagonal). We use the third case of (2.53) to obtain the equation

$$\left[\min\{f(x_{\min}), f(x_{\max})\}, f(x_c)\right] = [x_{\min}, x_{\max}] \tag{2.54}$$

The upper bound is thus the image of the critical point: $x_{\max} = f(x_c)$. The lower bound $x_{\min}$ satisfies the equation

$$x_{\min} = \min\{f(x_{\min}), f(x_{\max})\} = \min\{f(x_{\min}), f^2(x_c)\} \tag{2.55}$$

An obvious solution is $x_{\min} = f^2(x_c)$, which is valid provided that $f(f^2(x_c)) > f^2(x_c)$. This is always the case between the parameter value where the period-1 orbit is superstable and the one where bounded solutions cease to exist. The other possible solution is the fixed point $x_- = f(x_-)$. In the parameter region of interest, however, one has $x_- < f^2(x_c)$, and thus the smallest invariant interval is given by

$$I = \left[f^2(x_c), f(x_c)\right] \tag{2.56}$$

That it depends only on the orbit of the critical point x_c is remarkable. However, this merely prefigures Section 2.7.4, where we shall see that this orbit determines the dynamics completely. Note that $I \neq \emptyset$ as soon as $f(x_c) > x_c$, which is the only interesting parameter region from a dynamical point of view.

2.7.3.2 Existence of Forbidden Sequences

As shown previously, the set I_Λ of points whose symbolic sequence begins by the finite string Λ is given by $I_\Lambda = f_\Lambda^{-n}(I)$, where f_Λ^{-n} is defined by (2.51) and (2.47). It is easy to see that if $I_\Lambda = \emptyset$, the finite symbol sequence Λ is forbidden.

From the discussion above, the base intervals are

$$I_0 = \left[f^2(x_c), x_c\right] \quad I_1 = \left[x_c, f(x_c)\right] \tag{2.57}$$

which are nonempty for $f(x_c) > x_c$. The existence of symbolic sequences of length two is determined by the intervals

$$I_{00} = I_0 \cap f_0^{-1}(I_0) = \left[f^2(x_c), f_0^{-1}(x_c)\right] \tag{2.58a}$$

$$I_{01} = I_0 \cap f_0^{-1}(I_1) = \left[f_0^{-1}(x_c), x_c\right] \tag{2.58b}$$

$$I_{10} = I_1 \cap f_1^{-1}(I_0) = \left[f_1^{-1}(x_c), f(x_c)\right] \tag{2.58c}$$

$$I_{11} = I_1 \cap f_1^{-1}(I_1) = \left[x_c, f_1^{-1}(x_c)\right] \tag{2.58d}$$
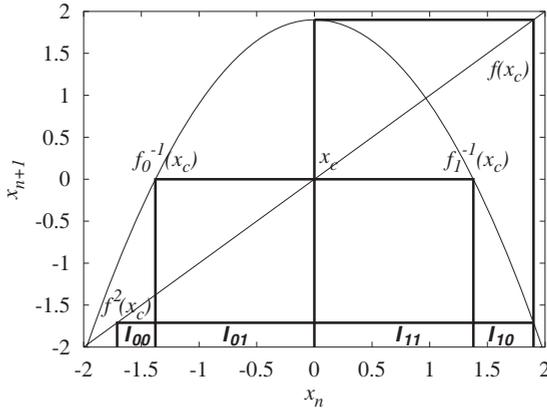
Figure 2.14 Intervals I_{00}, I_{01}, I_{11}, and I_{10} defined in (2.58). In each interval, the itineraries have the same leading two symbols.

which are computed by means of the interval arithmetics (2.53) but can also be obtained graphically (Figure 2.14). The last three do not provide useful information: They are nonempty whenever $f(x_c) > x_c$, i.e., as soon as I given by (2.56) is well defined.

By contrast, (2.58a) yields a nontrivial condition for I_{00} to be nonempty, namely that $f^2(x_c) < f_0^{-1}(x_c)$. This interval has zero width when its two bounds are equal; thus the string "00" becomes allowed when the critical point belongs to a period-3 orbit:

$$f^2(x_c) = f_0^{-1}(x_c) \Rightarrow f^3(x_c) = x_c \qquad (2.59)$$

which is then superstable since the derivative of f is zero at the critical point. For the logistic map (2.2), this occurs precisely at $a = a_{00} = 1.754\,877\,66\ldots$, inside the unique period-3 window that can be seen in the bifurcation diagram of Figure 2.3.

Since $I_{00} = \emptyset$ for $a < a_{00}$, we conclude that the symbolic string "00" never appears in the symbolic dynamics of regimes located at the left of the period-3 window. Thus the presence or absence of this string is sufficient to distinguish regimes located before and after this window.

In particular, this has consequences for the order of the appearance of periodic orbits. The first periodic orbit carrying the "00" string is the period-3 orbit $\overline{001}$. Therefore, all other periodic orbits whose names contain "00," for example $\overline{001011}$, must appear after $\overline{001}$. This shows that the geometrical structure of the map has a deep influence on the order of appearance of periodic orbits, as detailed later in a more systematic way.

Reproducing the calculation above for longer symbolic strings, we would find that new symbolic sequences always appeared when the critical point was part of a periodic orbit (i.e., at the parameter inside the periodic window where the orbit is superstable). This is not surprising if we note that the bounds of all the I_Λ intervals can be expressed in terms of the images and preimages of the critical point x_c.

As a result, the condition of zero width of these intervals can always be rewritten as an equation of the type $f_\Lambda^{-n}(x_c) = x_c$, expressing that x belongs to a periodic orbit of period n and of symbolic sequence Λ. For example, (2.59) corresponds to $f_{100}^{-3}(x_c) = x_c$. However, we will not proceed in this direction. The observation that the grammar of the symbolic dynamics is governed completely by the orbit of the critical point will lead us to a much more efficient framework for classifying symbolic sequences of orbits.

We conclude this section with the important remark that the symbolic dynamics of a chaotic dynamical system is in general *intimately related to its geometrical structure.* In the case of unimodal maps, the structure of the forbidden sequences depends only on the position of the image of the critical point organizing the dynamics. Thus, given an arbitrary symbolic sequence, it is in principle possible to determine whether it has been generated by a one-dimensional map. More generally, extracting the structure of a map from the grammar of the symbolic dynamics it generates is a fascinating problem. It has been much less explored for two-dimensional invertible maps than for maps of the interval, and even less for noninvertible maps of dimension 2 and higher.

2.7.4
Kneading Theory

Rather than solve algebraic equations such as (2.59) to determine forbidden sequences, it would be preferable to work completely in the space of symbolic sequences. Since the orbit of the critical point plays a crucial role in understanding which symbolic sequences are forbidden, it is natural to study more closely the distinguished symbolic sequence associated with the critical point.

Since the first symbol of this sequence does not carry any information (the critical point x_c is the border between intervals I_0 and I_1), we accordingly define the *kneading sequence* $K(f)$ as the itinerary of the image of x_c:

$$K(f) = \Sigma\left(f(x_c)\right) = \left\{s\left(f(x_c)\right), s\left(f^2(x_c)\right), \dots\right\} \tag{2.60}$$

Note that the first two symbols are constant inside the parameter region where $f^2(x_c) < x_c < f(x_c)$: $f^2(x_c)$ and $f(x_c)$ are the left and right ends of the invariant interval I defined in (2.56) and are thus associated with the symbols 0 and 1, respectively. Since the value of the third symbol depends on whether $f^3(x_c)$ is located to the left or right of the critical point, it changes when $f^3(x_c) = x_c$ (i.e., when the string "00" becomes allowed), and thus

$$a < a_{00} \Rightarrow K(f) = \{1, 0, 1, \dots\}$$
$$a > a_{00} \Rightarrow K(f) = \{1, 0, 0, \dots\} \tag{2.61}$$

This confirms the importance of the kneading sequence (2.60): The appearance of the symbolic string "00" in the symbolic dynamics of the logistic map coincides with its appearance in the kneading sequence.

To go beyond this observation, we need to be able to determine from $K(f)$ alone which sequences are allowed and which are not. The distinctive property of $f(x_c)$ is that it is the rightmost point of the invariant interval (2.56). To see that there is indeed a similar property for the kneading sequence, we first show that an order on itineraries can be defined.

2.7.4.1 **Ordering of Itineraries**

In the example of the $x_{n+1} = 2x_n \pmod 1$ map (Section 2.6.5), the itinerary of a point (i.e., its binary expansion) not only identifies it uniquely but also contains information about its position relative to the other points. In that case, the lexicographic order on symbolic sequences reflects exactly the order of the associated points on the interval. More generally, we would like to define for an arbitrary map an order relation $\prec$ on itineraries so that

$$\Sigma(x) \prec \Sigma(x') \Longleftrightarrow x < x' \tag{2.62}$$

Ordering two itineraries is easy when their leading symbols differ. If the base intervals I_α are numbered sequentially from left to right as in Figure 2.12, the itinerary with the smallest leading symbol is associated with the leftmost point and should be considered "smaller" than the other.

If the two itineraries have a common leading substring, one has to take into account the fact that map f can be orientation-reversing on some I_α intervals. For example, the two-symbol cylinders $I_{\alpha\alpha'}$ given by (2.58) and shown in Figure 2.14 appear left to right in the order I_{00}, I_{01}, I_{11}, and I_{10}.

Thus, $11 \prec 10$ for the logistic map, which differs markedly from the lexicographic order. This is because both strings have a leading 1, which is associated with the orientation-reversing branch f_1. Indeed, assume that $x_{11} \in I_{11}$, $x_{10} \in I_{10}$. From the second symbol we know that $f(x_{10}) < f(x_{11})$ because $0 < 1$. However, since f is orientation-reversing in I_1, this implies that $x_{11} < x_{10}$, hence $11 \prec 10$. With this point in mind, two arbitrary itineraries Σ, Σ' can be ordered as follows.

Assume that the two sequences $\Sigma = \Lambda s_m \ldots$ and $\Sigma' = \Lambda s'_m \ldots$ have a common leading symbolic string Λ of length m and first differ in symbols s_m and s'_m. Thus the corresponding points x and x' are such that $f^m(x)$ and $f^m(x')$ belong to different I_α intervals and, hence, can be ordered. As in the example above, it then suffices to determine whether the restriction f^m_Λ of f^m to the interval I_Λ is orientation-preserving or orientation-reversing (has a positive or a negative slope, respectively) to obtain the ordering of x and x', and thus that of Σ and Σ'. Define the branch parity

$$\epsilon(\alpha) = \begin{cases} +1 & \text{if} \quad f_\alpha \colon I_\alpha \to I \quad \text{is orientation-preserving} \\ -1 & \text{if} \quad f_\alpha \colon I_\alpha \to I \quad \text{is orientation-reversing} \end{cases} \tag{2.63}$$

The parity of the finite sequence $\Lambda = s_0 s_1 \ldots s_{m-1}$ is then given by

$$\epsilon(\Lambda) = \epsilon(s_0) \times \epsilon(s_1) \times \cdots \times \epsilon(s_{m-1}) \tag{2.64}$$
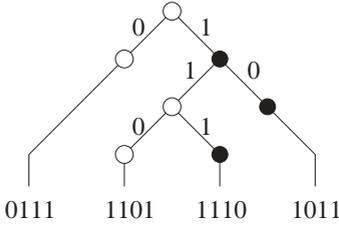
Figure 2.15 Determination of the relative order of symbolic sequences. White (resp. black) nodes correspond to positive (resp. negative) parity. The topmost node corresponds to the empty sequence, and sequences are formed by following edges carrying the symbols 0 or 1. When an edge "1" is followed, the parity of the node changes. A white node has an edge "0" on its left and an edge "1" on its right (this is the lexicographic order). At a black node, these two edges are in the opposite order because of the negative parity. To order a set of symbolic sequences, one follows the edges corresponding to the successive symbols of the sequence until no other sequence remains in the branch. The ordered sequences can then be read from left to right.

If the map $f_\Lambda^m = f_{s_{m-1}} \circ \ldots f_{s_1} \circ f_{s_0}$ (i.e., the restriction of f^m to the interval I_Λ) is orientation-preserving (resp. orientation-reversing), then $\epsilon(\Lambda) = +1$ (resp. -1). In the case of unimodal maps, $\epsilon(\Lambda) = +1$ if there is an even number of "1" (or of the symbol associated with the orientation-reversing branch), and -1 otherwise.

We can now define the order

$$\Sigma = \Lambda s \ldots \prec \Sigma' = \Lambda s' \ldots \Longleftrightarrow \begin{cases} s < s' \quad \text{and} \quad \epsilon(\Lambda) = +1 \\ \qquad\quad \text{or} \\ s > s' \quad \text{and} \quad \epsilon(\Lambda) = -1 \end{cases} \tag{2.65}$$

This order satisfies condition (2.62). Let us illustrate these rules with the example of period-4 orbit $\overline{0111}$ of the logistic map. The relative order of the four periodic orbits is

$$\overline{0111} \prec \overline{1101} \prec \overline{1110} \prec \overline{1011} \tag{2.66}$$

as detailed in Figure 2.15.

Another common technique for ordering symbolic sequences is to use *invariant coordinates*. Given a sequence $\Sigma = s_0 s_1 s_2 \ldots s_k \ldots \in \{0, \ldots, N-1\}^\mathbb{N}$, we define its invariant coordinate $\theta(\Sigma)$ by

$$\theta(\Sigma) = \sum_{i=0}^{\infty} \frac{t_i}{N^{i+1}} \quad t_i = \begin{cases} s_i & \text{if} \quad \epsilon(s_0 \ldots s_{i-1}) = +1 \\ (N-1) - s_i & \text{if} \quad \epsilon(s_0 \ldots s_{i-1}) = -1 \end{cases} \tag{2.67}$$

so that $0 \leq \theta(\Sigma) \leq 1$. By inspecting (2.65) and (2.67), one easily verifies that two symbolic sequences can be ordered by comparing their invariant coordinates:

$$\Sigma \prec \Sigma' \Longleftrightarrow \theta(\Sigma) < \theta(\Sigma') \tag{2.68}$$

As an example, the invariant coordinate of the periodic point $\overline{1011}$ of the logistic map is

$$
\theta\left(\overline{1011}\right) = \left(\frac{1}{2^1} + \frac{\mathbf{1}}{2^2} + \frac{\mathbf{0}}{2^3} + \frac{1}{2^4} + \frac{\mathbf{0}}{2^5} + \frac{0}{2^6} + \frac{1}{2^7} + \frac{\mathbf{0}}{2^8}\right)
$$
$$
\times \left(1 + \frac{1}{2^8} + \frac{1}{2^{16}} + \cdots\right) = \frac{105}{128} \times \frac{256}{255} = \frac{14}{17} \tag{2.69}
$$

where the digits in bold are those that have been inverted with respect to the original sequence. Because $\overline{1011}$ has negative parity, the binary digit sequence of $\theta\left(\Sigma\right)$ has period 8 instead of 4. The first factor in (2.69) corresponds to the basic pattern "11010010," and the second term comes from the infinite repetition of this pattern. Note that the fraction obtained is the position of the corresponding periodic point of the tent map defined on $[0, 1]$. The reader may verify as an exercise that $2 \times |1 - \theta\left(\overline{1011}\right)| = \theta\left(\overline{0111}\right)$ and that $2 \times \theta\left(\overline{0111}\right) = \theta\left(\overline{1110}\right)$.

2.7.4.2 Admissible Sequences

We showed earlier that each point x inside the invariant interval (2.56) satisfies $f^2(x_c) < x < f(x_c)$. Using (2.62), we can now translate this ordering relation between points into a ordering relation between symbolic sequences:

$$
\forall x \in I \quad \sigma K(f) \prec \Sigma(x) \prec K(f) \tag{2.70}
$$

since $K(f) = \Sigma(f(x_c))$, by definition. Moreover, the orbit of a point $x \in I$ is forever contained in I, by definition. A necessary condition for a sequence Σ to be the itinerary $\Sigma(x)$ of a point $x \in I$ is thus that (2.70) holds for any $\Sigma(f^n(x))$ and thus that

$$
\forall n \geq 0 \quad \sigma^n \Sigma \preceq K(f) \tag{2.71}
$$

One of the fundamental results of one-dimensional symbolic dynamics is that this is also a sufficient condition: Condition (2.71) completely determines whether a sequence occurs as the itinerary of a point [12, 52]. A sequence satisfying it is said to be *admissible* (equivalently, one can test whether $\sigma K(f) < \sigma^n \Sigma$ for all n).

Therefore, all the information about the symbolic dynamics of a map is contained in its kneading sequence $K(f)$. As a matter of fact, it can be shown that if two unimodal maps have the same kneading sequence, and that if this sequence is dense (i.e., the orbit of x_c is aperiodic), then the two maps are topologically conjugate.

Condition (2.71) is particularly simple to test when the symbolic sequence Σ is periodic since the shifts $\sigma^n \Sigma$ are finite in number. For example, let us assume that $K(f) = 1001001\ldots$ and that we want to know whether the periodic sequences $\overline{01101101}$ and $\overline{00101}$ are admissible. We first determine the rightmost periodic points (for which $\sigma^n \Sigma$ is maximal) of the two orbits: These are $\overline{10110110}$ and $\overline{10010}$. We then compare them to the kneading sequence $K(f)$ and find that

$$
\overline{10110110} \prec K(f) = 1001001\ldots \prec \overline{10010}
$$

Thus the period-8 sequence $\overline{10110110}$ is admissible, whereas the period-5 sequence $\overline{10010}$ is not. This indicates that the periodic orbit associated with the latter sequence does not exist in maps with the given $K(f)$. We also see that every map that has the second periodic orbit also has the first. Therefore, the order of appearance of periodic orbits is fixed, and the structure of the bifurcation diagram of Figure 2.3 is universal for unimodal maps. We investigate this universality in the next section.

2.7.5
Bifurcation Diagram of the Logistic Map Revisited

We are now in a position to understand the structure of the bifurcation diagram shown in Figure 2.3 using the tools of symbolic dynamics introduced in the previous sections. This bifurcation diagram displays two types of bifurcations: saddle–node and period-doubling bifurcations. Each saddle–node bifurcation creates a pair of periodic orbits of period p, one unstable (the saddle) and the other stable (the node). The latter is the germ of a period-doubling cascade with orbits of periods $p \times 2^n$.

As discussed in Section 2.7.4, the kneading sequence governs which symbolic sequences are admissible and which are forbidden, hence the order in which new sequences appear. Therefore, there must be a simple relation between the symbolic names of the orbits involved in a saddle–node or in a period-doubling bifurcation. Moreover, the different saddle–node bifurcations and their associated period-doubling cascades must be organized rigidly.

2.7.5.1 Saddle–Node Bifurcations
At a saddle–node bifurcation, the two newly born periodic orbits of period p are indistinguishable. Thus, they have formally the same symbolic name. This is not in contradiction with the one-to-one correspondence between orbits and itineraries that was shown to hold for chaotic regimes: At the bifurcation, the node is stable and there is no sensitivity to initial conditions.

When the two orbits are born, they have a multiplier of $+1$ (Section 2.4.1), which implies that f^p is orientation-preserving in the neighborhood of the orbit. Consequently, the common symbolic itinerary of the two orbits must contain an even number of symbols "1" (in the general case, an even number of symbols with negative parity).

A symbolic itinerary can change only if one of the periodic points crosses the critical point, which is the border of the partition. This happens to the stable node when it becomes superstable, changing its parity on its way to the period-doubling bifurcation where its multiplier crosses -1. Thus, its parity must be negative, and its final symbolic name (i.e., the one in the unstable regime) must differ from that of the saddle by a single symbol.

One can proceed as follows to see which symbol differs. Each periodic point is associated with a cyclic permutation of the symbolic name. For example, the orbit
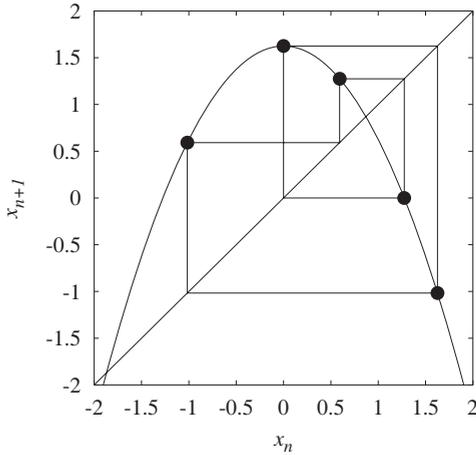
Figure 2.16 The 01111 orbit becomes 01011 at a superstable point.

$\overline{01111}$ has periodic points with sequences $\overline{01111}$, $\overline{11110}$, $\overline{11101}$, $\overline{11011}$, and $\overline{10111}$. These periodic points can be ordered using the kneading order (2.65), here

$$\overline{01111} \prec \overline{11110} \prec \overline{11011} \prec \overline{11101} \prec \overline{10111}$$

The symbol that is flipped at the superstable parameter value is obviously associated with the point that is then degenerate with the critical point. The image of this point is thus the rightmost periodic point, corresponding to the highest sequence in the kneading order (Figure 2.16). Consequently, a simple rule to obtain the symbolic name of the saddle–node partner of an orbit of given name is to flip the last symbol of the highest itinerary (i.e., of the itinerary of the rightmost point). Alternatively, one can flip the second to last symbol of the leftmost itinerary.

In the example above, the saddle–node partner of the 01111 orbit is thus 01101. Other examples of saddle–node pairs include $0_1^0 1$, $00_1^0 1$, and $01101_1^0 1$.

2.7.5.2 Period-Doubling Bifurcations

When applying the algorithm above on an arbitrary symbolic name, it can occur that the result is not valid because it is the repetition of a shorter name. For example, 0111 leads to $0101 = (01)^2$. This indicates that the long name corresponds to the period-doubled orbit (the daughter) of the orbit identified by the short name (the mother). The 0111 orbit is the daughter of the 01 orbit. The latter is itself the daughter of the 1 orbit.

This can be understood with the same arguments as for the saddle–node bifurcations. When the period-doubled orbit is born, its itinerary is a double copy of that of the mother. Its parity is thus positive, which is consistent with the fact that the orbit is born with a multiplier of $+1$. The symbolic name of the mother at the bifurcation is its final name. As the daughter orbit proceeds to its own period doubling (and thus to its unstable domain), the preimage of its rightmost point crosses the critical point, changing the associated symbol.

Hence we have a simple way to determine whether an orbit belongs to a period-doubling cascade and what the names of its mother and all its ancestors are. Other examples of mother–daughter pairs are (001, 001011) and (00101, 0010100111). We conclude with the period-doubling cascade originating from the period-1 orbit. The symbolic names of the successive period-doubled orbit can be constructed as

$$1 \xrightarrow{D} 11 \xrightarrow{F} \mathbf{01} \xrightarrow{D} 0101 \xrightarrow{F} \mathbf{0111} \xrightarrow{D} 01110111 \xrightarrow{F} \mathbf{01110101} \cdots \quad (2.72)$$

where D and F represent the action of doubling the word and flipping the second to last symbol, respectively.

2.7.5.3 Universal Sequence

Consider two periodic itineraries $\Sigma \prec \Sigma'$. For some parameter a, the kneading sequence $K(f_a)$ is such that $\Sigma \prec K(f_a) \prec \Sigma'$, so that Σ satisfies the admissibility condition (2.71) but not Σ'. Thus, the periodic orbit associated to Σ must be created before the one associated to Σ'.

This observation suffices to build a complete list of the successive bifurcations occurring in the bifurcation diagram of Figure 2.3. Using the rules derived in previous sections, we can classify all the symbolic names according to which series of bifurcations they belong to.

To this end, all periodic itineraries up to a given period p are sorted according to the kneading order, with saddle–node pairs and orbits of the same period-doubling

Table 2.1 Sequence of bifurcations in the logistic map up to period 8 (from top and to bottom and left to right). The notation P_i refers to the ith bifurcation of period P. We also give inside brackets an alternative classification that distinguishes between saddle– node and period-doubling bifurcations. In this scheme, the ith saddle–node bifurcation of period P is denoted s^i_P, and $s^i_P \times 2^k$ is the orbit of period $P \times 2^k$ belonging to the period-doubling cascade originating from s^i_P.

Name	Bifurcation	Name	Bifurcation	Name	Bifurcation
$^0_1$	$1_1[s_1]$	$00101^0_1 1$	$7_3[s^3_7]$	$0001^0_1 1$	$6_4[s^3_6]$
01	$2_1[s_1 \times 2^1]$	$001010^0_1 1$	$8_5[s^4_8]$	$000111^0_1 1$	$8_{11}[s^9_8]$
0111	$4_1[s_1 \times 2^2]$	$001^0_1 1$	$5_2[s^2_5]$	$00011^0_1 1$	$7_7[s^7_7]$
01010111	$8_1[s_1 \times 2^3]$	$001110^0_1 1$	$8_6[s^5_8]$	$000110^0_1 1$	$8_{12}[s^{10}_8]$
$0111^0_1 1$	$6_1[s^1_6]$	$00111^0_1 1$	$7_4[s^4_7]$	$000^0_1 1$	$5_3[s^3_5]$
$011111^0_1 1$	$8_2[s^1_8]$	$001111^0_1 1$	$8_7[s^6_8]$	$000010^0_1 1$	$8_{13}[s^{11}_8]$
$01111^0_1 1$	$7_1[s^1_7]$	$0011^0_1 1$	$6_3[s^2_6]$	$00001^0_1 1$	$7_8[s^8_7]$
$011^0_1 1$	$5_1[s^1_5]$	$001101^0_1 1$	$8_8[s^7_8]$	$000011^0_1 1$	$8_{14}[s^{12}_8]$
$01101^0_1 1$	$7_2[s^2_7]$	$00110^0_1 1$	$7_5[s^5_7]$	$0000^0_1 1$	$6_5[s^4_6]$
$011011^0_1 1$	$8_3[s^2_8]$	$00^0_1 1$	$4_2[s^1_4]$	$000001^0_1 1$	$8_{15}[s^{13}_8]$
$0^0_1 1$	$3_1[s_3]$	00010011	$8_9[s^1_4 \times 2^1]$	$00000^0_1 1$	$7_9[s^9_7]$
001011	$6_2[s_3 \times 2^1]$	$00010^0_1 1$	$7_6[s^6_7]$	$000000^0_1 1$	$8_{16}[s^{14}_8]$
$001011^0_1 1$	$8_4[s^3_8]$	$000101^0_1 1$	$8_{10}[s^8_8]$		

cascade grouped together. We denote the ith bifurcation creating period-P orbits as P_i, with the node being called $P_i f$ (for flip) and the saddle $P_i r$ (for regular). This is illustrated in Table 2.1, which lists the symbolic names of all periodic orbits of period up to 8 of the logistic map. These names are sorted by order of appearance, and the bifurcation in which they appear is indicated.

This sequence of symbolic names, often referred to as the *universal sequence*, was discovered by Metropolis, Stein, and Stein [5]. It is universal in that it depends only on the kneading order (2.65): The bifurcation diagram of any unimodal map will display exactly the same bifurcations in exactly the same order.

Note, however, that this holds only for one-dimensional maps. If a two-dimensional map is sufficiently dissipative so that its return map can be well approximated by a one-dimensional map, most of the bifurcation sequences will occur in the order predicted by the universal sequence. However, there will be a few discrepancies, and the order of many bifurcations will be reversed as one decreases dissipation [54].

2.7.5.4 Self-Similar Structure of the Bifurcation Diagram

In this section we mention briefly another surprising property of the bifurcation diagram of the logistic map that is unveiled by symbolic dynamics. Look at the period-3 window beginning at $a = 1.75$ in Figure 2.4. There is a whole parameter range where the attractor is contained in three disconnected pieces, before it expands suddenly. These pieces are visited successively in a fixed order. We call this parameter region the generalized period-3 window. Look more closely at, say, the middle branch: This is a complete copy of the whole bifurcation diagram! In particular, there is a period-9 window that is to the period-3 window what the period-3 is itself to the whole diagram.

To understand this, we note that the base symbols 0 and 1 can be viewed as the names of the period-1 orbits organizing the global dynamics. Similarly, let us denote by $X = 101$ and $Y = 100$ the symbolic names of the two period-3 orbits born in the saddle–node bifurcation initiating the period-3 window. All periodic orbits appearing in the generalized period-3 window can be written as words in the letters X and Y.

Indeed, since the attractor is split into three pieces visited successively, the dynamics can be simplified by considering the third iterate f^3. Each of the three pieces is a different attractor of f^3. The return map for each attractor is a unimodal map, with two "period-1" orbits that are in fact periodic points of the two period-3 orbits $\overline{100}$ and $\overline{101}$. Any pair of symbols X' and Y', such that the sequences $\overline{X'}$ and $\overline{Y'}$ correspond to periodic points that are degenerate at the period-3 saddle–node bifurcation, can thus be used to code orbits of this map. Because we chose X and Y above to be higher in the kneading order than all their cyclic permutations, they satisfy this condition as well as any pair $\sigma^k X, \sigma^k Y$.

Since the two words $X = 101$ and $Y = 100$ are such that (i) $\overline{X} \prec \overline{Y}$ and (ii) they have parities $\epsilon(X) = +1$ and $\epsilon(Y) = -1$, it is easy to see that the ordering of two sequences $W_1(X, Y)$ and $W_2(X, Y)$ will be exactly the same as for the corresponding

sequences $W_1(0, 1)$ and $W_2(0, 1)$. For example,

$$\overline{Y X Y Y X Y Y X} \prec \overline{Y X X Y X} \iff \overline{10110110} \prec \overline{10010}$$

This explains why the bifurcation diagram in the generalized period-3 window has exactly the same structure as the whole diagram. Using the names of the standard period-doubling cascade given in (2.72), we find that the orbits involved in the period-doubling cascade of this window are Y, XY, $XYYY$, $XYYYXYXY$. The first orbits to appear in the window are the X and Y orbits (naming them after their sequences in the unstable regime); the last is the YX^∞ orbit.

In fact, the results of this section could have been foreseen: They are a consequence of the qualitative universality of bifurcations in unimodal maps. Inside the period-3 window, the third return map is a unimodal map and therefore displays the same series of bifurcations as the first return map.

2.8
Shift Dynamical Systems, Markov Partitions, and Entropy

In Section 2.7, we saw how a chaotic system can be analyzed with the tools of symbolic dynamics. In particular, each regime of the logistic map is characterized by a different grammar (i.e., a set of forbidden symbolic sequences). Moreover, symbolic dynamics can be shown in some cases to provide a complete description of a dynamical system; for example, it is known that chaotic unimodal maps are conjugate if they have the same kneading sequence.

It is thus natural to study systems whose evolution laws are defined directly in a symbolic space by rules specifying which sequences are admissible. Such systems are usually referred to as *symbolic dynamical systems*, or as *shift dynamical systems* when they are based on the shift map [55]. Tools developed to characterize these systems can then be applied to any physical system for which a symbolic dynamical description has been obtained. This is illustrated by computations of entropy, an important measure of chaotic dynamics.

Here we limit ourselves to *shifts of finite type*, which are characterized by a finite set of forbidden sequences. The interest of finite shifts is twofold. First, there are dynamical systems, those for which a *Markov partition* exists, that can be shown to be equivalent to a finite shift. Second, systems whose grammar cannot be specified by a finite set of rules can always be approximated with increasing accuracy by a sequence of finite shifts of increasing order.

2.8.1
Shifts of Finite Type and Topological Markov Chains

The natural phase space of a symbolic dynamical system is the set of infinite or bi-infinite sequences of symbols from an alphabet $\mathcal{A}$. Here we assume that the alphabet is finite and choose $\mathcal{A} = \{0, \ldots, N-1\}$, where N is the number of symbols.

The systems we consider here share the same time-one map: the shift operator σ, which shifts symbols one place to the left (Section 2.7.1).

In the case of the logistic map, the symbolic space consisted of one-sided symbolic sequences. We noted in Section 2.7.1 that this makes σ noninvertible, since memory of the leading symbol is lost after each time step. If the shift operator has to be invertible, its action must not discard information. Thus sequences must be bi-infinite (two-sided), for example,

$$\Sigma = \dots s_{-3}s_{-2}s_{-1}.s_0s_1s_2 \tag{2.73}$$

with the dot separating the *forward sequence* $\Sigma_+ = s_0s_1\dots$ from the *backward sequence* $\Sigma_- = s_{-1}s_{-2}\dots$ These two sequences describe the future and the past of the point, respectively. The action of the shift operator on a sequence is then given by

$$\sigma(\dots s_{-1}.s_0s_1\dots) = \dots s_{-1}s_0.s_1\dots$$

The dot is merely moved to the right, which obviously preserves the information contained in the sequence. This is illustrated with the horseshoe map in Section 2.10.

The distinction between invertible and noninvertible dynamics is not made by most methods developed for characterizing symbolic dynamical systems. As we see below, they usually involve determining which finite blocks of symbols can appear in a typical sequence and which cannot. Thus, whether sequences are one- or two-sided is not relevant.

Full shifts are the simplest shift symbolic dynamical systems: Any sequence made of letters of the alphabet is allowed. Thus the symbolic space is $\mathcal{A}^{\mathbb{N}}$ or $\mathcal{A}^{\mathbb{Z}}$, depending on the invertibility of the dynamics. Two full shifts can differ only by the number of symbols of their alphabets. A full shift on r symbols is termed an r-shift.

In a general shift dynamical system, not all sequences are allowed: It is then called a *subshift*. We are interested only in subshifts whose set of allowed sequences $\mathcal{S}$ is shift-invariant (i.e., $\sigma\mathcal{S} = \mathcal{S}$). This implies that whether a finite symbol string can be found in a sequence $\Sigma \in \mathcal{S}$ depends not on its position in the sequence but only on the content of the string. Finite symbol strings are also often referred to as *blocks*, with an n-block containing n symbols, or as *words*.

One could therefore provide a complete description of a dynamical system $(\mathcal{S}, \sigma)$ by specifying its *language* (i.e., the list of all finite strings of symbols that can be extracted from infinite sequences). It is usually much more convenient to specify its set $\mathcal{F}$ of *irreducible forbidden words* (IFWs). An IFW never appears in a sequence of $\mathcal{S}$ and does not contain any other forbidden word. For example, assume that $\mathcal{F} = \{00\}$. The word 001 is a forbidden word but is not irreducible because it contains 00. More generally, any word of the form $u00v$, where u and v are arbitrary words, is not irreducible. By construction, any forbidden word of a language must have one of the elements of $\mathcal{F}$ as a substring or is itself an IFW. IFWs are of length $l \geq 2$ since length-1 forbidden words can be removed by reducing the alphabet.

Of particular importance are the *shifts of finite type* (SFTs), which are described by a finite number of IFWs. Indeed, they can be specified with a finite amount of

information, and invariant quantities such as the entropies described later can then be computed exactly. If the longest IFW of an SFT is of length $L + 1$, the *order* of the shift is L.

SFTs of order 1 are also called *topological Markov chains*. Since their set of IFWs contains only 2-blocks, the structure of this set can be described by a transition matrix M such that

$$M_{s_1,s_0} = \begin{cases} 0 & \text{if } s_0 s_1 \text{ is forbidden} \\ 1 & \text{if } s_0 s_1 \text{ is allowed} \end{cases} \tag{2.74}$$

That is, M_{s_1,s_0} is nonzero if s_1 is allowed to follow s_0 in a sequence (i.e., there is a transition $s_0 \to s_1$). A simple example is the transition matrix of the SFT with $\mathcal{F} = \{00\}$:

$$M = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.75}$$

which characterizes the symbolic dymamics of the logistic map immediately before the period-3 window (Section 2.7.3.2).

Markov chains are all the more important as any SFT of order L can be reformulated as a Markov chain by recoding sequences appropriately. Assume that there are N' allowed L-blocks, and denote by $\mathcal{A}_L$ the alphabet made of these N' symbols. Any sequence $s_0 s_1 \ldots s_k \ldots$ can then be recoded as $S_0 S_1 \ldots S_k \ldots$, where the new symbol $S_k \in \mathcal{A}_L$ is the L-block starting at position k: $S_k = s_k s_{k+1} \ldots s_{L-1+k}$. For example, $S_0 = s_0 s_1 \ldots s_{L-1}$ and $S_1 = s_1 s_2 \ldots s_L$.

Now, assume that $S = s_0 s_1 \ldots s_{L-1}$ and $S' = s'_0 s'_1 \ldots s'_{L-1}$ are two symbols of $\mathcal{A}_L$. The element $M_{S',S}$ of the new transition matrix is 1 if:

- The head of S coincides with the tail of S': $s'_0 s'_1 \ldots s'_{L-2} = s_1 \ldots s_{L-1}$;
- $s_0 s_1 \ldots s_{L-2} s'_{L-1}$ is an allowed L-block of the original shift

and is 0 otherwise. For example, assume that $\mathcal{F} = \{00, 0110\}$ and sequences are recoded using blocks of four symbols. Then one has $M_{101,011} = 0$ because 11 (tail of S) differs from 10 (head of S'), $M_{110,011} = 0$ because 0110 is an IFW, but $M_{111,011} = 1$ because 0111 does not contain any IFW.

Therefore, we see that any SFT can be described completely by a transition matrix M. In the following section we show how to extract from this matrix information about the spectrum of periodic orbits of the dynamical system and an important measure of chaos, topological entropy.

2.8.2
Periodic Orbits and Topological Entropy of a Markov Chain

As discussed previously, periodic points of a symbolic dynamical system correspond to periodic sequences Σ satisfying $\sigma^p \Sigma = \Sigma$. Since periodic orbits play a crucial role, we want to be able to compute the number of periodic sequences of

(a) (b)
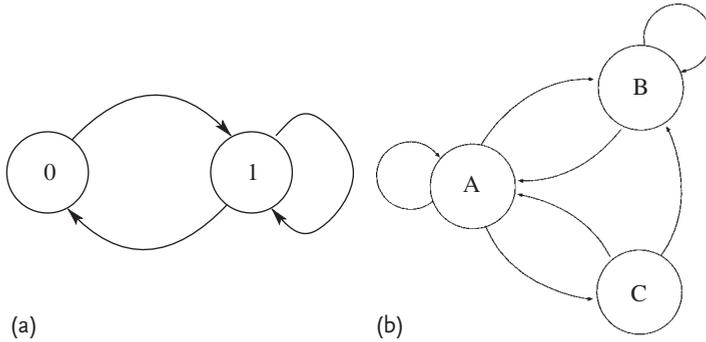
Figure 2.17 A transition matrix can be represented by a directed graph. Nodes correspond to possible states and edges indicate whether transition from one state to another is possible. (a) This graph corresponds to transition matrix (2.75); (b) this graph describes a grammar in which *C* must be preceded by *A*: The sequences *BC* and *CC* cannot occur.

period p of an arbitrary Markov chain, given its transition matrix M. This problem is part of a more general one, which is to determine the number of allowed symbol strings of length n. An important measure of chaos, *topological entropy*, characterizes how this number increases when $n \to \infty$.

A transition matrix is conveniently represented by a directed graph. To each symbol corresponds a node, which can be viewed as a state. When the transition from "state" i to "state" j is allowed (i.e., $M_{ji} \neq 0$), there is a directed edge going from node i to node j (Figure 2.17). The problems stated above can be reformulated as follows: How many distinct paths of length n does the graph have? How many of these paths are closed?

We first compute the number P_{ji}^n of paths connecting node i to node j in exactly n steps. This can be done inductively. Obviously, one can go from i to j in one step only if there is an edge between the two sites, thus $P_{ji}^1 = M_{ji}$. Then we note that each itinerary linking i to j in n steps goes from i to some site k in $n-1$ steps, then follows the edge from k to j in one step. By summing over all possible intermediary sites k, one obtains

$$P_{ji}^n = \sum_{k=0}^{N} M_{jk} P_{ki}^{n-1}$$

which is immediately seen to be the rule for matrix multiplication. Since $P_{ji}^1 = M_{ji}$, it follows that

$$P_{ji}^n = (M^n)_{ji} \tag{2.76}$$

Hence all the relevant information is contained in the successive powers of the transition matrix.

Periodic sequences of period p correspond to paths of length p that begin and end at the same node. Thus the number $N_f(p)$ of periodic points of period p is

$$N_f(p) = \sum_{k=0}^{N-1} M_{kk}^p = \operatorname{tr} M^p \tag{2.77}$$

Similarly, the number $N_s(p)$ of p-symbol strings equals the total number of paths of length $p - 1$ and thus is given by the sum of the elements of M^{p-1}. This can be formalized as follows. Let V^p be the vector whose ith component V_i^p is the number of length-p symbol strings beginning with symbol i. It is easy to see that the components of V^1 are all 1 and that $V^p = M^{p-1}V^1$. Thus,

$$N_s(p) = (V^1)^T V^p = (V^1)^T M^{p-1} V^1 \tag{2.78}$$

Expressions (2.77) and (2.78) show that $N_f(p)$ and $N_s(p)$ have the same asymptotic behavior. Indeed, the action of M^p for large p is determined by its largest eigenvalue λ_{max} and the associated eigenvector. It is easily shown that if $\lambda_{max} > 1$, then

$$\lim_{p \to \infty} \frac{\log N_f(p)}{p} = \lim_{p \to \infty} \frac{\log N_s(p)}{p} = \log \lambda_{max} \tag{2.79}$$

The growth rate of the number $N_s(p)$ of p-blocks

$$h_T = \lim_{p \to \infty} \frac{\log N_s(p)}{p} = \log \lambda_{max} \tag{2.80}$$

is called the *topological entropy*. It measures the average amount of information that is extracted by reading one symbol of a typical sequence. Equation 2.79 shows that the topological entropy of a Markov chain depends in a very simple way on the transition matrix. It also illustrates the fact that in general the growth rate of the number of periodic points is equal to the topological entropy, as noted in Section 2.6.4.1. More sophisticated techniques to compute topological entropy are presented in Section 2.8.6. Let us consider two examples:

- The eigenvalues of transition matrix (2.75) are $(1 \pm \sqrt{5})/2$. The largest one is $\lambda_{max} \sim 1.618\,033\,9$ and is known as the *golden mean*. This Markov chain, accordingly called the *golden mean shift*, has topological entropy $h_T \sim 0.481\,211\,8$.
- The transition matrix of the full N-shift is filled with 1s. Its largest eigenvalue is N, and the topological entropy is $h_T = \ln N$. In particular, the 2-shift has $h_T = 0.693\,147\,1$, which is greater than for the golden mean shift: Topological entropy increases as chaos becomes more developed.

2.8.3
Markov Partitions

Markov chains are interesting not only as model dynamical systems but also because there are some classical dynamical systems whose symbolic dynamics can

be represented exactly by a topological Markov chain. We have already encountered a few examples of such systems. The simplest ones are the tent and logistic maps associated with a full 2-shift: Every sequence of "0" and "1" is associated with a physical orbit. The logistic map at the beginning of the period-3 window is another example of a finite shift; the only forbidden sequence is the string "00." A natural question then is: Under what conditions is a dynamical system described faithfully by a Markov chain?

The symbolic coding of a dynamical system relies on the existence of a partition $\mathcal{P} = \{\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{N-1}\}$ of phase space into N disjoint regions $\mathcal{P}_i$. At each time step, the current system state is coded with the symbol i associated with the region $\mathcal{P}_i$ to which it belongs. Because the partition and the time-one map determine completely the symbolic dynamics, it is not surprising that the condition for being describable by a Markov chain involves the partition and the images of the members of the partition.

We now state this condition without a proof. Assume that there exists a partition $\mathcal{P} = \{\mathcal{P}_i\}$ such that the intersection of any member with the image of another is either itself or empty:

$$\forall i, j \quad \mathcal{P}_i \cap f\left(\mathcal{P}_j\right) = \begin{cases} \mathcal{P}_i \\ \emptyset \end{cases} \tag{2.81}$$

The structure of such a partition can be described concisely but faithfully by a transition matrix $M^{\mathcal{P}}$ defined by

$$M_{ij}^{\mathcal{P}} = \begin{cases} 1 & \text{if} \quad \mathcal{P}_i \cap f(\mathcal{P}_j) \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{2.82}$$

It can be shown that the dynamical system coded by the partition $\mathcal{P}$ is then completely equivalent to the Markov chain of transition matrix $M^{\mathcal{P}}$. In particular, the topological entropy of the original dynamical system is equal to that of the Markov chain, and both systems have the same spectrum of periodic orbits. Accordingly, a partition $\mathcal{P}$ satisfying (2.81) is called a *Markov partition*. Examples of Markov partitions in the logistic map at special values of parameter a are given in Section 2.9.2.

Note that even when there is a generating partition (such as the one based on critical points of a one-dimensional map) and it is not of a Markov type, the existence of a Markov partition is not precluded. If the system is equivalent to a shift of a finite type, then an analysis of the symbolic dynamics obtained with the generating partition should reveal that there are a finite number of irreducible forbidden words. As described in Section 2.8.1, a Markov chain can then be obtained with a suitable recoding and the topological entropy computed using the associated transition matrix.

2.8.4
Approximation by Markov Chains

In fact, only a small fraction of the regimes of a logistic map can be represented by a Markov chain exactly. Indeed, there is only a countable number of finite matrices of 0 and 1, whereas these regimes are indexed by parameter a, which is a real number [52]. Moreover, chaotic regimes are associated with kneading sequences that are not eventually periodic, which makes it generally impossible to describe the symbolic dynamics by a finite number of IFWs.

However, this does not make shifts of a finite type irrelevant. Indeed, it is not possible to analyze arbitrarily long symbol sequences. In practice, there is an upper bound on the length of the longest symbolic sequence that can be obtained in a reasonable time. This limits the search for forbidden symbol blocks to a maximal length. Otherwise, longer symbol blocks may be classified incorrectly as forbidden only because their probability of occurrence is too small. For example, assume that an orbit of 1 million ($\sim 2^{20}$) points has been recorded and coded on two symbols. It is certainly pointless to determine forbidden blocks longer than 20 symbols since the least probable one will occur at most once in the best case, where all blocks are equiprobable.

Therefore, Markov chains are still relevant for characterizing dynamical systems that are not conjugate to a shift of finite type, provided that a generating partition is known and a long symbolic sequence has been recorded. If the list of forbidden words has been determined for word lengths up to L, this gives a natural approximation of the system under study by a finite-type shift of order L, and hence by a Markov chain after a suitable recoding. Note that this systematically overestimates topological entropy estimates because higher-order forbidden sequences are neglected. The expression $h_T = \ln \lambda_{\max}$ for a Markov chain assumes that the number $N_s(p)$ of p-blocks can be determined for arbitrary p from the transition matrix. If there is more "pruning" than described by this transition matrix, the actual number of sequences will be lower, as well as the topological entropy. Carrying out this computation for increasing block lengths and comparing the results may help to estimate its accuracy.

2.8.5
Zeta Function

As shown in Section 2.8.2, the number of periodic points of period n, P_n, can be computed as $P_n = \operatorname{tr} M^n$, where M is the Markov transition matrix. The information contained in the transition matrix can be transformed into a generating function for P_n by defining

$$\zeta(t) = \exp\left(\sum_{n=1}^{\infty} \frac{P_n}{n} t^n\right) \tag{2.83}$$

With a little bit of algebraic calisthenics, it is possible to show that

$$\zeta(M, t) = \frac{1}{\det(\mathrm{Id} - tM)} \tag{2.84}$$

We illustrate one use of the zeta function in the example below.

Example 2.2

The spectrum of orbits forced by the period-3 orbit 3_1 of the logistic map is computed using the Markov transition matrix M; cf. (2.75). If the matrices

$$\text{Id} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$M = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

have been defined previously, as well as the positive integer $N \ (= 12$ below), the generating function $\sum_n P_n / n t^n$ is given up to degree N by the simple Maple call

$$> \text{taylor} \left(-\log \left(\det(\text{Id} - t * M) \right), t = 0, N + 1 \right)$$

$$t + \frac{3}{2}t^2 + \frac{4}{3}t^3 + \frac{7}{4}t^4 + \frac{11}{5}t^5 + \frac{18}{6}t^6 + \frac{29}{7}t^7 + \frac{47}{8}t^8 + \frac{76}{9}t^9$$

$$+ \frac{123}{10}t^{10} + \frac{199}{11}t^{11} + \frac{322}{12}t^{12} + \mathcal{O}\left(t^{13}\right) \tag{2.85}$$

We read these results as follows:

1. There is one period-1 point 1_1.
2. There are three period-2 points. One is the period-1 point 1_1; the other two belong to the single period-2 *orbit* 2_1.
3. There are four period-3 points. One is 1_1; the other three belong to the *degenerate* saddle–node pair 3_1 (001 and 011).
4. There are seven period-4 points, which belong to the orbits 1_1, 2_1, and 4_1 of the initial period-doubling cascade.
5. There are 11 period-5 points. One belongs to 1_1. The remaining 10 belong to two period-5 orbits, which comprise the saddle–node pair 01111 and 01101.

Continuing in this way, we construct the remaining results. These are summarized in Table 2.2. It is a simple matter to verify that the results of this table are consistent with the results of Table 2.1 up to period 8.

2.8.6
Dealing with Grammars

At the saddle–node bifurcation of the period-3 orbit, the adjacency matrix (Markov transition matrix) is given by (2.75). This matrix tells us that the symbol 0 must be followed by the symbol 1, and the symbol 1 can be followed by either of the symbols 0 or 1.

Table 2.2 Number of orbits up to period $p = 12$ forced by 3_1 computed using the zeta function based on the golden mean matrix (2.75).

p	$N(p)$	Lower-period orbits	Period-doubled	Number of saddle–node pairs
1	1	1_1		
2	3	$(1_1 +$	$2_1)$	
3	4	$1_1 + 3_1$		
4	7	$(1_1 + 2_1 +$	$4_1)$	0
5	11	1_1		1
6	18	$1_1 + 2_1 + 3_1$		1
7	29	1_1		2
8	47	$(1_1 + 2_1 + 4_1$	$8_1)$	2
9	76	$1_1 + 3_1$		4
10	123	$1_1 + 2_1 + (5_1 +$	$10_4)$	5
11	199	1_1		9
12	322	$(1_1 + 2_1 + 4_1) + 3_1 + (6_1 +$	$12_2)$	12

2.8.6.1 Simple Grammars

It is useful to introduce an alternative representation for the dynamics. This involves introducing two symbols (words) $A = 01$ and $B = 1$. These two words have lengths 2 and 1, respectively. In this representation, A can be followed by either A or B, as is true also for B. The transition matrix is full:

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.86}$$

and the grammar is simple (no transitions are forbidden).

The periodic orbits in the dynamics can be constructed as follows. Replace the nonzero elements in the first row of M by the symbol A and those in the second row by the symbol B (A and B do not commute).

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} A & A \\ B & B \end{bmatrix} = \mathcal{M} \tag{2.87}$$

Then the complete set of periodic orbits is constructed by computing $\operatorname{tr} \mathcal{M}^n$, $n = 1, 2, 3, \ldots$

Example 2.3

For $n = 1, 2$, and 3 we find

n	$\operatorname{tr} \mathcal{M}^n$	Orbits
1	$A + B$	$01 + 1$
2	$2A^2 + AB + B^2$	$2(01)^2 + 011 + (1)^2$
3	$2A^3 + 2ABA + 2A^2B + AB^2 + B^3$	$2(01)^3 + 4(01101) + 0111 + (1)^3$

Reduction to simple grammars is often useful in analyzing experimental data. For example, chaotic data generated by the Belousov–Zhabotinskii reaction (cf. Chapter 7) have been reduced to a symbolic code sequence involving the two symbols 0 and 1. The rules of grammar observed in the experimental data are as follows:

1. The symbol 0 must be followed by the symbol 1.
2. The symbol 1 can be followed by 0 or 1.
3. Symbol sequences $(11 \cdots 1)$ of length p can occur for $p = 1, 2, 3, 4$ but not for $p > 4$.

It appears that the vocabulary of this dynamics consists of the four words 01, 011, 0111, and 01111. It also appears that any of these words can be followed by any other word. The grammar is simple and represented by a 4×4 Markov transition matrix whose 16 elements are 1. The periodic orbits are obtained as described above for the golden mean case.

The topological entropy for dynamics consisting of a finite number of words of varying length obeying a simple grammar (full shift) is easily determined as follows. Assume that there are $w(p)$ words of length p, $p = 1, 2, 3, \ldots$ Then the number of ways, $N(T)$, of constructing a word of length T is determined by the difference equation

$$N(T) = w(1) N(T-1) + w(2) N(T-2) + \cdots = \sum_{p=1} w(p) N(T-p) \quad (2.88)$$

The number $N(T)$ behaves asymptotically like $N(T) \sim A(X_M)^T$, where A is some constant and X_M is the largest real root of the characteristic equation

$$X^T = \sum_{p=1} w(p) X^{T-p} \quad \text{or} \quad 1 = \sum_{p=1} \frac{w(p)}{X^p} \quad (2.89)$$

Example 2.4

For the full shift on two symbols 0 and 1,

$$1 = \frac{2}{X} \implies h_T = \log 2$$

Example 2.5

For golden mean dynamics, $w(1) = 1$, $w(2) = 1$, and

$$1 = \frac{1}{X} + \frac{1}{X^2} \implies X_M = \frac{1 + \sqrt{5}}{2} \quad \text{and} \quad h_T = 0.481\,212$$

Example 2.6

For the Belousov data described above, $w(p) = 1$ for $p = 2, 3, 4, 5$ and $w(p) = 0$ for $p = 1$ and $p > 5$, so that

$$1 = \frac{1}{X^2} + \frac{1}{X^3} + \frac{1}{X^4} + \frac{1}{X^5} \quad X_M = 1.534\,158 \quad \text{and} \quad h_T = 0.427\,982$$

2.8.6.2 Complicated Grammars

There are many cases in which the dynamics either consists of or is well approximated by a finite vocabulary with a nontrivial grammar. The two questions addressed above (description of periodic orbits, computation of entropy) are still of interest.

The spectrum of periodic orbits can be computed by following the algorithm used above.

1. Write out the Markov transition matrix M for the vocabulary.
2. Replace each nonzero matrix element 1 in row i by the noncommuting symbol w_i representing the ith word, effecting the transition $M \to \mathcal{M}$.
3. Compute $\operatorname{tr} \mathcal{M}^n$ for $n = 1, 2, \dots$
4. Replace each word sequence by the appropriate sequence of symbols from the original alphabet (i.e., 0 and 1).

The problem of computing the topological entropy for this dynamics is more subtle. It is isomorphic to the problem of computing the capacity of a transmission channel. The capacity of a transmission channel is (Shannon, [56, 57])

$$C = \lim_{T \to \infty} \frac{1}{T} \log N(T)$$

Here $N(T)$ is the number of allowed signals of duration T, and log is to base e.

In many grammars, not all symbol sequences are allowed (qu is OK, qv is not). In such cases, assume that there are m states $b_1, b_2, \dots, b_m$. For each state only certain symbols from the set $S_1, S_2, \dots, S_n$ can be transmitted (different subsets for different states). The transmission of symbol S_k from state b_i to state b_j (b_i may be the same as b_j) takes time $t_{ij}^{(k)}$, where k indexes all possible paths from b_i to b_j. This process is illustrated by a graph such as that shown in Figure 2.17.

Theorem 2.1

The channel capacity C is $\log W_0$, where W_0 is the largest real root of the $m \times m$ determinantal equation

$$\left| \sum_k W^{-t_{ij}^{(k)}} - \delta_{ij} \right| = 0 \tag{2.90}$$

For our purposes, we can regard each state as a word and t_{ij} is the length of word i.

Example 2.7

Assume that the vocabulary has three words A, B, and C or w_1, w_2, and w_3 of lengths p, q, and r and a grammar defined by Figure 2.17b. The Markov matrix is

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

The determinantal equation constructed from the Markov matrix and word lengths is

$$\begin{bmatrix} \frac{1}{W^p} - 1 & \frac{1}{W^p} & \frac{1}{W^p} \\ \frac{1}{W^q} & \frac{1}{W^q} - 1 & 0 \\ \frac{1}{W^r} & \frac{1}{W^r} & -1 \end{bmatrix} = 0$$

The characteristic equation for this dynamical system is

$$\frac{1}{W^p} + \frac{1}{W^q} + \frac{1}{W^{p+r}} = 1$$

The topological entropy is $h_T = \log W_0$, where W_0 is the largest real eigenvalue of this characteristic equation.

2.9
Fingerprints of Periodic Orbits and Orbit Forcing

2.9.1
Permutation of Periodic Points as a Topological Invariant

Using kneading theory, periodic points with given symbolic itineraries can be ordered along the interval. This is not only useful to determine the order in which periodic orbits appear but also to identify periodic orbits.

Indeed, consider the period-5 orbit with symbolic name 01011 of the logistic map (Figure 2.16). Its five periodic points are associated with the five cyclic permutations of the symbolic name, ordered by kneading theory as follows:

$$\overline{01101} \prec \overline{01011} \prec \overline{11010} \prec \overline{10101} \prec \overline{10110} \tag{2.91}$$

Label the sequences in (2.91) from left to right by Σ_i, $i = 1 \ldots 5$, and express them in terms of the leftmost sequence $\Sigma_1 = 01101$ and of powers of the shift operator. We have

$$\Sigma_1 \prec \Sigma_2 = \sigma^3 \Sigma_1 \prec \Sigma_3 = \sigma \Sigma_1 \prec \Sigma_4 = \sigma^2 \Sigma_1 \prec \Sigma_5 = \sigma^4 \Sigma_1 \tag{2.92}$$

We observe that under the action of the shift operator σ, these sequences are permuted: The lowest sequence becomes the third, the second one the last, and so on.

The corresponding permutation

$$\pi(01011) = (\pi_i) = (3, 5, 4, 2, 1) \tag{2.93}$$

such that $\sigma\Sigma_i = \Sigma_{\pi_i}$ provides crucial information about the orbit. Its dynamical relevance owes much to two fundamental properties.

First, *the permutation can be extracted directly from the periodic orbit without using any symbolic encoding and without having the graph of the map.* Indeed, consider the period-5 orbit in Figure 2.16. If we label the periodic points $x_1, x_2, \ldots, x_5$ from left to right, we can determine as above a permutation π such that $f(x_i) = x_{\pi_i}$. Obviously, this permutation is identical to (2.93) (e.g., the image of the first point is the third), which can easily be checked in Figure 2.16.

Second, permutation (2.93) *remains identical on the entire domain of existence of the orbit in parameter space.* As a parameter is varied, points x_i of a periodic orbit move along the interval, but they do so without ever becoming degenerate (otherwise, we would have one point with two images, in contradiction with the deterministic nature of the map). Thus the relative order of points is preserved, hence the corresponding permutation.

An important attribute of this invariance property is that the permutations associated with orbits interacting in a bifurcation will be strongly related. For example, saddle–node partners will have identical permutations since they are indistinguishable at the bifurcation. Similarly, the permutation associated with a period-doubled orbit can easily be obtained from that of its mother.

Since there is a definite relation between symbolic names and permutations on the one hand, and periodic orbits and permutations on the other hand, we see that the symbolic names of the orbits are more than a convenient labeling and that they carry topological information. To illustrate this, we now show that in one-dimensional maps, much about the symbolic name of an orbit can be recovered merely from the permutation. Consider the graphical representation in Figure 2.18 of permutation (2.93) extracted from Figure 2.16.

The global shape of Figure 2.18 is characteristic of unimodal permutations: The relative order of the leftmost points is preserved, while that of the rightmost points is reversed (in particular, the rightmost point is mapped to the leftmost). This is a signature of the existence of branches with different parities in the underlying map (assumed unknown). In fact, Figure 2.18 can be viewed as a topological representation of this map in a coordinate system where points x_i are equidistant.

Let us note the orientation-preserving (resp. orientation-reversing) branch 0 (resp. 1). It can be seen from Figure 2.18 that x_1 is necessarily on branch 0, while x_3, x_4, and x_5 must be on branch 1. The coding of x_2 is ambiguous: It can be on one branch or the other without modifying the permutation. Taking into account that the orbit of x_1 is $x_1 \rightarrow x_3 \rightarrow x_4 \rightarrow x_2 \rightarrow x_5$, the symbolic name of this orbit is thus $011_1^0 1$ (i.e., any of the two saddle–node partners born in the 5_1 saddle–node bifurcation listed in Table 2.1). If additional information is available, such as parity (is the multiplier of the orbit positive or negative?), it is possible to distinguish between these two orbits.
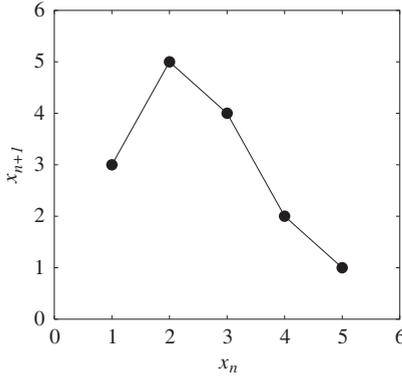
Figure 2.18 Graphical representation (i, π_i) of permutation (2.93). Orientation-preserving and orientation-reversing parts are easily distinguished.

By generalizing the example above, it is easy to see that every pair of saddle–node partners and every period-doubled orbit is associated with a different permutation. The permutation realized by a periodic orbit can thus be viewed *as a genuine fingerprint* of this orbit.

It is comforting to know that much of the discussion of this section is still relevant for orbits in three-dimensional flows. As discussed throughout this book, orbits in these systems will be associated with braids (a generalization of permutations deeply linked to knot theory). These braids will be characterized by topological invariants that do not depend on parameters and contain much information about the symbolic dynamics and the genealogy of periodic orbits. Just as the structure of unimodal maps governs that of the permutations (Figure 2.18), there is a systematic way to study the global organization of braids in three-dimensional systems.

2.9.2
Topological Entropy of a Periodic Orbit

The permutation associated with a periodic orbit not only provides qualitative information, but it can also provide estimates of fundamental quantitative measures of chaotic dynamics, as we show next. The key idea is that if the underlying map is continuous, the way in which the points of the orbit are mapped onto each other provides information on orbits in an extended neighborhood of the orbit.

For simplicity, let us consider a superstable periodic orbit such as the period-5 orbit of Figure 2.16. As previously, the points are numbered from left to right. Since x_2 is the critical point x_c, the leftmost and rightmost points x_1 and x_5 correspond to the lower and upper bounds of the invariant interval $I = [f(x_c), f^2(x_c)]$ where the relevant dynamics is confined. Using the fact that the x_i are mapped exactly onto each other, this will allow us to build a topological model of the dynamics.

To this end, consider the following partition of the invariant interval I:

$$I = [x_1, x_5] = I_1 \cup I_2 \cup I_3 \cup I_4 = [x_1, x_2] \cup [x_2, x_3] \cup [x_3, x_4] \cup [x_4, x_5] \quad (2.94)$$

Using the interval arithmetics (2.53) and permutation (2.93), we find easily that

$$f(I_1) = f([x_1, x_2]) = [x_3, x_5] = I_3 \cup I_4 \tag{2.95a}$$

$$f(I_2) = f([x_2, x_3]) = [x_4, x_5] = I_4 \tag{2.95b}$$

$$f(I_3) = f([x_3, x_4]) = [x_2, x_4] = I_2 \cup I_3 \tag{2.95c}$$

$$f(I_4) = f([x_4, x_5]) = [x_1, x_2] = I_1 \tag{2.95d}$$

The set of relations (2.95) is the analog of relations $f(I_1) = f(I_2) = I_1 \cup I_2$, which characterize the surjective logistic and tent maps (note that in a complete description of the map, the branch parities should also be specified). The key observation here is that both sets of relations define Markov partitions. In the example above, the Markov transition matrix as defined in (2.82) reads

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \tag{2.96}$$

where the nonzero entries correspond to pairs (I_i, I_j) such that $I_i \cap f(I_j) = I_i$. Even though the regime under study corresponds to a superstable orbit, matrix (2.96) is a signature of a chaotic dynamics. Its largest eigenvalue is $\lambda_{max} \sim 1.512\,876\,398$, which yields a topological entropy of $h_T = \ln \lambda_{max} \sim 0.414\,012\,738\,1$. Moreover, the matrix is transitive, which indicates that the associated Markov chain is topologically mixing.

In fact, the topological entropy as computed above characterizes the periodic orbit rather than the dynamical system to which it belongs: It is obtained from the permutation associated to the orbit,[4] not from the global structure of the system. However, the entire system cannot be less chaotic than implied by the periodic orbit. In particular, its topological entropy is necessarily greater than the entropy of the orbit. Thus, the observation of an orbit with a positive topological entropy (as obtained from its permutation), even in a window of stability, indicates the presence of chaos in the system under study, and in particular the existence of an infinity of periodic orbits. This is illustrated in the next section with the "period-3 implies chaos" theorem. A similar statement will be made later for flows: Some periodic orbits have knot types that can exist only in a chaotic system.

Note that the tools introduced here show why one-dimensional diffeomorphisms are not chaotic: Since they globally preserve or reverse the order of points, the associated transition matrices cannot have eigenvalues larger than 1. In fact, the reader may want to check that a one-dimensional diffeomorphism can only have a period-1 orbit, possibly a period-2 orbit if it globally reverses orientation.

4) For the sake of simplicity, we have indicated how to compute the topological entropy of an orbit only at a special parameter value, where the orbit is superstable. However, the topological entropy of an orbit depends only on the permutation associated with it and should be considered as a topological invariant of the orbit, defined on its entire domain of existence.

2.9.3
Period 3 Implies Chaos and Sarkovskii's Theorem

To illustrate the discussion above, we present now the famous statement "period 3 implies chaos," according to which the presence in a map of the interval of a periodic orbit of period 3 forces the presence of orbits of any other period [58].

If we carry out the same calculation for the superstable $0_1^0 1$ orbit as for the $011_1^0 1$ orbit in Section 2.9.2, we find a partition $I = [f^2(x_c), x_c] \cup [x_c, f(x_c)]$ whose transition matrix is the golden mean matrix (2.75) that we have already encountered:

$$A_{\mathrm{GM}} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.97}$$

It is easy to prove that $\forall n \geq 2$, $A_{ii}^n \neq 0$, hence there are fixed points for any period p. This is our first example of the general fact that the existence of some periodic orbits can force the existence of many other (here, an infinity) periodic orbits. This phenomenon is usually referred to as *orbit forcing*.

Note that the Markov partitions constructed from superstable orbits are refinements of the partition used for symbolic codings since the critical point is one of the border points. As a result, the transition matrices contain all the information needed to determine whether a given symbolic name is admissible. The example of the period-3 orbit is particularly simple since the Markov partition coincides with the coding partition: Transition matrix (2.97) indicates that all itineraries are allowed except those containing the string "00" (showing again that orbits of all periods exist). One may check in Table 2.1 that this is indeed what distinguishes orbits born before the 3_1 saddle–node bifurcation from orbits created afterward.

That the existence of an orbit of period 3 implies the existence of orbits of any other period can in fact be viewed as a particular case of a more general theorem due to Sarkovskii [59] (see also [52]). Consider the following ordering of the natural integers, written as the product $2^k \times (2n + 1)$ of a power of 2 by a prime number:

$$2^k \times 1 \lhd 2^l \times 1 \quad (k < l) \tag{2.98a}$$

$$2^k \times 1 \lhd 2^l \times (2n + 1) \quad (n > 0) \tag{2.98b}$$

$$2^l \times (2n + 1) \lhd 2^k \times (2n + 1) \quad (k < l, n > 0) \tag{2.98c}$$

$$2^k \times (2n + 1) \lhd 2^k \times (2m + 1) \quad (m < n) \tag{2.98d}$$

For example,

$$1 \lhd 2 \lhd 2^2 \lhd 2^3 \lhd 2^4 \times 7 \lhd 2^4 \times 3 \lhd 2^2 \times 5 \lhd 2^2 \times 3 \lhd 2 \times 3 \lhd 7 \lhd 5 \lhd 3$$

Sarkovskii's theorem states that if a continuous map of an interval into itself has an orbit of lowest period p and $q \lhd p$, it also has an orbit of lowest period q. It is easily seen that (i) if there are infinitely many different periods in a map, all the

periods corresponding to the period-doubling cascade of the period-1 orbit must be present; (2) since 3 comes last in (2.98), the presence of a period-3 orbit forces orbits of all other periods as shown above. It can be verified that the succession of bifurcations given in Table 2.1 satisfies the Sarkovskii theorem. In particular, the first period-7 orbit is created before the first period-5 orbit, which itself is created before the period-3 orbit. Moreover, all even periods are present when the first odd period appears.

2.9.4
Period 3 Does Not Always Imply Chaos: Role of Phase-Space Topology

Since so many properties of unimodal maps hold at least approximately for higher-dimensional systems, it might be troubling that numerous apparent counterexamples to the "period 3 implies chaos" theorem can be found in physical systems. In the modulated CO_2 lasers described in Chapter 1 and in Section 7.5.2, for example, it is quite common to observe multistability, with a period-3 orbit coexisting with the initial period-1 orbit, no other periodic orbit being present. Thus, a period-3 orbit does not necessarily imply chaos.

The clue to this paradox is that the modulated CO_2 laser can be described by a three-dimensional flow, hence by a two-dimensional Poincaré map, while the Sarkovskii theorem holds for a map of a one-dimensional interval into itself. It turns out that phase-space topology has a dramatic influence on orbit forcing.

The key topological difference between the two geometries is that in the two-dimensional case, it is possible to connect each of the three periodic points of the period-3 orbit to any other without encountering the third one. In the one-dimensional case, the two extreme points are isolated by the middle one.

Therefore, the three periodic points must be considered as arranged along a topological circle (obviously, this also applies to maps of a circle into itself; see Section 2.12). The three points divide this circle into three intervals versus two in the one-dimensional case. When the map is applied, the three points are cyclically permuted and the three intervals accordingly exchanged. As a result, the associated transition matrix is a simple permutation matrix, with zero topological entropy. The action of the map on the periodic orbit is equivalent to a pure rotation, which does not itself imply the existence of a chaotic dynamics. However, although it appears that orbit forcing can be modified dramatically when the topology of phase space is changed, it remains true in all cases that some orbits can force the existence of an infinity of other orbits.

2.9.5
Permutations and Orbit Forcing

The statements of the Li–Yorke theorem and of the Sarkovskii theorem seem to imply that period is the fundamental property in implication chains among orbits. However, forcing relations are more fine-grained. Admittedly, the existence of a period-5 orbit forces the existence of at least a period-7 orbit. However, an orbit of

the period-5 pair 5_1 of a logistic map does not force every period-7 orbit of this map. Actually, it forces the two 7_1 orbits *but is forced* by the two 7_2 orbits. As we will see below, it is in fact the permutation associated with a given orbit that determines which orbits it forces.

In Section 2.9.1, we noted that the permutation associated with a periodic orbit $\mathcal{O}$ could be represented graphically by a piecewise-linear map such as is shown in Figure 2.18. This map is conjugate to the Markov chain that describes the action of the map on the periodic points of the orbit (Section 2.9.2). If the Markov chain has positive topological entropy, it is chaotic, as is the piecewise-linear map, which has thus an infinite number of unstable periodic orbits. These orbits comprise the minimal set of orbits that must exist in a map having the original orbit $\mathcal{O}$ as one of its periodic orbits. The existence of these orbits is forced by that of $\mathcal{O}$.

As an example, we show that the period-5 orbit 01011 (born in bifurcation 7_1) forces exactly one pair of period-7 orbits and determines the symbolic code of this orbit. To this end, we apply the techniques outlined in Section 2.8.6 to transition matrix (2.96). Using the noncommutative symbols I_1, I_2, I_3, and I_4, we first rewrite it as

$$\mathcal{A} = \begin{bmatrix} 0 & 0 & 0 & I_1 \\ 0 & 0 & I_2 & 0 \\ I_3 & 0 & I_3 & 0 \\ I_4 & I_4 & 0 & 0 \end{bmatrix} \tag{2.99}$$

Then we compute

$$\operatorname{tr}\mathcal{A}^7 = [I_1\,I_3\,I_3\,I_2\,I_4\,I_1\,I_4] + [I_1\,I_3\,I_3\,I_3\,I_3\,I_2\,I_4] + (I_3)^7 \tag{2.100}$$

where $[W]$ represents the n cyclic permutations of a length-n word W. The minimal model compatible with the period-5 orbit 01011 has thus two period-7 orbits, whose itineraries in the Markov partition can be read from (2.100). Since interval I_1 is located to the left of the critical point (Section 2.9.2) and the intervals $I_2,\dots,I_4$ to the right of it, the canonical symbolic names of these orbits can be obtained through the recoding $I_1 \to 0$, $\{I_2, I_3, I_4\} \to 1$. The two period-7 orbits forced by the period-5 orbit under study are thus 0111101 and 0111101.

The reader can check in Table 2.1 that these are the two orbits born in saddle–node bifurcation 7_1 and that this bifurcation indeed occurs before 5_1 in the universal sequence. By repeating the above calculation for all low-period orbits of the logistic map, the whole universal sequence of Table 2.1 can be reproduced.

It is not a coincidence that the two period-7 orbits forced in the example are saddle–node partners and thus are associated with the same permutation. As was suggested by the simple example above, orbit forcing in one-dimensional maps is naturally expressed as an order on permutations rather than on periods. This is illustrated in Section 9.3.1. In unimodal maps, each bifurcation corresponds to a different permutation (Section 2.9.1) so that this order on permutations induces a total order on bifurcations, which corresponds to the universal sequence discussed in Section 2.7.5.3.

As is discussed in Chapter 9, there is no longer a total ordering of bifurcations in two-dimensional maps (and three-dimensional flows). The topological structure of a periodic orbit, then, is specified not by a permutation but by a braid type. There is a forcing relation on braid types. However, several saddle–node bifurcations can be associated with the same braid type, so that the induced order on bifurcations is only a partial order.

2.10
Two-Dimensional Dynamics: Smale's Horseshoe

2.10.1
Horseshoe Map

While one-dimensional maps display the most distinctive features of deterministic chaos, they lack a crucial property of time-one maps or Poincaré maps of flows: They are not invertible, as are the equations describing flows. Because many physical systems are described by differential equations, it is now time to turn to two-dimensional invertible maps to understand how the basic mechanisms studied in previous sections can be embedded in an invertible dynamics. This is not a trivial problem; many key points of our analysis of the tent and logistic map, such as property (2.31), rely heavily on noninvertibility. To help us in this task, our keystone will be the two-dimensional analog of the tent map, the paradigmatic Smale horseshoe.

The tent map example makes it clear that a key ingredient of chaos is expansion, or *stretching*. Yet many experimental and numerical examples of chaotic systems show that for a large number of them (i.e., dissipative systems), the asymptotic dynamics is confined to a small region of state space. Thus expansion must be balanced by contraction, or *squeezing*, so that trajectories remain forever in a bounded region. In continuous systems, these two antagonistic mechanisms can be combined by folding processes, as illustrated by Smale's horseshoe map f_S, defined as follows.

Take the unit square S, stretch it along one direction (the unstable direction) by a factor of approximately 2, while squeezing it in the transverse direction (the stable direction). Then fold the deformed rectangle so that it occupies roughly the same region as the original square, intersecting the latter in two disjoint strips (Figure 2.19a). Note that points located in the middle and the ends of S do not return to it.

The key topological property of the horseshoe map is that $f(S) \cap S$ consists of two disjoint components:

$$f(S) \cap S = H_0 \cup H_1 \tag{2.101}$$

which are the two horizontal strips H_0 and H_1 shown in Figure 2.19a.

To specify the action of the inverse map f^{-1}, we only need to determine $f^{-1}(H_0 \cup H_1)$ since $f^{-1}(f(S) \cap S) = S \cap f^{-1}(S)$. As shown in Figure 2.19b,
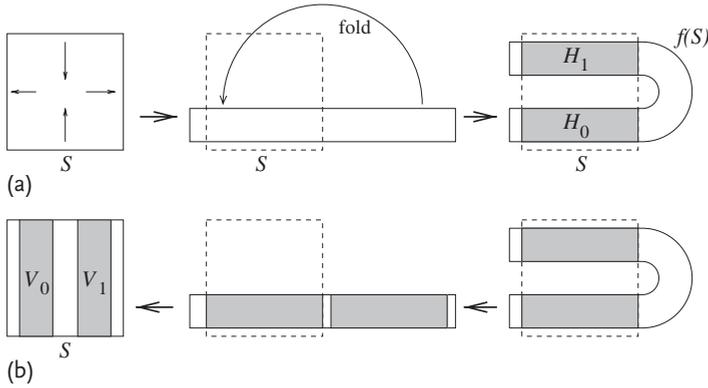
Figure 2.19 Graphical representation of Smale's horseshoe map. (a) The unit square S is stretched horizontally and squeezed vertically, then folded so that $f(S)$ intersects S in two horizontal strips H_i; (b) the map is iterated backward to obtain the preimages V_i of the strips H_i.

all the points sent by f to a given horizontal strip H_i come from a vertical strip $V_i = f^{-1}(H_i)$ and thus

$$f^{-1}(S) \cap S = V_0 \cup V_1 \tag{2.102}$$

2.10.2
Symbolic Dynamics of the Invariant Set

We want to study the structure of the invariant set $\Lambda = \cap_{-\infty}^{+\infty} f^n(S)$, which contains points whose orbits remain in S forever. As a first approximation, $\Lambda \subset f^{-1}(S) \cap S \cap f(S)$ or, by (2.101) and (2.102),

$$\Lambda \subset (V_0 \cup V_1) \cap (H_0 \cup H_1) \tag{2.103}$$

As illustrated in Figure 2.20, Λ is partitioned into four components, $H_i \cap V_j$, by the horizontal rectangles H_i and the vertical rectangles V_j. Since $H_i = f(V_i)$, whether a point X belongs to a square $H_i \cap V_j$ depends only on the vertical rectangles V_j its orbit successively visits:

$$X \in H_i \cap V_j \Leftrightarrow X \in f(V_i) \cap V_j \Leftrightarrow f^{-1}(X) \in V_i \text{ and } X \in V_j \tag{2.104}$$

that is, on the symbolic sequence $\Sigma(X)$ obtained by encoding the orbit of X with the coding function

$$\forall X \in \Lambda, \quad s(X) = \begin{cases} 0 & \text{if } X \in V_0 \\ 1 & \text{if } X \in V_1 \end{cases} \tag{2.105}$$

corresponding to the partition $\Lambda = (\Lambda \cap V_0) \cup (\Lambda \cap V_1)$. Because f is invertible, the action of the shift operator on symbolic sequences must not discard information,
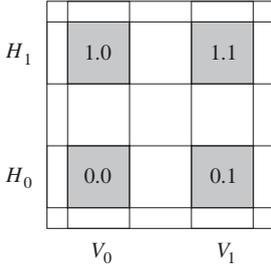
Figure 2.20 The invariant set Λ is contained in the intersections of the vertical rectangles V_i with the horizontal rectangles H_i. As explained in the text, $H_i \cap V_j$ is labeled $i.j$.

unlike in the one-dimensional case. Thus the sequence $\Sigma(X)$ associated to X must be bi-infinite:

$$\Sigma(X) = \ldots s_{-3} s_{-2} s_{-1}.s_0 s_1 s_2 \ldots \quad s_i = s(f^i(X)) \quad i \in \mathbb{Z} \tag{2.106}$$

with the dot separating the *forward sequence* $\Sigma_+(X) = s_0 s_1 \ldots$ from the *backward sequence* $\Sigma_-(X) = s_{-1} s_{-2} \ldots$ These two sequences describe the future and the past of the point, respectively. Applying the shift operator to a sequence simply amounts to shifting the dot: $\sigma(\ldots s_{-1}.s_0 s_1 \ldots) = \ldots s_{-1} s_0.s_1 \ldots$

As in (2.46), let us define cylinders as sets of points whose symbolic sequences contain a common substring, such as

$$\mathcal{C}[s_{-2} s_{-1}.s_0 s_1] \equiv \{X \in \Lambda; \Sigma(X) = \ldots s_{-2} s_{-1}.s_0 s_1 \ldots\} \tag{2.107}$$

It is easily seen from (2.104) that the squares in Figure 2.20 are cylinders:

$$(H_i \cap V_j) \cap \Lambda = \mathcal{C}[i.j] \tag{2.108}$$

which explains the labeling of the squares in Figure 2.20. To show that the partition based on (2.105) is generating, we now detail the structure of Λ and of higher-order cylinders.

Start from the cylinders $\mathcal{C}[i.j] = H_i \cap V_j$ shown in Figures 2.20 and 2.21a and compute their images $\mathcal{C}[ij.] = f(\mathcal{C}[i.j])$. These are the horizontal rectangles $H_{ij} = f(H_i) \cap H_j$ shown in Figure 2.21b, which are seen to be the components of $f^2(S) \cap S$. A dual decomposition of Λ is based on the preimages $\mathcal{C}[.ij] = f^{-1}(\mathcal{C}[i.j])$, which are the vertical rectangles $V_{ij} = V_i \cap f^{-1}(V_j)$ covering $f^{-2}(S) \cap S$ (Figure 2.21c).

Figure 2.21d illustrates the fact that the four-symbol cylinders $\mathcal{C}[ij.kl]$ can be obtained as the intersections of the vertical and the horizontal two-symbol rectangles:

$$\mathcal{C}[ij.kl] = \mathcal{C}[ij.] \cap \mathcal{C}[.kl] = H_{ij} \cap V_{kl} = f^2(V_i) \cap f(V_j) \cap V_k \cap f^{-1}(V_l) \tag{2.109}$$

which follows directly from definition (2.108). The forward (resp. backward) sequence $\Sigma_+(X)$ (resp. $\Sigma_-(X)$) specifies in which of the vertical (resp. horizontal) rectangles V_{ij} (resp. H_{ij}) point X lies.

Note that if one takes into account that $\Sigma_-(X) = s_{-1} s_{-2} \ldots$ should be read from the labels of Figure 2.21b from right to left, forward and backward sequences are
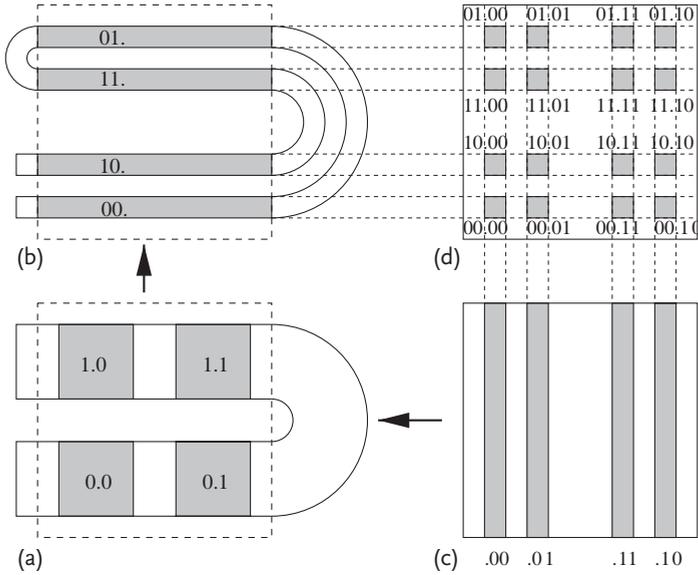
Figure 2.21 (a) Decomposition of $f(S) \cap S \cap f^{-1}(S)$ in squares $H_i \cap V_j$. These are mapped by the horseshoe map f to (b) the horizontal rectangles $H_{ij} = f(H_i) \cap H_j$, which are the components of $f^2(S) \cap S$, and by f^{-1} to (c) the vertical rectangles $V_{ij} = V_i \cap f^{-1}(V_j)$, which cover $f^{-2}(S) \cap S$. Arrows indicate the action of the horseshoe map; (d) by intersecting the horizontal with the vertical rectangles, one obtains a cover of the invariant set Λ by cylinders $C[ij.kl]$.

ordered in exactly the same way along the horizontal and vertical sides of S, respectively. In fact, they follow the unimodal order that governs the tent and logistic maps: This is a signature of the folding process (cf. Figure 2.14).

In general, cylinders of the type $C[s_{-N} \ldots s_{-1}.s_0 \ldots s_{N-1}]$ will be contained in the 2^{2N} intersections of 2^N vertical rectangles with 2^N horizontal rectangles. Without loss of generality, the action of the horseshoe map can be chosen to be piecewise linear in the regions of interest, with expansion and contraction rates $\lambda_u > 2$ and $\lambda_s < 1/2$. In that case, the vertical rectangles $V_{s_0 \ldots s_{N-1}}$ will be of width $1/\lambda_u^N$ while the height of the horizontal rectangles $H_{s_{-N} \ldots s_{-1}}$ will be λ_s^N. It is easy to see that both sizes shrink to zero in the limit $N \to \infty$, and thus that the cylinder $C[s_{-N} \ldots s_{-1}.s_0 \ldots s_{N-1}]$ converges to a unique point, as required for a generating partition.

This shows that the invariant set Λ is the product of a Cantor set of vertical lines with a Cantor set of horizontal lines, and that it is in 1 : 1 correspondence with the set of bi-infinite binary sequences. It can even be shown that this correspondence is a homeomorphism if the space of symbolic sequences is equipped with the following metric:

$$d\left(\{s_k\}_{k\in\mathbb{Z}}, \{s_k'\}_{k\in\mathbb{Z}}\right) = \sum_{i=-\infty}^{\infty} \frac{\delta_i}{2^{|i|}} \quad \delta_i = \begin{cases} 0 & \text{if} \quad s_i = s_i' \\ 1 & \text{if} \quad s_i \neq s_i' \end{cases} \qquad (2.110)$$

which means that the closer two sequences are according to this metric, the closer the associated points are. This property will be useful in what follows.

2.10.3
Dynamical Properties

A few dynamical properties of the horseshoe map can be deduced from the discussion of previous sections. Some of them follow directly from the correspondence with a full shift, as discussed in Section 2.6.5: sensitivity to initial conditions, existence of a dense set of periodic orbits, existence of dense orbits, and so on. Others are linked to the coexistence of two complementary mechanisms, stretching and squeezing, and to the horseshoe map's being two-dimensional. In particular, a foliation of Λ can be constructed in the following way.

Consider a point X and the vertical line $\mathcal{V}(X)$ passing through X. From the results of Section 2.10.2 it is easy to see that X and any point $Y \in \mathcal{V}(X)$ have the same forward sequence $\Sigma_+(X) = \Sigma_+(Y)$ (the vertical line is the limit of a sequence of vertical rectangles). This implies that X and Y have the same future, more precisely that their orbits converge to each other in forward time: $d(f^n(X), f^n(Y)) \to 0$ as $n \to \infty$. Indeed, the two corresponding symbolic sequences converge to each other according to metric (2.110) as the discrepancies between them are pushed farther and farther in the backward sequence.

Let us define the stable and unstable manifolds $W_s(X)$ and $W_u(X)$ of a point X as the sets of points whose orbits converge to that of X in the future and in the past, respectively:

$$W_s(X) = \left\{ Y \in \Lambda, \lim_{n \to \infty} d\left(f^n(X), f^n(Y)\right) = 0 \right\} \tag{2.111a}$$

$$W_u(X) = \left\{ Y \in \Lambda, \lim_{n \to \infty} d\left(f^{-n}(X), f^{-n}(Y)\right) = 0 \right\} \tag{2.111b}$$

According to this definition, the vertical line $\mathcal{V}(X)$ is a segment of the stable manifold of X: $\mathcal{V}(X) \subset W_s(X)$.

Similarly, two points X' and Y' located along the same horizontal line $\mathcal{H}(X')$ have the same backward sequence, hence the same past. Their orbits converge to each other backward in time: $d(f^{-n}(X'), f^{-n}(Y')) \to 0$ as $n \to \infty$. Thus they belong to the same segment of unstable manifold $W_u(X') = W_u(Y')$.

Hence Λ is foliated by horizontal and vertical lines that are segments of stable and unstable manifolds, respectively. The astute reader may have noticed that this could have been obtained directly from the geometrical description of the horseshoe map. However, it is in general not straightforward to determine these segments for a general two-dimensional invertible map. Because the discussion above is based on symbolic dynamical properties only, it can be extended directly to a large class of dynamical systems: As soon as a symbolic encoding of the invariant set is available, segments of stable and unstable manifolds can be obtained simply by following points with identical forward or backward sequences, respectively.

We conclude this section with an important remark regarding how noninvertibility can be embedded in an invertible dynamics. As discussed above, each vertical line $\mathcal{V}$ is associated with a unique forward sequence $\Sigma_+(\mathcal{V})$ shared by all the points in the line. Obviously, the images of all these points will have the same forward sequence: The image of the line by the horseshoe map is thus another vertical line $f(\mathcal{V})$ whose forward sequence is $\Sigma_+(f(\mathcal{V})) = \sigma\Sigma_+(\mathcal{V})$.

If $\mathcal{V}(\Lambda)$ represents the complete set of vertical lines (or, equivalently, of classes of points with identical forward sequences), the horseshoe map induces a map $f_V : \mathcal{V}(\Lambda) \to \mathcal{V}(\Lambda)$. Because forward sequences are one-sided, map f_V is intrinsically noninvertible; in fact, it is conjugate to the tent map since we know that forward sequences are ordered according to the unimodal order. In the invertible horseshoe map, there is thus *an underlying noninvertible map that organizes the dynamics.*

The clue to this paradox is that the shift operator acts on symbol sequences by transferring one symbol from the forward sequence to the backward sequence. The forward sequence loses information (i.e., the associated dynamics is noninvertible), gained by the backward sequence (whose dynamics thus depends on the forward sequence), while the net information flow is zero for the complete sequence (i.e., the global dynamics is invertible). Again, we see that this observation applies to a large class of dynamical systems, and it will be one of the building blocks of this book.

To conclude, let us emphasize that the two-sidedness of symbolic sequences for the horseshoe map is not due directly to the dynamics' being two-dimensional. Rather, it originates in the distinction between stable and unstable subspaces, regardless of their dimension. For example, one could imagine that forward (resp. backward) sequences parameterize a two (resp. one)-dimensional subspace.

2.10.4
Variations on the Horseshoe Map: Baker Maps

The tent map is the simplest one-dimensional map that is conjugate to a full one-sided 2-shift. To illustrate horseshoe dynamics, we now present examples of piecewise-linear maps that are conjugate to *two-sided* full shifts. These maps are the two-dimensional counterparts of the tent map. As we shall see, the stretching and squeezing mechanisms at play in the horseshoe map can be organized in a few inequivalent ways.

For definiteness, we restrict ourselves to bijections of the unit square into itself. Our first example, shown in Figure 2.22a, is a deformation of the horseshoe map of Figure 2.19. It is easily seen that this map can be described by two linear maps defined on the domains $x < 1/2$ and $x > 1/2$. More precisely, the coordinates (x', y') of the image of a point (x, y) are given by

$$x' = 1 - 2\left|x - \frac{1}{2}\right|, \quad y' = \begin{cases} \frac{y'}{2} & \text{if} \quad x \le \frac{1}{2} \\ 1 - \frac{y'}{2} & \text{if} \quad x > \frac{1}{2} \end{cases} \tag{2.112}$$
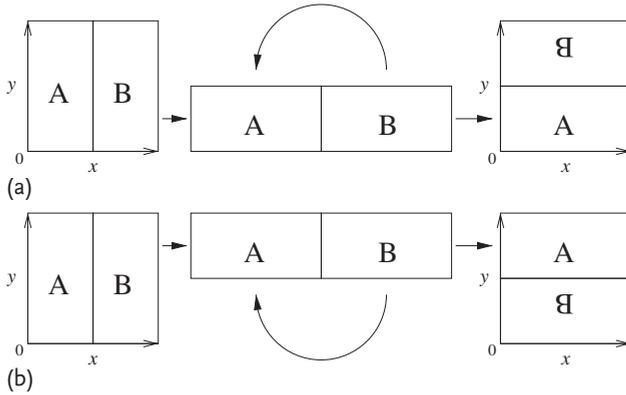
Figure 2.22 Two topologically distinct horseshoelike maps that differ in the folding process. Rotation of rectangle *B* can occur in the (a) counterclockwise or in the (b) clockwise directions.

Note that while the one-dimensional tent map is continuous but not differentiable at the critical point, the two-dimensional piecewise-linear map (2.112) has necessarily a discontinuity, here at the line $x = 1/2$. More importantly, (2.112) are our first example of a *reducible* system. Indeed, they have the following structure:

$$x' = f(x) \tag{2.113a}$$

$$y' = g(x, y) \tag{2.113b}$$

The key property of (2.113) is that it has a lower-dimensional subsystem (2.113a) that can be iterated independently: The dynamics of the x variable depends only on x itself. From the discussion of previous sections, this is not surprising. Indeed, the y direction is the stable direction, and vertical lines $x = c$ are segments of the stable manifold. Since the vertical lines parameterized by x are mapped to other vertical lines, the original two-dimensional map induces a one-dimensional map $x' = f(x)$. As discussed earlier, this map is intrinsically noninvertible: An invertible chaotic dynamics in the global state space is associated with a noninvertible dynamics in the unstable space.

This structure allows one to represent the action of map (2.112) by plotting the graphs (x, x') and (y, y'), as shown in Figure 2.23. We see that the chaotic two-dimensional map (2.112) is associated with two singular one-dimensional maps (f_u, f_s) such that $x' = f_u(x)$ and $y = f_s(y')$. The existence of f_u has been discussed previously and is a signature of the folding process. The existence of f_s follows from the fact that map (2.112) is invertible and that for the inverse map, stable and unstable spaces are exchanged. Hence the inverse map is associated with one-dimensional maps (f_s, f_u). Note that f_s is not arbitrary: Each monotonic branch of f_s corresponds to a given branch of f_u. Therefore, both maps have the same number of branches.

Figures 2.22b and 2.23b display another map that is extremely similar to our first example, except that folding takes place in the other direction. We note that the two
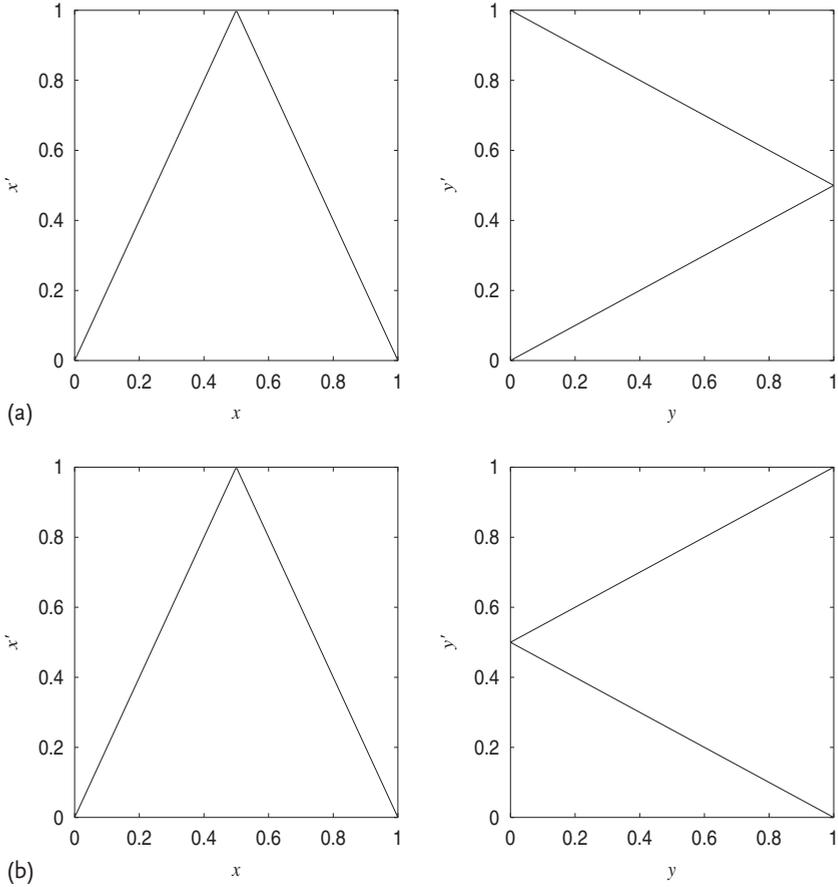
Figure 2.23 Graphs of the noninvertible one-dimensional maps associated with the invertible maps of Figure 2.22; (a) corresponds to the map in Figure 2.22a and (b) corresponds to the map in Figure 2.22b.

maps are associated with the same map in the unstable space, *but not in the stable space*. They are topologically distinct. These two topologically classes will later be referred to as the horseshoe and the *reverse* (or *twisted*) horseshoe. We shall see in Chapter 7 that both classes can be observed experimentally.

In these examples, reducibility is clearly apparent because the unstable and stable directions are fixed over the state space. This makes it easy to find coordinate systems whose axes are parallel to these directions. As the example of the Hénon map in the next section will show, this is not the case in general. However, it is tempting to conjecture that the equations describing a chaotic system can always be brought, at least approximately, to a reducible form (2.113) by a suitable change of coordinates. As noted previously, this should be the case if a good symbolic description of the system is available: Points that have identical forward (resp. back-
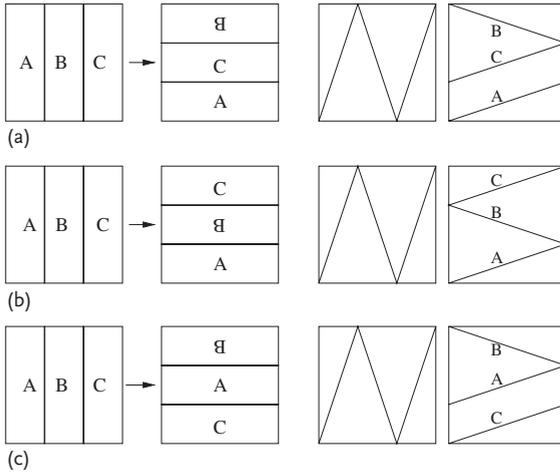
Figure 2.24 Examples of piecewise-linear maps on three domains that are topologically distinct. The associated one-dimensional maps are shown on the right-hand side.

ward) sequences should be on the same segment of the stable (unstable) manifold.

To elaborate on this classification, we display in Figure 2.24 a few examples of piecewise-linear maps with three branches. The associated pairs of noninvertible maps are also shown. We note that all three maps have the same map in the unstable space. Maps in Figure 2.24a,c have apparently the same map in the stable space. However, the way in which branches of the stable map are associated to branches of the unstable map differ. Thus, we see that a two-dimensional piecewise-linear map such as shown in Figures 2.22 and 2.24 can be specified by (i) the map f_u in the unstable space and (ii) the permutation indicating the order in which different branches are encountered along the stable direction (how they are stacked). This observation will help us to understand the construction of templates in Chapter 5.

2.11
Hénon Map

2.11.1
A Once-Folding Map

The Smale horseshoe is the classical example of a *structurally stable* chaotic system: Its dynamical properties do not change under small perturbations, such as changes in control parameters. This is due to the horseshoe map's being *hyperbolic* (i.e., the stable and unstable manifolds are transverse at each point of the invariant set). Recall that the study of the symbolic dynamics of the horseshoe map involved horizontal and vertical rectangles used to define cylinders (Section 2.10.2). If these

horizontal and vertical rectangles intersect transversely everywhere, intersections are preserved if the rectangles are slightly deformed. Thus the symbolic dynamics remains described by a full shift. Consequently, every map sufficiently close to the standard horseshoe map has the same spectrum of periodic orbits.

This is usually not the case for real systems, where periodic orbits are created or destroyed as a control parameter is varied. This is well illustrated by the example of the logistic map studied in Section 2.7. As the control parameter a is increased from 0 to 2, this one-dimensional system displays a gradual transition from a perfectly ordered state to a chaotic behavior that is associated with a binary symbolic dynamics and thus is asymptotically as random as coin tossing.

Similarly, the Hénon map [60]

$$x_{n+1} = a - x_n^2 + y_n \tag{2.114a}$$

$$y_{n+1} = b x_n \tag{2.114b}$$

is undoubtedly the most widely studied invertible two-dimensional map. It displays bifurcation sequences that lead to a complete horseshoe at large values of parameter a. Since the Jacobian $\partial(x_{n+1}, y_{n+1})/\partial(x_n, y_n) = -b$, parameter b admits of a simple interpretation: The area of an arbitrary region of the plane is reduced by a factor of $-b$ under the action of the map. When $|b| = 1$, areas are preserved and the map is said to be *conservative*. When $|b|$ is decreased to zero, the map becomes more and more dissipative. At $b = 0$, the y variable is forced to zero. The system then reduces to the logistic map, thereby becoming noninvertible.

If (2.114) are iterated at the usual parameter values $(a, b) = (1.4, 0.3)$ (which will be used throughout this section), and if successive points of a typical orbit are plotted in the (x, y) plane, the familiar picture shown in Figure 2.25a is obtained. It has been widely used to illustrate the fractal structure of chaotic attractors. While few other chaotic systems have been more deeply studied, a number of questions about its dynamics have not received a final answer, such as: How does one construct a symbolic encoding? Or: What are all the possible routes to a complete horseshoe? These questions are discussed in Section 2.11.2 and Chapter 9.

Interestingly, the Hénon map was originally introduced as a simplified model of a Poincaré map but is orientation-reversing at the parameters given above (i.e., its Jacobian is negative). However, this is impossible for the Poincaré map of a flow, which can be viewed as resulting from a smooth deformation of the plane continuously connected to the identity. For most properties (fractal dimensions, Lyapunov exponents), this makes no difference. However, there are topological measures that distinguish the orientation-preserving Hénon map from the orientation-reversing one [61].

The action of the Hénon map on points of a plane is represented graphically in Figure 2.25b. The folded dark curve is the image of the enclosing square and shows that the Hénon map is a once-folding map. The lighter curve corresponds to the second iterate of this square and is accordingly folded twice. This figure illustrates the similarity of the Hénon map with the hyperbolic horseshoe but also shows a crucial difference. At the parameter values considered, the intersection of
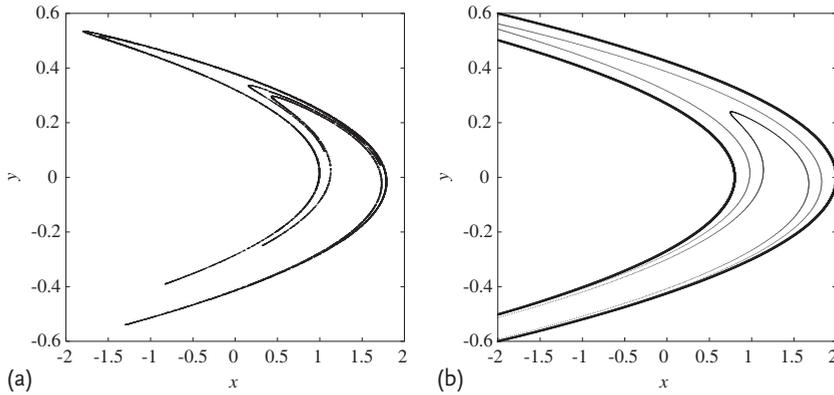
Figure 2.25 (a) The Hénon attractor obtained with (2.114) for $a = 1.4$, $b = 0.3$; (b) to illustrate the action of the map, the first and second iterates of the square $[-2, 2] \times [-0.6, 0.6]$ have been plotted, showing that the Hénon map is a once-folding map.

the square enclosing the invariant set and of its image is in one piece, whereas it had two disjoint components in the case of the horseshoe map (Figure 2.19).

Since this natural partition of state space in two regions was the keystone of our study of the symbolic dynamics of the horseshoe map, it is not obvious whether there is a symbolic coding for the Hénon map and how it can be constructed. We discuss this important question in Section 2.11.2. We shall see that it is directly connected to the nonhyperbolicity of the Hénon attractor (i.e., to the lack of transversality of certain intersections between the stable and unstable manifolds).

2.11.2
Symbolic Dynamics of the Hénon Map: Coding

The relevance of symbolic codings relies heavily on the mixing property of chaotic dynamics (Section 2.7.2). Assume that we have an arbitrary partition $M = M_1 \cup M_2$ of phase space. If a map f is mixing, there exists for any region $N \in M$ an integer $n_0(M)$ such that $f^{n_0}(N) \cap M_1 \neq \emptyset$ and $f^{n_0}(N) \cap M_2 \neq \emptyset$. However small a region N is, there exist points in N with different symbolic itineraries, and thus N can always be divided into smaller regions. Since this operation can in principle be repeated indefinitely, it could be tempting to believe that well-defined symbolic codings exist for any choice of the partition.

However, the example of the logistic map in Section 2.7 has shown that this is not true. The weakness in the argument above is that mixing cannot be invoked if the orbits of two points are strongly correlated for arbitrarily long times. In particular, we noted in our study of the logistic map that the different preimages of a given point must be associated with different symbols. This rule is indeed necessary to prevent the symbolic itineraries of such points from being identical since their orbits differ only in their initial conditions. If the map under study is expansive,

separating preimages is actually sufficient for obtaining generating partitions, as was shown in Section 2.7.2.

However, this criterion cannot be applied to the Hénon map; since it is invertible, every point has one and only one preimage. Thus we have to extend the simple rule used for the logistic map to handle the case of invertible chaotic maps. The extension should also be consistent with the coding of the hyperbolic horseshoe map.

The key point is that while the orbits of two different points cannot collapse onto each other in finite time if the map is invertible, they can become indiscernible after an infinite number of iterations: Two different points can have the same asymptotic future or the same asymptotic past (or both), as we see below.

In Figure 2.26a, we have plotted segments of the local stable and unstable manifolds of the fixed point $X^* \sim (0.884, 0.265)$. We denote these segments by W_s^l and W_u^l, respectively. Recall that the stable manifold $W_s(X^*)$ and unstable manifold $W_u(X^*)$ consist of the points whose orbits converge to the fixed point X^* under repeated action of the Hénon map f and of the inverse map f^{-1}, respectively:

$$X \in W_s(X^*) \Leftrightarrow \lim_{n \to +\infty} || f^n(X) - X^* || = 0 \tag{2.115a}$$

$$X \in W_u(X^*) \Leftrightarrow \lim_{n \to +\infty} || f^{-n}(X) - X^* || = 0 \tag{2.115b}$$

Obviously, the stable and unstable manifolds are invariant under both f and f^{-1}, hence the restrictions of these maps to the invariant manifolds behave like one-dimensional diffeomorphisms. Since the eigenvalues of the Jacobian at X^* along the stable and unstable directions are $0.156 > 0$ and $-1.92 < 0$, f and f^{-1} preserve (reverse) orientation along the stable (unstable) manifold. The action of f^{-1} on the stable segment W_s^l and that of f on the unstable segment W_u^l are illustrated in Figure 2.26b–f, which display $f^{-n}(W_s^l)$ and $f^n(W_u^l)$ for $n = 1, 2, 3, 4$, and 8, respectively. The two segments are stretched at each step and span a longer and longer part of the two invariant manifolds.[5] They are also folded at each iteration, a signature of the horseshoelike process that organizes the Hénon attractor. For example, Figure 2.26c can be compared with Figure 2.21 (keeping in mind that, unlike the classical horseshoe map, the Hénon map is orientation-reversing along the unstable manifold).

An extremely important consequence of the folding mechanism is that the stable and unstable manifolds of the fixed point X^* intersect themselves not only at X^* but also at other locations, called *homoclinic points*. Since homoclinic points belong to both the unstable and the stable manifolds, their orbits converge to the fixed point both forward and backward in time. The stable and unstable manifolds being invariant, the set of homoclinic points is itself invariant: All images and preimages of a homoclinic point are themselves homoclinic points. More generally, two points are said to be homoclinic to each other if their orbits converge to each other

---

5) This is how segments W_s^l and W_u^l of this example were themselves obtained, by iterating repeatedly from extremely small linear segments located around the fixed points and aligned along the stable and unstable direction at that point.
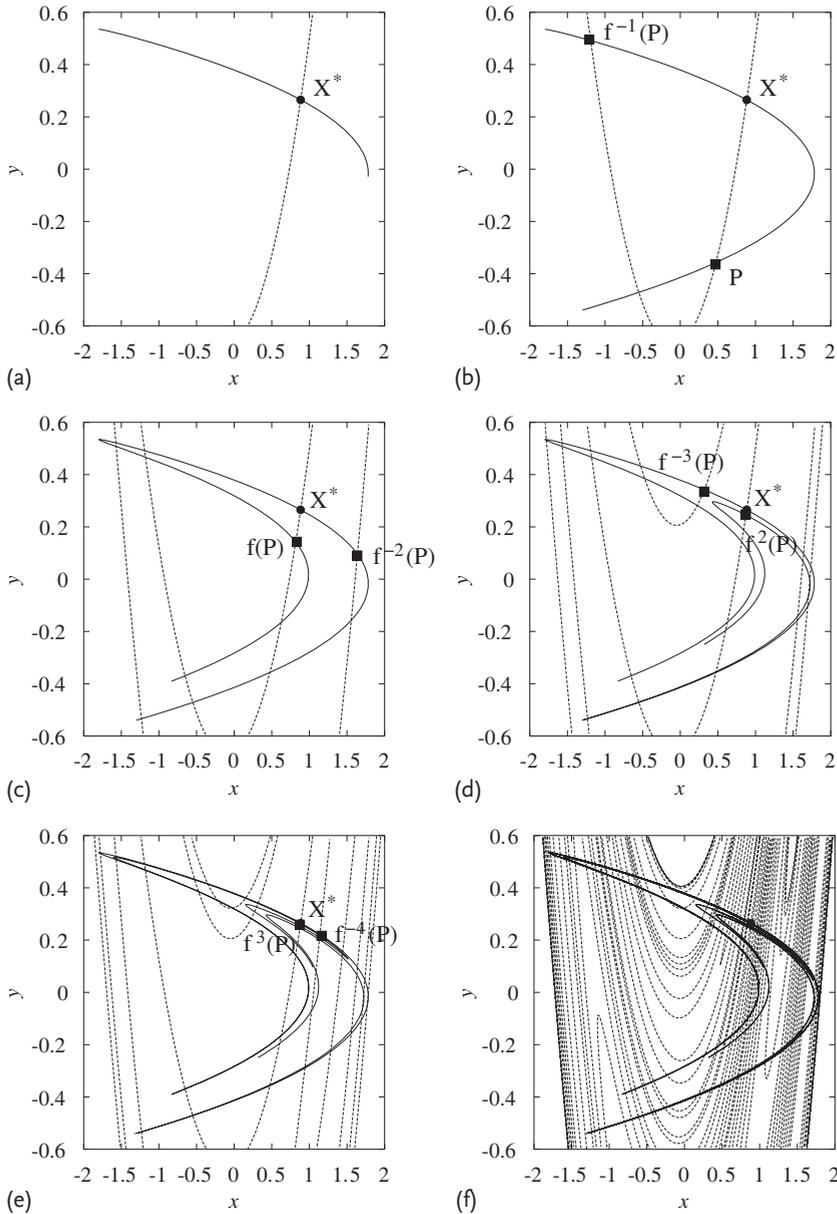
Figure 2.26 (a) The solid (dotted) line represents a segment W_u^l (W_s^l) of the unstable (stable) manifold of the fixed point located at their intersection. The orbits of points belonging to this line converge to the fixed point backward (forward) in time; (b–e) images $f^n(W_u^l)$ and preimages $f^{-n}(W_s^l)$ of the segments W_u^l and W_s^l for $n = 1, 2, 3, 4$. They are folded at each iteration and the number of intersections grows exponentially; (f) this is the same picture after eight iterations. The image of the unstable segment becomes indiscernible from the strange attractor. One clearly notices locations where the stable and unstable manifolds are tangent. These degenerate intersections are called homoclinic tangencies.

both forward and backward in time, with "homoclinic points" denoting points that are homoclinic to a fixed point. The loop formed by the pieces of the stable and unstable manifolds joining X^* and P is called a homoclinic loop.

Figure 2.26 also shows that, since the segments $f^{-n}(W_s^l)$ and $f^n(W_u^l)$ are increasingly folded as the iteration number n increases, the number of intersections $N_i(n)$ between the two segments (and thus of apparent homoclinic points) increases exponentially with n. However, it does not grow as fast as could have been expected from the study of the horseshoe map: $N_i(1) = 4$, $N_i(2) = 12$ and $N_i(3) = 30$ instead of $N_i(n) = 4^n$ for the horseshoe. This is an indication that the symbolic dynamics is not described by a full 2-shift.

In Figure 2.26b, we have distinguished two of the three homoclinic points present at that stage. By studying how f and f^{-1} act on the segments shown in Figure 2.26a,b, it is easy to see that one point, denoted by P, is the image of the other, indicated as $f^{-1}(P)$. Images and preimages of P are easily located in Figure 2.26b–e. Since they are themselves homoclinic points, they can be found among the intersections that appear in the consecutive plots of Figure 2.26. For example, $P = W_s^l \cap (f(W_u^l) \setminus W_u^l)$, hence $f(P) \in W_s^l \cap (f^2(W_u^l) \setminus f(W_u^l))$, which consists of two points. These two points are the images of P and of the unlabeled homoclinic point shown in Figure 2.26b. Since the latter is further from X^* than P and since f is continuous, $f(P)$ is the point of $W_s^l \cap (f^2(W_u^l) \setminus f(W_u^l))$ that is closest to X^* (Figure 2.26c). Determining the positions of all the images and preimages $f^i(P)$ shown in Figure 2.26b–e is an interesting exercise left to the reader.

We see that the sequence $\{P, f(P), f^2(P), f^3(P), \ldots\}$ converges to the fixed point X^* along the stable manifold, approaching it from one side. Similarly, the sequence $\{P, f^{-1}(P), f^{-2}(P), f^{-3}(P), \ldots\}$ converges to the fixed point along the unstable manifold, alternating between the two sides of the fixed point and with a smaller convergence rate since the unstable eigenvalue is negative and has a modulus closer to one than the stable eigenvalue. Thus, the orbits of X^* and P are strongly correlated both in the past and in the future, and some care has to be taken in order to prevent the two points from being associated with the same symbolic itinerary.

Before addressing this problem, a few remarks are in order. First, we note that the existence of homoclinic points forces the unstable manifold to be infinitely folded. On the one hand, all the $f^i(P)$ iterates belong to the unstable manifold, by definition. On the other hand, it can be seen in Figure 2.26b–e that they all belong for $i > 0$ to the segment of the stable manifold joining P and X^* (f is contracting along the stable manifold). Thus, the unstable manifold intersects this segment infinitely many times and is infinitely folded, which is consistent with what Figure 2.26 shows. The same argument can be applied to the stable manifold. This suggests that homoclinic points are associated with a complicated dynamics. As a matter of fact, it can be proven that if a dynamical system displays homoclinic points, then it has an invariant set on which the dynamics is chaotic and there are periodic orbits arbitrarily close to the homoclinic point [52].

Figure 2.26f shows the segments $f^8(W_u^l)$ and $f^{-8}(W_s^l)$. At the resolution of the plot, they provide good approximations of the unstable and stable manifolds of the

fixed point. Two important properties are illustrated by this picture. The first one is that the unstable segment $f^8(W_u^l)$ is indistinguishable from the strange attractor shown in Figure 2.25. Indeed, it is believed that the strange attractor is nothing but the closure of the unstable manifold $W_u(X^*)$. Figure 2.26 thus provides an illustration of how strange attractors are built and of their hierarchical structure. The second property is that, although the intersections between the unstable and stable manifolds are mostly transverse, places where the two manifolds are tangent to each other can be clearly discerned. These points, where it is not possible to define distinct stable and unstable directions, are called *homoclinic tangencies*. Their existence is a crucial difference between the horseshoe map and the Hénon map: The former is hyperbolic, the latter is not. Note that images or preimages of a homoclinic tangency are themselves homoclinic tangencies. Homoclinic tangencies play a foremost role in the problem of coding the Hénon attractor, to which we now return.

Assume that fixed point X^* lies in the region $\mathcal{P}_k$ of the partition. Its symbolic itinerary $\Sigma(X^*)$ then consists of the symbol s_k infinitely repeated. A necessary and sufficient condition for points P and X^* to have different symbolic itineraries is thus that at least one iterate $f^i(P)$ is separated from X^* by the border of the partition, so that it is associated with a symbol $s_{k'} \neq s_k$. Note also that if $\Sigma(P) \neq \Sigma(X^*)$, then obviously $\Sigma(f^i(P)) \neq \Sigma(X^*)$, so that the problem is solved at once for the whole orbit of homoclinic points based on P.

For the sake of simplicity, assume that we search for a partition where P itself and X^* belong to different regions. This is a natural choice because P is more distant from X^* than other points of its orbit, most of them being extremely close to X^*. Furthermore, P is one of the three intersections in Figure 2.26b, a first-generation homoclinic point so to speak. To design a simple criterion for placing the border of the partition, we need a quantity that distinguishes P from X^* (and more generally two points that are homoclinic to each other), that is related in a natural way to the geometry of the invariant manifolds, and that makes it easy to define a "middle point."

We note that since X^* and P in Figure 2.26b form a homoclinic loop, the unstable manifold crosses the stable manifold in different directions at X^* and P. This can be used to distinguish the two points in a robust way. To this end, we consider at each point X the angle $\beta(X)$ between the unstable and stable manifolds. This is a signed quantity because the two manifolds can be oriented. For example, assume that the two stable and unstable segments going from X^* to P are positively oriented. Then, an oriented frame $(\boldsymbol{v}_s(X), \boldsymbol{v}_u(X))$ can be defined at each point, with the two vectors being tangent to the stable and unstable manifold, and positively oriented. The quantity $\beta(X)$ is then defined as the oriented angle between $\boldsymbol{v}_s(X)$ and $\boldsymbol{v}_u(X)$, with $\beta(X) \in [-\pi, \pi]$.
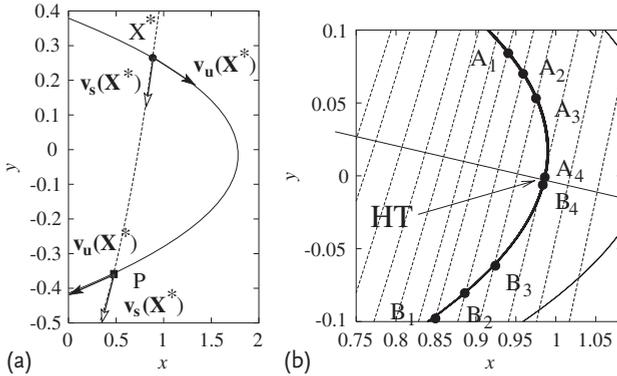
Figure 2.27 (a) Illustration of the definition of the oriented angle $\beta(X)$. A two-dimensional frame $\nu_s(X)$, $\nu_u(X)$ is attached to each point of the plane, each vector being tangent with one of the two invariant manifolds and positively oriented. The angles at two points connected by a homoclinic loop have necessarily opposite signs; (b) magnified view of the neighborhood of a homoclinic tangency HT. One can find pairs (A_i, B_i) of points homoclinic to each other arbitrarily close to the homoclinic tangency. Consequently, the border of the partition (shown here as a straight line) must pass through the homoclinic tangency in order to separate each pair.

The key observation is that $\beta(X)$ varies continuously in the plane, as do the unstable and stable directions,[6] and that $\beta(P)$ and $\beta(X^*)$ have necessarily opposite signs because X^* and P are connected by a homoclinic loop (Figure 2.27a). Thus, any path going from X^* to P has a point where $\beta(X) = 0$. Such a point is nothing but one of the homoclinic tangencies mentioned earlier. This observation is a two-dimensional version of the theorem stating that two zeroes of a continuous function $g\colon \mathbb{R} \to \mathbb{R}$ must be separated by a zero of its derivative, i.e., by a singularity.

Now we note that around a homoclinic tangency, there are arbitrarily small homoclinic loops (Figure 2.27). Thus, all the corresponding pairs of homoclinic points will be encoded with unique itineraries if and only if the partition border passes exactly through the homoclinic tangency.

Accordingly, it has been conjectured by Grassberger and Kantz that a generating partition for a nonhyperbolic system can be obtained by connecting homoclinic tangencies [62, 63]. This generalizes the procedure for coding one-dimensional maps, with the singularity located at the homoclinic tangency replacing the singularity located at the critical point, and pairs of homoclinic points replacing preimages of a given point. This conjecture has not been proven yet, but extensive numerical evidence that it yields generating partitions has been given over the years [63–69]. The simplest partition that can be obtained in this way for the Hénon at-

---

6) In the pictures shown in this section, we have shown only the unstable manifold of the fixed point X^*. This manifold is confined to the strange attractor. However, a stable and an unstable direction can be determined at each point of the plane using definitions (2.115). Thus, $\beta(X)$ is well defined everywhere, inside or outside of the strange attractor.
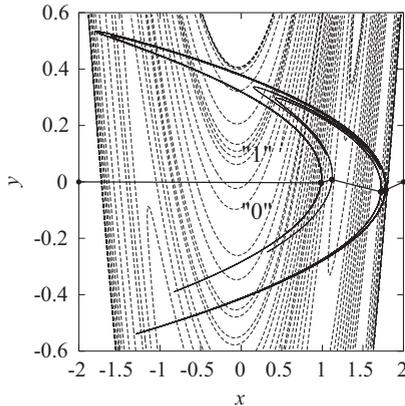
Figure 2.28 A generating partition for the Hénon map constructed by connecting homoclinic tangencies.

tractor is shown in Figure 2.28. Note that with this partition, the itinerary of P is $\Sigma(P) = \cdots 111.0111 \cdots$ and differs from $\Sigma(X^*)$ by only one symbol.

However, there are some difficulties with this approach. Individual lines defined by the equation $\beta(X) = 0$ usually do not make a satisfactory partition border. In fact, one generally finds that several such lines have to be followed, the connection between two different lines occurring outside the attractor. As a result, it is not always obvious which homoclinic tangencies to connect. Various criteria have been proposed to overcome this ambiguity (see, e.g., [66, 69]), but to our knowledge, no definitive solution has been found so far. Another problem is that there is a dramatic noise amplification precisely at homoclinic tangencies [68], which can make their extraction from experimental time series difficult. However, the folding and squeezing processes that build the strange attractor not only create localized singularities (homoclinic tangencies) but also determine the global structure of the attractor. We shall see later that analyzing the topological organization of a chaotic attractor provides detailed information about its symbolic dynamics. In particular, the symbolic name of many low-period orbits can be determined directly from its topological invariants. This fact has been used to design methods for contructing generating partitions that are based on topological analysis and do not rely on the structure of homoclinic tangencies [61, 70]. These methods will be outlined in Section A.4.

2.11.3
Symbolic Dynamics of the Hénon Map: Grammar

In the previous section, we discussed how to code trajectories of the Hénon map so that they can be associated in a one-to-one way with bi-infinite sequences of 0s and 1s. The purpose of such a coding is to characterize the dynamics by determining the grammar that determines which symbol sequences are allowed and

which never appear. We have already encountered this question when discussing the symbolic dynamics of the logistic map. In this case, there was a simple answer: Admissible sequences are such that none of their iterates is larger than the kneading sequence (Section 2.7.4). For the Hénon map and, more generally, diffeomorphisms of a plane, the problem is much more complicated, and in fact has not been completely solved, even if many profound results have been obtained. In the present section, we will only touch on it, with more details given in Chapter 9, where orbit forcing will be discussed. Indeed, understanding which symbol sequences can be realized involves understanding in what order they appear.

As was noted in Section 2.11.2, the stable and unstable manifolds of a Hénon map have fewer intersections than a full horseshoe map. While there are four intersections in both Figures 2.26b and 2.21a, there are only 12 in Figure 2.26c, which is to be contrasted to the $16 = 4^2$ that exist in Figure 2.21d. In Figure 2.26d, there are 30 instead of 64. The key point is that a Hénon map evolves from parameter values where there are no transverse intersections between invariant manifolds, and thus no chaos, to a fully chaotic state where it is topologically equivalent to a Smale horseshoe. As it evolves, intersections gradually appear through series of bifurcations associated with tangencies between invariant manifolds. Each new intersection corresponds to a new homoclinic orbit similar to point P in Figure 2.26b but with a distinct symbolic itinerary, so that new symbol sequences become allowed. Conversely, one could see a given Hénon map as a full horseshoe map from which some orbits and symbol sequences have been "pruned away." The key question is how to characterize the pruned symbol sequences in the most simple and logical way.

The term "pruning" was introduced in a pioneering paper by Cvitanović, Gunaratne, and Procaccia [64] in which the "pruning front conjecture" was formulated. The key idea is that since the unstable manifold is folded into infinitely many "leaves" located in different parts of the phase plane, stacked in the stable direction, a single kneading sequence is no longer sufficient. Rather, infinitely many are needed, one for each "leaf." More precisely, complexity arises because the kneading sequence, which determines the admissible *forward* sequences, depends on the *backward* sequence, which indicates what leaf the current iterate is located in. The kneading sequence changes monotonically from leaf to leaf, hence the term "pruning front."

Due to dissipation, however, only a finite number of kneading sequences are relevant in practice since only a finite number of leaves can be distinguished for a given resolution. For example, there is little difference to the naked eye between the unstable manifolds drawn in Figure 2.26e and f, whereas we clearly see the finer structure in Figure 2.26d compared to Figure 2.26c.

Accordingly, Hansen and Cvitanović were able to reproduce much of the structure of the Hénon map bifurcation diagram by considering bi-unimodal and four-unimodal maps [71]. An n-unimodal map is defined by n different unimodal maps, each associated with a symbol sequence. At each iteration, the map corresponding to the symbol or symbol sequence last seen (which indicates position along

(a)   (b)
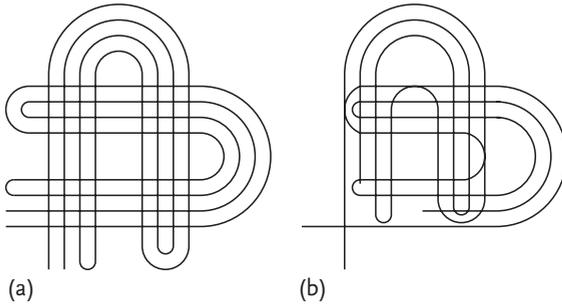
Figure 2.29 (a) A complete homoclinic tangle with all the intersections between the unstable and stable manifolds allowed by a binary grammar; (b) an incomplete tangle where some of the intersections have been removed. Intersections disappear through tangencies in sequences that are constrained by the geometry of the tangle.

the stable manifold) is applied. An *n*-unimodal map is naturally characterized by *n* kneading sequences, which can serve as canonical parameters for the map.

In particular, Hansen and Cvitanović could reproduce specific codimension-2 bifurcation structures in the two-parameter space of bi-unimodal maps, which are ubiquitous in invertible maps and flows. These structures involve two pairs of saddle–node twin orbits that exchange partners depending on parameter values. Termed "swallowtails" by Hansen and Cvitanović [71], these structures are also known as "shrimps" in the literature [72, 73].

More generally, the complex structure formed by the intertwined unstable and stable manifolds of a fixed point or periodic orbit is known as a *homoclinic tangle* (Figure 2.26) (or heteroclinic if it involves several orbits). As is well known, their importance was first recognized by Poincaré in his analysis of the three-body problem [48]. More recently, characterizing the structure of tangles has attracted much interest [74–81]. As parameter values are changed, the stable and unstable manifolds change. Their intersections, which correspond to homoclinic or heteroclinic orbits, appear or disappear through tangencies. As suggested by Figures 2.26 and 2.27, the order in which this occurs is severely constrained by the geometry of the tangle, which induces a partial order between homoclinic orbits: the presence of some orbits implies infinitely many others [77, 78], as illustrated by Figure 2.29. Comparing this picture to Figure 2.21 also provides insight into the pruning front conjecture and the property that the kneading sequence (i.e., the rightmost admissible forward sequence) depends on the backward sequence. The monotonicity of the pruning results from the fact that inner loops must disappear first.

Invariant manifolds are unbounded objects, and thus difficult to study, but their structure can be hierarchically approached with finite portions of increasing length of the invariant manifolds, *trellises*, which are amenable to combinatorial analysis [77]. Using trellises, the forcing order between homoclinic orbits can be computed and important information about the dynamics of the system and its symbolic dynamics extracted [75, 78, 81, 82]. Thus, homoclinic orbits are very important for understanding a chaotic dynamics. However, they are much more difficult to

extract from experimental time series than periodic orbits, which explains why the emphasis will be put on the latter in this book.

2.12
Circle Maps

2.12.1
A New Global Topology

In Section 2.9.4 we saw that the global topology of phase space can have dramatic consequences: A period-3 orbit forces orbits of every period if it belongs to a map of an interval into itself, none if the state space is two-dimensional or is the unit circle. This indicates that qualitatively different behaviors can appear when phase space topology is changed. Accordingly, this section is devoted to a brief review of dynamical properties of maps from the unit circle S^1 into itself. If S^1 is parameterized with an angular variable $\theta \in [0, 1]$, these maps can be written as $\theta_{n+1} = f(\theta_n)$ (mod 1).

Physically, the study of circle maps is motivated by the problem of coupled oscillators. Assume that we have two systems oscillating on periodic cycles at frequencies ν and ν', respectively. The state of each oscillator can be described by an angular variable $\theta(t) = \nu t$ (mod 1). In the spirit of Poincaré sections, let us sample these angles stroboscopically at the frequency ν' so that we need only measure the successive samples $\theta_n = \theta(t_0 + n/\nu')$ of the first angle, given by

$$\theta_{n+1} = (\theta_n + w) \quad (\text{mod } 1) \tag{2.116}$$

where $w = \nu/\nu'$. Map (2.116) describes a rotation by a fraction w of a full turn per sampling period and is denoted $R(w)$ in what follows.

Two different qualitative behaviors can occur depending on the value of w. If w is a rational p/q with $p, q \in \mathbb{Z}$, we have that $\theta_{n+q} = \theta_n + qw$ (mod 1) $= \theta_n$: The dynamical regime is a periodic orbit, and θ_n takes only a finite number of different values. If w is irrational, the sequence $\{\theta_n\}$ fills densely the interval $[0, 1]$. This is a *quasiperiodic* regime and corresponds to the superposition of two incommensurate frequencies.

2.12.2
Frequency Locking and Arnold Tongues

It is known that the set of rational numbers is dense in $[0, 1]$ but that it has zero measure: The frequency ratio of two uncoupled oscillators is irrational with a probability of 1, even if one can find rational values arbitrarily close. However, this changes as soon as a coupling is introduced. One then observes *frequency locking*: The frequency ratio of the two oscillators remains fixed at a rational value p/q in a finite range $w \in [p/q - \Delta\rho_1, p/q + \Delta\rho_2]$.
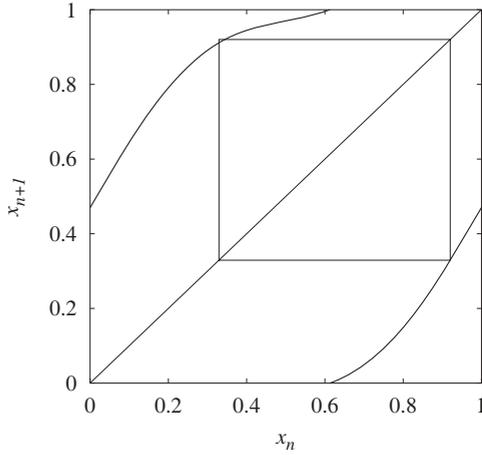
Figure 2.30 Graph of circle map (2.117) for $w = 0.47$ and $K = 0.8$. A period-2 orbit is also represented.

To study this phenomenon, the following circle map was introduced by Arnold [83]:

$$\theta_{n+1} = \left[\theta_n + w + \frac{K}{2\pi} \sin\left(2\pi\theta_n\right) \right] \quad (\text{mod } 1) \tag{2.117}$$

which features a nonlinear coupling characterized by its strength K. Figure 2.30 displays the graph of the map obtained for $(w, K) = (0.47, 0.8)$.

To describe the asymptotic regimes of (2.117), one introduces the *rotation number* [12, 52, 84]

$$\rho = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \Delta\theta_n \quad \text{with} \quad \Delta\theta_n = \left[w + \frac{K}{2}\pi \sin\left(2\pi\theta_n\right) \right] \tag{2.118}$$

Note that $\rho = w$ in the limit $K = 0$. The structure of the function $\rho(w, K)$ thus provides insight into the phenomenon of frequency locking as the nonlinear coupling is increased.

When the circle map (2.117) is a homeomorphism of S^1 into itself (i.e., for $K \le 1$), the following properties of the rotation number $\rho(w, K)$ can be established:

- The rotation number (2.118) does not depend on the orbit used to compute it.
- If $\rho(w, K)$ is irrational, the circle map is equivalent to the pure rotation $R(\rho)$; the motion is quasiperiodic.
- If $\rho(w, K) = p/q$ with p and q relatively prime integers, the asymptotic regime is a periodic orbit of period q. The periodic points of this orbit are ordered along the unit circle as with the pure rotation $R(p/q)$.

Thus, the classification of dynamical behaviors of the Arnold map for $K \le 1$ amounts to determining the parameter regions in the (w, K) plane, where $\rho(w, K)$ is rational.
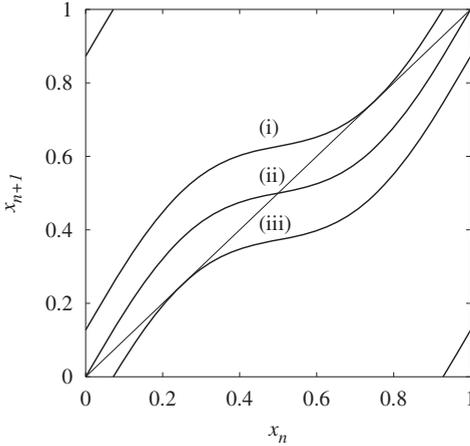
Figure 2.31 Graph of the circle map for $K = 0.8$ and (i) $w = 0.8/2\pi$; (ii) $w = 0$; (iii) $w = -0.8/2\pi$.

As a simple example, let us consider the region $\rho(w, K) = 0$, where the oscillator frequencies are locked to each other in a 1 : 1 ratio. The corresponding asymptotic regime is a fixed point $\theta_{n+1} = \theta_n$ whose location, according to (2.117), is given by the equation $w = -(K/2\pi)\sin\theta$. It is easy to see that for $w \in [-K/2\pi, +K/2\pi]$, there are two solutions, one of which is stable in the whole domain $0 \le K \le 1$, at least (Figure 2.31). Indeed, the slope of the graph at the two intersections is positive (the function is monotonic) and must be lower than 1 at one of the intersections. Hence, there is a periodic orbit having multiplier $0 \le \mu \le 1$. For $w = \pm K/2\pi$, the graph of the map is tangent to the diagonal, indicating that the stable and unstable periodic orbits are created and destroyed through saddle–node bifurcations. Note that the width of the frequency-locking interval $\rho = 0$ increases linearly with K and corresponds at $K = 1$ to almost one-third of the possible values of w.

By determining which regions of the (w, K) plane correspond to rotation numbers $\rho(w, K) = p/q$ with a small denominator q, the diagram shown in Figure 2.32 is obtained for $q \le 8$. The regions of frequency locking are called the *Arnold tongues*. Each of them corresponds to a different rational p/q, which governs the order in which they are encountered as w is increased at fixed K, since $\rho(w, K)$ is a monotonic function of w. As discussed above, tongues are bounded on both sides by saddle–node bifurcations where periodic orbits of the corresponding rotation number are created or destroyed.

It is interesting to note that the rotation numbers corresponding to the most important tongues can be classified according to a hierarchy based on an arithmetic operation on fractions. Indeed, it turns out that the principal tongue located between two tongues of rotation numbers p_1/q_1 and p_2/q_2 that satisfy $p_1 q_2 - p_2 q_1 = \pm 1$ is the one associated with the *Farey sum* of these two fractions, defined as follows: $p_1/q_1 \oplus p_2/q_2 = (p_1 + p_2)/(q_1 + q_2)$. Starting from the fundamental tongues $0/1$ and $1/1$, one first obtains the $1/2$ tongue. The latter is then separated from $0/1$
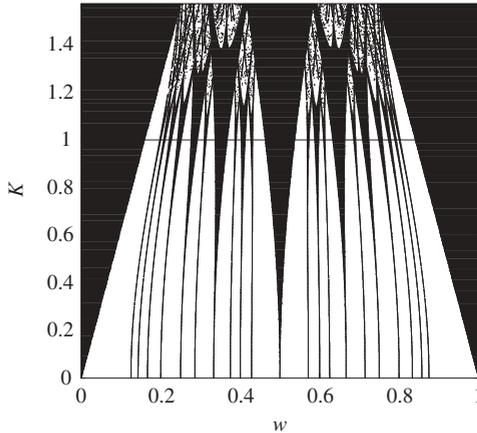
Figure 2.32 Arnold tongues for circle map (2.117) corresponding to rational rotation numbers $\rho(w, K) = p/q$ with $q \leq 8$.
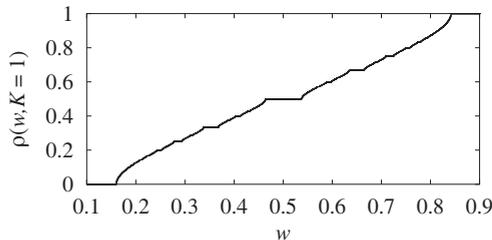
Figure 2.33 The graph of the rotation number $\rho(w, K = 1)$ is a devil's staircase.

by 1/3 and from 1/1 by 2/3. At the third level, one obtains 1/4, 2/5, 3/5, 3/4, and so on. Tongues at a given level are wider than those at the next levels, as can be verified by visual inspection of Figure 2.32.

As K is increased from 0 to 1, the relative proportions of the quasiperiodic and periodic regimes are exchanged. At $K = 0$, quasiperiodic regimes have a probability of 1, as mentioned above. Since there are an infinite number of tongues, it might not be obvious that the total length in w of the frequency-locked intervals goes to zero as $K \to 0$. That this is the case is due to the width $\Delta w(p/q)$ of the $\rho = p/q$ tongue decreasing sufficiently fast as $K \to 0$, more precisely as $\Delta w(p/q) \sim K^q$ or $\Delta w(p/q) \sim K^{q-1}$, depending on p/q.

At $K = 1$, values of w yielding quasiperiodic regimes are confined to a Cantor set of measure 0 and of fractal dimension $D \sim 0.87$; frequency-locked regimes have measure 1. The graph of the function $\rho(w, K = 1)$, shown in Figure 2.33, has a very peculiar structure, known as a *devil's staircase*. It is continuous and monotonic but increases only where ρ is irrational: Each rational value occurs on a finite interval. Moreover, it is self-similar: Any part of the graph contains a reduced copy of the entire graph. Incomplete devil's staircases are observed for $K < 1$ (i.e., the set of parameters yielding irrational rotation numbers then has positive measure).

2.12.3
Chaotic Circle Maps as Limits of Annulus Maps

The $K = 1$ line in the phase diagram of Figure 2.32 is called the *critical line*. Beyond it, circle map (2.117) has a point with zero derivative and hence is no longer invertible, which has dramatic consequences on the dynamics. On the one hand, there are no longer quasiperiodic regimes. Indeed, the latter are equivalent to a pure rotation with an irrational rotation number, which cannot be conjugate to a noninvertible map. On the other hand, more complex behavior can then appear, including chaos. Since map (2.117) has branches with negative slope, the stable periodic orbit can now have a negative multiplier and undergo a period doubling when the latter crosses -1. Most of the analysis carried out for the logistic map applies here: One observes cascades of period doubling leading to chaos. The white zones in the $K > 1$ part of Figure 2.32 correspond to chaotic regimes or to periodic regimes of high period.

Another consequence of noninvertibility is that the rotation number (2.118) now depends on the initial condition. Accordingly, a given regime is characterized by a *rotation interval* $[\rho^-, \rho^+]$ rather than by a single number. This corresponds to Arnold tongues gradually overlapping as K is increased above 1, as can be seen in Figure 2.32.

As discussed in the introduction to this section, invertible circle maps can be obtained rigorously as a first-return map when the dynamics is confined to a two-dimensional torus T^2. Obviously, this interpretation is not valid for noninvertible circle maps. However, just as the one-dimensional logistic map can be viewed as the infinitely dissipative of a two-dimensional horseshoelike invertible map, noninvertible circle maps can be thought of as limits of maps of an annulus into itself. Not that this interpretation is limited to circle maps having a degree of 1 (the image of the annulus winds once around the center). This is illustrated in Section 10.8 with the important example of the forced van der Pol oscillator.

2.13
Annulus Maps

Annulus maps have been represented in a number of different ways. One such map is shown in Figure 2.34. The expression for the map $(\rho, \theta) \rightarrow (\rho'', \theta'')$ is

$$\rho' = \rho_0 \left[\cos(\pi\theta) + 2\theta \right] \qquad \rho'' = \rho' + d\,(\rho - \rho_0)\,\frac{d\theta'}{d\theta}$$

$$\theta' = \omega + \theta + \frac{K}{2\pi}\sin(2\pi\theta) \qquad \theta'' = \theta' - d\,(\rho - \rho_0)\,\frac{d\rho'}{d\theta} \qquad (2.119)$$

In this map the two state variables are ρ and θ, $0 \leq \theta \leq 1$ (mod 1) with the endpoints identified. The latter is essentially the same as that for the circle map, while the new state variable ρ is the distance of a point from the origin in the phase space R^2. The two control parameters ω (rotation angle) and K (strength
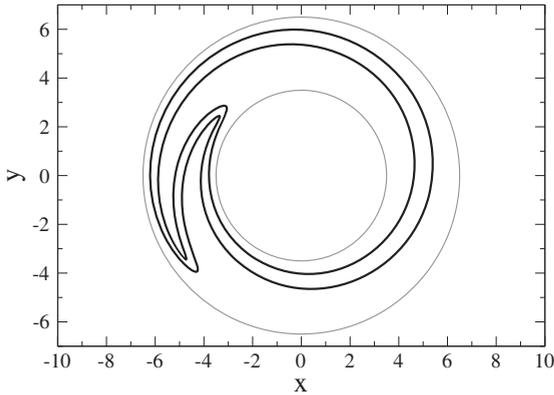
Figure 2.34 The transformation $(\rho, \theta) \to (\rho'', \theta'')$ invertibly maps the annular set (circles with radii $\rho_\pm = \rho_0 \pm \Delta\rho = 3.5, 6.5$) onto the deformed set, shown with darker lines inside the annulus. Parameter values: $(\rho_0, \omega, K/2\pi, d) = (5.0, 0.0, 2.0, 0.06)$.

of nonlinearity) have the same meaning as for the circle map. The new control parameter d is a measure of the dissipation: as $d \to 0$ we recover the circle map.

Annulus maps exhibit a variety of exciting behaviors. Some of this behavior is described by the Afraimovich–Shilnikov torus breakdown theorem [85, 86]. This theorem is local in nature, describing a spectrum of behaviors in the neighborhood of an Arnold tongue [83]. Annulus maps can support multiple coexisting basins of attraction. This is made manifest by the occurrence of hysteresis in bifurcation diagrams under some conditions. This type of global behavior is not covered by the Afraimovich–Shilnikov theorem.

We describe the Afraimovich–Shilnikov theorem as it applies to a period-3 Arnold tongue in an annulus map. The tongue in a two-control parameter plane is indicated in Figure 2.35. The Arnold tongue is bounded by two saddle–node bifurcation curves, labeled SNB_1 and SNB_2. Crossing these curves involves a direct or an inverse saddle–node bifurcation. For the period-3 tongue, this involves the creation or annihilation of a period-3 pair of orbits, one saddle and one node. "Opposite" the vertex of the Arnold tongue is a curve indicating the beginning of a period-doubling cascade to chaos. The larger the period q of the p/q tongue, the faster the transition through this region to chaos. Curves Ho_1^- and Ho_2^- indicate where there is a homoclinic connection between the stable and unstable manifolds of the period-3 saddle orbit. These curves connect the saddle–node curves with the period-doubling curve. Outside the tongue curves He_1 and He_2 locate the heteroclinic connection, the curve in control parameter space along which the unstable manifold of the saddle is tangent to the stable manifold of the node. Below this exterior region the behavior is quasiperiodic; above it is chaotic.

We now illustrate this behavior with the study by Letellier, Messager, and Gilmore [87] of the dynamics of an annulus map introduced by Curry and Yorke [88]. The Curry–Yorke map is defined as follows. Let Ψ_1 and Ψ_2 be two
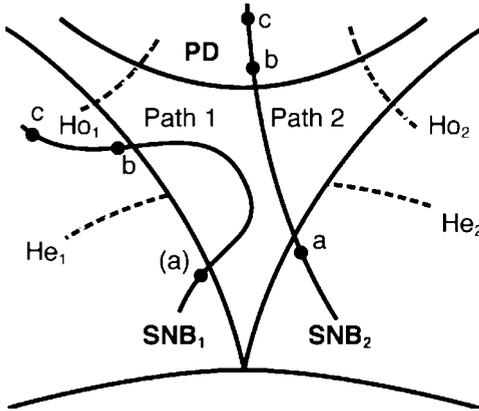
Figure 2.35 Schematic of an Arnold tongue in a control parameter plane and the kinds of behavior in its vicinity. SNB_1, SNB_2 are saddle–node bifurcation curves. PD is the first period-doubling bifurcation curve. Ho_1^-, Ho_2^- are curves indicating homo- clinic connections, tangencies of the stable, and unstable manifolds of the saddle cycle. He_1, He_2 are curves defining the heteroclinic connections of the unstable and stable manifolds of the saddle and node period-3 orbits. Path 1: Figure 2.36a–c. Path 2: Figure 2.36d–f.

homeomorphisms of the plane. Ψ_1 is defined in polar coordinates (ρ, θ) by

$$\Psi_1(\rho, \theta) = \left[\epsilon \ln(1 + \rho), \theta + \theta_0 \right] \tag{2.120}$$

and Ψ_2 is defined in cartesian coordinates by

$$\Psi_2(x, y) = \left(x, x^2 + y \right) \tag{2.121}$$

The Curry–Yorke map Ψ is obtained by composing these two maps: $\Psi = \Psi_2 \circ \Psi_1$. A complicated dynamics, typical of an annulus map, can be observed when this map is iterated.

In the left part of Figure 2.36 we show snapshots of the attractor at three points along Path 1 in Figure 2.35. In Figure 2.36a the motion is quasiperiodic on the torus, which has been deformed to triangular shape in anticipation of the impending saddle–node bifurcation that creates the stable period-3 node. The circles show the location of the real part of the saddle node pairs just before the bifurcation. After the bifurcation there is a stable period-3 orbit. It is not shown here, but the locations of the three points in this orbit are clear from the positions of their analytic continuations in Figure 2.36a (just before entering the tongue) and Figure 2.36b (just after leaving the tongue). The torus is wrinkled (see insets, Figure 2.36b) on leaving the tongue above the curve He_1. There is a "hard" transition to chaos. As the control parameter moves further away from the boundary of the tongue, the attractor becomes increasingly wrinkled, as shown in Figure 2.36c.

Path 2 enters the tongue by crossing the other saddle–node boundary curve SNB_2 below the heteroclinic curve He_2. Before entering the tongue the behavior is quasiperiodic. Once again the torus becomes deformed to a triangular shape
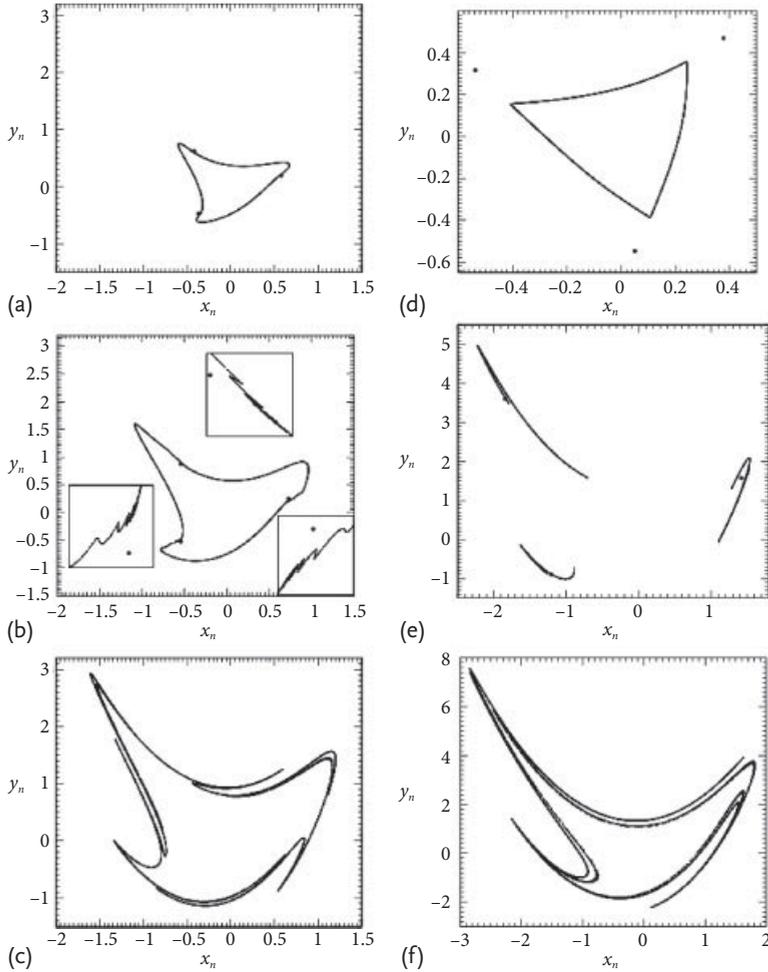
Figure 2.36 (a–c) Snapshots of the attractor at three points along Path 1 of Figure 2.35; (a) quasiperiodicity is evident just outside the left-hand boundary of the period-3 tongue and below the heteroclinic curve He_1; (b) toroidal chaotic behavior is exhibited after crossing the tongue boundary above He_1. Wrinkling is evident close to the remnants of the period-3 orbit (see insets): the real parts of the coordinates are shown here; (c) toroidal chaos becomes increasingly evident and the folding more pronounced as the path recedes from the tongue boundary; (d–f) snapshots of the attractor at three points along Path 2 of Figure 2.35; (a) quasiperiodicity is evident just outside the right-hand boundary of the period-3 tongue and below the heteroclinic curve He_2; (b) a banded period-3 chaotic attractor occurs after passage through the period-doubling cascade and beyond the inverse noisy period-halving bifurcations; (c) toroidal chaos occurs after one or more crises increase the size of the attractor. After Letellier, Messager, and Gilmore [87].

in anticipation of the impending saddle–node bifurcation. The dots shown in Figure 2.36d indicate the locations of the real part of the period-3 orbit. The path exits the tongue through the period-doubling region. After the accumulation curve is

crossed, a series of noisy period-halving bifurcations ensues, creating a banded (three bands) chaotic attractor. This is shown in Figure 2.36e. The locations of the period-3 orbit, stable inside the tongue but now unstable, are shown as dots. After one or more crises have taken place, a chaotic toroidal attractor is formed. This is shown in Figure 2.36f.

2.14
Summary

Although maps are very simple dynamical systems, they display most of the key features of chaos. This has allowed us to become familiar with concepts that will appear throughout this book, without excessive mathematical complexity.

Even the simplest dynamical system that one can think of, the logistic map, is able to reproduce surprisingly well qualitative behaviors that are observed experimentally in the laser system described in Chapter 1. As a control parameter is varied, it experiences bifurcations, in particular a period-doubling cascade leading to chaos, and a variety of chaotic regimes.

The simple structure of the logistic map makes it possible to study one of the basic mechanisms responsible for chaotic behavior, namely stretching. In its most chaotic regime, the logistic map is basically a "multiply by two" machine. This has far-reaching consequences: sensitivity to initial conditions, existence of an infinite number of periodic orbits that are dense in the invariant set, and so on.

Stretching is at the root of an extremely powerful tool for unfolding chaos, symbolic dynamics. Thanks to the unlimited magnification provided by sensitivity to initial conditions, a series of coarse-grained measurements of the system state suffices to specify it with arbitrary accuracy. Symbolic dynamics not only allows us to classify orbits (e.g., how many periodic orbits of period p?) but also to study their genealogies (e.g.: In what order do orbits appear? Which orbit is a period-doubled orbit from another?). By studying the grammar of a chaotic system, we can classify regimes and compute quantitative invariants such as topological entropy. Not all of the results obtained (e.g., the universal sequence) can be directly extended to higher dimensions. That there are topological invariants (e.g., permutations) both of which are deeply related to symbolic dynamics and play a major role in forcing relations will later prove to be a key property.

The logistic map is a noninvertible system. Many physical systems are described by ODEs and thus are associated with invertible maps, such as the Hénon map. A chaotic invertible map shares many properties with noninvertible ones. In particular, the dynamics in the unstable space is associated with a noninvertible map, as the example of the horseshoe map shows. There are also new problems, such as constructing relevant symbolic encodings in that case. Finally, global phase space topology can have a profound influence on phenomena observed, as exemplified by circle and annulus maps.