CHAPTER 7

Simple Systems with Complicated Orbit Structure

This chapter presents a rich array of properties of a collection of examples. Its coherence derives from the fact that it is part of a general theory we outline in Chapter 10. The examples (other than the quadratic map f_4) are instances of hyperbolic dynamical systems (or symbolic dynamical systems), and the properties we derive here are largely properties common to hyperbolic and symbolic dynamical systems.

7.1 GROWTH OF PERIODIC POINTS

Periodic orbits represent the most distinctive special class of orbits. So far we have mostly encountered maps with few periodic orbits or, as in the case of a rational rotation, only periodic orbits. In these basic examples different periods did not appear for the same map. Even the most complex situations so far still involve periodic orbits neatly organized by period in families such as invariant curves in plane rotations, linear twists, the time-1 map for the mathematical pendulum, or billiards. There we placed more emphasis on coherence than complexity. Now we encounter the first examples with a different periodic pattern. In these cases, when periodic points of different periods are present, we want to count them.

Definition 7.1.1 For a map $f\colon X \to X$, let $P_n(f)$ be the number of periodic points of f with (not necessarily minimal) period n, that is, the number of fixed points for f^n.

This section introduces numerous new examples of dynamical systems. For now they are introduced with a view to their periodic orbit structure, but in due time numerous other fascinating features of their orbit structure will emerge.

7.1.1 Linear Expanding Maps

Consider the noninvertible map E_2 of the circle given in multiplicative notation by

$$E_2(z) = z^2, \qquad |z| = 1,$$

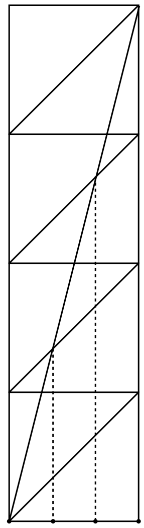

Figure 7.1.1. Periodic points for an expanding map.

and in additive notation by

(7.1.1) $$E_2(x) = 2x \pmod 1.$$

Proposition 7.1.2 $P_n(E_2) = 2^n - 1$ *and periodic points for* E_2 *are dense in* S^1.

Proof If $E_2^n(z) = z$, then $z^{2^n} = z$, and $z^{2^n-1} = 1$. Thus every root of unity of order $2^n - 1$ is a periodic point for E_2 of period n. There are exactly $2^n - 1$ of these, and they are uniformly spread over the circle with equal intervals. In particular, when n becomes large these intervals become small. (See Figure 7.1.1) □

We see from Proposition 7.1.2 that a natural measure of asymptotic growth of the number of periodic points is the exponential growth rate $p(f)$ for the sequence $P_n(f)$:

(7.1.2) $$p(f) = \overline{\lim_{n\to\infty}} \frac{\log_+ P_n(f)}{n},$$

where $\log_+(x) = \log(x)$ for $x \geq 1$, 0 otherwise. In particular, Proposition 7.1.2 shows that $p(E_2) = \overline{\lim}_{n\to\infty}(\log 2^n + \log(1 - 2^{-n}))/n = \log 2$.

The maps

$$E_m \colon x \mapsto mx \pmod 1,$$

where m is an integer of absolute value greater than one, represent a straightforward generalization of the map E_2. Not surprisingly, these maps also have dense sets of periodic orbits. The proof of Proposition 7.1.2 holds verbatim with the replacement of 2 by m:

Proposition 7.1.3 $P_n(E_m) = |m^n - 1|$ *and periodic points for* E_m *are dense.*

Proof $z = E_m^n(z) = z^{m^n}$ has $|m^n - 1|$ solutions that are evenly spaced. □

See also Section 7.1.3.

Another property of the maps E_m worth noticing is preservation of length similar to the property of preservation of phase volume discussed in Section 6.1.

Naturally, the length of an image of any arc increases; however, if one considers the *complete preimage* of an arc Δ under E_m, one immediately sees that it consists of $|m|$ arcs of length $l(\Delta)/|m|$ each, placed along the circle at equal distances. The analysis in Section 6.1.2 can be extended to noninvertible volume-preserving maps, so recurrent points are dense in this situation as well.

7.1.2 Quadratic and Quadratic-Like Maps

For $\lambda \in \mathbb{R}$, let $f_\lambda \colon \mathbb{R} \to \mathbb{R}$, $f_\lambda(x) := \lambda x(1-x)$. For $0 \le \lambda \le 4$, the f_λ map the unit interval $I = [0, 1]$ into itself. The family f_λ is referred to as the *quadratic family.* For $\lambda \le 3$, this family was discussed in detail in Section 2.5, and the asymptotic behavior for any such λ is fairly simple and changes with λ only a few times. As it turns out, for the remaining interval of parameter values the quadratic family develops a bewildering array of complicated but different types of behavior, which change with caleidoscopic speed (see Figure 7.1.2 and Chapter 11). Note that $P_n(f_\lambda) \le 2^n$ because the nth iterate of f_λ is a polynomial of degree 2^n, and hence the equation $(f_\lambda)^n(x) = x$ has at most 2^n solutions. While one may expect that in the complex plane this equation would indeed have exactly 2^n solutions for most values of the parameter λ, this is certainly not the case for real solutions.

Here we consider the behavior of the quadratic family for large values of the parameter, namely, $\lambda \ge 4$. While for $\lambda > 4$ the interval $[0, 1]$ is not preserved, the set of points that remains in that interval is still quite interesting.

The analysis of the behavior of the quadratic family on the unit interval for $0 \le \lambda \le 3$ carried out in Section 2.5 showed simple periodic patterns: Only points of periods 1 and 2 appear, and their number is small. With moderate effort this analysis can be extended as far as $\lambda = 1 + \sqrt{6}$ (Proposition 11.2.1). On the other hand, we have:

Proposition 7.1.4 *For $\lambda \ge 4$ we have $P_n(f_\lambda) = 2^n$.*

Proof Since $P_n(f_\lambda) \le 2^n$, it suffices to prove the reverse inequality. To that end we use the following observation: If $f \colon \mathbb{R} \to \mathbb{R}$ is continuous and $\Delta \subset [0, 1]$ is an interval such that one endpoint is mapped to 0 and the other to 1, then by the Intermediate-Value Theorem there is a fixed point of f in Δ. Now $[0, 1] \subset [f_\lambda(0), f_\lambda(1/2)]$ and $[0, 1] \subset [f_\lambda(1/2), f_\lambda(1)]$, so there are intervals $\Delta_0 \subset [0, 1/2]$ and $\Delta_1 \subset [1/2, 1]$ whose images under f_λ are exactly $[0, 1]$, giving us two fixed points for f. The nonzero fixed

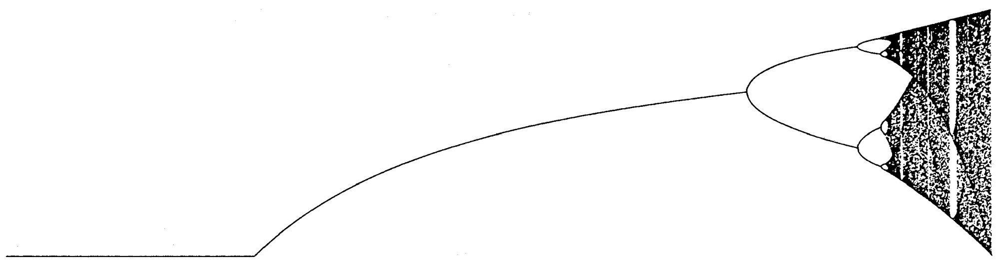

Figure 7.1.2. Bifurcation diagram.

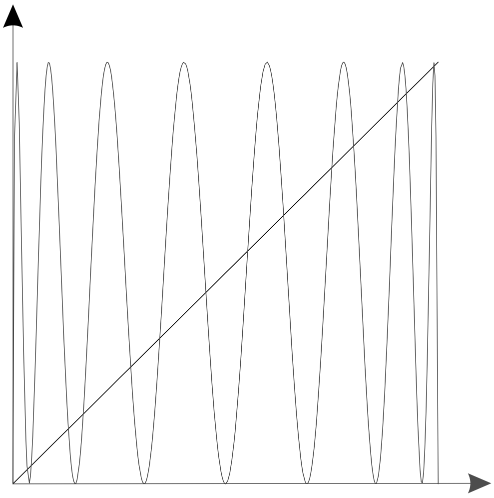

Figure 7.1.3. Periodic points of f_4.

point is indeed in the interior of Δ_1 because the right endpoint of Δ_1 is 1 and hence is mapped to 0, so the other endpoint is mapped to 1 and therefore neither are fixed.

Furthermore, the preimages of Δ_0 and Δ_1 under f consist of two intervals each, so there are four intervals whose images under f^2 are exactly $[0, 1]$. Each contains a fixed point of f_λ^2, again every one except 0 being in the interior of the corresponding interval, so no two of these fixed points coincide.

Repeating this argument successively for higher iterates of f_λ we obtain 2^n intervals whose images under f_λ^n are $[0, 1]$, and each of which therefore contains at least one fixed point, giving 2^n distinct orbits of period n for f_λ. □

It is useful that the argument to show that $P_n(f_\lambda) \geq 2^n$ applies to any continuous map $f\colon [0, 1] \to \mathbb{R}$ with $f(0) = f(1) = 0$ and such that there is a $c \in [0, 1]$ with $f(c) \geq 1$. In this more general case it is somewhat more convenient, however, to talk about intervals whose images under f^n contain $[0, 1]$ rather than being exactly $[0, 1]$.

In the quadratic case (for $\lambda > 4$) one can refine the preceding argument slightly to show that there are exactly 2^n periodic points (rather than using that the degree of f^n is 2^n). This also works for some continuous maps f of this more general nature, which are monotone on $[0, c]$ as well as $[c, 1]$. A continuous map defined on an interval that is increasing to the left of an interior point and decreasing thereafter is said to be *unimodal*. Thus we have found

Proposition 7.1.5 *If $f\colon [0, 1] \to \mathbb{R}$ is continuous, $f(0) = f(1) = 0$, and there exists $c \in [0, 1]$ such that $f(c) > 1$, then $P_n(f) \geq 2^n$. If, in addition, f is unimodal and expanding (that is, $|f(x) - f(y)| > |x - y|$) on each interval of $f^{-1}((0, 1))$, then $P_n(f) = 2^n$.*

The heart of the proof is the following lemma:

Lemma 7.1.6 *Denote by $\mathcal{M}_k$ the collection of continuous maps $f\colon [0, 1] \to \mathbb{R}$ such that $f^{-1}((0, 1)) = \bigcup_{i=1}^k I_i$ with $I_i \subset [0, 1]$ open intervals, f monotonic on I_i, and $f(I_i) = (0, 1)$. Then $f \circ g \in \mathcal{M}_{kl}$ whenever $f \in \mathcal{M}_k$ and $g \in \mathcal{M}_l$.*

Proof If $f \in \mathcal{M}_k$ and $g \in \mathcal{M}_l$, then $f^{-1}((0,1)) = \bigcup_{i=1}^k I_i$ and $g^{-1}(I_i) = \bigcup_{j=1}^l J_{ij}$ with $\{J_{ij} \mid 1 \le i \le k,\ \ 1 \le j \le l\}$ pairwise disjoint and $(f \circ g)^{-1}((0,1)) = \bigcup_{ij} J_{ij}$. The composition $f \circ g$ is monotonic on J_{ij} and $f \circ g(J_{ij}) = (0,1)$. □

Proof of Proposition 7.1.5 The lemma shows that $P_n(f) \ge k^n$ for $f \in \mathcal{M}_k$ because $f^n \in \mathcal{M}_{k^n}$. If f is expanding on every interval of $f^{-1}((0,1))$, then the same holds for iterates of f. This shows that on each of those intervals there is at most one solution of $f^n(x) = x$. Therefore, $P_n(f) \le k^n$, proving equality. □

7.1.3 Expanding Maps and Degree

Next we consider a nonlinear generalization of the expanding maps E_m. We use additive notation for circle maps. In this notation derivatives of maps can be expressed as real-valued functions.

Definition 7.1.7 A continuously differentiable map $f\colon S^1 \to S^1$ is said to be an *expanding* map if $|f'(x)| > 1$ for all $x \in S^1$.

Since f' is continuous and periodic, the minimum of $|f'|$ is attained and hence is greater than 1.

Proposition 4.3.1 gives us a function $F\colon \mathbb{R} \to \mathbb{R}$ that satisfies $[F(x)] = f([x])$ and $F(s+1) = F(s) + \deg(f)$, where $\deg(f)$ is the *degree* of f. It has the following simple property:

Lemma 7.1.8 *If $f, g\colon S^1 \to S^1$ are continuous, then* $\deg(g \circ f) = \deg(f)\deg(g)$, *in particular* $\deg(f^n) = \deg(f)^n$.

Proof If F, G are lifts of f and g, respectively, then $G(s+k) = G(s+k-1) + \deg(g) = \cdots = G(s) + k\deg(g)$ and $G(F(s+1)) = G(F(s) + \deg(f)) = G(F(s)) + \deg(g)\deg(f)$. □

This property is useful for counting periodic points.

Proposition 7.1.9 *If $f\colon S^1 \to S^1$ is an expanding map, then $|\deg(f)| > 1$ and $P_n(f) = |\deg(f)^n - 1|$.*

Proof $|f'| > 1$ implies $|F'| > 1$ for any lift, so, by the Mean-Value Theorem A.2.3, $|\deg(f)| = |F(x+1) - F(x)| > 1$. By the chain rule an iterate of an expanding map is itself expanding, so by Lemma 7.1.8 it suffices to consider the case $n = 1$. Take a lift F of f and consider it on the interval $[0,1]$. The fixed points of f are the projections of the points x for which $F(x) - x \in \mathbb{Z}$. The function $g(x) := F(x) - x$ satisfies $g(1) = g(0) + \deg(f) - 1$, so by the Intermediate-Value Theorem there are at least $|\deg(f) - 1|$ points x where $g(x) \in \mathbb{Z}$. If $g(0) \in \mathbb{Z}$, then there are $|\deg(f) - 1| + 1$ such points, but 0 and 1 project to the same point on S^1. Now $g'(x) \ne 0$, so g is strictly monotone and hence takes every value at most once. Thus there are exactly $|\deg(f) - 1|$ fixed points on S^1. □

This proposition in particular establishes part of the analog of Proposition 7.1.2 for E_m.

Similarly to quadratic maps, the argument that shows $P_n(f) \geq |\deg(f)^n - 1|$ works for any continuous map. It is trivial for maps of degree 1 because the assertion is vacuous. Indeed, irrational rotations do not have any fixed or periodic points. For maps of degree 0 it merely guarantees a fixed point. For maps f with $|\deg(f)| > 1$, however, this result gives exponential growth of the number of periodic points: $p(f) \geq \log_+(|\deg(f)|)$.

7.1.4 Hyperbolic Linear Map of the Torus

The previous examples were all one-dimensional, but the patterns of the growth and distribution of periodic points observed in those examples also appear in higher dimension.

A convenient model to demonstrate this is built from the following linear map of $\mathbb{R}^2$:

$$L(x, y) = (2x + y, x + y) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

If two vectors (x, y) and (x', y') represent the same element of $\mathbb{T}^2$, that is, if $(x - x', y - y') \in \mathbb{Z}^2$, then $L(x, y) - L(x', y') \in \mathbb{Z}^2$, so $L(x, y)$ and $L(x', y')$ also represent the same element of $\mathbb{T}^2$. Thus L defines a map $F_L \colon \mathbb{T}^2 \to \mathbb{T}^2$:

$$F_L(x, y) = (2x + y, x + y) \pmod 1.$$

The map F_L is, in fact, an automorphism of the torus viewed as an additive group. It is invertible because the matrix $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ has determinant one, so L^{-1} also has integer entries [in fact $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 2 \end{smallmatrix}\right)$] and hence defines a map $F_{L^{-1}} = F_L^{-1}$ on $\mathbb{T}^2$ by the same argument. The eigenvalues of L are

$$\lambda_1 = \frac{3 + \sqrt{5}}{2} > 1 \quad \text{and} \quad \lambda_1^{-1} = \lambda_2 = \frac{3 - \sqrt{5}}{2} < 1. \tag{7.1.3}$$

Figure 7.1.4 gives an idea of the action of F_L on the fundamental square $I = \{(x, y) \mid 0 \leq x < 1,\ 0 \leq y \leq 1\}$. The lines with arrows are the eigendirections. For any matrix L with determinant ± 1, the map F_L preserves the area of sets on the torus.

Proposition 7.1.10 *Periodic points of F_L are dense and $P_n(F_L) = \lambda_1^n + \lambda_1^{-n} - 2$.*

Proof To obtain density we show that points with rational coordinates are periodic points. Let $x, y \in \mathbb{Q}$. Taking the common denominator write $x = s/q$, $y = t/q$, where $s, t, q \in \mathbb{Z}$. Then $F_L(s/q, t/q) = ((2s + t)/q, (s + t)/q)$ is a rational point whose coordinates also have denominator q. But there are only q^2 different points on $\mathbb{T}^2$ whose coordinates can be represented as rational numbers with denominator q, and all iterates $F_L^n(s/q, t/q)$, $n = 0, 1, 2 \ldots$, belong to that finite set. Thus they must repeat, that is, $F_L^n(s/q, t/q) = F_L^m(s/q, t/q)$ for some $n, m \in \mathbb{Z}$. But since F_L is invertible, $F_L^{n-m}(s/q, t/q) = (s/q, t/q)$ and $(s/q, t/q)$ is a periodic point, as required. This gives density. (By contrast, not all rational points are periodic for E_m. See Exercise 7.1.1.)

Next we show that points with rational coordinates are the only periodic points for F_L. Write $F_L^n(x, y) = (ax + by,\ cx + dy) \pmod 1$, where $a, b, c, d \in \mathbb{Z}$. If

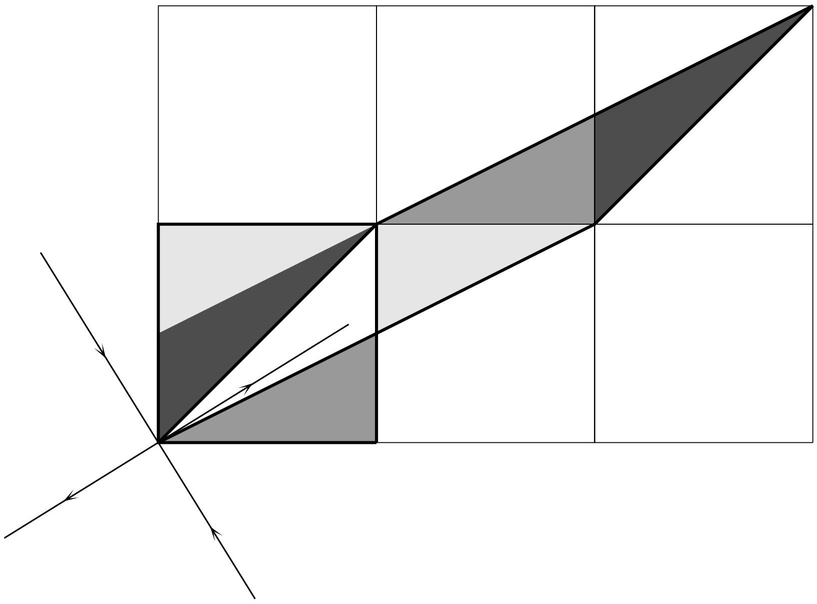

Figure 7.1.4. The hyperbolic toral map.

$F_L^n(x, y) = (x, y)$, then

$$ax + by = x + k,$$

$$cx + dy = y + l$$

for $k, l \in \mathbb{Z}$. Since 1 is not an eigenvalue for L^n, we can solve for (x, y):

$$x = \frac{(d-1)k - bl}{(a-1)(d-1) - cb}, \qquad y = \frac{(a-1)l - ck}{(a-1)(d-1) - cb}.$$

Thus $x, y \in \mathbb{Q}$.

Now we calculate $P_n(F_L)$. The map

$$G = F_L^n - \mathrm{Id}\colon (x, y) \mapsto ((a-1)x + by,\ cx + (d-1)y) \pmod 1$$

is a well-defined noninvertible map of the torus onto itself. As before, if $F_L^n(x, y) = (x, y)$, then $(a-1)x + by$ and $cx + (d-1)y$ are integers; hence $G(x, y) = 0 \pmod 1$, that is, the fixed points of F_L^n are exactly the preimages of the point $(0, 0)$ under G. Modulo 1 these are exactly the points of $\mathbb{Z}^2$ in $(L^n - \mathrm{Id})([0, 1) \times [0, 1))$. We presently show that their number is given by the area of $(L^n - \mathrm{Id})([0, 1) \times [0, 1))$, which is $|\det(L^n - \mathrm{Id})| = |(\lambda_1^n - 1)(\lambda_1^{-n} - 1)| = \lambda_1^n + \lambda_1^{-n} - 2$. □

Lemma 7.1.11 *The area of a parallelogram with integer vertices is the number of lattice points it contains, where points on the edges are counted as half, and all vertices count as a single point.*

Proof Denote the area of the parallelogram by A. Adding the number of lattice points it contains in the prescribed way gives an integer N, which is the same for any translate of the parallelogram.

Now consider the canonical tiling of the plane by copies of this parallelogram translated by integer multiples of the edges. Denote by l the longest diagonal. The area of the tiles can be determined in a backward way by determining how many tiles lie in the square $[0, n) \times [0, n)$ for $n > 2l$. Those that lie inside cover the smaller square $[l, n-l) \times [l, n-l)$ completely, so there are at least

$$\frac{(n-2l)^2}{A} \geq \frac{n^2}{A}\left(1-\frac{4l}{n}\right).$$

Since any tile that intersects the square is contained in $[l, n-l) \times [l, n-l)$, there are at most

$$\frac{(n+2l)^2}{A} = \frac{n^2}{A}\left(1+\frac{4l}{n}\left(1+\frac{l}{n}\right)\right) < \frac{n^2}{A}\left(1+\frac{6l}{n}\right).$$

The number n^2 of integer points in the square is at least the number of points in tiles in the square and at most the number of points in tiles that intersect the square. Therefore

$$N \cdot \frac{n^2}{A}\left(1-\frac{4l}{n}\right) \leq n^2 \leq N \cdot \frac{n^2}{A}\left(1+\frac{6l}{n}\right) \qquad \text{and} \qquad 1-\frac{4l}{n} \leq \frac{A}{N} \leq 1+\frac{6l}{n}$$

for all $n > 2l$.

This shows that $N = A$. □

7.1.5 Inverse Limits

The closest invertible analog to E_2 so far is the toral automorphism induced by $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. We digress briefly to describe a general construction that "makes a map invertible", that is, it that produces an invertible map from a noninvertible one in a standard way. The way to overcome noninvertibility is to replace the points of the given space by sequences $(x_n)_{n\in\mathbb{Z}}$ with $f(x_n) = x_{n+1}$. This way the ambiguity about preimages is resolved by listing the entire orbit explicitly. Indeed, the map $F((x_n)_{n\in\mathbb{Z}}) := (x_{n+1})_{n\in\mathbb{Z}}$ is clearly invertible.

Definition 7.1.12 If X is a metric space and $f\colon X \to X$ continuous, then the *inverse limit* is defined on the space

$$X' := \{(x_n)_{n\in\mathbb{Z}} \mid x_n \in X \text{ and } f(x_n) = x_{n+1} \text{ for all } n \in \mathbb{Z}\}$$

by $F((x_n)_{n\in\mathbb{Z}}) := (x_{n+1})_{n\in\mathbb{Z}}$.

Consider explicitly $f = E_2$ on S^1. Then the inverse limit is the space

$$\mathbb{S} := \{(x_n)_{n\in\mathbb{Z}} \mid x_n \in X \text{ and } f(x_n) = x_{n+1} \text{ for all } n \in \mathbb{Z}\}$$

with the map $F((x_n)_{n\in\mathbb{Z}}) := (x_{n+1})_{n\in\mathbb{Z}} = (2x_n)_{n\in\mathbb{Z}} \pmod 1$. This is called the *solenoid*.

Compared to listing the whole sequence, there is a more economical way to identify a point in $\mathbb{S}$. Once an entry, such as x_0, is specified, all subsequent entries are uniquely determined (by the orbit of $x_0 \in S^1$ under E_2). In order to specify all previous members of the sequence, one need only choose (recursively) one of two preimages at each step. For any given x_0 this can be coded by a one-sided 0-1-sequence. Since the space Ω_2 of these is a Cantor set (Section 7.3.5), the solenoid $\mathbb{S}$ is locally the product of an interval (points on S^1 near x_0) and a Cantor set.

There is a beautiful way to visualize the inverse limit construction. Beginning with a circle of "initial conditions" x_0, there are "twice as many" possible preimages x_{-1}, so the circle has to be doubled up like a rubber band around a newspaper. But there are twice as many second preimages, and so on, so it is necessary to double up ad infinitum. This is analogous to the construction of the ternary Cantor set, where an interval becomes two, then four, and so on.

The definitive geometric realization is carried out in Section 13.2 and illustrated in Figure 13.2.1 and on the cover of this book. This picture is representative of a great wealth of ideas in dynamics and deserves to be an icon for chaotic dynamics. Together with the horseshoe and linear toral automorphisms, the expanding map E_2 and the solenoid are the most tractable representatives of hyperbolic dynamical systems, and these have provided the framework of concepts and techniques within which each chaotic dynamical system is studied and described. This framework is developed in this chapter and the next, and it is described further in Chapter 10.

EXERCISES

Exercise 7.1.1 Prove that for the expanding map $E_m(|m| \geq 2)$ rational points are preimages of periodic points ("eventually periodic").

Exercise 7.1.2 Find a necessary and sufficient condition for a rational point to be periodic under E_m.

Exercise 7.1.3 Carry out the proof of Proposition 7.1.3 for the case $m < -1$.

Exercise 7.1.4 Prove that for any $n \in \mathbb{N}$ and $\lambda \geq 4$ the quadratic map f_λ has a periodic point whose *minimal* period (Definition 2.2.6) is n.

Exercise 7.1.5 Give an example of a continuous map $f\colon [0,1] \to \mathbb{R}$ with $f(0) = f(1) = 0$ for which there exists $c \in [0,1]$ such that $f(c) > 1$, and such that $P_n(f) > 2^n$.

Exercise 7.1.6 Give an example of a smooth unimodal map f such that $P_n(f) < 2^n$.

Exercise 7.1.7 Show that a continuous map f of S^1 can be deformed to $E_{\deg(f)}$, that is, that there is a continuous map $F\colon [0,1] \times S^1 \to S^1$ with $F(0,\cdot) = E_{\deg(f)}$ and $F(1,\cdot) = f$.

Exercise 7.1.8 Show that maps of different degrees cannot be deformed into each other, that is, that there is no continuous map $F\colon [0,1] \times S^1 \to S^1$ such that $\deg(F(0,\cdot)) \neq \deg(F(1,\cdot))$.

Exercise 7.1.9 Suppose $f\colon S^1 \to S^1$ has degree 2 and 0 is an attracting fixed point. Show that $P_n(f) > 2^n$.

Exercise 7.1.10 Consider the Fibonacci sequence from Section 1.2.2, Example 2.2.9, and Section 3.1.9. Show that the sequence obtained from taking the last digit of each Fibonacci number is periodic.

Exercise 7.1.11 Apply the inverse limit construction to a homeomorphism and prove that the result is naturally equivalent to the original system.

PROBLEMS FOR FURTHER STUDY

Problem 7.1.12 Prove that the solenoid in Section 7.1.5 is connected but not path-connected.

7.2 TOPOLOGICAL TRANSITIVITY AND CHAOS

We will show that some of the examples considered in the previous section are topologically transitive in the sense of Definition 4.1.3, that is, they have dense orbits. That there are at the same time infinitely many periodic points makes these examples different from irrational rotations and the other topologically transitive examples of Chapter 4 and Chapter 5. In expanding maps and hyperbolic linear maps of the torus we even found that the union of the periodic points is dense, which means that dense and periodic orbits are inextricably intertwined.

Thus, the global orbit structure is far more complex in these examples. This intertwining of density and periodicity is an essential feature of the complexity of the orbit structure. It causes sensitive dependence of any orbit on its initial conditions (see Definition 7.2.11 and Theorem 7.2.12), which is regarded as an essential ingredient of *chaos*.

Definition 7.2.1 A continuous map $f\colon X \to X$ of a metric space is said to be *chaotic* if it is topologically transitive and its periodic points are dense.[1]

Circle rotations show that neither condition alone gives much complexity.

We will show presently that expanding and hyperbolic maps are chaotic. In fact, we show the stronger property of topological mixing (Definition 7.2.5), which is absent in the minimal examples of Chapter 4 and Chapter 5. Before introducing the mixing property, we give an alternative definition of topological transitivity.

7.2.1 A Criterion for Topological Transitivity

We defined topological transitivity as the existence of a dense orbit. However, it is useful to have an alternate characterization in terms of subsets of phase space. In order to include noninvertible maps, we say that a sequence $(x_i)_{i\in\mathbb{Z}}$ is an orbit of f if $f(x_i) = x_{i+1}$ for all $i \in \mathbb{Z}$. However, we simply write $f^i(x)$ for $i \in \mathbb{Z}$ anyway to keep the notations more familiar.

Proposition 7.2.2 *Let X be a complete separable (that is, there is a countable dense subset) metric space with no isolated points. If $f\colon X \to X$ is a continuous map, then the following four conditions are equivalent:*

(1) *f is topologically transitive, that is, it has a dense orbit.*
(2) *f has a dense positive semiorbit.*
(3) *If $\varnothing \neq U, V \subset X$, then there exists an $N \in \mathbb{Z}$ such that $f^N(U) \cap V \neq \varnothing$.*
(4) *If $\varnothing \neq U, V \subset X$, then there exists an $N \in \mathbb{N}$ such that $f^N(U) \cap V \neq \varnothing$.*

Of course, the implications (4) $\Rightarrow$ (3) and (2) $\Rightarrow$ (1) are clear. To show which hypotheses are needed for which of the remaining directions, we prove Proposition 7.2.2 in the following form.

[1] There is no universally accepted definition of chaos, but this definition is equivalent to the one most commonly found in expository literature, which was put forward by Robert Devaney.

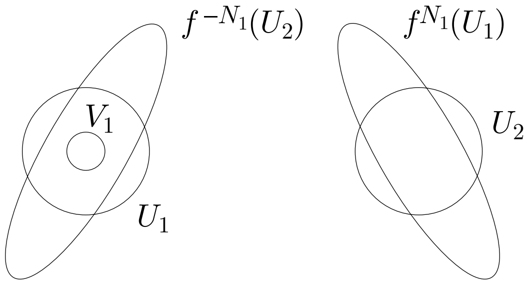

Figure 7.2.1. Construction for the proof.

Lemma 7.2.3 *Let X be a metric space and $f\colon X \to X$ a continuous map. Then* (1) *implies* (3). *If X has no isolated points, then* (1) *implies* (4). *If X is separable, then* (3) *implies* (1) *and* (4) *implies* (2).

Proof Let f be topologically transitive and suppose the orbit of $x \in X$ is dense. Then there exists an $n \in \mathbb{Z}$ such that $f^n(x) \in U$, and there is an $m \in \mathbb{Z}$ such that $f^m(x) \in V$; hence $f^{m-n}(U) \cap V \neq \varnothing$. This implies (3).

If we can choose $m > n$, then by taking $N := m - n$ we have even established (4). Otherwise we use the assumption that X has no isolated points, so $f^m(x)$ is not an isolated point and therefore there are $n_k \in \mathbb{Z}$ such that $|n_k| \to \infty$, $f^{n_k}(x) \in V$, and $f^{n_k}(x) \to f^m(x)$ as $k \to \infty$. Indeed, $n_k \to -\infty$ since $n_k \le n$ by assumption (otherwise we are in the first case), so we can choose an $m' < 2m - n$ from among the n_k such that $f^{m'}(x) \in f^{m-n}(U)$. Then $x' := f^{n-m}(f^{m'}(x)) \in U$ and $f^{2m-n-m'}(x') = f^m(x) \in V$, so $f^N(U) \cap V \neq \varnothing$ with $N := 2m - n - m' \in \mathbb{N}$. Thus (1) $\Rightarrow$ (4) if X has no isolated points.

Now assume separability and one of the intersection conditions (3) and (4). We give one argument to prove both that (3) implies (1) and (4) implies (2). For a countable dense subset $S \subset X$, let $U_1, U_2, \dots$ be the countable collection of balls centered at points of S with rational radius. We need to construct an orbit or semiorbit that intersects every U_n. By (3) there exists $N_1 \in \mathbb{Z}$ such that $f^{N_1}(U_1) \cap U_2 \neq \varnothing$. If (4) holds, we can take $N_1 \in \mathbb{N}$. Let V_1 be an open ball of radius at most $1/2$ such that $\bar{V}_1 \subset U_1 \cap f^{-N_1}(U_2)$. (See Figure 7.2.1.) There exists $N_2 \in \mathbb{Z}$ such that $f^{N_2}(V_1) \cap U_3$ is nonempty, and, if (4) holds, we can take $N_2 \in \mathbb{N}$. Again, take an open ball V_2 of radius at most $1/4$ such that $\bar{V}_2 \subset V_1 \cap f^{-N_2}(U_3)$. By induction, we construct a nested sequence of open balls V_n of radii at most 2^{-n} such that $\bar{V}_{n+1} \subset V_n \cap f^{-N_{n+1}}(U_{n+2})$. The centers of these balls form a Cauchy sequence whose limit x is the unique point in the intersection $V = \bigcap_{n=1}^{\infty} \bar{V}_n = \bigcap_{n=1}^{\infty} V_n$. Then $f^{N_{n-1}}(x) \in U_n$ for every $n \in \mathbb{N}$, and all $N_n \in \mathbb{N}$ if (4) holds.

If f is noninvertible, the last step may involve choices for negative values of N_n: Take i_k such that $N_{i_k} < 0$ for all k and $N_{i_{k+1}} < N_{i_k}$. Choose $x_0 = x$ and $x_{N_{i_k}} \in U_{i_k+1}$. Together with $f(x_k) = x_{k+1}$, this defines an orbit of x. □

Corollary 7.2.4 *A continuous open (Definition A.1.16) map f of a complete separable metric space without isolated points is topologically transitive if and only if there are no two disjoint open nonempty f-invariant sets.*

Proof If $U, V \subset X$ are open, then the invariant sets $\tilde{U} := \bigcup_{n\in\mathbb{Z}} f^n(U)$ and $\tilde{V} := \bigcup_{n\in\mathbb{Z}} f^n(V)$ are open because f is an open map, and therefore not disjoint by assumption, so $f^n(U) \cap f^m(V) \neq \varnothing$ for some $n, m \in \mathbb{Z}$. Then $f^{n-m}(U) \cap V \neq \varnothing$ and f is topologically transitive by Proposition 7.2.2. The other direction is obvious: A dense orbit visits every open set. □

7.2.2 Topological Mixing

There is a property of a dynamical system that immediately implies this criterion but is indeed much stronger:

Definition 7.2.5 A continuous map $f : X \to X$ is said to be *topologically mixing* if for any two nonempty open sets $U, V \subset X$ there is an $N \in \mathbb{N}$ such that $f^n(U) \cap V \neq \varnothing$ for every $n > N$.

By Proposition 7.2.2, every topologically mixing map is topologically transitive. On the other hand, our simple examples are not mixing. No translation T_γ, in particular no circle rotation, is topologically mixing. This follows from the fact that translations preserve the natural metric on the torus induced by the standard Euclidean metric on $\mathbb{R}^n$ and from the following general criterion.

Lemma 7.2.6 *Isometries are not topologically mixing.*

Proof Let $f : X \to X$ be an isometry (that is, a map that preserves the metric on X). Take distinct points $x, y, z \in X$, and let $\delta := \min(d(x, y), d(y, z), d(z, x))/4$. Let U, V_1, V_2 be δ-balls around x, y, z correspondingly. Since f preserves the diameter of any set, the diameter of $f^n(U)$ is at most 2δ whereas the distance between any $p \in V_1$ and $q \in V_2$ is greater than 2δ. Thus for each n either $f^n(U) \cap V_1 = \varnothing$ or $f^n(U) \cap V_2 = \varnothing$. □

7.2.3 Expanding Maps

For expanding maps we prove topological mixing by showing the stronger fact that, for any open set, its image under some iterate of the map contains S^1. For the linear expanding maps E_m this is obvious: Every open set contains an interval of the form $[l/|m|^k, (l+1)/|m|^k]$ for some integers k and $l \leq |m|^k$. The image of this interval under E_m^k is S^1.

Proposition 7.2.7 *Expanding maps of S^1 are topologically mixing.*

Proof Let $f : S^1 \to S^1$ such that $|f'(x)| \geq \lambda > 1$ for all x. Consider a lift F of f to $\mathbb{R}$. Then $|F'(x)| \geq \lambda$ for $x \in \mathbb{R}$. If $[a, b] \subset \mathbb{R}$ is an interval, then by the Mean-Value Theorem A.2.3 there exists a $c \in (a, b)$ such that $|F(b) - F(a)| = |F'(c)(b - a)| \geq \lambda(b - a)$ and so the length of any interval is increased by a factor at least λ^n under F^n. Consequently, for every interval I there exists $n \in \mathbb{N}$ such that the length of $F(I)$ exceeds 1. Thus the image of the projection of I to S^1 under f^n contains S^1. Since every open set of S^1 contains an interval, this shows that every open set has an image under an iterate of f that contains S^1. □

Corollary 7.2.8 *Linear expanding maps of S^1 are chaotic.*

Proof Transitivity follows from Proposition 7.2.7 and the density of periodic points from Proposition 7.1.3. □

For nonlinear expanding maps, this result also holds by invoking Theorem 7.4.3 (which is only stated for degree 2 in the following, but holds for any expanding map).

7.2.4 Hyperbolic Linear Map on the Torus

The hyperbolic linear map F_L of the torus induced by the linear map L with matrix $\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ was introduced in Section 7.1.4. The eigenvectors corresponding to the first eigenvalue belong to the line $y = (\sqrt{5} - 1/2)x$. The family of lines parallel to it is invariant under L, and L uniformly expands distances on those lines by a factor λ_1. Similarly, there is an invariant family of contracting lines $y = (-\sqrt{5} - 1/2)x + \text{const}$.

Proposition 7.2.9 *The automorphism F_L is topologically mixing.*

Proof Fix open sets $U, V \subset \mathbb{T}^2$. The L-invariant family of lines

$$y = \frac{\sqrt{5} - 1}{2} x + \text{const.} \tag{7.2.1}$$

projects to $\mathbb{T}^2$ as an F_L-invariant family of orbits of the linear flow T^t_ω with irrational slope $\omega = (1, (\sqrt{5} - 1)/2)$. By Proposition 5.1.3, this flow is minimal. Thus the projection of each line is everywhere dense on the torus, and hence U contains a piece J of an expanding line; furthermore, for any $\epsilon > 0$, there exists $T = T(\epsilon)$ and a segment of an expanding line of length T that intersects any ϵ-ball on the torus. Since all segments of a given length are translations of one another, this property holds for all segments. Now take ϵ such that V contains an ϵ-ball and $N \in \mathbb{N}$ such that $f^N(J)$ has length at least T. Then $f^n(J) \cap V \neq \varnothing$ for $n \geq N$ and thus $f^n(U) \cap V \neq \varnothing$ for $n \geq N$. □

Corollary 7.2.10 *The automorphism F_L is chaotic.*

Proof Combine Proposition 7.2.9, and Proposition 7.1.10. □

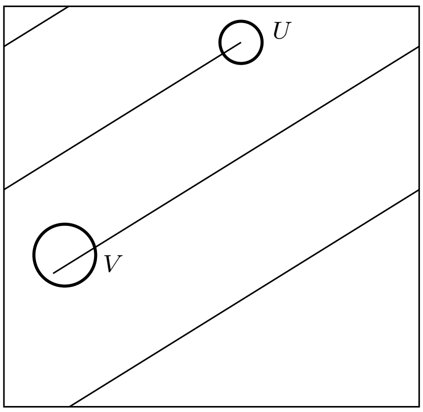

Figure 7.2.2. Topological mixing.

7.2.5 Chaos

At the outset of this section we motivated our definition of a chaotic map by saying that it implies sensitive dependence on initial conditions. We now justify this claim by defining and verifying sensitive dependence.

Definition 7.2.11 A map $f\colon X \to X$ of a metric space is said to exhibit *sensitive dependence* on initial conditions if there is a $\Delta > 0$, called a sensitivity constant, such that for every $x \in X$ and $\epsilon > 0$ there exists a point $y \in X$ with $d(x, y) < \epsilon$ and $d(f^N(x), f^N(y)) \geq \Delta$ for some $N \in \mathbb{N}$.

This means that the slightest error (ϵ) in any initial condition (x) can lead to a macroscopic discrepancy (Δ) in the evolution of the dynamics. Accordingly, Δ tells us at what scale these errors show up. Suppose I start a dynamical system in a state x, let it evolve for a while, and try to reproduce this experiment. Even if I reproduce x to within a billionth of an inch, the initial minuscule error may magnify to a large difference in behavior in finite (often relatively short) time, that is, I may find that the second run of the same experiment bears little resemblance to the first. This is what Poincaré meant by his comment quoted in Section 1.1.1.

For linear expanding maps this property is clearly true: Any initial error of an orbit for E_m grows exponentially (by a factor of $|m|$ at every iteration) until it has grown to more than $1/2|m|$. In particular, $\delta = 1/2|m|$ is a sensitivity constant. On the other hand, this property clearly fails for isometries because points do not move apart at all under iteration.

It is important for the definition that Δ does not depend on x, nor on ϵ, but only on the system. Thus, the *smallest* error *anywhere* can lead to discrepancies of size Δ eventually.[2]

Theorem 7.2.12 *Chaotic maps exhibit sensitive dependenceon initial conditions, except when the entire space consists of a single periodic orbit.*

Proof Unless the entire space consists only of a single periodic orbit, the density of periodic points implies that there are two distinct periodic orbits. Since they have no common point, there are periodic points p, q such that $\Delta := \min\{d(f^n(p), f^m(q)) \mid n, m \in \mathbb{Z}\}/8 > 0$. (Note that n and m need not agree.) We now show that Δ is a sensitivity constant.

If $x \in X$, the orbit of one of these two points keeps a distance at least 4Δ from x: If they were both within less than 4Δ of x, then their mutual distance would be less than 8Δ. Suppose this point is q.

Take any $\epsilon \in (0, \Delta)$. By the density of periodic points, there is a periodic point $p \in B(x, \epsilon)$ whose period we call n. Then the set

$$V := \bigcap_{i=0}^{n} f^{-i}(B(f^i(q), \Delta))$$

[2] The meteorologist Edward Lorenz described this as the “butterfly effect”: Weather appears to be a chaotic dynamical system, so it is conceivable that a butterfly that flutters by in Rio may cause a typhoon in Tokyo a few days later.

of points whose first n iterates track those of q up to Δ is an open neighborhood of q. By Proposition 7.2.2 (used in the direction that does not require completeness) there exists a $k \in \mathbb{N}$ such that $f^k(B(x, \epsilon)) \cap V \neq \varnothing$, that is, there exists a $y \in B(x, \epsilon)$ such that $f^k(y) \in V$. If $j := \lfloor k/n \rfloor + 1$, then $k/n < j \leq (k/n) + 1$ and

$$k = n \cdot \frac{k}{n} < nj \leq n\left(\frac{k}{n} + 1\right) = k + n.$$

If we take $N := nj$, then this shows that $0 < N - k \leq n$. Since $f^N(p) = p$, the triangle inequality gives

$$\begin{aligned} d(f^N(p), f^N(y)) &= d(p, f^N(y)) \\ &\geq d(x, f^{N-k}(q)) - d(f^{N-k}(q), f^N(y)) - d(p, x) \\ &\geq 4\Delta - \Delta - \Delta = 2\Delta \end{aligned} \tag{7.2.2}$$

because $p \in B(x, \epsilon) \subset B(x, \Delta)$ and

$$f^N(y) = f^{N-k}(f^k(y)) \in f^{N-k}(V) \subset B(f^{N-k}(q), \Delta)$$

by definition of V. Both p and y are in $B(x, \epsilon)$ and either $d(f^N(p), f^N(x)) \geq \Delta$ or $d(f^N(y), f^N(x)) \geq \Delta$ by (7.2.2). □

Remark 7.2.13 There are maps exhibiting sensitive dependence that are not chaotic, such as the linear twist from Section 6.1.1. Here, any point x has arbitrarily nearby points (on a vertical segment through x) that move a considerable distance away after sufficiently many iterates. The set of periodic points consists of those points whose second coordinate is rational and is hence dense. On the other hand, this map is clearly not topologically transitive.

Sensitive dependence can be derived from topological mixing alone, without an assumption on periodic points:

Proposition 7.2.14 *A topologically mixing map (on a space with more than one point) has sensitive dependence.*

Proof Take $\Delta > 0$ such that there are points x_1, x_2 with $d(x_1, x_2) > 4\Delta$. We show that Δ is a sensitivity constant.

Let $V_i = B_\Delta(x_i)$ for $i = 1, 2$. Suppose $x \in X$ and U is a neighborhood of x. By topological mixing there are $N_1, N_2 \in \mathbb{N}$ such that $f^n(U) \cap V_1 \neq \varnothing$ for $n \geq N_1$ and $f^n(U) \cap V_2 \neq \varnothing$ for $n \geq N_2$. For $n \geq N := \max(N_1, N_2)$, there are points $y_1, y_2 \in U$ with $f^n(y_1) \in V_1$ and $f^n(y_2) \in V_2$; hence $d(f^n(y_1), f^n(y_2)) \geq 2\Delta$. By the triangle inequality $d(f^n(y_1), f^n(x)) \geq \Delta$ or $d(f^n(y_2), f^n(x)) \geq \Delta$. □

EXERCISES

Exercise 7.2.1 Find the maximal sensitivity constant for E_2.

Exercise 7.2.2 Find the supremum of sensitivity constants for F_L in Section 7.2.4.

Exercise 7.2.3 Prove that, for a topologically mixing map, any number less than the diameter $\sup\{d(x, y) \mid x, y \in X\}$ of the space X is a sensitivity constant.

■ **Exercise 7.2.4** Consider the linear twist $T\colon S^1 \times [0, 1] \to S^1 \times [0, 1]$, $T(x, y) = (x + y, y)$ from Section 6.1.1 that was remarked upon in Remark 7.2.13. Prove that it has the following property of *partial topological mixing*: Let $U, V \subset S^1$ be nonempty open sets. Then there exists $N(U, V) \in \mathbb{N}$ such that $T^n(U \times [0, 1]) \cap (V \times [0, 1]) \neq \varnothing$ for any $n \geq N$.

■ **Exercise 7.2.5** Show that for a compact space sensitive dependence is a topological invariant (see Section 7.3.6).

■ **Exercise 7.2.6** Prove that for any two periodic points of F_L the set of heteroclinic points (see Definition 2.3.4) is dense.

■ **Exercise 7.2.7** Consider a 2×2 integer matrix L without eigenvalues of absolute value 1 and with $|\det L| > 1$. Prove that the induced noninvertible hyperbolic linear map $F_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ is topologically mixing.

7.3 CODING

One of the most important ideas for studying complicated dynamics sounds strange at first. It involves throwing away some information by tracking orbits only approximately. The idea is to divide the phase space into finitely many pieces and to follow an orbit only to the extent of specifying which piece it is in at a given time. This is a bit like the itinerary of the harried tourist in Europe, who decides that it is Tuesday, so the place must be Belgium. A more technological analogy would be to look at the records of a cell phone addict and track which local transmitters were used at various times.

In these analogies one genuinely loses information, because the sequence of European countries or of local cellular stations does not pinpoint the traveller at any given moment. However, orbits in a dynamical system do not move around at whim, and the deterministic nature of the dynamics has the effect that a complete *itinerary* of this sort may (and often does) give all the information about a point. This is the process of *coding* of a dynamical system.

7.3.1 Linear Expanding Maps

The linear expanding maps

$$E_m\colon S^1 \to S^1,\ E_m(x) = mx \pmod 1$$

from Section 7.1.1 are chaotic (Corollary 7.2.8), that is, they exhibit coexistence of dense orbits (Proposition 7.2.7) with a countable dense set of periodic orbits (Proposition 7.1.3). Thus the orbit structure is both complicated and highly nonuniform. Now we look at these maps from a different point of view, which in turn gives a deeper appreciation of just how complicated their orbit structure really is. To simplify notations, assume as before that $m = 2$.

Consider the binary intervals

$$\Delta_n^k := \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right] \qquad \text{for } n = 1, \dots \quad \text{and} \quad k = 0, 1, \dots, 2^n - 1.$$

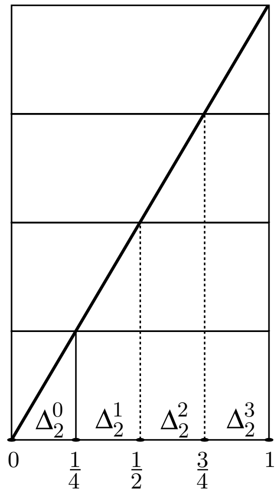

Figure 7.3.1. Linear coding.

Figure 7.3.1 illustrates this for $n = 2$. Let $x = 0.x_1x_2\dots$ be the binary representation of $x \in [0, 1]$. Then $2x = x_1.x_2x_3\dots = 0.x_2x_3\dots \pmod 1$. Thus

$$E_2(x) = 0.x_2x_3\dots \pmod 1. \tag{7.3.1}$$

This is the first and easiest example of *coding*, which we will discuss in greater detail shortly.

7.3.2 Implications of Coding

We briefly derive a few new facts about linear expanding maps that are best seen via this coding.

1. Proof of Transitivity via Coding. First, we use this representation to give another proof of topological transitivity by describing explicitly the binary representation of a number whose orbit under the iterates of E_2 is dense. Consider an integer k, $0 \le k \le 2^n - 1$. Let $k_0 \dots k_{n-1}$ be the binary representation of k, maybe with several zeroes at the beginning. Then $x \in \Delta_n^k$ if and only if $x_i = k_i$ for $i = 0, \dots n-1$. Therefore we write $\Delta_{k_0\dots k_{n-1}} := \Delta_n^k$ from now on. Now put the binary representations of all numbers from 0 to $2^n - 1$ (with zeroes in front if necessary) one after another and form a finite sequence, which we denote by ω_n, that is, ω_n is obtained by concatenating all 2^n binary sequences of length n. Having done this for every $n \in \mathbb{N}$, put the sequences ω_n, $n = 1, 2, \dots$ in that order, call the resulting infinite sequence ω, and consider the number x with the binary representation $0.\omega$. Since by construction moving ω to the left and cutting off the first digits produces at various moments binary representations of any n-digit number, this means that the orbit of the point x under the iterates of the map E_2 intersects every interval $\Delta_{k_0\dots k_{n-1}}$ and hence is dense.

This construction extends to any $m \ge 2$. To construct a dense orbit for E_m with $m \le -2$, we notice that $E_m^2 = E_{m^2}$. Obviously the orbit of any point under the iterates of a square of a map is a subset of the orbit under the iterates of the map itself; thus if the former is dense, so is latter. So we apply our construction to the map E_{m^2} and obtain a point with dense orbit under E_m.

2. Exotic Asymptotics. Next we use this approach to show that besides periodic and dense orbits there are other types of asymptotic behavior for orbits of expanding maps. One can construct such orbits for E_2, but the simplest and most elegant example appears for the map E_3.

Proposition 7.3.1 *There exists a point $x \in S^1$ such that the closure of its orbit with respect to the map E_3 in additive notation coincides with the standard middle-third Cantor setK. In particular, K is E_3-invariant and contains a dense orbit.*

Proof The middle-third Cantor set K can be described as the set of all points on the unit interval that have a representation in base 3 with only 0's and 2's as digits (see Section 2.7.1). Similarly to (7.3.1), the map E_3 acts as the shift of digits to the left in the base 3 representation. This implies that K is E_3-invariant. It remains to show that E_3 has a dense orbit in K.

Every point in K has a unique representation in base 3 without 1's. Let $x \in K$ and

$$0.x_1x_2x_3\dots \tag{7.3.2}$$

be such a representation. Let $h(x)$ be the number whose representation in base 2 is

$$0.\frac{x_1}{2}\frac{x_2}{2}\frac{x_3}{2}\dots,$$

that is, it is obtained from (7.3.2) by replacing 2's by 1's. Thus we have constructed a map $h\colon K \to [0, 1]$ that is continuous, nondecreasing [that is, $x > y$ implies $h(x) \geq h(y)$], and one-to-one, except for the fact that binary rationals have two preimages each (compare Section 2.7.1 and Section 4.4.1). Furthermore, $h \circ E_3 = E_2 \circ h$. Let $D \subset [0, 1]$ be a dense set of points that does not contain binary rationals. Then $h^{-1}(D)$ is dense in K because, if Δ is an open interval such that $\Delta \cap K \neq \varnothing$, then $h(\Delta)$ is a nonempty interval open, closed, or semiclosed and hence contains points of D. Now take any $x \in [0, 1]$ whose E_2-orbit is dense; the E_3-orbit of $h^{-1}(x) \in K$ is dense in K. □

3. Nonrecurrent Points. Another interesting example is the construction of a nonrecurrent point, that is, such a point x that for some neighborhood U of x all iterates of x avoid U (see Definition 6.1.8). In fact, there is a dense set of nonrecurrent points for the map E_2.

Pick any fixed sequence $(\omega_0, \dots, \omega_{n-1})$ of 0's and 1's and add a tail of 0's if $\omega_{n-1} = 1$, or of 1's if $\omega_{n-1} = 0$. Call the resulting infinite sequence ω. As before, let x be the number with binary representation $0.\omega$. Thus, x lies in a prescribed interval $\Delta_{\omega_0\dots\omega_{n-1}}$ and by construction $x \neq 0$. On the other hand, $E_2^n x = 0$ and hence $E_2^m x = 0$ for all $m \geq n$, so x is a nonrecurrent point.

Thus, we have found that E_m is chaotic and topologically mixing, that its periodic and nonrecurrent orbits are dense, and that E_3 has orbits whose closure is a Cantor set.

7.3.3 A Two-Dimensional Cantor Set

We now describe a map in the plane that naturally gives rise to a two-dimensional Cantor set (previously encountered in Problem 2.7.5) on which ternary expansion of the coordinates provides all information about the dynamics. This *horseshoe map* plays a central role in our further development.

Consider a map defined on the unit square $[0, 1] \times [0, 1]$ by the following construction: First apply the linear transformation $(x, y) \mapsto (3x, y/3)$ to get a horizontal strip whose left and right thirds will be rigid in the next transformation. Holding the left third fixed, bend and stretch the middle third such that the right third falls rigidly on the top third of the original unit square. This results in a "G"-shape. For points that are in and return to the unit square, this map is given analytically by

$$(x, y) \mapsto \begin{cases} (3x, y/3) & \text{if } x \leq 1/3 \\ (3x - 2, (y + 2)/3) & \text{if } x \geq 2/3. \end{cases}$$

The inverse can be written as

$$(x, y) \mapsto \begin{cases} (x/3, 3y) & \text{if } y \leq 1/3 \\ ((x + 2)/3, 3y - 2) & \text{if } y \geq 2/3. \end{cases}$$

Geometrically, the inverse looks like an "e"-shape rotated counterclockwise by 90°.

To iterate this map one triples the x-coordinate repeatedly and always assumes that the resulting value is either at most $1/3$ or else at least $2/3$, that is, that the first ternary digit is 0 or 2, but not 1. (If the expansion is not unique, one requires such a choice to be possible.) Comparing with the construction of the ternary Cantor set in Section 2.7.1, one sees that the x-coordinate lies in the ternary Cantor set C. Looking at the inverse one sees likewise that, in order for all preimages to be defined, the y-coordinate lies in the Cantor set as well. Therefore this map is defined for all positive and negative iterates on the two-dimensional Cantor set $C \times C$. There is a straightforward way of using ternary expansion to code the dynamics. For a point (x, y) the map shifts the ternary expansion of x one step to the left, dropping the first term, and shifts the ternary expansion of y to the right. It is natural to fill in the now-ambiguous first digit of the shifted y-coordinate with the entry from the x-coordinate that was just dropped. This retains all information, and the best way of vizualizing the result is to write the expansion of the y-coordinate in reverse and in front of that of the x-coordinate. This gives a bi-infinite string of 0's and 2's (remember, no 1's allowed), which is shifted by the map. Of course, one should verify that the inverse acts by shifting in the opposite direction.

7.3.4 Sequence Spaces

Now we are ready to discuss the concept of coding in general. We mean by coding a representation of points in the phase space of a discrete-time dynamical system or an invariant subset by sequences (not necessarily unique) of symbols from a certain "alphabet," in this case the symbols $0, \dots, N - 1$. So we should acquaint ourselves with these spaces.

Denote by Ω_N^R the space of sequences $\omega = (\omega_i)_{i=0}^{\infty}$ whose entries are integers between 0 and $N - 1$. Define a metric by

$$d_\lambda(\omega, \omega') := \sum_{i=0}^{\infty} \frac{\delta(\omega_i, \omega_i')}{\lambda^i}, \tag{7.3.3}$$

where $\delta(k,l) = 1$ if $k \neq l$, $\delta(k,k) = 0$, and $\lambda > 2$. The same definition can be made for two-sided sequences by summing over $i \in \mathbb{Z}$:

$$d_\lambda(\omega, \omega') := \sum_{i\in\mathbb{Z}} \frac{\delta(\omega_i, \omega'_i)}{\lambda^{|i|}}, \tag{7.3.4}$$

for some $\lambda > 3$. This means that two sequences are close if they agree on a long stretch of entries around the origin.

Consider the symmetric *cylinder* defined by

$$C_{\alpha_{1-n}\dots\alpha_{n-1}} := \{\omega \in \Omega_N \mid \omega_i = \alpha_i \text{ for } |i| < n\}.$$

Fix a sequence $\alpha \in C_{\alpha_{1-n}\dots\alpha_{n-1}}$. If $\omega \in C_{\alpha_{1-n}\dots\alpha_{n-1}}$, then

$$d_\lambda(\alpha, \omega) = \sum_{i\in\mathbb{Z}} \frac{\delta(\alpha_i, \omega_i)}{\lambda^{|i|}} = \sum_{|i|\geq n} \frac{\delta(\alpha_i, \omega_i)}{\lambda^{|i|}} \leq \sum_{|i|\geq n} \frac{1}{\lambda^{|i|}} = \frac{1}{\lambda^{n-1}} \frac{2}{\lambda - 1} < \frac{1}{\lambda^{n-1}}.$$

Thus $C_{\alpha_{1-n}\dots\alpha_{n-1}} \subset B_{d_\lambda}(\alpha, \lambda^{1-n})$, the λ^{1-n}-ball around α. If $\omega \notin C_{\alpha_{1-n}\dots\alpha_{n-1}}$, then

$$d_\lambda(\alpha, \omega) = \sum_{i\in\mathbb{Z}} \frac{\delta(\alpha_i, \omega_i)}{\lambda^{|i|}} \geq \lambda^{1-n}$$

because $\omega_i \neq \alpha_i$ for some $|i| < n$. Thus $\omega \notin B_{d_\lambda}(\alpha, \lambda^{1-n})$, and the symmetric cylinder is the ball of radius λ^{1-n} around any of its points:

$$C_{\alpha_{1-n}\dots\alpha_{n-1}} = B_{d_\lambda}(\alpha, \lambda^{1-n}). \tag{7.3.5}$$

Therefore, balls in Ω_N are described by specifying a symmetric stretch of entries around the initial one.

For one-sided sequences this discussion works along the same lines [one only needs $\lambda > 2$ in (7.3.4)] and λ^{1-n}-balls are described by specifying a string of n initial entries.

Our examples [see (7.3.1)] suggest to represent points in the phase space by sequences in such a way that the sequences representing the image of a point are obtained from those representing the point itself by the shift (translation) of the symbols. In this way the given transformation corresponds to the *shift transformation*

$$\begin{aligned} \sigma\colon \Omega_N \to \Omega_N, &\qquad (\sigma\omega)_i = \omega_{i+1} \\ \sigma^R\colon \Omega_N^R \to \Omega_N^R, &\qquad (\sigma^R\omega)_i = \omega_{i+1}. \end{aligned} \tag{7.3.6}$$

We often write σ_N for the shift σ on Ω_N and likewise σ_N^R for σ^R on Ω_N^R. For invertible discrete-time systems, any coding involves sequences of symbols extending in both directions; while for noninvertible systems, one-sided sequences do the job. Section 7.3.7 studies these shifts as dynamical systems.

Among the shift transformations that arise from coding there is also a new kind of combinatorial model for a dynamical system that is described by the possibility or impossibility of certain successions of events.

Definition 7.3.2 Let $A = (a_{ij})_{i,j=0}^{N-1}$ be an $N \times N$ matrix whose entries a_{ij} are either 0's or 1's. (We call such a matrix a 0-1 matrix.) Let

$$\Omega_A := \left\{\omega \in \Omega_N \mid a_{\omega_n\omega_{n+1}} = 1 \text{ for } n \in \mathbb{Z}\right\}. \tag{7.3.7}$$

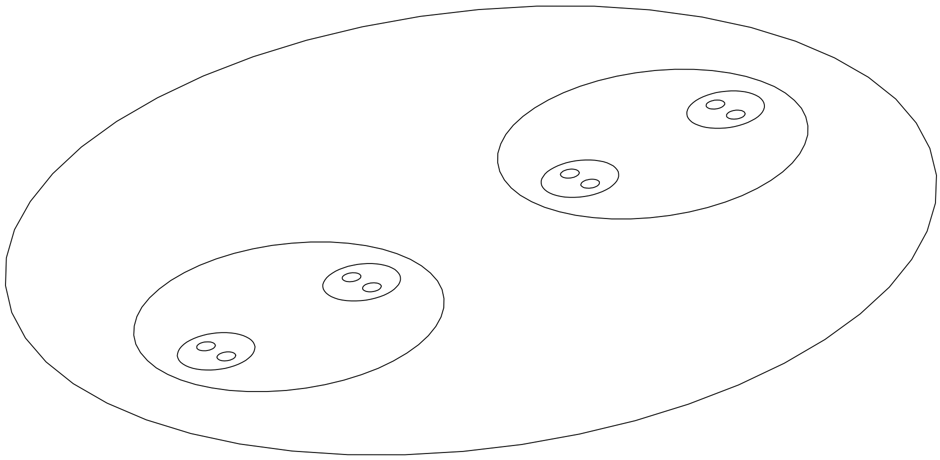

Figure 7.3.2. Obtaining a Cantor set.

The space Ω_A is closed and shift-invariant, and the restriction

$$\sigma_N\restriction_{\Omega_A} =: \sigma_A$$

is called the *topological Markov chain*determined by A.

This is a particular case of a *subshift of finite type*.

7.3.5 Coding

Sequences representing a given point of the phase space are called the *codes* of that point. We have several examples of coding: for the map E_m on the whole circle by sequences from the *alphabet* $\{0, \dots, |m| - 1\}$; for the restriction of the map E_3 to the middle-third Cantor set K by one-sided sequences of 0's and 1's; and for the ternary horseshoe in Section 7.3.3 by bi-infinite sequences of 0's and 2's. In both cases we used one-sided sequences, all sequences appeared as codes of some points, and each code represented only one point. There was, however, an important difference: In the first case, which involved for positive m a representation in base m, a point could have either one or two codes; in the latter there was only one code.

This shows that the space of binary sequences is a Cantor set (Definition 2.7.4). In fact, this also holds for the other sequence spaces.

7.3.6 Conjugacy and Factors

This situation can be roughly described by saying that the shift (Ω_2^R, σ^R) "contains" the map f up to a continuous coordinate change. (We already encountered such a situation in Theorem 4.3.20.)

Definition 7.3.3 Suppose that $g\colon X \to X$ and $f\colon Y \to Y$ are maps of metric spaces X and Y and that there is a continuous surjective map $h\colon X \to Y$ such that $h \circ g = f \circ h$. Then f is said to be a *factor* of g via the *semiconjugacy* or *factor map h*. If this h is a homeomorphism, then f and g are said to be *conjugate* and h is said to be a *conjugacy*.

These notions made a brief appearance in Section 4.3.5 in connection with modeling an arbitrary homeomorphism of the circle by a rotation. The notion of conjugacy is natural and central; two conjugate maps are obtained from one another by a continuous change of coordinates. Hence all properties that are independent of such changes of coordinates are unchanged, such as the numbers of periodic orbits for each period, sensitive dependence (Exercise 7.2.5), topological transitivity, topological mixing, and hence also being chaotic. Such properties are said to be topological invariants. Later in this book we will encounter further important topological invariants such as topological entropy (Definition 8.2.1).

7.3.7 Dynamics of Shifts and Topological Markov Chains

We now study the properties of shifts and topological Markov chains introduced in (7.3.6) and Definition 7.3.2 in more detail. These are important because many interesting dynamical systems are coded by shifts or topological Markov chains. To such dynamical systems the results of this section have immediate applications.

Proposition 7.3.4 *Periodic points for the shifts σ_N and σ_N^R are dense in Ω_N and Ω_N^R, correspondingly, $P_n(\sigma_N) = P_n(\sigma_N^R) = N^n$, and both σ_N and σ_N^R are topologically mixing.*

Proof Periodic orbits for a shift are periodic sequences, that is, $(\sigma_N)^m\,\omega = \omega$ if and only if $\omega_{n+m} = \omega_n$ for all $n \in \mathbb{Z}$. In order to prove density of periodic points, it is enough to find a periodic point in every ball (symmetric cylinder), because every open set contains a ball. To find a periodic point in $C_{\alpha_{-m},\dots,\alpha_m}$, take the sequence ω defined by $\omega_n = \alpha_{n'}$ for $|n'| \le m$, $n' = n \pmod{2m+1}$. It lies in this cylinder and has period $2m+1$.

Every periodic sequence ω of period n is uniquely determined by its coordinates $\omega_0, \dots, \omega_{n-1}$. There are N^n different finite sequences $(\omega_0, \dots, \omega_{n-1})$.

To prove topological mixing, we show that $\sigma_N^n(C_{\alpha_{-m},\dots,\alpha_m}) \cap C_{\beta_{-m},\dots,\beta_m} \ne \varnothing$ for $n > 2m+1$, say, $n = 2m+k+1$ with $k > 0$. Consider any sequence ω such that

$$\omega_i = \alpha_i \text{ for } |i| \le m, \qquad \omega_i = \beta_{i-n} \quad \text{for } i = m+k+1, \dots, 3m+k+1.$$

Then $\omega \in C_{\alpha_{-m},\dots,\alpha_m}$ and $\sigma_N^n(\omega) \in C_{\beta_{-m},\dots,\beta_m}$.

The arguments for the one-sided shift are analogous. □

There is a useful geometric representation of topological Markov chains. Connect i with j by an arrow if $a_{ij} = 1$ to obtain a *Markov graph* G_A with N vertices and several oriented edges. We say that a finite or infinite sequence of vertices of G_A is an *admissible path* or *admissible sequence* if any two consecutive vertices in the sequence are connected by an oriented arrow. A point of Ω_A corresponds to a doubly infinite path in G_A with marked origin; the topological Markov chain σ_A corresponds to moving the origin to the next vertex. The following simple combinatorial lemma is a key to the study of topological Markov chains:

Lemma 7.3.5 *For every $i, j \in \{0, 1, \dots, N-1\}$, the number N_{ij}^m of admissible paths of length $m+1$ that begin at x_i and end at x_j is equal to the entry a_{ij}^m of the matrix A^m.*

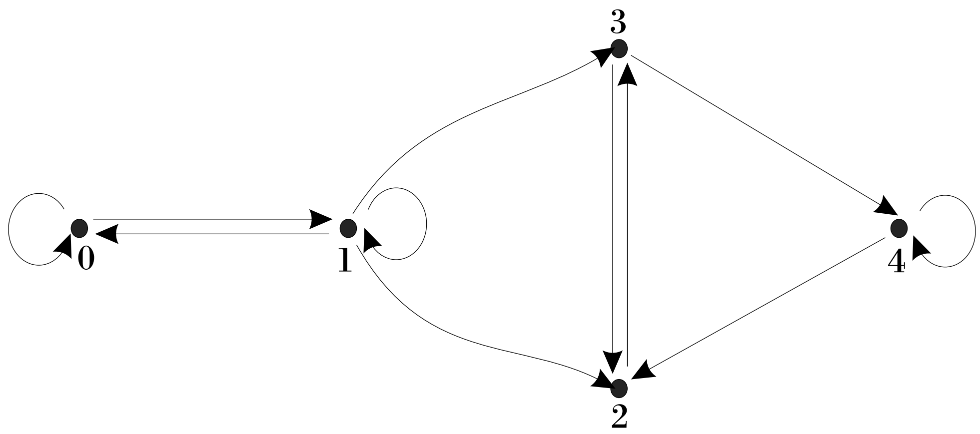

Figure 7.3.3. A Markov graph.

Proof We use induction on m. First, it follows from the definition of the graph G_A that $N^1_{ij} = a_{ij}$. To show that

(7.3.8)
$$N^{m+1}_{ij} = \sum_{k=0}^{N-1} N^m_{ik} a_{kj},$$

take $k \in \{0, \dots, N-1\}$ and an admissible path of length $m+1$ connecting i and k. It can be extended to an admissible path of length $m+2$ connecting i to j (by adding j) if and only if $a_{kj} = 1$. This proves (7.3.8). Now, assuming by induction that $N^m_{ij} = a^m_{ij}$ for all ij, we obtain $N^{m+1}_{ij} = a^{m+1}_{ij}$ from (7.3.8). $\square$

Corollary 7.3.6 $P_n(\sigma_A) = \operatorname{tr} A^n$.

Proof Every admissible closed path of length $m+1$ with marked origin, that is, a path that begins and ends at the same vertex of G_A, produces exactly one periodic point of σ_A of period m. $\square$

Because the eigenvalue of largest absolute value dominates the trace, it determines the exponential growth rate:

Proposition 7.3.7 *$p(\sigma_A) = r(A)$, where $r(A)$ is the spectral radius.*

Proof "$\le$" is clear. To show "$\ge$" we need to avoid cancellations: If $\lambda_j = re^{2\pi i\varphi_j} (1 \le j \le k)$ are the eigenvalues of maximal absolute value then there is a sequence $m_n \to \infty$ such that $m_n\varphi_j \to 0 \,(\mathrm{mod}\ 1)$ for all j (recurrence for toral translations, Section 5.1), so $\sum \lambda_i^{m_n} \sim r^{m_n}$. $\square$

Example 7.3.8 The Markov graph in Figure 7.3.3 produces three fixed points, $\overline{0}$, $\overline{1}$, and $\overline{4}$. $\overline{01}$ and $\overline{23}$ give four periodic points with period 2. The period-3 orbits are generated by $\overline{011}$, $\overline{001}$, $\overline{234}$.

Topological Markov chains can be classified according to the recurrence properties of various orbits they contain. Now we concentrate on those topological Markov chains that possess the strongest recurrence properties.

Definition 7.3.9 A matrix A is said to be *positive* if all its entries are positive. A 0-1 matrix A is said to be *transitive* if A^m is positive for some $m \in \mathbb{N}$. A topological Markov chain σ_A is said to be *transitive* if A is a transitive matrix.

Lemma 7.3.10 *If A^m is positive, then so is A^n for any $n \geq m$.*

Proof If $a_{ij}^n > 0$ for all i, j, then for each j there is a k such that $a_{kj} = 1$. Otherwise, $a_{ij}^n = 0$ for every n and i. Now use induction. If $a_{ij}^n > 0$ for all i, j, then $a_{ij}^{n+1} = \sum_{k=0}^{N-1} a_{ik}^n a_{kj} > 0$ because $a_{kj} = 1$ for at least one k. $\square$

Lemma 7.3.11 *If A is transitive and $\alpha_{-k}, \dots, \alpha_k$ is admissible, that is, $a_{\alpha_i \alpha_{i+1}} = 1$ for $i = -k, \dots, k-1$, then the intersection $\Omega_A \cap C_{\alpha_{-k},\dots,\alpha_k} =: C_{\alpha_{-k},\dots,\alpha_k,A}$ is nonempty and moreover contains a periodic point.*

Proof Take m such that $a_{\alpha_k,\alpha_{-k}}^m > 0$. Then one can extend the sequence α to an admissible sequence of length $2k + m + 1$ that begins and ends with α_{-k}. Repeating this sequence periodically, we obtain a periodic point in $C_{\alpha_{-k},\dots,\alpha_k,A}$. $\square$

Proposition 7.3.12 *If A is a transitive matrix, then the topological Markov chain σ_A is topologically mixing and its periodic orbits are dense in Ω_A; in particular, σ_A is chaotic and hence has sensitive dependence on initial conditions.*

Proof The density of periodic orbits follows from Lemma 7.3.11. To prove topological mixing, pick open sets $U, V \subset \Omega_A$ and nonempty symmetric cylinders $C_{\alpha_{-k},\dots,\alpha_k,A} \subset U$ and $C_{\beta_{-k},\dots,\beta_k,A} \subset V$. Then it suffices to show that $\sigma_A^n(C_{\alpha_{-k},\dots,\alpha_k,A}) \cap C_{\beta_{-k},\dots,\beta_k,A} \neq \varnothing$ for any sufficiently large n. Take $n = 2k + 1 + m + l$ with $l \geq 0$, where m is as in Definition 7.3.9. Then $a_{\alpha_k \beta_{-k}}^{m+l} > 0$ by Lemma 7.3.10, so there is an admissible sequence of length $4k + 2 + m + l$ whose first $2k + 1$ symbols are identical to $\alpha_{-k}, \dots, \alpha_k$ and the last $2k + 1$ symbols to $\beta_{-k}, \dots, \beta_k\beta$. By Lemma 7.3.11, this sequence can be extended to a periodic element of Ω_A which belongs to $\sigma_A^n(C_{\alpha_{-k},\dots,\alpha_k,A}) \cap C_{\beta_{-k},\dots,\beta_k,A}$. $\square$

Example 7.3.13 The matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is not transitive because all its powers are upper triangular and hence there is no path from 1 to 0. In fact, the space Ω_A is countable and consists of two fixed points $(\dots, 0, \dots, 0, \dots)$ and $(\dots, 1, \dots, 1, \dots)$, and a single heteroclinic orbit connecting them (consisting of the sequences that are 1 up to some place and 0 thereafter).

Example 7.3.14 For the matrix

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

every orbit alternates between entries from the first group $\{0, 1\}$ on the one hand and from the second group $\{2, 3\}$ on the other hand, that is, the parity (even–odd) must alternate. Therefore no power of the matrix has all entries positive.

EXERCISES

Exercise 7.3.1 Prove that E_2 has a nonperiodic orbit all of whose even iterates lie in the left half of the unit interval.

Exercise 7.3.2 Prove that E_2 has a uncountably many orbits for which no segment of length 10 has more than one point in the left half of the unit interval.

Exercise 7.3.3 Prove that linear maps that are conjugate in the sense of linear algebra are topologically conjugate in the sense of Definition 7.3.3.

Exercise 7.3.4 Write down the Markov matrix for Figure 7.3.3 and check Corollary 7.3.6 up to period 3.

Exercise 7.3.5 Consider the metric

$$d'_\lambda(\alpha, \omega) := \sum_{i\in\mathbb{Z}} \frac{|\alpha_i - \omega_i|}{\lambda^{|i|}} \tag{7.3.9}$$

on Ω_N. Show that for $\lambda > 2N - 1$ the cylinder $C_{\alpha_{1-n}\cdots\alpha_{n-1}}$ is a λ^{1-n}-ball for d'_λ.

Exercise 7.3.6 Repeat the previous exercise for one-sided shifts (with $\lambda > N$).

Exercise 7.3.7 Consider the metric

$$d''_\lambda(\alpha, \omega) := \lambda^{-\max\{n\in\mathbb{N} \mid \alpha_i=\omega_i \text{ for } |i|\le n\}} \tag{7.3.10}$$

[and $d''_\lambda(\alpha, \alpha) = 0$] on Ω_N. Show that the cylinder $C_{\alpha_{1-n}\cdots\alpha_{n-1}}$ is a ball for d''_λ.

Exercise 7.3.8 Find the supremum of sensitivity constants for a transitive topological Markov chain with respect to the metric d''_λ.

Exercise 7.3.9 Find the supremum of sensitivity constants for a transitive topological Markov chain with respect to the metric d'_λ.

Exercise 7.3.10 Show that for $m < n$ the shift on Ω_m is a factor of the shift on Ω_n.

Exercise 7.3.11 Prove that the quadratic map f_4 on $[0, 1]$ is not conjugate to any of the maps f_λ for $\lambda \in [0, 4)$.

Exercise 7.3.12 Show that the topological Markov chains determined by the matrices

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

are conjugate.

Exercise 7.3.13 Find the smallest positive value of $p(\sigma_A)$ for a transitive topological Markov chain with two states (that is, with a 2×2 matrix A).

PROBLEMS FOR FURTHER STUDY

Problem 7.3.14 Find all factors of an irrational rotation R_α of the circle.

Problem 7.3.15 Find the smallest value of $p(\sigma_A)$ for a transitive topological Markov chain with three states (that is, with a 3×3 matrix A).

7.4 MORE EXAMPLES OF CODING

We now carry out a coding construction for several familiar dynamical systems.

7.4.1 Nonlinear Expanding Maps

There is a correspondence between general (not necessarily linear) expanding maps of the circle (Section 7.1.3) and a shift on a sequence space. The construction is similar to the one from Section 7.3.1. There is some effort involved, but there is a beautiful prize at the end: We obtain a complete classification of a large class of maps in terms of a simple invariant.

To keep notations simple, we consider an expanding map $f\colon S^1 \to S^1$ of degree 2. By Proposition 7.1.9, f has exactly one fixed point p. (For maps of higher degree, we could pick any one of the fixed points.) Since $\deg(f) = 2$, there is exactly one point $q \neq p$ such that $f(q) = p$. The points p and q divide the circle into two arcs. Starting from p in the positive direction, denote the first arc by Δ_0 and the second arc by Δ_1. Define the coding for $x \in S^1$ as follows: x is represented by the sequence $\omega \in \Omega_2^R$ for which

$$f^n(x) \in \Delta_{\omega_n}. \tag{7.4.1}$$

This representation is unique unless $f^n(x) \in \{p, q\} = \Delta_0 \cap \Delta_1$. This lack of uniqueness is similar to the case of binary rationals for the map E_2. Suppose a point x has an iterate in $\{p, q\}$. Then either $x = p$ and $f^n(x) = p$ for all $n \in \mathbb{N}$, or else the point q must appear before p in the sequence of iterates, that is, $f^n(x) \notin \{p, q\}$ for all n less than some k and then $f^k(x) = q$ and $f^{k+1}(x) = p$. In this case we make the following convention. p has two codes, all 0's and all 1's, and q has two codes, $01111111\dots$ and $1000000\dots$, and any x such that $F^k(x) = q$ has two codes given by the first $k - 1$ digits uniquely defined by (7.4.1), followed by either of the codes for q.

Actually, going the other way around is better:

Proposition 7.4.1 *If $f\colon S^1 \to S^1$ is an expanding map of degree 2, then f is a factor of σ^R on Ω_2^R (Definition 7.3.3), that is, there is a surjective continuous map $h\colon \Omega_2^R \to S^1$ such that $f^n(h(\omega)) \in \Delta_{\omega_n}$ for all $n \in \mathbb{N}_0$, that is, $h \circ \sigma^R = f \circ h$.*

Proof That the domain of h is Ω_2^R requires that every sequence of 0's and 1's appears as the code of some point. First, f maps each of the two intervals Δ_0 and Δ_1 onto S^1 almost injectively, the only identification being at the ends. Let

$$\Delta_{00} \text{ be the core of } \Delta_0 \cap f^{-1}(\Delta_0),$$

$$\Delta_{01} \text{ be the core of } \Delta_0 \cap f^{-1}(\Delta_1),$$

$$\Delta_{10} \text{ be the core of } \Delta_1 \cap f^{-1}(\Delta_0),$$

$$\Delta_{11} \text{ be the core of } \Delta_1 \cap f^{-1}(\Delta_1).$$

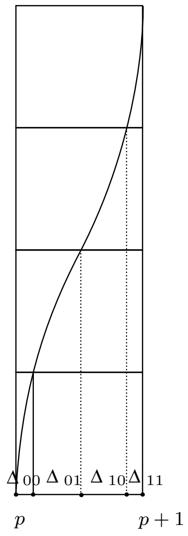

Figure 7.4.1. Nonlinear coding.

What we mean by "core" is that each indicated intersection consists of an interval as well as an isolated point (p or q), and we discard this extraneous point. Each of these four intervals is mapped onto S^1 by f^2, again the only identification being at the ends. By definition, any point from Δ_{ij} has ij as the first two symbols of its code. Proceeding inductively we construct for any finite sequence $\omega_0, \dots, \omega_{n-1}$ the interval

(7.4.2) $$\Delta_{\omega_0,\dots,\omega_{n-1}} := \text{ the core of } \Delta_{\omega_0} \cap f^{-1}(\Delta_{\omega_1}) \cdots \cap f^{1-n}(\Delta_{\omega_{n-1}}),$$

which is mapped by f^n onto S^1 with identification of the endpoints. Now take any infinite sequence $\omega = \omega_1, \dots \in \Omega_2^R$. The intersection $\bigcap_{n=1}^{\infty} \Delta_{\omega_0,\dots,\omega_{n-1}}$ of the nested closed intervals $\Delta_{\omega_0,\dots,\omega_{n-1}}$ is nonempty, and any point in this intersection has the sequence ω as its code.

So far we have only used the fact that f is a monotone map of degree 2. To show that h is well defined, we use the expanding property to check that $\bigcap_{n=1}^{\infty} \Delta_{\omega_0,\dots,\omega_{n-1}}$ consists of a single point, hence a point with a given code is unique.

If $g\colon I \to S^1$ is an injective map of an open interval I with a nonnegative derivative, then by the Mean-Value Theorem A.2.3 $l(g(I)) = \int_I g'(x)\,dx = g'(\xi)l(I)$ for some $\xi \in I$. Thus, in our case, there is a ξ_n such that

$$1 = l(S^1) = \int_{\Delta_{\omega_0,\dots,\omega_{n-1}}} (f^n)'(x)\,dx = (f^n)'(\xi_n) \cdot l(\Delta_{\omega_0,\dots,\omega_{n-1}}).$$

Since f is expanding $|(f^n)'| > \lambda^n$ for some $\lambda > 1$, hence $l(\Delta_{\omega_0,\dots,\omega_{n-1}}) < \lambda^{-n} \to 0$ as $n \to \infty$ and $\bigcap_{n=1}^{\infty} \Delta_{\omega_0,\dots,\omega_{n-1}}$ consists of a single point x_ω. This gives a well-defined surjective map $h\colon \Omega_2^R \to S^1$, $\omega \mapsto x_\omega$.

Give Ω_2^R the metric d_4 from (7.3.3). We showed in Section 7.3.4 that if $\epsilon = \lambda^{-n}$ and $\delta = 4^{-n}$, then $d(\omega, \omega') < \delta$ implies that $\omega_i = \omega'_i$ for $i < n$ and hence $|x_\omega - x_{\omega'}| \le l(\Delta_{\omega_0,\dots,\omega_{n-1}}) < \lambda^{-n} = \epsilon$. Thus h is continuous.

That $h(\sigma^R(\omega)) = f(h(\omega))$ is clear from the construction. $\square$

7.4.2 Classification via Coding

Proposition 7.4.1 and the discussion preceding it established a semiconjugacy between the one-sided 2-shift and the expanding map f on S^1, that is,

Proposition 7.4.2 *Let $f\colon S^1 \to S^1$ be an expanding map of degree 2. Then f is a factor of the one-sided 2-shift (Ω_2^R, σ_R) via a semiconjugacy $h\colon \Omega_2^R \to S^1$. If $h(\omega) = h(\omega') =: x$, then there exists an $n \in \mathbb{N}_0$ such that $f^n(x) \in \{p, q\}$, where $p = f(p) = f(q)$, $q \neq p$.*

The last sentence of this proposition says that h is "very close" to being a conjugacy: There are only countably many image points where injectivity fails.

An important feature of this coding is that it is obtained in a uniform way for all expanding maps, and that the absence of injectivity occurs at points defined by their dynamics, namely, the fixed point and its preimages. This leads us to the prize promised at the beginning:

Theorem 7.4.3 *If $f, g\colon S^1 \to S^1$ are expanding maps of degree 2, then f and g are topologically conjugate; in particular, every expanding map of S^1 of degree 2 is conjugate to E_2.*

Proof We have semiconjugacies $h_f, h_g\colon \Omega_2^R \to S^1$ for f and g. For $x \in S^1$, consider the set $H_x := h_g(h_f^{-1}(\{x\}))$. If x is a point of injectivity of h_f, that is, $h_f^{-1}(\{x\})$ is a single point, then so is H_x. Otherwise, x is a preimage of the fixed point under some iterate of f and $h_f^{-1}(\{x\})$ consists of a collection of sequences that are mapped under h_g to a single point. Therefore, H_x always consists of precisely one point $h(x)$. The bijective map $h\colon S^1 \to S^1$ thus defined is clearly a conjugacy: $h \circ f = g \circ h$. It is continuous because h_f sends open sets to open sets, that is, the image of a sequence and all sufficiently closeby sequences contains a small interval. Exchanging f and g shows that h^{-1} is also continuous. □

This holds for any degree via an appropriate coding. It is the first major conjugacy result that establishes conjugacy with a specific model for all maps from a certain class. The Poincaré Classification Theorem 4.3.20 comes close, but requires extra assumptions (such as the existence of the second derivative; see Section 4.4.3) to produce a conjugacy with a rotation.

7.4.3 Quadratic Maps

For $\lambda > 4$ consider the quadratic map

$$f\colon \mathbb{R} \to \mathbb{R}, \quad x \to \lambda x(1-x).$$

If $x < 0$, then $f(x) < x$ and $f'(x) > \lambda > 4$, so $f^n(x) \to -\infty$. When $x > 1$, $f(x) < 0$ and hence $f^n(x) \to -\infty$. Thus the set of points with bounded orbits is $\bigcap_{n\in\mathbb{N}_0} f^{-n}([0,1])$.

Proposition 7.4.4 *If $\lambda > 2 + \sqrt{5}$ and $f\colon \mathbb{R} \to \mathbb{R}$, $x \to \lambda x(1-x)$, then there is a homeomorphism $h\colon \Omega_2^R \to \Lambda := \bigcap_{n\in\mathbb{N}_0} f^{-n}([0,1])$ such that $h \circ \sigma^R = f \circ h$, that is, $f\restriction_\Lambda$ is conjugate to the 2-shift.*

Proof Let

$$\Delta_0 = \left[0, \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\lambda}}\right] \qquad \text{and} \qquad \Delta_1 = \left[\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{\lambda}}, 1\right].$$

Then $f^{-1}([0,1]) = \Delta_0 \cup \Delta_1$ by solving the quadratic equation $f(x) = 1$. Likewise, $f^{-2}([0,1]) = \Delta_{00} \cup \Delta_{01} \cup \Delta_{11} \cup \Delta_{10}$ consists of four intervals, and so forth. Consider the partition of Λ by Δ_0 and Δ_1. These pieces do not overlap and

$$\begin{aligned} |f'(x)| = |\lambda(1-2x)| = 2\lambda \left|x - \frac{1}{2}\right| &\geq 2\lambda\sqrt{\frac{1}{4} - \frac{1}{\lambda}} \\ &= \sqrt{\lambda^2 - 4\lambda} > \sqrt{(2+\sqrt{5})^2 - 4(2+\sqrt{5})} = 1 \end{aligned}$$

on $\Delta_0 \cup \Delta_1$. Thus, for any sequence $\omega = (\omega_0, \omega_1, \dots)$, the diameter of the intersections

$$\bigcap_{n=0}^{N} f^{-n}(\Delta_{\omega_n})$$

decreases (exponentially) as $N \to \infty$. This shows that for a sequence $\omega = (\omega_0, \omega_1, \dots)$ the intersection

$$h(\{\omega\}) = \bigcap_{n \in \mathbb{N}_0} f^{-n}(\Delta_{\omega_n}) \tag{7.4.3}$$

consists of exactly one point and this map $h\colon \Omega_2^R \to \Lambda$ is a homeomorphism. □

Remark 7.4.5 It turns out that Proposition 7.4.4 holds whenever $\lambda > 4$ (Proposition 11.4.1), but this is significantly less straightforward to prove than the present result. The situation present in either case, where a map folds an interval entirely over itself, is referred to as a one-dimensional *horseshoe*, in analogy to the geometry seen in the next subsection.

7.4.4 Linear Horseshoe

We now describe Smale's original "*horseshoe*," which provides one of the best examples of perfect coding. (In Section 7.3.3 a special case was constructed, in which ternary expansion provides the coding.)

Let Δ be a rectangle in $\mathbb{R}^2$ and $f\colon \Delta \to \mathbb{R}^2$ a diffeomorphism of Δ onto its image such that the intersection $\Delta \cap f(\Delta)$ consists of two "horizontal" rectangles Δ_0 and Δ_1 and the restriction of f to the components $\Delta^i := f^{-1}(\Delta_i)$, $i = 0, 1$, of $f^{-1}(\Delta)$ is a hyperbolic linear map, contracting in the vertical direction and expanding in the horizontal direction. This implies that the sets Δ^0 and Δ^1 are "vertical" rectangles. One of the simplest ways to achieve this effect is to bend Δ into a "horseshoe", or rather into the shape of a permanent magnet (Figure 7.4.2), although this method produces some inconveniences with orientation. Another way, which is better from the point of view of orientation, is to bend Δ roughly into a paper clip shape (Figure 7.4.3). This is an exaggerated version of the ternary horseshoe in Section 7.3.3, which also leaves some extra margin. If the horizontal and vertical rectangles lie strictly inside Δ, then the maximal invariant subset $\Lambda = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta)$ of Δ is contained in the interior of Δ.

Proposition 7.4.6 $f\restriction_{\Lambda}$ *is topologically conjugate to* σ_2.

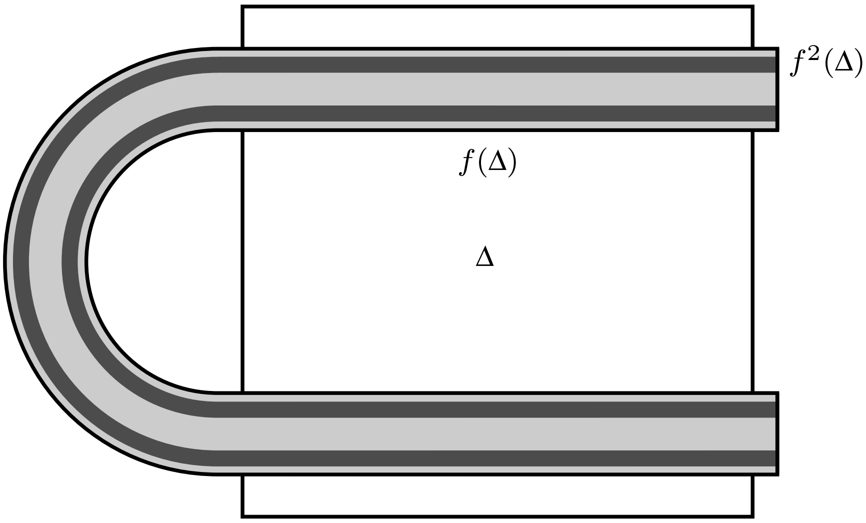

Figure 7.4.2. The horseshoe.

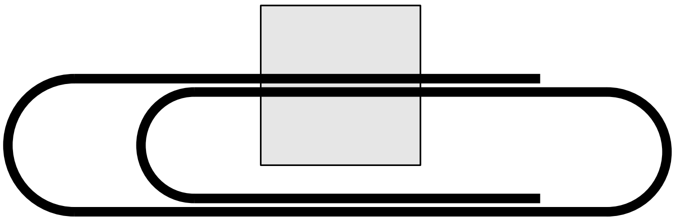

Figure 7.4.3. The paper clip.

Proof We use Δ^0 and Δ^1 as the "pieces" in the coding construction and start with positive iterates. The intersection $\Delta \cap f(\Delta) \cap f^2(\Delta)$ consists of four thin horizontal rectangles: $\Delta_{ij} = \Delta_i \cap f(\Delta_j) = f(\Delta^i) \cap f^2(\Delta^j)$, $i, j \in \{0, 1\}$ (see Figure 7.4.2). Continuing inductively, one sees that $\bigcap_{i=0}^n f^i(\Delta)$ consists of 2^n thin disjoint horizontal rectangles whose heights are exponentially decreasing with n. Each such rectangle has the form $\Delta_{\omega_1,\dots,\omega_n} = \bigcap_{i=1}^n f^i(\Delta^{\omega_i})$, where $\omega_i \in \{0, 1\}$ for $i = 1, \dots, n$. Each infinite intersection $\bigcap_{n=1}^\infty f^n(\Delta^{\omega_n})$, $\omega_n \in \{0, 1\}$, is a horizontal segment, and the intersection $\bigcap_{n=1}^\infty f^n(\Delta)$ is the product of the horizontal segment with a Cantor set in the vertical direction. Similarly, one defines and studies vertical rectangles $\Delta^{\omega_0,\dots,\omega_{-n}} = \bigcap_{i=0}^n f^{-i}(\Delta^{\omega_{-i}})$, the vertical segments $\bigcap_{n=0}^\infty f^{-n}(\Delta^{\omega_{-n}})$, and the set $\bigcap_{n=0}^\infty f^{-n}(\Delta)$, which is the product of a segment in the vertical direction with a Cantor set in the horizontal direction.

The desired invariant set $\Lambda = \bigcap_{n=-\infty}^\infty f^{-n}(\Delta)$ is the product of two Cantor sets and hence is a Cantor set itself (Problem 2.7.5), and the map

$$h\colon \Omega_2 \to \Lambda, \qquad h(\{\omega\}) = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta^{\omega_n})$$

is a homeomorphism that conjugates the shift σ_2 and the restriction of the diffeomorphism f to the set Λ. □

Since periodic points and topological mixing are invariants of topological conjugacy, Proposition 7.4.6 and Proposition 7.3.4 immediately give substantial information about the behavior of f on Λ.

Corollary 7.4.7 *Periodic points of f are dense in Λ, $P_n(f\restriction_\Lambda) = 2^n$, and the restriction of f to Λ is topologically mixing.*

Remark 7.4.8 Any map for which there is a perfect coding is defined on a Cantor set, because the perfect coding establishes a homeomorphism between the phase space and a sequence space, which is a Cantor set.

7.4.5 Coding of the Toral Automorphism

The idea of coding can be applied to hyperbolic toral automorphisms. To simplify notations and keep the construction more visual, we consider the standard example. Among our examples, this is the first where the coding is ingenious, even though it is geometrically simple. Section 10.3 describes a construction whose dynamical implications are quite similar to those obtained here, but where the geometry is complicated and almost always fractal.

Theorem 7.4.9 *For the map*

$$F(x, y) = (2x + y,\ x + y) \qquad (\mathrm{mod}\ 1)$$

of the 2-torus from Section 7.1.4 there is a semiconjugacy $h\colon \Omega_A \to \mathbb{T}^2$ with

$$F \circ h = h \circ \sigma_5\restriction_{\Omega_A}, \qquad \text{where}$$

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}. \tag{7.4.4}$$

Proof Draw segments of the two eigenlines at the origin until they cross sufficiently many times and separate the torus into disjoint rectangles. Specifically, extend a segment of the contracting line in the fourth quadrant until it intersects the segment of the expanding line twice in the first quadrant and once in the third quadrant (see Figure 7.4.4). The resulting configuration is a decomposition of the torus into two rectangles $R^{(1)}$ and $R^{(2)}$. Three pairs among the seven vertices of the plane configuration are identified, so there are only four different points on the torus that serve as vertices of the rectangles; the origin and three intersection points. Although $R^{(1)}$ and $R^{(2)}$ are not disjoint, one can apply the method used for the horseshoe, using $R^{(1)}$ and $R^{(2)}$ as basic rectangles. The expanding and contracting eigendirections play the role of the "horizontal" and "vertical" directions, correspondingly. Figure 7.4.5 shows that the image $F(R^{(i)})$ $(i = 1, 2)$ consists of several "horizontal" rectangles of full length. The union of the boundaries $\partial R^{(1)} \cup \partial R^{(2)}$ consists of the segments of the two eigenlines at the origin just described. The image of the contracting segment is a part of that segment. Thus, the images of $R^{(1)}$ and $R^{(2)}$ have to be "anchored" at parts of their "vertical" sides; that is, once one of the images "enters" either $R^{(1)}$ or $R^{(2)}$, it has to stretch all the way through it. By matching things up along the contracting direction one sees that $F(R^{(1)})$ consists of three components, two in $R^{(1)}$ and one in $R^{(2)}$. The image of $R^{(2)}$ has two components, one in each

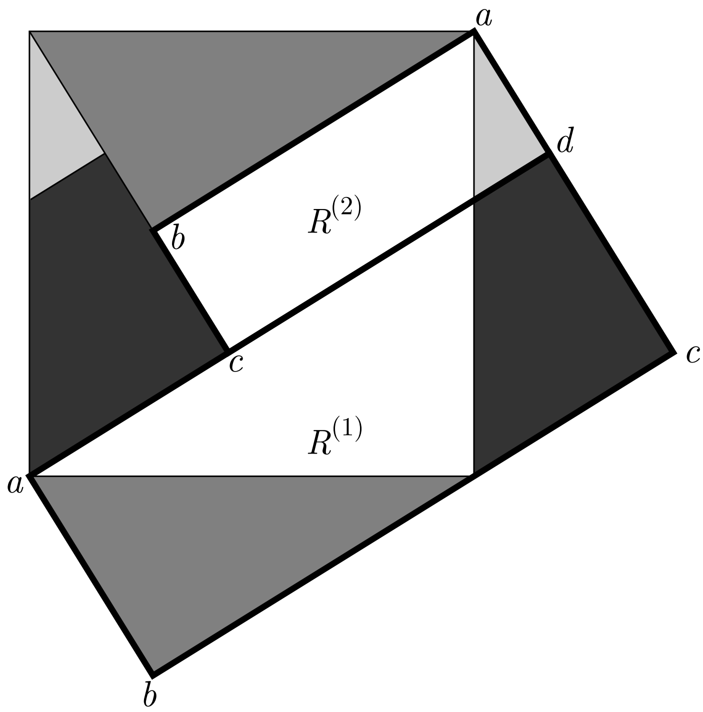

Figure 7.4.4. Partitioning the torus.

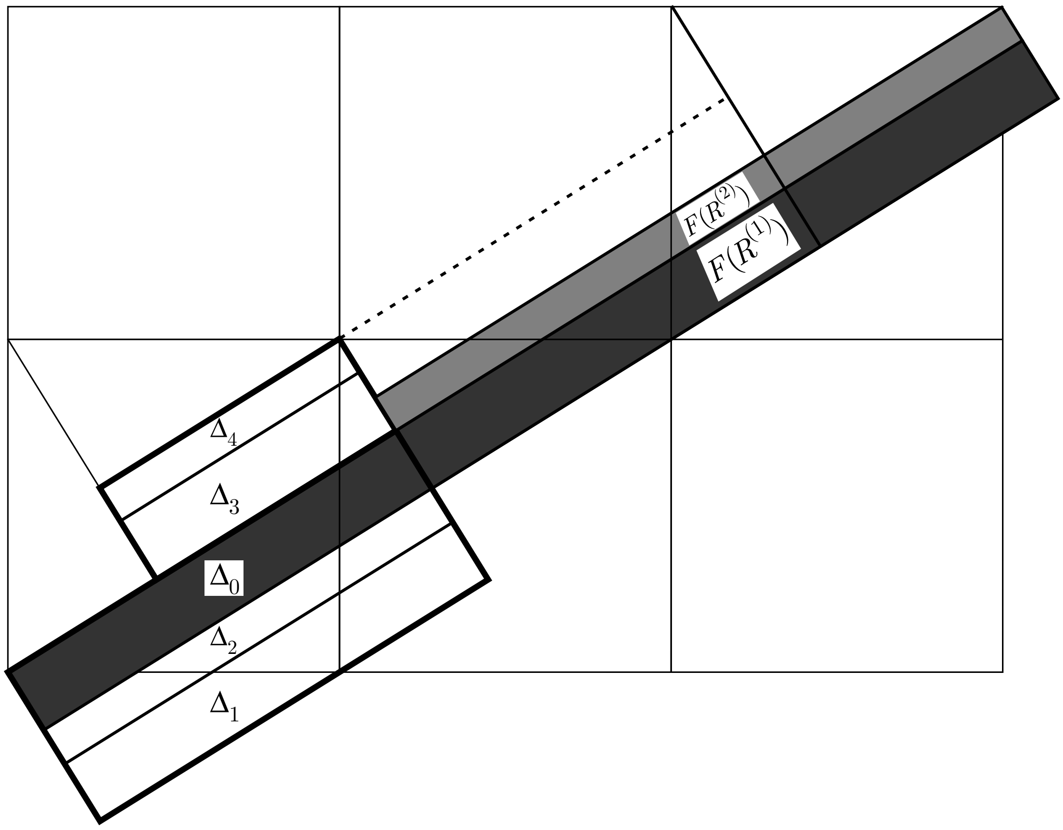

Figure 7.4.5. The image of the partition.

rectangle (see Figure 7.4.5). The fact that $F(R^{(1)})$ has two components in $R^{(1)}$ would cause problems if we were to use $R^{(1)}$ and $R^{(2)}$ for coding construction (more than one point for some sequences), but we use these five components Δ_0, Δ_1, Δ_2, Δ_3, Δ_4 (or their preimages) as the pieces in our coding construction. There is exactly one rectangle $\Delta_{\omega_{-\ell}\ldots\omega_0,\omega_1\ldots\omega_k}$ defined by $\bigcap_{n=-\ell}^{k} F^{-n}(\Delta_{\omega_n})$, not several. (As in the case of expanding maps in Section 7.3.1, we have to discard extraneous pieces, in this case line segments.) Due to the contraction of F in the "vertical" direction, $\Delta_{\omega_{-\ell}\ldots\omega_0,\omega_1\ldots\omega_k}$ has "height" less than $((3-\sqrt{5})/2)^{\ell}$, and due to the contraction of F^{-1} in the "horizontal" direction $\Delta_{\omega_{-\ell}\ldots\omega_0,\omega_1\ldots\omega_k}$ has "width" less than $((3-\sqrt{5})/2)^{k}$. These go to zero as $\ell \to \infty$ and $k \to \infty$, so the intersection $\bigcap_{n\in\mathbb{Z}} F^{-n}(\Delta_{\omega_n})$ defines at most one point $h(\omega)$. On the other hand, because of the "Markov" property described previously, that is, the images going full length through rectangles, the following is true: If $\omega \in \Omega_5$ and $F^{-1}(\Delta_{\omega_n})$ overlaps $\Delta_{\omega_{n+1}}$ for all $n \in \mathbb{Z}$, then there is such a point $h(\omega)$ in $\bigcap_{n\in\mathbb{Z}} F^{-n}(\Delta_{\omega_n})$. Thus, we have a coding, which, however, is not defined for all sequences of Ω_5.

Instead, we have to restrict attention to the subspace Ω_A of Ω_5 that contains only those sequences where any two successive entries constitute an "allowed transition", that is, 0, 1, 2 can be followed by 0, 1, or 3, and 3 and 4 can be followed by 2 or 4. This is exactly the topological Markov chain (Definition 7.3.2) for (7.4.4). □

Theorem 7.4.10 *The semiconjugacy between σ_A and F is one-to-one on all periodic points except for the fixed points. The number of preimages of any point not negatively asymptotic to the fixed point is bounded.*

Proof We describe carefully the identifications arising from our semiconjugacy, that is, what points on the torus have more than one preimage. First, obviously, the topological Markov chain σ_A has three fixed points, namely, the constant sequences of 0's, 1's, and 4's, whereas the toral automorphism F has only one, the origin. It is easy to see that all three fixed points are indeed mapped to the origin. As we have seen in Proposition 7.1.10, $P_n(F) = \lambda_1^n + \lambda_1^{-n} - 2$, and accordingly $P_n(\sigma_A) = \operatorname{tr} A^n = \lambda_1^n + \lambda_1^{-n} = P_n(F) + 2$ (Corollary 7.3.6), where $\lambda_1 = (3+\sqrt{5})/2$ is the maximal eigenvalue for both the 2×2 matrix $\left(\begin{smallmatrix}2&1\\1&1\end{smallmatrix}\right)$ and for the 5×5 matrix (7.4.4). To see that the eigenvalues are the same, consider $A - \lambda\,\mathrm{Id}$, subtract column 4 from the first two columns and column 5 from the third, and then add rows 1 and 2 to row 4 and row 3 to row 5:

$$\begin{pmatrix} 1-\lambda & 1 & 0 & 1 & 0 \\ 1 & 1-\lambda & 0 & 1 & 0 \\ 1 & 1 & -\lambda & 1 & 0 \\ 0 & 0 & 1 & -\lambda & 1 \\ 0 & 0 & 1 & 0 & 1-\lambda \end{pmatrix} \to \begin{pmatrix} -\lambda & 0 & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 1 & 0 \\ 0 & 0 & -\lambda & 1 & 0 \\ \lambda & \lambda & 0 & -\lambda & 1 \\ 0 & 0 & \lambda & 0 & 1-\lambda \end{pmatrix}$$

$$\to \begin{pmatrix} -\lambda & 0 & 0 & 1 & 0 \\ 0 & -\lambda & 0 & 1 & 0 \\ 0 & 0 & -\lambda & 1 & 0 \\ 0 & 0 & 0 & 2-\lambda & 1 \\ 0 & 0 & 0 & 1 & 1-\lambda \end{pmatrix}.$$

Furthermore, one can see that every point $q \in \mathbb{T}^2$ whose positive and negative iterates avoid the boundaries $\partial R^{(1)}$ and $\partial R^{(2)}$ has a unique preimage, and vice versa. In particular, periodic points other than the origin (which have rational coordinates) fall into this category. The points of Ω_A whose images are on those boundaries or their iterates under F fall into three categories corresponding to the three segments of stable and unstable manifolds through 0 that define parts of the boundary. Thus sequences are identified in the following cases: They have a constant infinite right (future) tail consisting of 0's or 4's, and agree otherwise – this corresponds to a stable boundary piece – or else an infinite left (past) tail (of 0's and 1's, or of 4's), and agree otherwise – this corresponds to an unstable boundary piece. □

EXERCISES

Exercise 7.4.1 Prove that for $\lambda \geq 1$ every bounded orbit of the quadratic map f_λ is in $[0, 1]$.

Exercise 7.4.2 Give a detailed argument that (7.4.3) defines a homeomorphism.

Exercise 7.4.3 Construct a Markov partition for $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ that consists of two squares.

Exercise 7.4.4 Construct a Markov partition and describe the corresponding topological Markov chain for the automorphism F_L, where $L = \left(\begin{smallmatrix} 1 & 1 \\ 2 & 1 \end{smallmatrix}\right)$.

Exercise 7.4.5 Given a 0-1 $n \times n$-matrix A, describe a system of n rectangles $\Delta_1, \dots, \Delta_n$ in $\mathbb{R}^2$ and map $f\colon \Delta := \bigcup_{i=1}^{n} \Delta_i \to \mathbb{R}^2$ such that the restriction of f to the set of points that stay inside Δ for all iterates of f is topologically equivalent to the topological Markov chain σ_A.

Exercise 7.4.6 Check that the process (7.4.2) of discarding extraneous points in the coding construction amounts to taking $\Delta_{\omega_0,\dots,\omega_{n-1}} = \overline{\bigcap_{i=0}^{n-1} \operatorname{Int}(f^{-i}(\Delta_{\omega_i}))}$, and $\{h(\omega)\} := \bigcap_{n \in \mathbb{N}} \Delta_{\omega_0,\dots,\omega_{n-1}}$.

PROBLEMS FOR FURTHER STUDY

Problem 7.4.7 Show that the assertion of Theorem 7.4.3 remains true for any map f of degree 2 such that $f' \geq 1$ and $f' = 1$ only at finitely many points.

Problem 7.4.8 Prove the assertion of Theorem 7.4.9 for some 0-1 matrix A for any automorphism

$$F_L\colon \mathbb{T}^2 \to \mathbb{T}^2, x \mapsto Lx \pmod 1,$$

where L is an integer 2×2 matrix with determinant $+1$ or -1 and with real eigenvalues different from ± 1.

7.5 UNIFORM DISTRIBUTION

We now investigate whether the notion of the uniform distribution of orbits that appeared in previous chapters for rotations of the circle and translations of the torus has any meaning for the group of examples discussed in the present chapter, such as linear or nonlinear expanding maps of the circle, shifts, and automorphisms of the torus.