Moreover, the a_i's are chosen to reflect the dynamics—the orbit structure—of f. In this way we have reduced the study of the mapping to an essentially combinatorial problem involving the symbols $\{1, 2\}$.

EXERCISE 5.0.2. Show that there is a countable infinity of periodic orbits and asymptotically periodic orbits for g.

EXERCISE 5.0.3. Work out the symbolic dynamics for the cubic map $x \to f(x) = ax - x^3$ and describe the structure of its invariant set for "large" a. (Hint: Find a_c such that for $a > a_c$ there is a subinterval $I = [-\sqrt{1 + a}, \sqrt{1 + a}]$ such that $I \cap f^{-1}(I)$ has three components on which $|f'(x)| > 1$.)

5.1. The Smale Horseshoe: An Example of a Hyperbolic Limit Set

The one-dimensional mapping g described above is closely related to another example, the *Smale horseshoe*, which is a hyperbolic limit set that has been a principal motivating example for the development of the modern theory of dynamical systems. We shall now describe this example in detail, using its symbolic dynamics. The example is described in terms of an invertible planar map which can be thought of as a Poincaré map arising from a three-dimensional autonomous differential equation or a forced oscillator problem. In Section 5.3, below, we will see how the horseshoe arises whenever one has transverse homoclinic orbits, as in the Duffing equation. In Section 2.4 we have already met an example in which maps and horseshoes arise directly.

We begin with the unit square $S = [0, 1] \times [0, 1]$ in the plane and define a mapping $f: S \to \mathbb{R}^2$ so that $f(S) \cap S$ consists of two components which are mapped rectilinearly by f. See Figure 5.1.1.

One can think of the map as performing a linear vertical expansion and a horizontal contraction of S, by factors μ and λ, respectively, followed by a folding, the latter being done so that the folded portion falls outside S. Figure 5.1.1(b). Thus, restricted to $S \cap f^{-1}(S)$, the map is linear.

Reversing the folding and stretching, one can easily see that the inverse image, $f^{-1}(S \cap f(S)) = S \cap f^{-1}(S)$, is two horizontal bands $H_1 = [0, 1] \times [a, a + \mu^{-1}]$ and $H_2 = [0, 1] \times [b, b + \mu^{-1}]$, on each of which f will have a constant Jacobian

$$\begin{pmatrix} \pm\lambda & 0 \\ 0 & \pm\mu \end{pmatrix}, \qquad (+ \text{ on } H_1, - \text{ on } H_2), \tag{5.1.1}$$

with $0 < \lambda < \frac{1}{2}$ and $\mu > 2$. On each H_i, the map f compresses horizontal segments by a factor of λ and stretches vertical segments by a factor of μ.

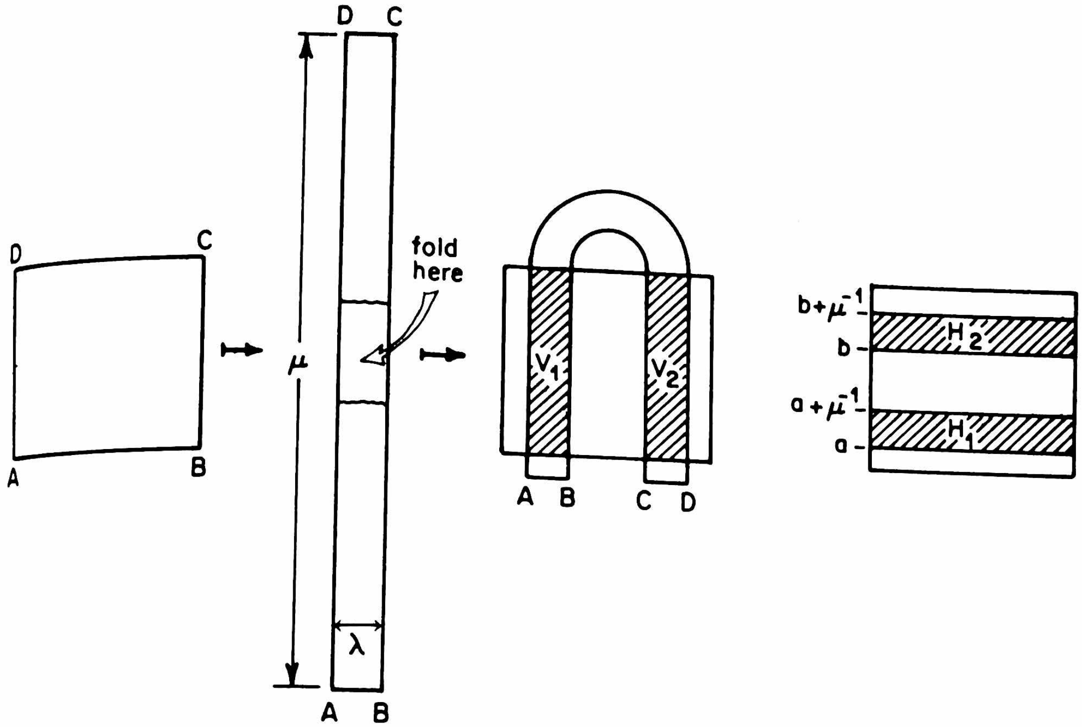

Figure 5.1.1. The Smale horseshoe.

As f is iterated, most points either leave S or are not contained in an image $f^i(S)$. Those points which do remain in S for all time form a set $\Lambda = \{x \mid f^i(x) \in S, -\infty < i < \infty\}$. This set Λ has a complicated topological structure that we now describe. Each horizontal band H_i is stretched by f to a rectangle $V_i = f(H_i)$ which intersects both H_1 and H_2. Since f is rectilinear on H_i, those points that end up in H_i after applying f come from thinner horizontal strips in H_i. See Figure 5.1.2.

Now $H_1 \cup H_2 = f^{-1}(S \cap f(S))$, so the four thinner strips constitute $f^{-2}(S \cap f(S) \cap f^2(S))$. If we continue this argument inductively, we find that $f^{-n}(S \cap f(S) \cap \cdots \cap f^n(S))$ is the union of 2^n horizontal strips. The thickness of each of these is μ^{-n} since $|\partial f/\partial y| = \mu$ at all points of $H_1 \cup H_2$ and the first $(n-1)$ iterates of the horizontal strips remain inside $H_1 \cup H_2$.

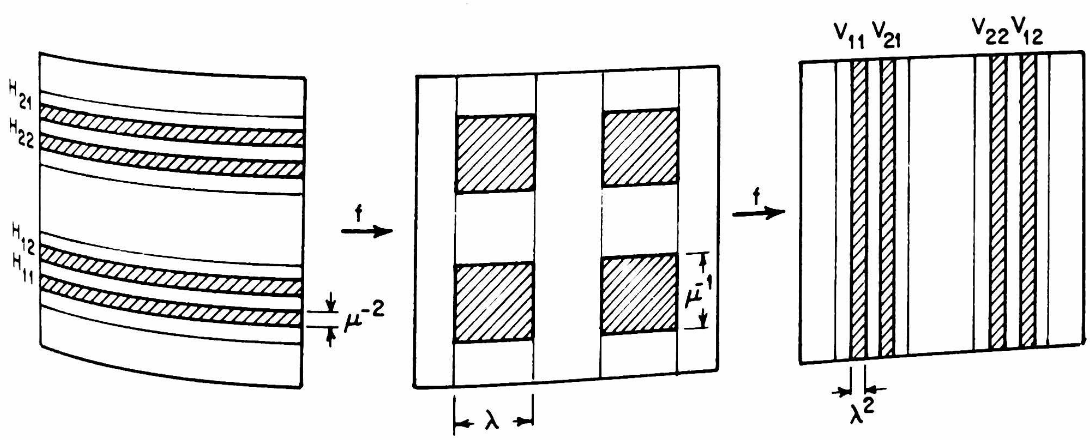

Figure 5.1.2. Iteration of $f: V_{ij} = f^2(H_{ij})$.

The intersection of all these horizontal strips (as $n \to \infty$) forms a Cantor set of horizontal segments. We shall discuss this structure below.

Consider now the image under f^n of one of the 2^n horizontal strips in $f^{-n}(S \cap f(S) \cap \cdots \cap f^n(S))$. Using the chain rule, we have

$$Df^n = \begin{pmatrix} \pm\lambda^n & 0 \\ 0 & \pm\mu^n \end{pmatrix}$$

at these points, so the image is a rectangle of horizontal width λ^n which extends vertically from the top to the bottom of the square. The map f^n is 1–1, so the images of the horizontal strips are distinct. We conclude that $S \cap f(S) \cap \cdots \cap f^n(S)$ is the union of 2^n vertical strips, each of width λ^n. The intersection of these sets over all $n \geq 0$ is a Cantor set of vertical segments composed of those points which are in the images of all the f^n. To be in Λ, a point x must be in both a vertical segment and a horizontal segment from the collection described above. Therefore, topologically, Λ is itself a Cantor set: its components are each points and each point of Λ is an accumulation point for Λ. In Figure 5.1.3 we show the sixteen components of $f^{-2}(S) \cap f^{-1}(S) \cap S \cap f(S) \cap f^2(S)$, to give an idea of the structure of Λ.

So far we have essentially repeated the informal description of the horseshoe in Section 2.4. We can, however, achieve a more complete description which contains information about the dynamics of each point by a construction similar to the one we used for the one-dimensional mappings in the previous section. In this construction, we note which horizontal band H_1 or H_2 each iterate of a point $x \in \Lambda$ visits, and use this information as an actual characterization of the point. Each point $x \in \Lambda$ will be characterized by a *bi-infinite* sequence, since here the map is invertible, unlike $x \to 2x$ (mod 1).

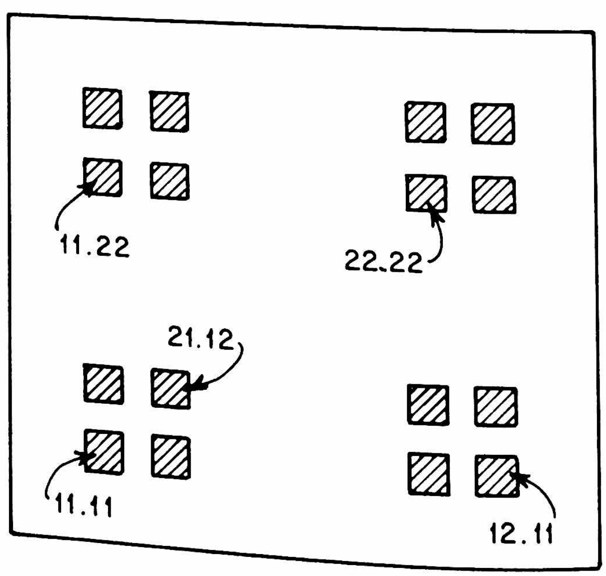

Figure 5.1.3. The small black rectangles of width λ^2 and height μ^{-2} are the components of $\bigcap_{n=-2}^{2} f^n(S)$. The four symbol sequences $\{a_{-2}a_{-1} \cdot a_0 a_1\}$ refer to Exercise 5.1.1 below.

A bi-infinite sequence is one whose index set is all of $\mathbb{Z}$: we use the notation $\mathbf{a} = \{a_i\}_{i=-\infty}^{\infty}$.

Theorem 5.1.1. *There is a 1–1 correspondence ϕ between Λ and the set Σ of bi-infinite sequences of two symbols such that the sequence $\mathbf{b} = \phi(f(x))$ is obtained from the sequence $\mathbf{a} = \phi(x)$ by shifting indices one place: $b_i = a_{i+1}$. The set Σ has a metric defined by*

$$d(a, b) = \sum_{i=-\infty}^{\infty} \delta_i 2^{-|i|}, \qquad \delta_i = \begin{cases} 0 & \text{if } a_i = b_i, \\ 1 & \text{if } a_i \neq b_i. \end{cases} \tag{5.1.2}$$

The map ϕ is a homeomorphism from Λ to Σ endowed with this metric.

Proof. The proof of this theorem provides a basic illustration of how symbolic dynamics works. Take the two symbols of the theorem to be 1 and 2. The map is defined by the recipe

$$\phi(x) = \{a_i\}_{i=-\infty}^{\infty}, \quad \text{with } f^i(x) \in H_{a_i}. \tag{5.1.3}$$

In words, x is in Λ if and only if $f^i(x)$ is in $H_1 \cup H_2$ for each i, and we associate to x the sequence of indices that tells us which of H_1 and H_2 contains each $f^i(x)$. Unlike the map $f = 2x$ (mod 1), this definition of ϕ is unambiguous because H_1 and H_2 are disjoint. This description of ϕ leads immediately to the shift property required by the theorem: since $f^{i+1}(x) = f^i(f(x))$, it follows that $\phi(f(x))$ is obtained from $\phi(x)$ by shifting indices. To see that ϕ is both 1–1 and continuous, we look at the set of x's which each possess a given central string of symbols. Specifying $b_{-m}, b_{-m+1}, \ldots b_0, \ldots, b_n$ we denote as $R(b_{-m}, b_{-m+1}, \ldots b_0, \ldots, b_n)$ the set of x's for which $f^i(x) \in H_{b_i}$ for $-m \le i \le n$. We observe inductively that $R(b_{-m}, \ldots, b_n)$ is a rectangle of height $\mu^{-(n+1)}$ and width λ^m, obtained from the intersection of a horizontal and a vertical strip. As one lets $m, n \to \infty$, the diameter of the sets $R(b_{-m}, \ldots, b_n) \to 0$. Consequently, ϕ is both 1–1 and continuous.

The final point is that ϕ is onto. This is crucial for the applications of symbolic dynamics. The reason that ϕ is onto is that for each choice of $b_{-m}, \ldots, b_n$, the set $R(b_{-m}, \ldots, b_n)$ is nonempty. To see this, reference to Figure 5.1.2 is helpful. Note that $R(b_0, \ldots, b_n)$ is a horizontal strip mapped vertically from top to bottom of the square S by f^{n+1}. Therefore, $f^{n+1}(R(b_0, \ldots, b_n))$ intersects each H_i and $R(b_0, \ldots, b_n, b_{n+1})$ is a nonempty horizontal strip extending across S. Similarly, we have already observed that $S \cap f(S) \cap \cdots \cap f^m(S)$ consists of 2^m vertical strips. Each of these is a set of the form $R(b_{-m}, \ldots, b_{-1})$ and all sequences $(b_{-m}, \ldots, b_{-1})$ occur. Finally, $R(b_{-m}, \ldots, b_n)$ is nonempty because every vertical strip $R(b_{-m}, \ldots, b_{-1})$ intersects every horizontal strip $R(b_0, \ldots, b_n)$ and $R(b_{-m}, \ldots, b_n) = R(b_{-m}, \ldots, b_{-1}) \cap R(b_0, \ldots, b_n)$. □

Exercise 5.1.1. Verify that the central parts of the symbol sequences attached to the rectangles of Figure 5.1.3 are correct and label the remaining shaded rectangles by their (finite) symbol sequences.

The correspondence ϕ between Λ and Σ imparts to Λ a symbolic description which is an extraordinarily useful tool for understanding the dynamics of Λ. It is helpful to give a formal name to the process of "shifting indices." Thus

$$\sigma: \Sigma \to \Sigma, \tag{5.1.4}$$

the *shift map*, is defined by $\sigma(\mathbf{a}) = \mathbf{b}$ with $b_i = a_{i+1}$. The basic property of the theorem is now restated as the equation

$$\phi \circ (f|_\Lambda) = \sigma \circ \phi. \tag{5.1.5}$$

This equation expresses the topological conjugacy of $f|_\Lambda$ and σ. Written as $f|_\Lambda = \phi^{-1} \circ \sigma \circ \phi$, it has the immediate consequence that

$$f^n|_\Lambda = \phi^{-1} \circ \sigma^n \circ \phi, \tag{5.1.6}$$

so that ϕ maps orbits of f in Λ to orbits of σ in Σ. The description of σ is explicit enough that many dynamical properties are readily determined. For example, a periodic orbit of period n for σ consists of a sequence which is periodic: $a_i = a_{i+n}$ for all i in the sequence $\mathbf{a}$. Fixing n, we readily count the sequences with the property $a_i = a_{i+n}$ and find that f^n has 2^n fixed points in Λ. This set includes all points which are periodic with period n or a divisor of n.

Exercise 5.1.2. Show that all the periodic orbits in Λ are of saddle type. Show that Λ contains a countable infinity of heteroclinic and homoclinic orbits. Show that Λ contains orbits which are not asymptotically periodic. List the first few periodic orbits (with periods, say ≤ 5) and locate them on Figure 5.1.3. Show that Λ contains an uncountable collection of nonperiodic orbits and describe their symbol sequences.

Exercise 5.1.3. Display a point of Σ whose σ-orbit is dense in Σ. (Hint: Two points of Σ are close if they agree in a long "central block": find a sequence which contains all finite strings of 1's and 2's.)

The description of the horseshoe we have just given is "robust" with respect to small changes in the mapping f. The reader should try to imagine what changes will take place if the assumption that f is rectilinear on $H_1 \cup H_2$ is dropped. We imagine perturbing f to a mapping $\tilde{f}$ in such a way that the Jacobian derivative of $\tilde{f}$ can be nonconstant but still is close to that of f. Qualitatively, nothing changes. The sets $S \cap \tilde{f}(S) \cap \cdots \cap \tilde{f}^n(S)$ will still consist of 2^n "vertical" strips (which are, however, no longer exactly rectangles). Similarly, the set $\tilde{f}^{-n}(S) \cap \cdots \cap S$ will consist of 2^n "horizontal" strips which are no longer exactly rectangles. Nonetheless, the set of points all of whose $\tilde{f}$-iterates remain in S form a set $\tilde{\Lambda}$ which is topologically conjugate to the shift Σ. This result of Smale [1963, 1967] is an early example of a *structural stability* theorem.

We can summarize the results of this section as follows:

Theorem 5.1.2. *The horseshoe map f has an invariant Cantor set Λ such that:*

(a) *Λ contains a countable set of periodic orbits of arbitrarily long periods.*
(b) *Λ contains an uncountable set of bounded nonperiodic motions.*
(c) *Λ contains a dense orbit.*

Moreover, any sufficiently C^1 close map $\tilde{f}$ has an invariant Cantor set $\tilde{\Lambda}$ with $\tilde{f}|_{\tilde{\Lambda}}$ topologically equivalent to $f|_{\Lambda}$.

We shall take up the question of nonlinear maps possessing horseshoes in the next section.

5.2. Invariant Sets and Hyperbolicity

The example described above, the Smale horseshoe, provides a good intuitive basis for the way in which orbits of mappings, and hence of ordinary differential equations, can be chaotic. Later in this chapter, we shall describe a general theory of "Axiom A" dynamical systems which builds upon this example. The concept of structural stability makes this generalization a very natural one, but the class of Axiom A systems is not adequate to encompass the various examples described in Chapter 2. There are many unresolved issues about the details of the dynamics in these examples, so that our discussion becomes more tentative toward the end of the chapter when we apply the theory developed here to these examples. Of particular interest will be the question of when a "typical" trajectory can be expected to have the chaotic dynamical features of the horseshoe. Our aim being to describe what is known, observed, and suspected, we include only details about the proofs of results which we feel are illuminating.

A few topological definitions are necessary at the beginning of our discussion. The horseshoe Λ described in Section 5.1 has a rather complicated topological structure, but it cannot be further split into closed invariant subsets because there are orbits which are dense in Λ. We want to focus on sets like these, which carry most of the interesting dynamical information of a flow. If ϕ is a discrete or continuous flow, then a fundamental property of all sets we consider is that they be invariant. (Recall from Section 1.6 that the set S is *invariant* if $\phi_t(S) = S$ for all t.) There are various kinds of invariant sets: the ones which interest us the most will be composed of asymptotic limit sets of points. Again let ϕ_t be a discrete or continuous flow, and recall from Section 1.6:

Definition 5.2.1. The *α limit set of x for ϕ_t* is the set of accumulation points of $\phi_t(x)$, $t \to -\infty$. The *ω limit set of x for ϕ_t* is the set of accumulation points of $\phi_t(x)$, $t \to \infty$. The α and ω limits of x are its asymptotic limit sets. (y is an

Theorem 5.1.2. *The horseshoe map f has an invariant Cantor set Λ such that:*

(a) *Λ contains a countable set of periodic orbits of arbitrarily long periods.*
(b) *Λ contains an uncountable set of bounded nonperiodic motions.*
(c) *Λ contains a dense orbit.*

Moreover, any sufficiently C^1 close map $\tilde{f}$ has an invariant Cantor set $\tilde{\Lambda}$ with $\tilde{f}|_{\tilde{\Lambda}}$ topologically equivalent to $f|_{\Lambda}$.

We shall take up the question of nonlinear maps possessing horseshoes in the next section.

5.2. Invariant Sets and Hyperbolicity

The example described above, the Smale horseshoe, provides a good intuitive basis for the way in which orbits of mappings, and hence of ordinary differential equations, can be chaotic. Later in this chapter, we shall describe a general theory of "Axiom A" dynamical systems which builds upon this example. The concept of structural stability makes this generalization a very natural one, but the class of Axiom A systems is not adequate to encompass the various examples described in Chapter 2. There are many unresolved issues about the details of the dynamics in these examples, so that our discussion becomes more tentative toward the end of the chapter when we apply the theory developed here to these examples. Of particular interest will be the question of when a "typical" trajectory can be expected to have the chaotic dynamical features of the horseshoe. Our aim being to describe what is known, observed, and suspected, we include only details about the proofs of results which we feel are illuminating.

A few topological definitions are necessary at the beginning of our discussion. The horseshoe Λ described in Section 5.1 has a rather complicated topological structure, but it cannot be further split into closed invariant subsets because there are orbits which are dense in Λ. We want to focus on sets like these, which carry most of the interesting dynamical information of a flow. If ϕ is a discrete or continuous flow, then a fundamental property of all sets we consider is that they be invariant. (Recall from Section 1.6 that the set S is *invariant* if $\phi_t(S) = S$ for all t.) There are various kinds of invariant sets; the ones which interest us the most will be composed of asymptotic limit sets of points. Again let ϕ_t be a discrete or continuous flow, and recall from Section 1.6: