Chapter IV

Chaotic Dynamics on Fractals

1 The Addresses of Points on Fractals

We begin by considering informally the concept of the *addresses* of points on the attractor of a hyperbolic IFS. Figure IV.92 shows the attractor of the IFS:

$$\{\mathbb{C};\ w_1(z) = (0.13 + 0.64i)z,\ w_2(z) = (0.13 + 0.64i)z + 1\}.$$

This attractor, A, is the union of two disjoint sets, $w_1(A)$ and $w_2(A)$, lying to the left and right, respectively, of the dotted line ab. In turn, each of these two sets is made of two disjoint sets:

$$w_1(A) = w_1(w_1(A)) \cup w_2(w_1(A)),\ w_2(A) = w_2(w_2(A)) \cup w_2(w_2(A)).$$

This leads to the idea of addressing points in terms of the sequences of transformations, applied to A, which lead to them. All points belonging to A, in the subset $w_1(w_1(A))$, are situated on the piece of the attractor that lies below dc and to the left of ab, and their addresses all begin $11\ldots$. Clearly, the more precisely we specify geometrically where a point in A lies, the more bits to the address we can provide. For example, every point to the right of ab, below ef, to the left of gh, has an address that begins $212\ldots$. In Theorem 2.1 we prove that, in examples such as this one, it is possible to assign a unique address to every point of A. In such cases we say that the IFS is "totally disconnected."

Here is a different type of example. Consider the IFS

$$\{\mathbb{C};\ w_1(z) = \frac{1}{2}z,\ w_2(z) = \frac{1}{2}z + \frac{1}{2},\ w_3(z) = \frac{1}{2}z + \frac{1}{2}i\}.$$

The attractor, A, of this IFS is a Sierpinski triangle with vertices at $(0, 0)$, $(1, 0)$, and $(0, 1)$. Again we can address points on A according to the sequences of transformations that lead to them. This time there are at least three points in A that

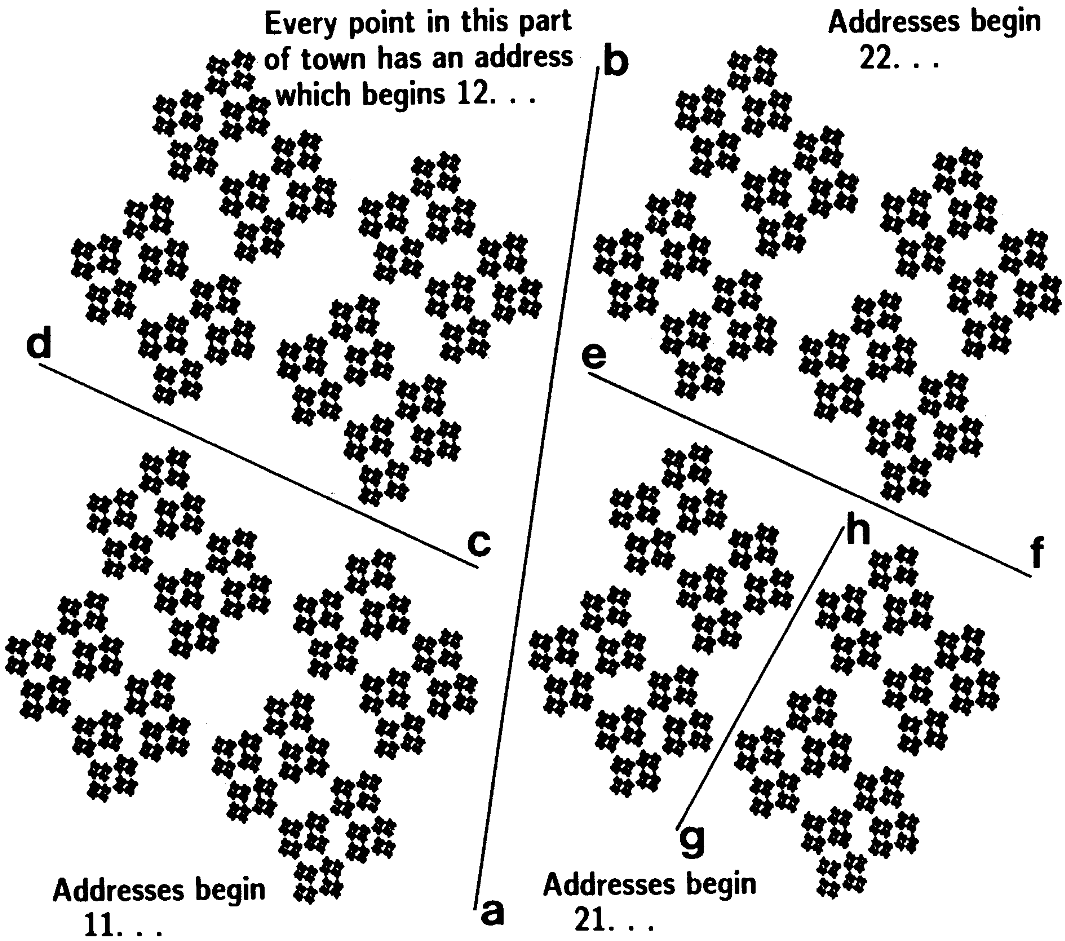


Figure IV.92. Addresses of points on an attractor. The lines ab, cd, ef, and gh are not part of the attractor.

have two addresses, because there is a point in each of the sets $w_1(A) \cap w_2(A)$, $w_2(A) \cap w_3(A)$, and $w_3(A) \cap w_1(A)$, as illustrated in Figure IV.93.

On the other hand, some points on the Sierpinski triangle have only one address, such as the three vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$. Although the attractor is connected, the proportion of points with multiple addresses is "small," in a sense we do not yet make precise. In such cases as this we say that the IFS is "just-touching." Notice that this terminology refers to the IFS itself rather than to its attractor.

Let us look at a third, fundamentally different example. Consider the hyperbolic IFS

$$\{[0, 1]; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}.$$

The attractor is $A = [0, 1]$, but now

$$w_1(A) \cap w_2(A) = [0, \frac{1}{2}] \cap [\frac{1}{4}, 1] = [\frac{1}{4}, \frac{1}{2}];$$

so $w_1(A) \cap w_2(A)$ is a significant piece of the attractor. The attractor would look very different if the overlapping piece $[\frac{1}{4}, \frac{1}{2}]$ were missing. Now observe that every

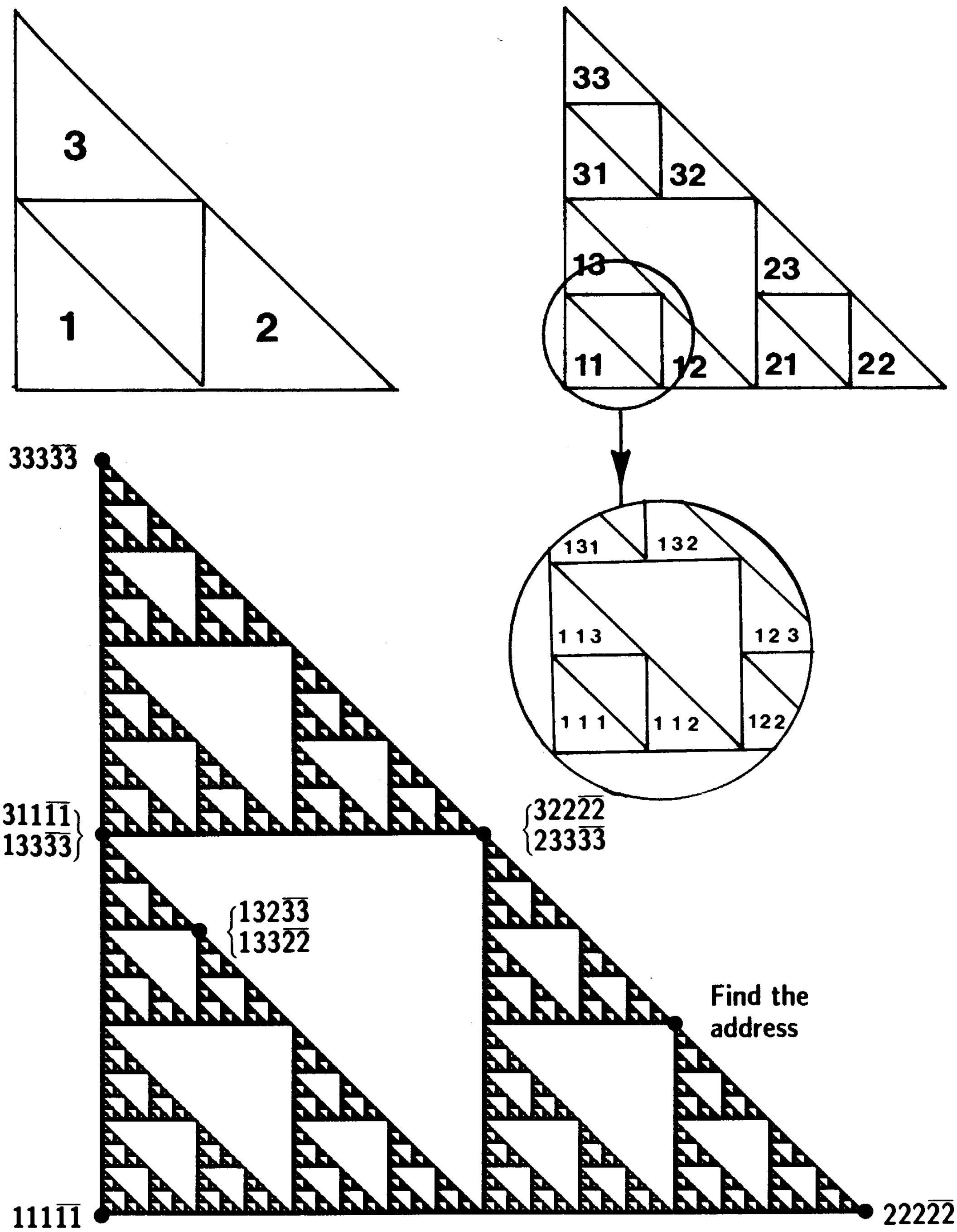


Figure IV.93. Some points on this Sierpinski triangle have two addresses, while others have only one address. Overlining on the last symbols, in an expression such as $311\overline{11}$, means that the overlined symbols are repeated endlessly. For example, $311\overline{11}$ = 31111111111111111 1111..., and $31\overline{123}$ = 31123123123123...

point in $[\frac{1}{4}, \frac{1}{2}]$ has at least two addresses. On the other hand, the points 0 and 1 have only one address each. Nonetheless, it appears that the proportion of points with multiple addresses is large. In such cases we say that the IFS is "overlapping."

The terminologies "totally disconnected," "just-touching," and "overlapping" refer to the IFS itself rather than to the attractor. The reason for this is that the same set may be the attractor of several different hyperbolic IFS's. Consider, for example,

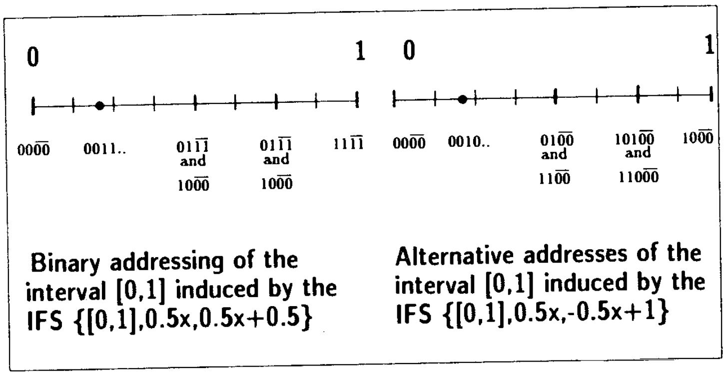


Figure IV.94. Different IFS's with the same attractor provide different addressing schemes. Here the symbols $\{0, 1\}$ are used in place of $\{1, 2\}$ for obvious reasons.

the two IFS's

$$\{[0, 1]; w_1(x) = \frac{1}{2}x, w_2(x) = \frac{1}{2}x + \frac{1}{2}\}$$

and

$$\{[0, 1]; w_1(x) = \frac{1}{2}x, w_2(x) = -\frac{1}{2}x + 1\}.$$

The attractor of each one is the real interval $[0, 1]$. We can obtain two different addressing schemes for the points in $[0, 1]$, as illustrated in Figure IV.94.

These two IFS are just-touching. However, the IFS

$$\{[0, 1]; w_1(x) = \frac{1}{2}x, w_2(x) = \frac{3}{4}x + \frac{1}{4}\}$$

is overlapping, while its attractor is also $[0, 1]$.

Examples & Exercises

1.1. Figure IV.95 shows the attractor of an IFS of the form $\{\mathbb{R}^2; w_n, n = 1, 2, 3\}$, where each of the transformations $w_n : \mathbb{R}^2 \to \mathbb{R}^2$ is affine. The addresses of several points are given. Find the addresses of a, b, and c.

1.2. In Figure IV.95 locate the point whose address is $111\overline{11}$.

1.3. A *quadtree* is an addressing scheme used in computer science for addressing small squares in the unit square $\blacksquare = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$ as follows. The square is broken into four quarters. Points in the first quarter have addresses that begin 0, points in the second quarter have addresses that begin 1, and so on, as illustrated in Figure IV.96. Find an IFS that gives rise to the addressing scheme suggested in Figure IV.96. Is this a totally disconnected, just-touching, or overlapping IFS?

1.4. Addresses are assigned to the Sierpinski triangle, as in Figure IV.93. Characterize the addresses of the set of points that lie on the outermost boundary, the triangle with vertices $\overline{11}$, $\overline{22}$, and $\overline{33}$.

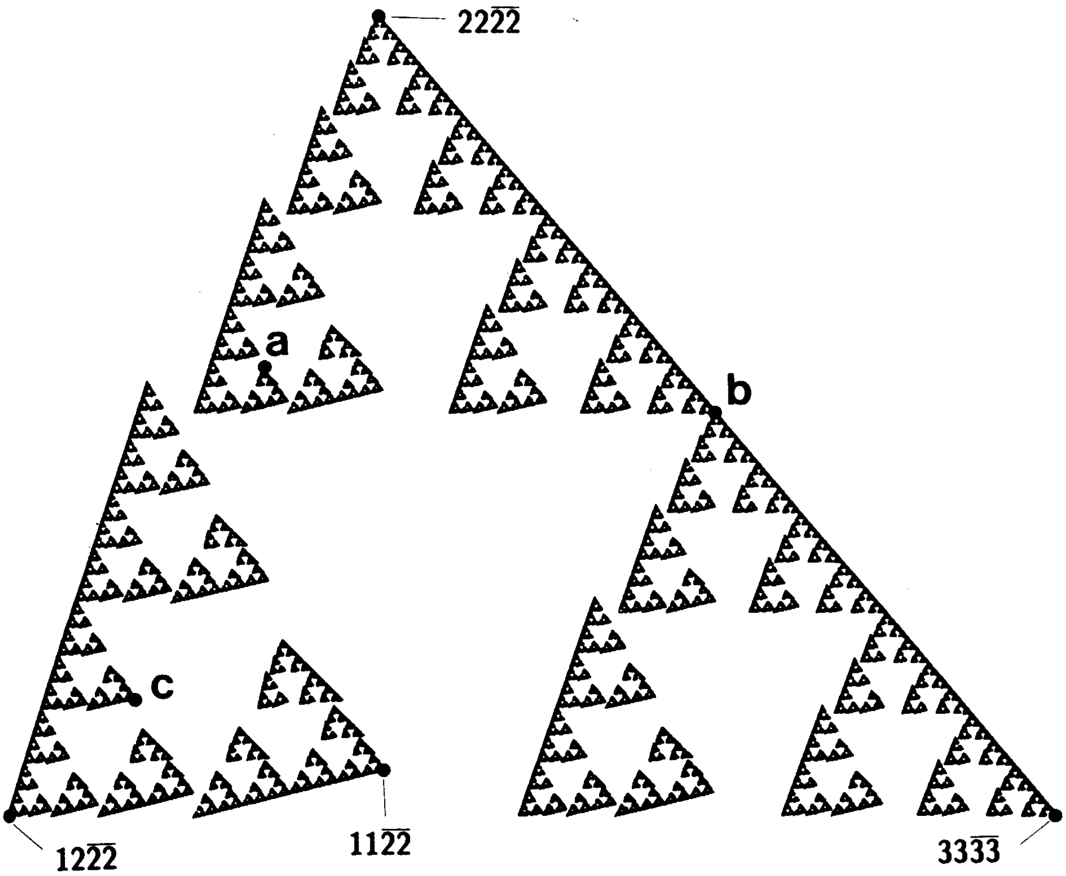


Figure IV.95. Can you find the addresses of a, b, and c?

1.5. Characterize the addresses of points belonging to the boundary of the largest hole in Figure IV.97.

1.6. Consider a hyperbolic IFS with condensation set C. Suppose the condensation set is itself the attractor of another hyperbolic IFS. Design an addressing scheme for the attractor of the IFS with condensation. Can all possible addresses occur?

1.7. Figure IV.98 shows an "overlapping" IFS attractor, for two affine transformations in $\mathbb{R}^2$. Choose one point in each of the marked regions on the attractor. Find the first four numbers in two different addresses for each of these points. The first few numbers in the addresses of some points on the attractor are included in the figure to remove possible ambiguities.

1.8. ⋆ Identify the set of addresses of points on the attractor, A, of a hyperbolic IFS with code space. Argue that nearby codes correspond to points on A which are nearby.

1.9. Address the real number 0.7513 in each of the two coding schemes given in Figure IV.94.

33	32	23	22
30	31	20	21
03	02	13	12
00	01	10	11

Figure IV.96. Addresses at depth two in a quadtree.

In thinking about the addresses of points on fractals, already we have been led to try to compare "how many" points have a certain property to how many have another property. For example, in the case of the addressing scheme on the Sierpinski triangle described above, we wanted to compare the number of points with multiple addresses to the number of points with single addresses. It turns out that both numbers are infinite. Yet still we want to compare their numbers. One way in which this may be done is through the concept of countability.

Definition 1.1 *Let S be a set. S is* countable *if it is empty or if there is an onto transformation $c : I \to S$, where I is either one of the sets*

$$\{1\}, \{1,2\}, \{1,2,3\}, \ldots, \{1,2,3,\ldots,n\}, \ldots,$$

or the positive integers $\{1,2,3,4,\ldots\}$. S is uncountable *if it is not countable.*

We think of an uncountable set as being larger than a countable set.

We are going to make fundamental use of code space to formalize the concept of addresses. How many points does code space contain?

Theorem 1.1 *Code space on two or more symbols is uncountable.*

Proof We prove it here for the code space on the two symbols $\{1, 2\}$. Denote an element of code space Σ by $\omega = \omega_1\omega_2\omega_3\ldots$, where each $\omega_i \in \{1, 2\}$. Define

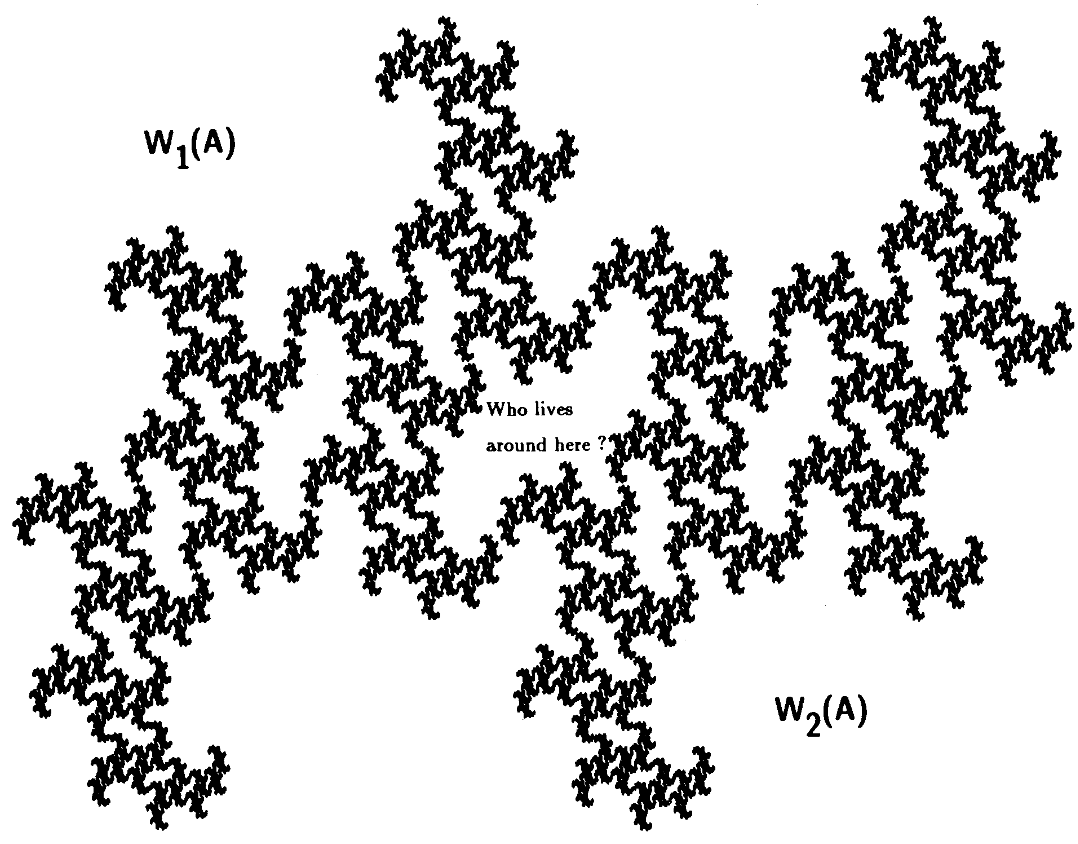


Figure IV.97. Can you describe the addresses of the points on the boundary of the central white region?

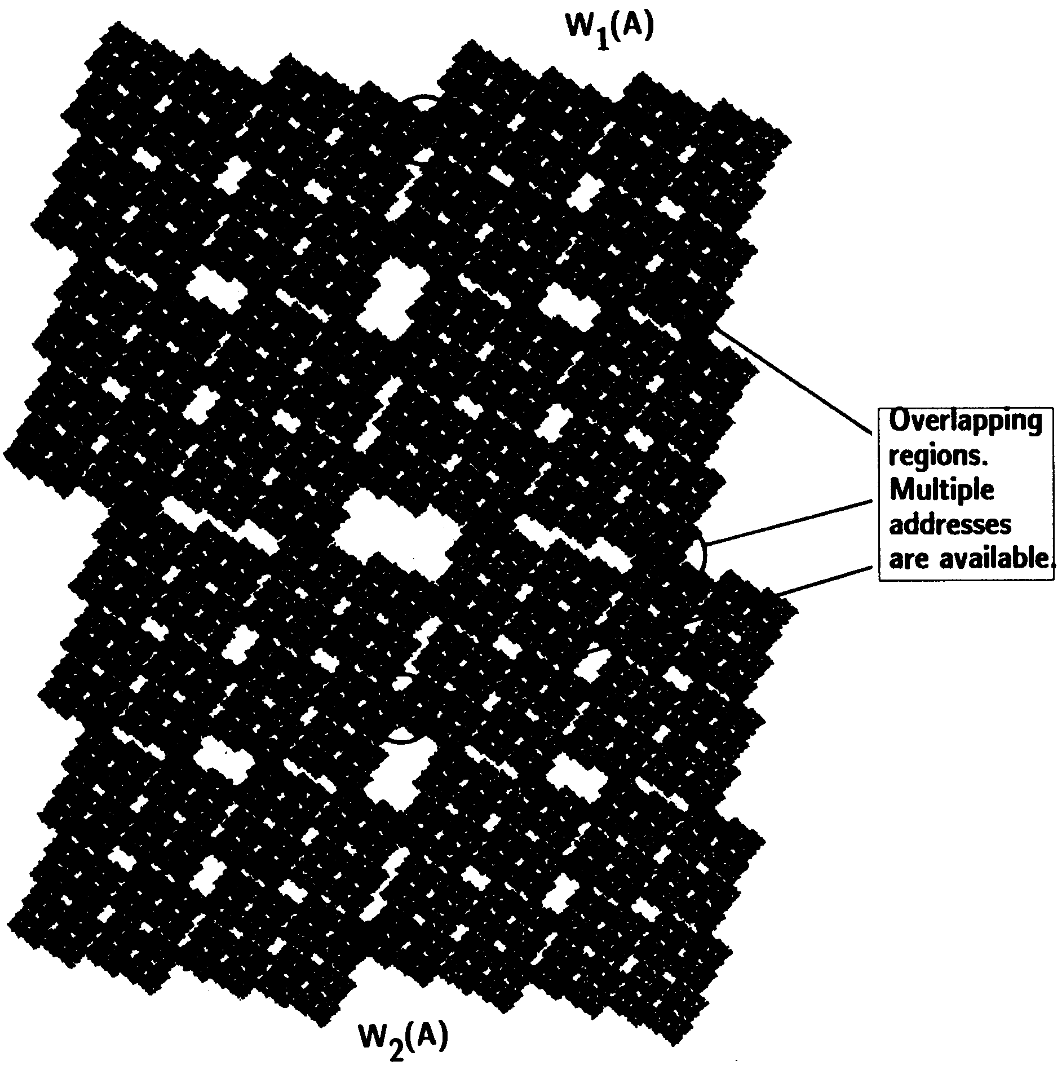


Figure IV.98. Attractor of a hyperbolic IFS in the overlapping case. In the overlapping regions multiple addresses are available.

$\rho : \{1, 2\} \to \{1, 2\}$ by $\rho(1) = 2$ and $\rho(2) = 1$. Suppose code space is countable. Let the counting function be $c : \{1, 2, 3, \ldots\} \to \Sigma$. Consider the point $\sigma \in \Sigma$ defined by

$$\sigma = \sigma_1\sigma_2\sigma_3\ldots,$$

where $\sigma_n = \rho((c(n))_n)$, and $(c(n))_n$ means the nth symbol of $c(n)$. When does the counting function reach σ? Never! For example, $c(3) \neq \sigma$ because their third symbols are different! This completes the proof.

Examples & Exercises

1.10. The set of integers $\mathbb{N} = \{0, \pm1, \pm2, \ldots, \}$ is countable. Define $c : N \to \mathbb{N}$ by $c(z) = (z - 1)/2$ if z is odd, $c(z) = -z/2$ if z is even.

1.11. Prove that a countable set of countable sets is countable. Show that an uncountable set, take away a countable set, is uncountable.

1.12. The *rational* numbers are countable. A rational number can be written in the form p/q, where p and q are integers with $q \neq 0$. Figure IV.99 shows how to count the positive ones, some numbers being counted more than once. Make a rule that gets rid of the redundant countings. Also, show how to include the negative rationals in the scheme.

1.13. Show that a Sierpinski triangle contains countably many triangles.

1.14. Let S be a perfect subset of a metric space. Suppose that S contains more than one point. Prove that S is uncountable.

1.15. Characterize the addresses of the missing pieces in Figure IV.100.

2 Continuous Transformations from Code Space to Fractals

Definition 2.1 *Let* $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ *be a hyperbolic IFS. The* code space *associated with the IFS,* (Σ, d_C), *is defined to be the code space on N symbols* $\{1, 2, \ldots, N\}$, *with the metric* d_C *given by*

$$d_C(\omega, \sigma) = \sum_{n=1}^{\infty} \frac{|\omega_n - \sigma_n|}{(N+1)^n} \qquad \text{for all } \omega, \sigma \in \Sigma.$$

Our goal is to construct a continuous transformation ϕ from the code space associated with an IFS onto the attractor of the IFS. This will allow us to formalize our notion of addresses. In order to make this construction we will need two lemmas. The first lemma tells us that if we have a hyperbolic IFS acting on a complete metric space, but we are only interested in studying how the IFS acts in relation to a fixed compact subset of $\mathbf{X}$, then we can treat the IFS as though it were defined on a compact metric space.

Lemma 2.1 *Let* $\{\mathbf{X}; w_n : n = 1, 2, \ldots, N\}$ *be a hyperbolic IFS, where* $(\mathbf{X}, d)$ *is a complete metric space. Let* $K \in \mathcal{H}(\mathbf{X})$. *Then there exists* $\tilde{K} \in \mathcal{H}(\mathbf{X})$ *such that* $K \subset \tilde{K}$ *and* $w_n : \tilde{K} \to \tilde{K}$ *for* $n = 1, 2, \ldots, N$. *In other words,* $\{\tilde{K}; w_n : n = 1, 2, 3, \ldots, N\}$ *is a hyperbolic IFS where the underlying space is compact.*

Proof Define $W : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ by

$$W(B) = \cup_{n=1}^{N} w_n(B) \qquad \text{for all } B \in \mathcal{H}(\mathbf{X}).$$

To construct $\tilde{K}$ consider the IFS with condensation $\{\mathbf{X}; w_n; n = 0, 1, 2, \ldots, N\}$, where the condensation map w_0 is associated with the condensation set K. By Theorem 7.1 in Chapter III the attractor of this IFS belongs to $\mathcal{H}(\mathbf{X})$. By exercise 9.1 in Chapter III it can be written

$$\tilde{K} = \text{Closure of } (K \cup W^{\circ 1}(K) \cup W^{\circ 2}(K) \cup W^{\circ 3}(K) \cup \ldots \cup W^{\circ n}(K) \cup \ldots\ldots).$$

It is readily seen that $K \subset \tilde{K}$ and that $W(\tilde{K}) \subset \tilde{K}$. This completes the proof.

The next lemma makes the first step in linking code space to IFS attractors, by introducing a certain transformation ϕ, which maps the Cartesian product space $\Sigma \times N \times \mathbf{X}$ into $\mathbf{X}$. By taking appropriate limits, in Theorem 2.1 below, we will eliminate the dependence on N and $\mathbf{X}$ to provide the desired connection between Σ and $\mathbf{X}$.

Lemma 2.2 *Let $\{\mathbf{X}; w_n : n = 1, 2, \ldots, N\}$ be a hyperbolic IFS of contractivity s, where $(\mathbf{X}, d)$ is a complete metric space. Let (Σ, d_C) denote the code space associated with the IFS. For each $\sigma \in \Sigma$, $n \in N$, and $x \in \mathbf{X}$, define*

$$\phi(\sigma, n, x) = w_{\sigma_1} \circ w_{\sigma_2} \circ \ldots \circ w_{\sigma_n}(x).$$

Let K denote a compact nonempty subset of $\mathbf{X}$. Then there is a real constant D such that

$$d(\phi(\sigma, m, x_1), \phi(\sigma, n, x_2)) \leq D s^{m \wedge n}$$

for all $\sigma \in \Sigma$, all $m, n \in N$, and all $x_1, x_2 \in K$.

Proof Let σ, m, n, x_1, and x_2 be as stated in the lemma. Construct $\tilde{K}$ from K as in Lemma 2.1. Without any loss of generality we can suppose that $m < n$. Then observe that

$$\phi(\sigma, n, x_2) = \phi(\sigma, m, \phi(\omega, n - m, x_2)),$$

where

$$\omega = \sigma_{n-m+1}\sigma_{n-m+2} \ldots \sigma_n \ldots \in \Sigma.$$

Let $x_3 = \phi(\omega, n - m, x_2)$. Then x_3 belongs to $\tilde{K}$. Hence we can write

$$\begin{aligned} d(\phi(\sigma, m, x_1), \phi(\sigma, n, x_2)) &= d(\phi(\sigma, m, x_1), \phi(\sigma, m, x_3)) \\ &\leq s d(w_{\sigma_2} \circ \ldots \circ w_{\sigma_m}(x_1), w_{\sigma_2} \circ \ldots \circ w_{\sigma_m}(x_3)) \\ &\leq s^2 d(w_{\sigma_3} \circ \ldots \circ w_{\sigma_m}(x_1), w_{\sigma_3} \circ \ldots \circ w_{\sigma_m}(x_3)) \\ &\leq s^m d(x_1, x_3) \leq s^m D, \end{aligned}$$

where $D = \max\{d(x_1, x_3) : x_1, x_3 \in \tilde{K}\}$. D is finite because $\tilde{K}$ is compact. This completes the proof.

Theorem 2.1 *Let $(\mathbf{X}, d)$ be a complete metric space. Let $\{\mathbf{X}; w_n : n = 1, 2, \ldots, N\}$ be a hyperbolic IFS. Let A denote the attractor of the IFS. Let (Σ, d_C) denote the code space associated with the IFS. For each $\sigma \in \Sigma$, $n \in N$, and $x \in \mathbf{X}$, let*

$$\phi(\sigma, n, x) = w_{\sigma_1} \circ w_{\sigma_2} \circ \ldots \circ w_{\sigma_n}(x).$$

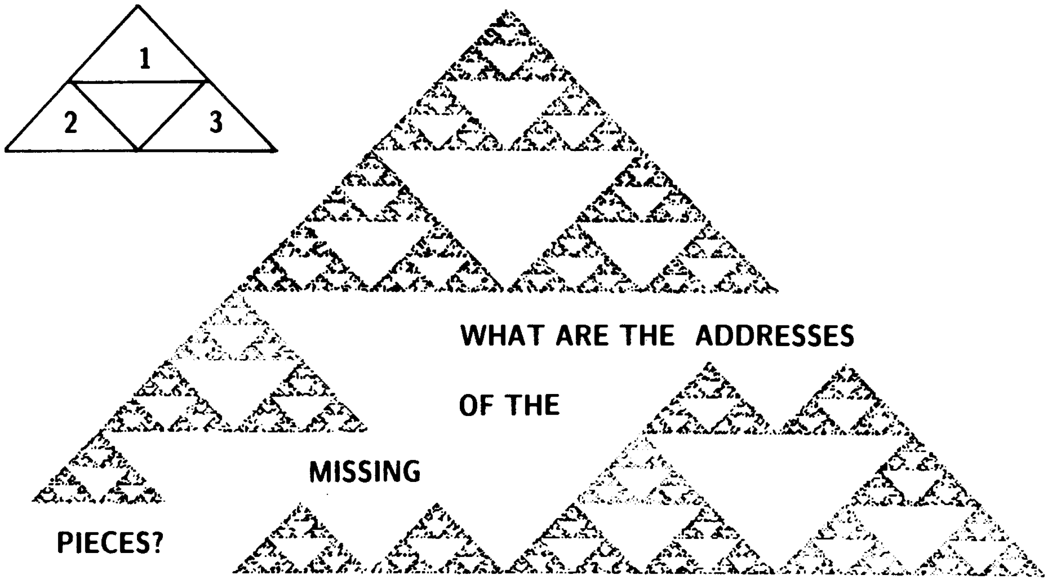
Then

$$\phi(\sigma) = \lim_{n \to \infty} \phi(\sigma, n, x)$$

c(1) = 1/1
c(2) = 1/2
c(3) = 2/1
c(4) = 3/1
c(5) = 2/2
c(6) = 1/3
c(7) = 1/4
c(8) = 2/3
c(9) = 3/2
c(10) = 4/1
c(11) = 5/1
c(12) = 4/2
c(13) = 3/3
c(14) = 2/4
c(15) = 1/5
.
.

Figure IV.99. How to count the positive rational numbers. What is $c(24)$?

Figure IV.100. Characterize the addresses of the missing pieces.



exists, belongs to A, and is independent of $x \in \mathbf{X}$. If K is a compact subset of $\mathbf{X}$, then the convergence is uniform over $x \in K$. The function $\phi : \Sigma \to A$ thus provided is continuous and onto.

Proof Let $x \in \mathbf{X}$. Let $K \in \mathcal{H}(\mathbf{X})$ be such that $x \in K$. Construct $\tilde{K}$ as in Lemma 2.1. Define $W : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ in the usual way. By Theorem 7.1 in Chapter III, W is a contraction mapping on the metric space $(\mathcal{H}(\mathbf{X}), h(d))$; and we have

$$A = \lim_{n\to\infty} \{W^{\circ n}(K)\}.$$

In particular $\{W^{\circ n}(K)\}$ is a Cauchy sequence in $(\mathcal{H}, h)$. Notice that $\phi(\sigma, n, x) \in W^{\circ n}(K)$. It follows from Theorem 7.1 in Chapter II that if $\lim_{n\to\infty} \phi(\sigma, n, x)$ exists, then it belongs to A.

That the latter limit does exist follows from the fact that, for fixed $\sigma \in \Sigma$, $\{\phi(\sigma, n, x)\}_{n=1}^{\infty}$ is a Cauchy sequence: by Lemma 2.2

$$d(\phi(\sigma, m, x), \phi(\sigma, n, x)) \leq Ds^{m\wedge n} \qquad \text{for all } x \in K,$$

and the right-hand side here tends to zero as m and n tend to infinity. The uniformity of the convergence follows from the fact that the constant D is independent of $x \in K$.

Next we prove that $\phi : \Sigma \to A$ is continuous. Let $\epsilon > 0$ be given. Choose n so that $s^n D < \epsilon$, and let $\sigma, \omega \in \Sigma$ obey

$$d_C(\sigma, \omega) < \sum_{m=n+2}^{\infty} \frac{N}{(N+1)^m} = \frac{1}{(N+1)^{n+1}}.$$

Then one can verify that σ must agree with ω through n terms; that is, $\sigma_1 = \omega_1, \sigma_2 = \omega_2, \ldots, \sigma_n = \omega_n$. It follows that for each $m \geq n$ we can write

$$d(\phi(\sigma, m, x), \phi(\omega, m, x)) = d(\phi(\sigma, n, x_1), \phi(\sigma, n, x_2)),$$

for some pair $x_1, x_2 \in \tilde{K}$. By Lemma 2.2 the right-hand side here is smaller than $s^n D$ which is smaller than ϵ. Taking the limit as $m \to \infty$ we find

$$d(\phi(\sigma), \phi(\omega)) < \epsilon.$$

Finally, we prove that ϕ is onto. Let $a \in A$. Then, since

$$A = \lim_{n \to \infty} W^{\circ n}(\{x\}),$$

it follows from Theorem 7.1 in Chapter II that there is a sequence $\{\omega^{(n)} \in \Sigma : n = 1, 2, 3, \ldots\}$ such that

$$\lim_{n \to \infty} \phi(\omega^{(n)}, n, x) = a.$$

Since (Σ, d_C) is compact, it follows that $\{\omega^{(n)} : n = 1, 2, 3, \ldots\}$ possesses a convergent subsequence with limit $\omega \in \Sigma$. Without loss of generality assume $\lim_{n \to \infty} \omega^{(n)} = \omega$. Then the number of successive initial agreements between the components of $\omega^{(n)}$ and ω increases without limit. That is, if

$$\alpha(n) = \text{ number of elements in } \{j \in N : \omega_k^{(n)} = \omega_k \text{ for } 1 \leq k \leq j\},$$

where $N = \{1, 2, 3, \ldots\}$, then $\alpha(n) \to \infty$ as $n \to \infty$. It follows that

$$d(\phi(\omega, n, x), \phi(\omega^{(n)}, n, x)) \leq s^{\alpha(n)} D.$$

By taking the limit on both sides as $n \to \infty$ we find $d(\phi(\omega), a) = 0$, which implies $\phi(\omega) = a$. Hence $\phi : \Sigma \to A$ is onto. This completes the proof.

Definition 2.2 *Let* $\{\mathbf{X}; w_n, n = 1, 2, 3, \ldots, N\}$ *be a hyperbolic IFS with associated code space* Σ. *Let* $\phi : \Sigma \to A$ *be the continuous function from code space onto the attractor of the IFS constructed in Theorem 1. An* address *of a point* $a \in A$ *is any member of the set*

$$\phi^{-1}(A) = \{\omega \in \Sigma : \phi(\omega) = a\}.$$

This set is called the set of addresses *of* $a \in A$. *The IFS is said to be* totally disconnected *if each point of its attractor possesses a unique address. The IFS is said to be* just-touching *if it is not totally disconnected yet its attractor contains an open set* $\mathcal{O}$ *such that*

(1) $w_i(\mathcal{O}) \cap w_j(\mathcal{O}) = \emptyset \forall i, j \in \{1, 2, \ldots, N\}$ with $i \neq j$;
(2) $\bigcup^N \lim_{i=1} w_i(\mathcal{O}) \subset \mathcal{O}$.

An IFS whose attractor obeys (i) and (ii) is said to obey the open set *condition. The IFS is said to be* overlapping *if it is neither just-touching nor disconnected.*

Theorem 2.2 *Let* $\{\mathbf{X}; w_n, n = 1, 2, \ldots, N\}$ *be a hyperbolic IFS with attractor* A. *The IFS is* totally disconnected *if and only if*

$$w_i(A) \cap w_j(A) = \emptyset \forall i, \quad j \in \{1, 2, \ldots, N\} \quad \text{with } i \neq j. \tag{1}$$

Proof If the IFS is totally disconnected, then each point on its attractor possesses a unique address. This implies Equation 3. If the IFS is not totally disconnected, then some point on its attractor possesses two different addresses. These must disagree at some first place: choose inverse images to get this place out front, to produce a contradiction to Equation 3. This completes the proof.

Examples & Exercises

2.1. Show that the attractor of the IFS $\{\mathbb{R}; \frac{1}{2}x, \frac{1}{2}x + \frac{1}{2}\}$ is just-touching. Classify the attractor for the IFS $\{\mathbb{R}; \frac{1}{2}x, 1\}$.

2.2. Prove that the attractor of the IFS $\{\mathbb{R}; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}$ is overlapping.

2.3. Consider the IFS $\{[0, 1], w_n(x) = \frac{n-1}{10} + \frac{1}{10}x, n = 1, 2, 3, \ldots, 10\}$ and for the associated code space use the symbols $\{0, 1, 2, \ldots, 9\}$. Show that the attractor of the IFS is $[0, 1]$ and that it is just-touching. Identify the addresses of points with multiple addresses. Show that the address of a point is just its decimal representation. Comment on the fact that some numbers have two decimal representations.

2.4. Prove that the attractor to the IFS $\{[0, 1]; w_1(x) = \frac{1}{3}x, w_2(x) = \frac{1}{3}x + \frac{2}{3}\}$ is totally disconnected.

2.5. Prove that the IFS that generates the *Black Spleenwort* fern, given in Chapter 2, is just-touching.

2.6. Show that the IFS $\{[0, 1]; w_1(x) = \frac{1}{2}, w_2(x) = \frac{1}{2}\}$ is overlapping.

We need to understand the structure of code space. Theorem 2.1 told us that the code space on N symbols is the mother of all hyperbolic IFS consisting of N maps. We will use the following theorem to show that the mother is metrically equivalent to a classical Cantor set.

Theorem 2.3 *Let Σ denote the code space of the N symbols, $\{1, 2, \ldots, N\}$, and define two different metrics on Σ by*

$$d_1(x, y) = \sum_{i=1}^{\infty} \frac{|x_i - y_i|}{(N+1)^i}, \qquad d_2(x, y) = \left| \sum_{i=1}^{\infty} \frac{x_i - y_i}{(N+1)^i} \right|.$$

Then (Σ, d_1) and (Σ, d_2) are equivalent metric spaces.

Proof We give the proof for the case $N = 10$. Let $x, y \in \Sigma$ be given. Clearly we have $d_2(x, y) \leq d_1(x, y)$. We must show that there is a constant C so that $Cd_1(x, y) \leq d_2(x, y)$, where C is independent of x and y. Here we pick $C = \frac{1}{19}$ and show that it works.

We can suppose that for some $k \in \{1, 2, 3, \ldots\}$, $x_1 = y_1, x_2 = y_2, \ldots, x_{k-1} = y_{k-1}, x_k \neq y_k$. Then

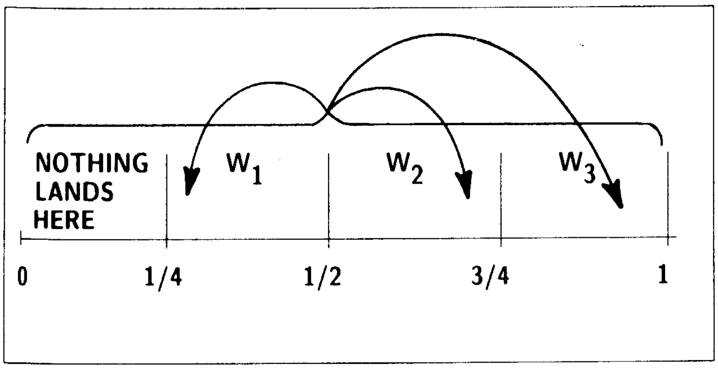


Figure IV.101. Nothing lands here.

$$d_2(x,y) = |\sum_{i=k}^{\infty} \frac{x_i - y_i}{11^i}| \geq \frac{|x_k - y_k|}{11^k} - \sum_{i=k+1}^{\infty} \frac{|x_i - y_i|}{11^i}$$

$$\geq \frac{|x_k - y_k|}{11^k} - \sum_{i=k+1}^{\infty} \frac{9}{11^i} = (|x_k - y_k| - \frac{9}{10})\frac{1}{11^k}$$

$$\geq \frac{1}{19}(|x_k - y_k| + \frac{9}{10})\frac{1}{11^k},$$

(verify this by checking it for $|x_k - y_k| \in \{1, 2, \ldots, 9\}$),

$$\geq \frac{1}{19}(\frac{|x_k - y_k|}{11^k} + \sum_{i=k+1}^{\infty} \frac{9}{11^i}) \geq \frac{1}{19}(\frac{|x_k - y_k|}{11^k} + \sum_{i=k+1}^{\infty} \frac{|x_i - y_i|}{11^i})$$

$$\geq \frac{1}{19}\sum_{n=1}^{\infty} \frac{|x_i - y_i|}{11^i} = \frac{1}{19} d_1(x,y).$$

This completes the proof.

We now show that code space is metrically equivalent to a totally disconnected Cantor subset of $[0, 1]$. Define a hyperbolic IFS by $\{[0,1]; w_n(x) = \frac{1}{(N+1)}x + \frac{n}{N+1} : n = 1, 2, \ldots, N\}$. Thus

$$w_n([0,1]) = [\frac{n}{N+1}, \frac{n+1}{N+1}] \qquad \text{for } n = 1, 2, \ldots, N,$$

as illustrated for $N = 3$ in Figure IV.101.

The attractor for this IFS is totally disconnected, as illustrated in Figure IV.102 for $N = 3$.

In the case $N = 3$, the attractor is contained in $[\frac{1}{3}, 1]$. The fixed points of the three transformations $w_1(x) = \frac{1}{4}x + \frac{1}{4}$, $w_2(x) = \frac{1}{4}x + \frac{1}{2}$, $w_2(x) = \frac{1}{4}x + \frac{3}{4}$ are $\frac{1}{3}$, $\frac{2}{3}$, and 1, respectively. Moreover, the address of any point on the attractor is exactly the same as the string of digits that represents it in base $N + 1$. What is happening here is this. At level zero we begin with all numbers in $[0, 1]$ represented in base $(N + 1)$. We remove all those points whose first digit is 0. For example, in the case $N = 3$ this eliminates the interval $[0, \frac{1}{4}]$. At the second level we remove from the remaining points all those that have digit 0 in the second place. And so on. We end up with

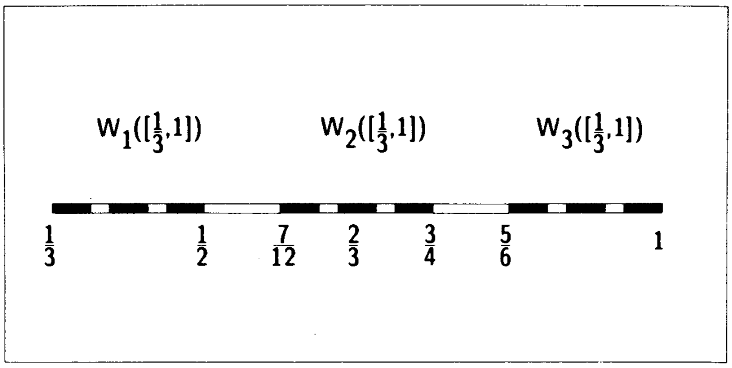


Figure IV.102. A special ternary Cantor set in the making.

those numbers whose expansion in base $(N+1)$ does not contain the digit 0. Now consider the continuous transformation $\phi : (\Sigma, d_C) \to (A, \text{ Euclidean})$. It follows from Theorem 2.3 that the two metric spaces are equivalent. ϕ is the transformation that provides the equivalence. Thus, we have a realization, a way of picturing code space.

Examples & Exercises

2.7. Find the figure analogous to Figure IV.102, corresponding to the case $N = 9$.

2.8. What is the smallest number in $[0, 1]$ whose decimal expansion contains no zeros?

We continue to discuss the relationship between the attractor A of a hyperbolic IFS $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ and its associated code space Σ. Let $\phi : \Sigma \to \mathbf{X}$ be the code space map constructed in Theorem 2.1. Let $\omega = \omega_1\omega_2\omega_3\omega_4\ldots$ be an address of a point $x \in A$. Then

$$\tilde{\omega} = j\omega_1\omega_2\omega_3\omega_4\ldots$$

is an address of $w_j(x)$, for each $j \in \{1, 2, \ldots, N\}$.

Definition 2.3 *Let $\{\mathbf{X}, w_1, w_2, \ldots, w_N\}$ be a hyperbolic IFS with attractor A. A point $a \in A$ is called a* periodic point *of the IFS if there is a finite sequence of numbers $\{\sigma(n) \in \{1, 2, \ldots, N\}\}_{n=1}^{P}$ such that*

$$a = w_{\sigma(P)} \circ w_{\sigma(P-1)} \circ \ldots \circ w_{\sigma(1)}(a). \tag{2}$$

If $a \in A$ is periodic, then the smallest integer P such that the latter statement is true is called the period *of a.*

Thus, a point on an attractor is periodic if we can apply a sequence of w_n's to it, in such a way as to get back to exactly the same point after finitely many steps. Let $a \in A$ be a periodic point that obeys (2). Let σ be the point in the associated code space, defined by

$$\begin{aligned} \sigma &= \sigma(P)\sigma(P-1)\ldots\sigma(1)\sigma(P)\sigma(P-1)\ldots\sigma(1)\sigma(P)\sigma(P-1)\ldots \\ &= \overline{\sigma(P)\sigma(P-1)\ldots\sigma(1)}. \end{aligned} \tag{3}$$

Then, by considering $\lim_{n\to\infty}\phi(\sigma, n, a)$, we see that $\phi(\sigma) = a$.

Definition 2.4 *A point in code space whose symbols are periodic, as in (3), is called a* periodic address. *A point in code space whose symbols are periodic after a finite initial set is omitted is called* eventually periodic.

Examples & Exercises

2.9. An example of a periodic address is

$$121212121212121212121212121212121212121212121212121212\ldots,$$

where 12 is repeated endlessly. An example of an eventually periodic address is:

$$112111111211111211111211111212212112121212121212121212121\ldots,$$

where 21 is repeated endlessly.

2.10. Prove the following theorem: "Let $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ be a hyperbolic IFS with attractor A. Then the following statements are equivalent:

(1) $x \in A$ is a periodic point;
(2) $x \in A$ possesses a periodic address;
(3) $x \in A$ is a fixed point of an element of the semigroup of transformations generated by $\{w_1, w_2, \ldots, w_N\}$."

2.11. Show that a point $x \in [0, 1]$ is a periodic point of the IFS

$$\{[0, 1]; \frac{1}{2}x, \frac{1}{2}x + \frac{1}{2}\}$$

if and only if it can be written $x = p/(2^N - 1)$ for some integer $0 \le p \le 2^N - 1$ and some integer $N \in \{1, 2, 3, \ldots\}$.

2.12. Let $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ denote a hyperbolic IFS with attractor A. Define $W(S) = \cup_{n=1}^{N} w_n(S)$ when S is a subset of $\mathbf{X}$. Let P denote the set of periodic points of the IFS. Show that $W(P) = P$.

2.13. Locate all the periodic points of period 3 for the IFS $\{\mathbb{R}^2; \frac{1}{2}z, \frac{1}{2}z + \frac{1}{2}, \frac{1}{2}z + \frac{i}{2}\}$. Mark the positions of these points on $\mathbb{A}$.

2.14. Locate all periodic points of the IFS $\{\mathbb{R}; w_1(x) = 0, w_2(x) = \frac{1}{2}x + \frac{1}{2}\}$.

Theorem 2.4 *The attractor of an IFS is the closure of its periodic points.*

Proof Code space is the closure of the set of periodic codes. Lift this statement to A using the code space map $\phi : \Sigma \to A$. (ϕ is a continuous mapping from a metric space Σ onto a metric space A. If $S \subset \Sigma$ is such that its closure equals Σ, then the closure of $f(S)$ equals A.)

Examples & Exercises

2.15. Prove that the attractor of a totally disconnected hyperbolic IFS of two or more maps is uncountable.

2.16. Under what conditions does the attractor of a hyperbolic IFS contain uncountably many points with multiple addresses? Do not try to give a complete answer; just some conditions: think about the problem.

2.17. Under what conditions do there exist points in the attractor of a hyperbolic IFS with uncountably many addresses? As in 2.16, do not try to give a full answer.

2.18. In the standard construction of the classical Cantor set $\mathcal{C}$, described in exercise 1.5 in Chapter III, a succession of open subintervals of [0, 1] is removed. The endpoints of each of these intervals belong to $\mathcal{C}$. Show that the set of such interval endpoints is countable. Show that $\mathcal{C}$ itself is uncountable. $\mathcal{C}$ is the attractor of the IFS $\{[0, 1]; \frac{1}{3}x, \frac{1}{3}x + \frac{2}{3}\}$. Characterize the addresses of the set of interval endpoints in $\mathcal{C}$.

3 Introduction to Dynamical Systems

We introduce the idea of a dynamical system and some of the associated terminology.

Definition 3.1 *A* dynamical system *is a transformation $f : \mathbf{X} \to \mathbf{X}$ on a metric space $(\mathbf{X}, d)$. It is denoted by $\{\mathbf{X}; f\}$. The* orbit *of a point $x \in \mathbf{X}$ is the sequence $\{f^{\circ n}(x)\}_{n=0}^{\infty}$.*

As we will discover, dynamical systems are sources of deterministic fractals. The reasons for this are deeply intertwined with IFS theory, as we will see. Later we will introduce a special type of dynamical system, called a shift dynamical system, which can be associated with an IFS. By studying the orbits of these systems we will learn more about fractals. One of our goals is to learn why the Random Iteration Algorithm, used in Program 2 in Chapter III, successfully calculates the images of attractors of IFS. More information about the deep structure of attractors of IFS will be discovered.

Examples & Exercises

3.1. Define a function on code space, $f : \Sigma \to \Sigma$, by

$$f(x_1x_2x_3x_4\ldots) = x_2x_3x_4x_5\ldots.$$

Then $\{\Sigma; f\}$ is a dynamical system.

3.2. $\{[0, 1]; f(x) = \lambda x(1 - x)\}$ is a dynamical system for each $\lambda \in [0, 4]$. We say that we have a *one-parameter family* of dynamical systems.

3.3. Let $w(x) = Ax + t$ be an affine transformation in $\mathbb{R}^2$. Then $\{\mathbb{R}^2; w\}$ is a dynamical system.

3.4. Define $T : C[0, 1] \to C[0, 1]$ by

$$(Tf)(x) = \frac{1}{2}f(\frac{1}{2}x) + \frac{1}{2}f(\frac{1}{2}x + \frac{1}{2}).$$

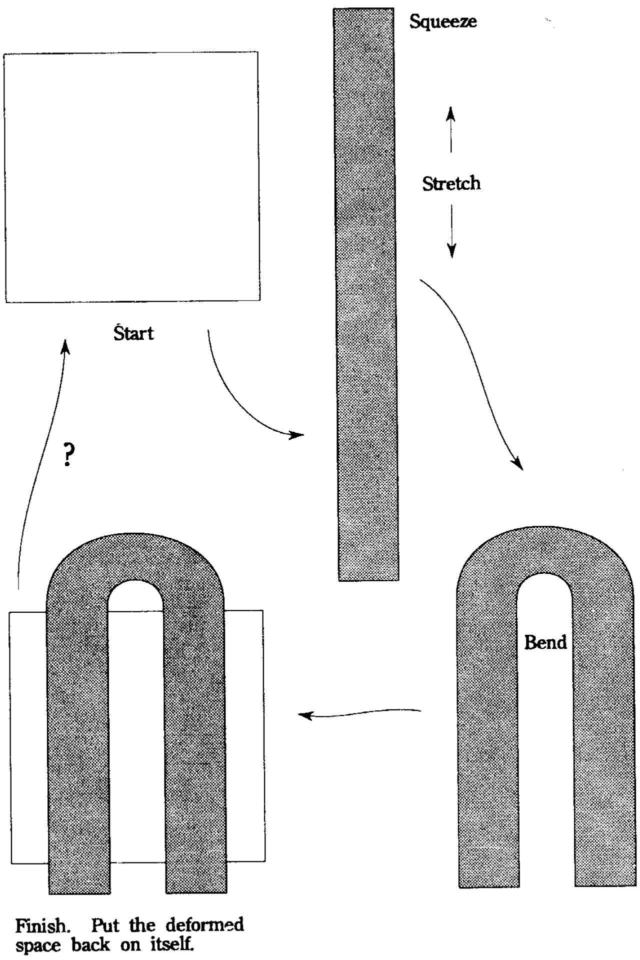


Figure IV.103. An example of a "stretch, squeeze, and bend" dynamical system (Smale horseshoe function).

Then $\{C[0, 1]; T\}$ is a dynamical system.

3.5. Let $w: \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be a Möbius transformation. That is $w(z) = (az + b)/(cz + d)$, where $a, b, c, d \in \mathbb{C}$, and $(ad - bc) \neq 0$. Then $\{\hat{\mathbb{C}}; w(z)\}$ is a dynamical system.

3.6. $\{[0, 1]; 2x mod 1\}$ is a dynamical system. Here $2x mod 1 = 2x - [2x]$, where $[2x]$ denotes the greatest integer less than or equal to $2x$.

3.7. Define a transformation $f : \blacksquare \to \blacksquare$ as illustrated in Figure IV.103. $\{\blacksquare; f\}$ is a dynamical system.

In dynamical systems theory one is interested in what happens when one follows a typical orbit: is there some kind of attractor that usually occurs? Dynamical systems become interesting when the transformations involved are *not* contraction mappings, so that a single transformation suffices to produce interesting behavior. The orbit of a single point may be a geometrically complex set. Some thought about horizontal slices through Figure IV.104 will quickly suggest to the inquisitive student that there

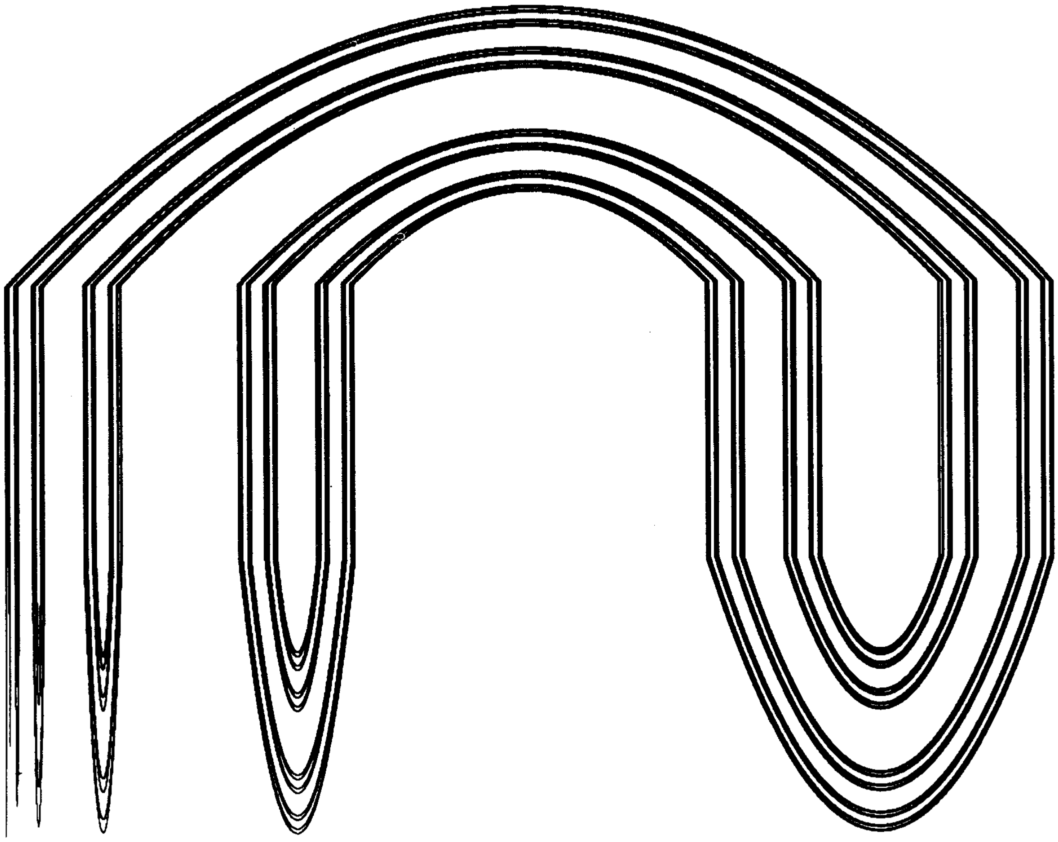


Figure IV.104. One million iterations of a small black square in a "stretch, squeeze, and bend" dynamical system. Can you find a relationship to IFS theory?

is a close relationship between this noncontractive dynamical system and a hyperbolic IFS.

Definition 3.2 *Let* $\{\mathbf{X}; f\}$ *be a dynamical system. A* periodic point *of* f *is a point* $x \in \mathbf{X}$ *such that* $f^{\circ n}(x) = x$ *for some* $n \in \{1, 2, 3, \ldots\}$. *If* x *is a periodic point of* f, *then an integer* n *such that* $f^{\circ n}(x) = x$, $n \in \{1, 2, 3, \ldots\}$ *is called a period of* x. *The least such integer is called the* minimal period *of the periodic point* x. *The orbit of a periodic point of* f *is called a* cycle *of* f. *The minimal period of a cycle is the number of distinct points it contains. A period of a cycle of* f *is a period of a point in the cycle.*

Definition 3.3 *Let* $\{\mathbf{X}; f\}$ *be a dynamical system and let* $x_f \in \mathbf{X}$ *be a fixed point of* f. *The point* x_f *is called an* attractive *fixed point of* f *if there is a number* $\epsilon > 0$ *so that* f *maps the ball* $B(x_f, \epsilon)$ *into itself, and moreover* f *is a contraction mapping on* $B(x_f, \epsilon)$. *Here* $B(x_f, \epsilon) = \{y \in \mathbf{X} : d(x_f, y) \leq \epsilon\}$. *The point* x_f *is called a* repulsive *fixed point of* f *if there are numbers* $\epsilon > 0$ *and* $C > 1$ *such that*

$$d(f(x_f), f(y)) \geq C d(x_f, y) \qquad \text{for all } y \in B(x_f, \epsilon).$$

A periodic point of f of period n is *attractive* if it is an attractive fixed point of $f^{\circ n}$. A cycle of period n is an *attractive cycle* of f if the cycle contains an attractive periodic point of f of period n. A periodic point of f of period n is *repulsive* if it

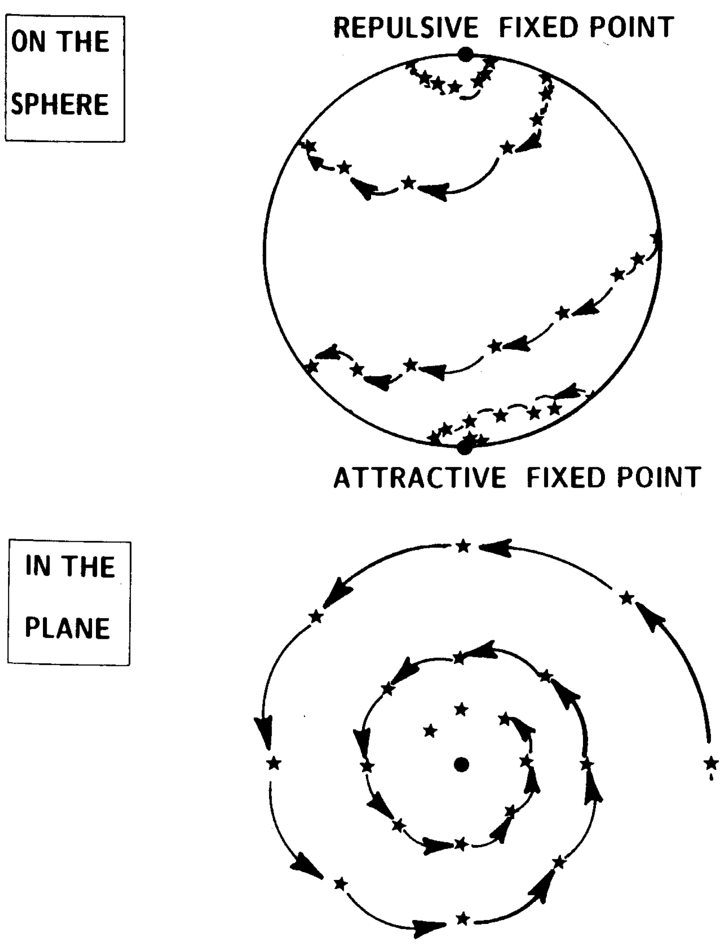


Figure IV.105. The dynamics of a simple Möbius transformation. Points spiral away from one fixed point and spiral in toward the other. What happens if the fixed points coincide?

is a repulsive fixed point of $f^{\circ n}$. A cycle of period n is a *repulsive cycle* of f if the cycle contains a repulsive periodic point of f of period n.

Definition 3.4 *Let* $\{\mathbf{X}, f\}$ *be a dynamical system. A point* $x \in \mathbf{X}$ *is called an* eventually periodic *point of* f *if* $f^{\circ m}(x)$ *is periodic for some positve integer* m.

Remark: The definitions given here for attractive and repulsive points are consistent with the definitions we use for metric equivalence and will be used throughout the text. The definitions used in dynamical systems theory are usually more topological in nature. These are given later in exercises 5.4 and 5.5.

Examples & Exercises

3.8. The point $x_f = 0$ is an attractive fixed point for the dynamical system $\{\mathbb{R}; \frac{1}{2}x\}$, and a repulsive fixed point for the dynamical system $\{\mathbb{R}; 2x\}$.

3.9. The point $z = 0$ is an attractive fixed point, and $z = \infty$ is a repulsive fixed point, for the dynamical system

$$\{\hat{\mathbb{C}}; (\cos 10° + i \sin 10°)(0.9)z\}.$$

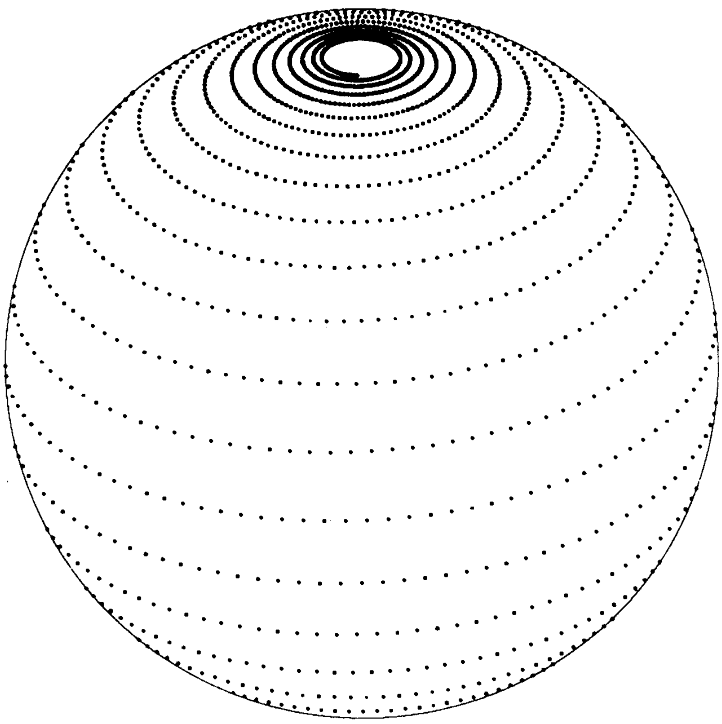


Figure IV.106. Points belonging to an orbit of a Möbius transformation on a sphere.

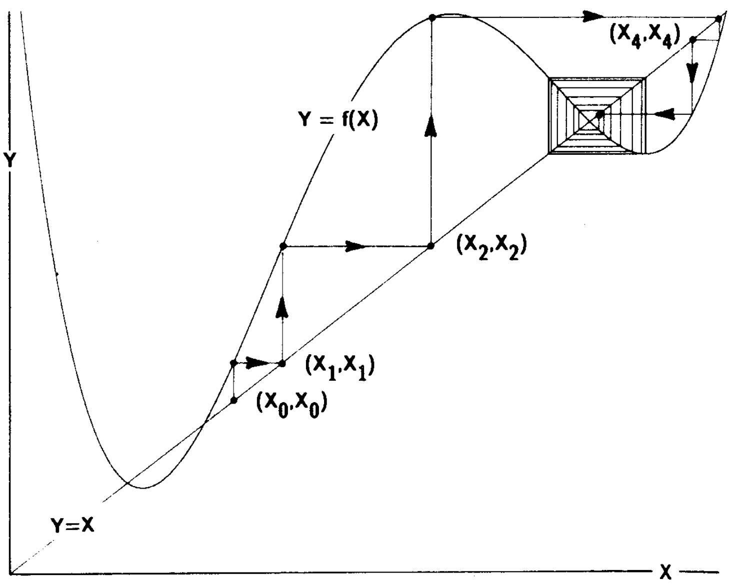


Figure IV.107. This shows an example of a web diagram. A web diagram is a means for displaying and analyzing the orbit of a point $x_0 \in \mathbb{R}$ for a dynamical system $(\mathbb{R}, f)$. The geometrical construction of a web diagram makes use of the graph of $f(x)$.

A typical orbit, starting from near the point of infinity on the sphere, is shown in Figures IV.105 and IV.106.

3.10. The point $x_f = 111\overline{111}$ is a repulsive fixed point for the dynamical system $\{\Sigma; f\}$ where $f : \Sigma \to \Sigma$ is defined by

$$f(x_1 x_2 x_3 x_4 x_5 \ldots) = x_2 x_3 x_4 x_5 \ldots.$$

Show that $x = 121\overline{212}$ is a repulsive fixed point of period 2, and that $\{12\overline{12}, 21\overline{21}\}$ is a repulsive cycle of period 2.

3.11. The dynamical system $\{[0, 1]; \frac{1}{2}x(1-x)\}$ possesses the attractive fixed point $x_f = 0$. Can you find a repulsive fixed point for this system?

There is a delightful construction for representing orbits of a dynamical system of the special form $\{\mathbb{R}; f(x)\}$. It utilizes the graph of the function f : $\mathbb{R} \to \mathbb{R}$. We describe here how it is used to represent the orbit $\{x_n = f^{\circ n}(x_0)\}_{n=1}^{\infty}$ of a point $x_0 \in \mathbb{R}$.

For simplicity we suppose that $f : [0, 1] \to [0, 1]$. Draw the square $\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ and sketch the graphs of $y = f(x)$ and $y = x$ for $x \in [0, 1]$. Start at the point (x_0, x_0) and connect it by a straight-line segment to the point $(x_0, x_1 = f(x_0))$. Connect this point by a straight-line segment to the point (x_1, x_1). Connect this point by a straight-line segment to the point $(x_1, x_2 = f(x_1))$; and continue. The orbit itself shows up on the 45° line $y = x$, as the sequence of points $(x_0, x_0), (x_1, x_1), (x_2, x_2), \ldots$. We call the result of this geometrical construction a *web diagram.*

It is straightforward to write computergraphical routines that plot web diagrams on the graphics display device of a microcomputer. The following program is written in BASIC. It runs without modification on an IBM PC with Color Graphics Adaptor and Turbobasic. On any line the words preceded by a ' are comments: they are not part of the program.

Program 1.

```
l=3.79 : xn=0.95            'parameter value 3.79, orbit starts
                             at 0.95
def fnf(xn)=l*xn*(1-xn)     'change this function f(x) for other
                             dynamical systems
screen 1 : cls              'initialize computer graphics
window (0,0)-(1,1)          'set plotting window to 0 < x < 1 ,
                             0 < y < 1
for k=1 to 400              'plot the graph of the f(x)
pset(k/400, fnf(k/400))
next k
do                          'the main computational loop
n=n+1                       'increment the counter, $n$
y=fnf(xn)                   'compute the next point on the orbit
line (xn,xn)-(xn,y), n      'draw a line from (xn,xn) to (xn,y)
                             in color n
line (xn,y)-(y,y), n        'draw a line segment from (xn,y) to
                             (y,y) in color n
xn=y                        'set xn to be the most recently computed
                             point on the orbit

loop until instat : end     'stop running if a key is pressed.
```

Two examples of some web diagrams computed using this program are shown in Figure IV.108. The dynamical system used in this case is $\{[0, 1]; f(x) = 3.79x(1 - x)\}$.

Examples & Exercises

3.12. Rewrite Program 1 in a form suitable for your own computer environment. Use the resulting system to study the dynamical systems $\{[0, 1]; \lambda x(1 - x)\}$ for $\lambda = 0.55, 1.3, 2.225, 3.014, 3.794$. Try to classify the various species of web diagrams that occur for this one-parameter family of dynamical systems.

3.13. Divide $[0, 1]$ into 16 subintervals $[0, \frac{1}{16}), [\frac{1}{16}, \frac{1}{16}), \ldots, [\frac{14}{16}, \frac{15}{16}), [\frac{15}{16}, 1]$. Let $f : [0, 1] \to [0, 1]$ be defined by $f(x) = \lambda x(1 - x)$, where $\lambda \in [0, 4]$ is a parameter. Compute $\{f^{\circ n}(\frac{1}{2}) : n = 0, 1, 2, \ldots, 5000\}$ and keep track of the *frequency* with which $f^{\circ n}(\frac{1}{2})$ falls in the kth interval for $k = 1, 2, 4, 8, 16$, and $\lambda = 0.55, 1.3, 2.225, 3.014, 3.794$. Make histograms of your results.

3.14. Describe the behavior for the one-parameter family of dynamical system $s\{\mathbb{R} \cup \{\infty\}; \lambda x\}$, where λ is a real parameter, in the cases (i) $\lambda = 0$; (ii) $0 < |\lambda| < 1$; (iii) $\lambda = -1$; (iv) $\lambda = 1$; (v) $1 < \lambda < \infty$.

3.15. Analyze possible behaviors of $\{\mathbb{R}^2; Ax + t\}$, where $Ax + t$ is an affine transformation.

3.16. Study possible behaviors of orbits for the dynamical system $\{\hat{\mathbb{C}};$Möbius transformation$\}$. You should make appropriate changes of coordinates to simplify the discussion.

3.17. Show that all points are eventually periodic for the slide-and-fold dynamical system $\{\mathbb{R}; f\}$, where

$$f(x) = \begin{cases} x + 1 & \text{if } x \leq 0, \\ -x + 1 & \text{if } x \geq 0. \end{cases}$$

This system is illustrated in Figure IV.109.

3.18. Let $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ be a hyperbolic IFS. Then $\{\mathcal{H}(\mathbf{X}); W\}$ is a dynamical system, where

$$W(B) = \cup_{n=1}^{N} w_n(B) \text{ for all } B \in \mathcal{H}(X).$$

Dynamical systems that act on sets in place of points are sometimes called *set dynamical systems*. Show that the attractor of the IFS is an attractive fixed point of the dynamical system $\{\mathcal{H}(\mathbf{X}); W\}$. You should quote appropriate results from earlier theorems.

3.19. We consider again our two-dimensional code space, having both past and future, called the space of shifts (see exercise 1.12 in Chapter II). In this space, the operation of the shift transformation is a homeomorphism of the space to itself (it is frequently called the *shift automorphism*). There is a very geometrical interpretation

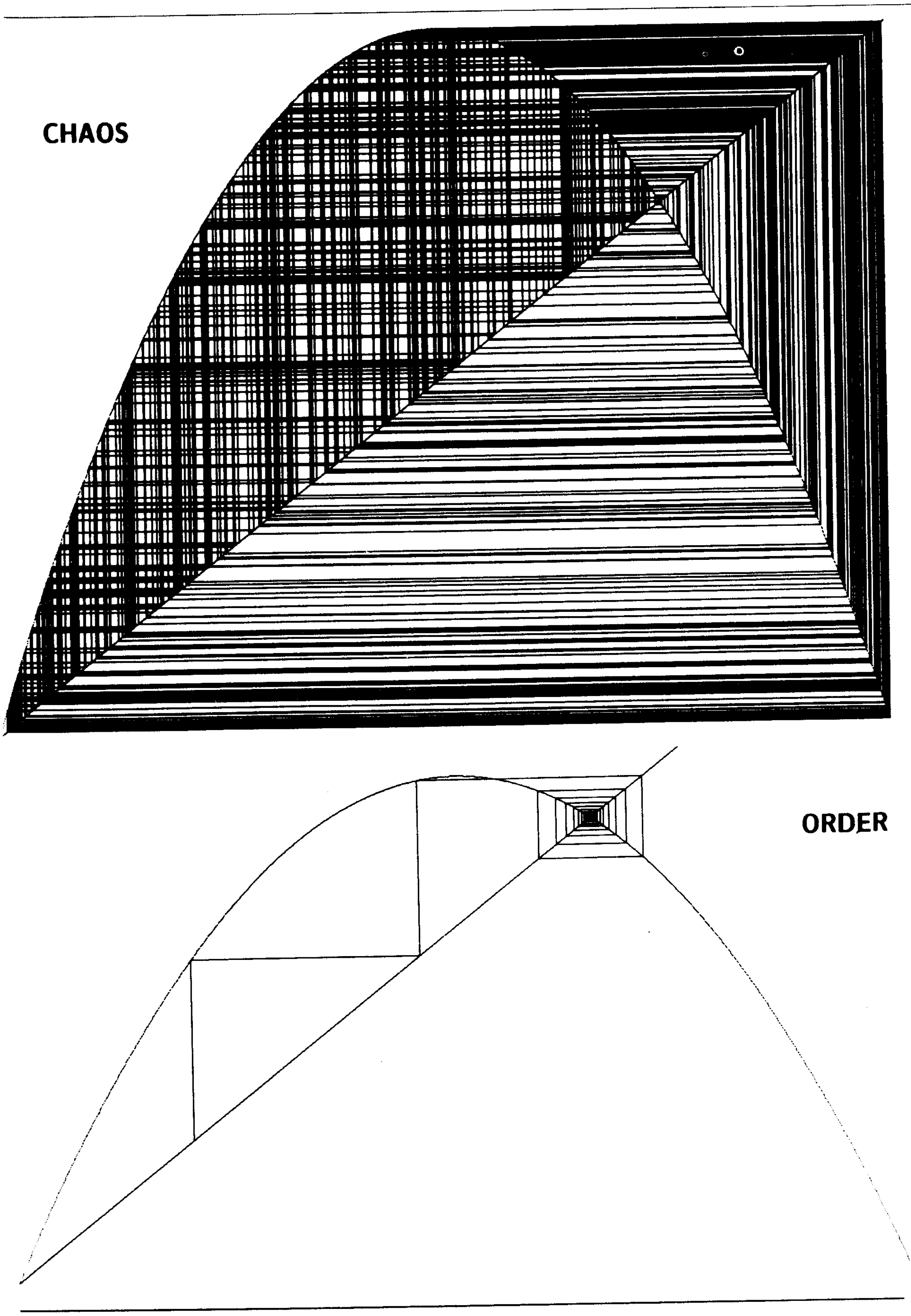


Figure IV.108. Two examples of web diagrams computed using Program 1. The dynamical system in this case is $\{[0, 1]; f(x) = \lambda x(1 - x)\}$, for two different values of $\lambda \in (0, 4)$. The system corresponding to the lower value of λ is orderly; the other is close to being chaotic.

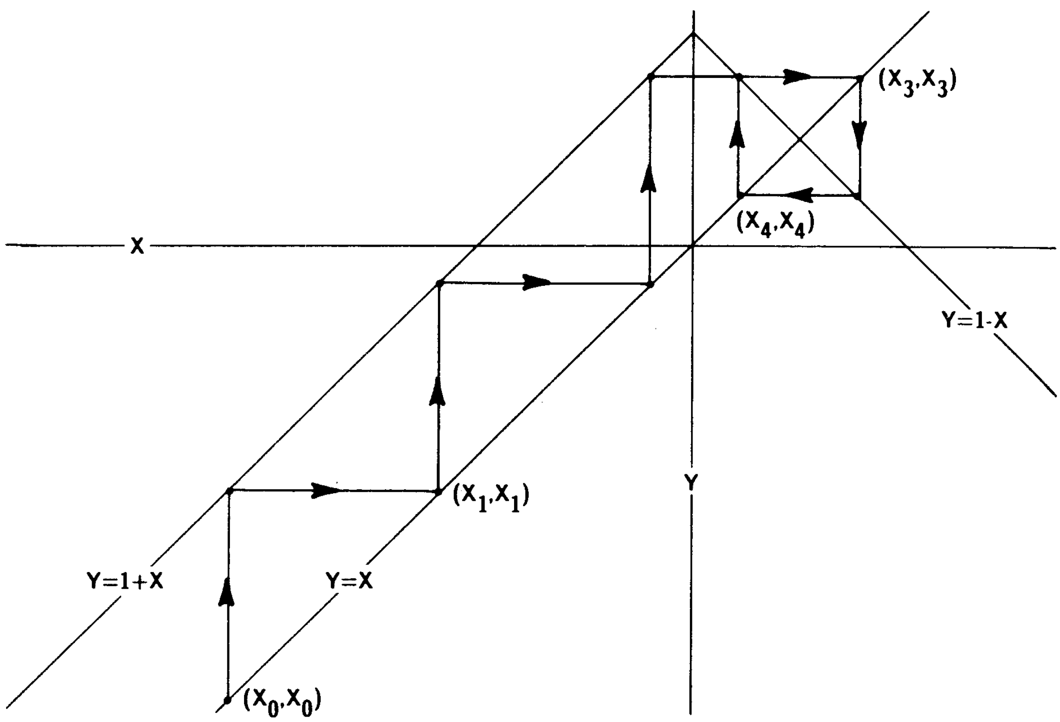


Figure IV.109. An orbit of the "slide-and-fold" dynamical system described in example 3.17. Can you prove that all orbits are eventually periodic?

of the shift automorphism here. We arrive at this by looking at the action of the shift transformation with the metric d_k with $k = N$, as in exercise 2.19, Chapter II. To simplify the discussion, assume $N = 2$. The space of shifts is a two dimensional code space with points

$$(x, y) = (x_1x_2x_3\ldots, y_1y_2y_3\ldots)$$

on which we put the "Euclidean" metric (see exercise 2.6 in Chapter II),

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{\left(\sum_{i=1}^{\infty} \frac{x_{1_i} - x_{2_i}}{2^i}\right)^2 + \left(\sum_{i=1}^{\infty} \frac{y_{1_i} - y_{2_i}}{2^i}\right)^2}.$$

The shift transformation here is best described by writing

$$(x, y) = \ldots y_3y_2y_1.x_1x_2x_3\ldots.$$

We now shift by moving the dot one place to the right, to get

$$T(x, y) = \ldots y_2y_1x_1.x_2x_3x_4\ldots = (x_2x_3\ldots, x_1y_1y_2\ldots).$$

With the metric just mentioned, we can relate it to the square $[0, 1] \times [0, 1]$. Each point in this square has a binary expansion in terms of ones and zeros, so that a point (x, y) can be written $(.x_1x_2\ldots, .y_1y_2\ldots)$, with precisely the same symbols and metric (Euclidean). The shift operation can now be seen as doing the following:

stretch x**:** double x. This shifts the first digit up so that it is in the ones place.

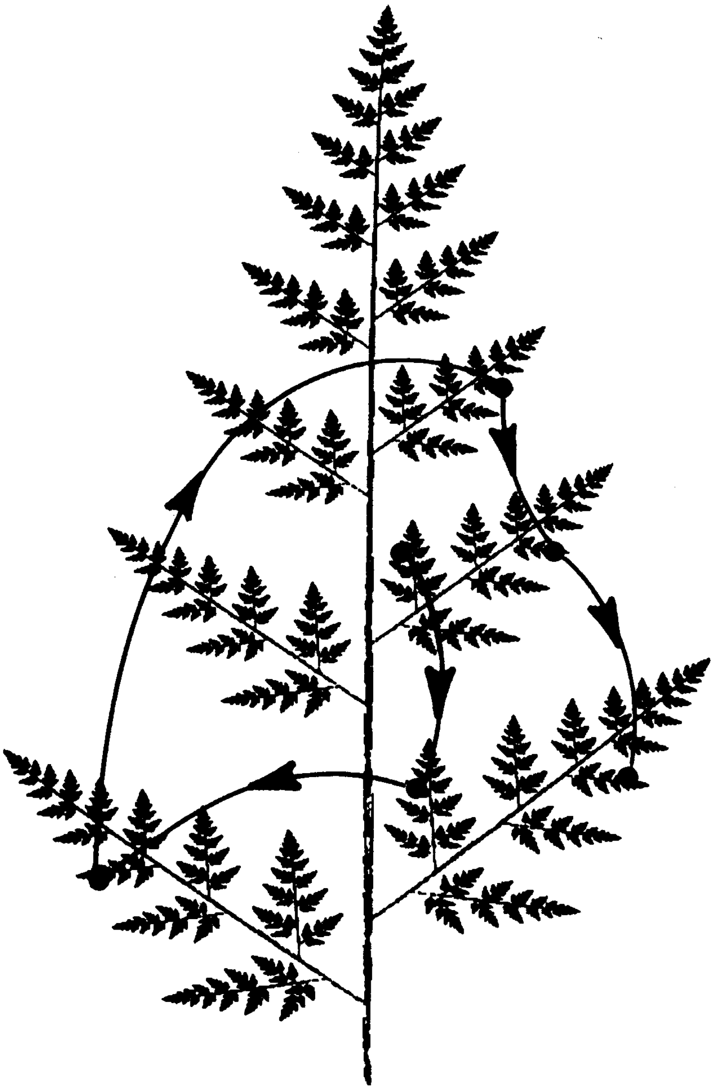


Figure IV.110. A sign of things to come.

squeeze y**:** halve y. This shifts a zero into the first digit and shifts all other digits down.

raise half the interval: If the digit now in the ones place for x is a 1, replace the new 0 first digit of y with a 1. This adds a half to y.

put it on top: If the digit in the ones place for x is a 1, discard it. This brings the x values to those between 0 and 1, so this half of the points is put above the other half.

What we have done is stretch the square to twice its width (double x) and half its height (halve y), cut the rectangle into two pieces at $x = 1$, and put the right half on top of the bottom half (add $1/2$ to y if the new x is greater than 1). This operation of stretching out the square, cutting it, and stacking the pieces is called a *baker's transformation*, because it resembles a baker rolling, cutting, and stacking dough (to make pastry, for example). It is identical to the shift transformation on the space of shifts so long as the dough remains in distinct layers (unlike $\mathbb{R}^2$, $0\bar{1} \neq 1\bar{0}$).

Definition 3.5 *This transformation is famous because it is the heart of any invertible "mixing" function (one has to allow any number of cuts and for uneven rolling). A mixing function is a function f such that given any set A (from some class of sets; here let's say sets with an interior), and any other set B from the same class, there is an N such that $f^n(A) \cap B \neq \emptyset$ for any $n > N$.*

The term *mixing* is appropriate: if A is red and B is blue, then eventually they are both somewhat purple (have both red and blue in them). A nice property of this mixing business is that there is at least one point in the space such that $\{f^n(x) : n = 1, 2, \ldots\}$ is dense, that is, given an open set $\mathcal{O}$, there is an n such that $f^n(x) \in \mathcal{O}$. When f has this property, we say that it has a *dense orbit*. T is mixing on the space of shifts and on code space, and it has a dense orbit as a result.

Examples & Exercises

3.20. Prove that for any code space Σ on N symbols, there is a point $\sigma \in \Sigma$, such that σ has a dense orbit under the shift transformation, that is $\{T^n(\sigma) : n = 1, 2, 3, \ldots\}$ is dense in Σ.

3.21. Show that T is mixing on code space for the class of open sets.

4 Dynamics on Fractals: Or How to Compute Orbits by Looking at Pictures

We continue with the main theme for this chapter, namely dynamical systems on fractals. We will need the following result.

Lemma 4.1 *Let $\{X; w_n, n = 1, 2, \ldots, N\}$ be a hyperbolic IFS with attractor A. If the IFS is totally disconnected, then for each $n \in \{1, 2, \ldots, N\}$, the transformation w_n: $A \to A$ is one-to-one.*

Proof We use a code space argument. Suppose that there is an integer $n \in \{1, 2, \ldots, N\}$ and distinct points $a_1, a_2 \in A$ so that $w_n(a_1) = w_n(a_2) = a \in A$. If a_1 has address ω and a_2 has address σ, then a has the two addresses $n\omega$ and $n\sigma$. This is impossible because A is totally disconnected. This completes the proof.

Lemma 4.1 shows that the following definition is good.

Definition 4.1 *Let $\{X; w_n, n = 1, 2, \ldots, N\}$ be a totally disconnected hyperbolic IFS with attractor A. The associated* shift transformation *on A is the transformation $S : A \to A$ defined by*

$$S(a) = w_n^{-1}(a) \qquad \text{for } a \in w_n(A),$$

where w_n is viewed as a transformation on A. The dynamical system $\{A; S\}$ is called the shift dynamical system *associated with the IFS.*

Examples & Exercises

4.1. Figure IV.111 shows the attractor of the IFS

$$\left\{\mathbb{R}^2; 0.47\begin{pmatrix}x_1\\x_2\end{pmatrix}, 0.47\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}1\\0\end{pmatrix}, 0.47\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}1\\0\end{pmatrix}\right\}.$$

Figure IV.111 also shows an eventually periodic orbit $\{a_n = S^{\circ n}(a_0)\}_{n=0}^{\infty}$ for the associated shift dynamical system. This orbit actually ends up at the fixed point $\phi(22\overline{22})$. The orbit reads $a_0 = \phi(131322\overline{22})$, $a_1 = \phi(313122\overline{22})$, $a_2 = \phi(1322\overline{22})$, $a_3 = \phi(322\overline{22})$, $a_4 = \phi(\overline{22}22)$, where $\phi : \Sigma \to A$ is the associated code space map. $a_4 \in A$ is clearly a repulsive fixed point of the dynamical system. Notice how one can read off the orbit of the point a_0 from its address. Start from another point very close to a_0 and see what happens. Notice how the dynamics depend not only on A itself, but also on the IFS. A different IFS with the same attractor will in general lead to different shift dynamics.

4.2. Both Figures IV.112 and IV.113 show attractors of IFS's. In each case the implied IFS is the obvious one. Give the addresses of the points $\{a_n = S^{\circ n}(a_0)\}_{n=0}^{\infty}$ of the eventually periodic orbit in Figure IV.112. Show that the cycle to which the

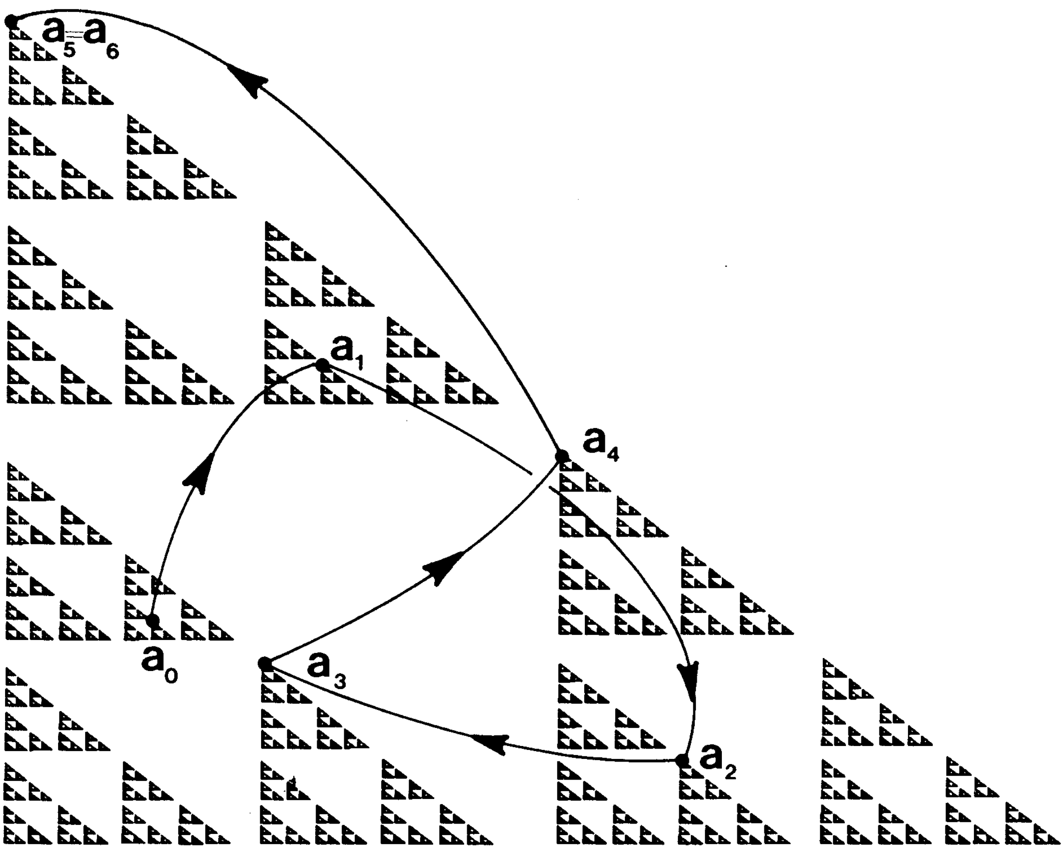


Figure IV.111. An orbit of a shift dynamical system on a fractal.

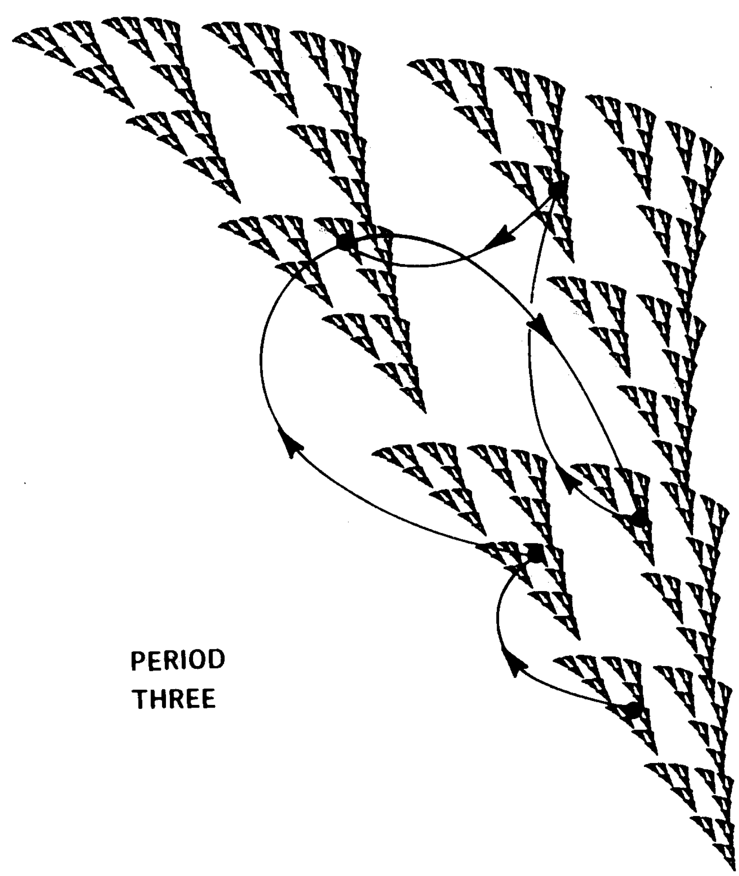


Figure IV.112. This orbit ends up in a cycle of period 3.

orbit converges is a repulsive cycle of period 3. The orbit in Figure IV.113 is either very long or infinitely long: why is it hard for us to know which?

4.3. Figure IV.114 shows an orbit of a point under the shift dynamical system associated with a certain IFS $\{\mathbb{R}^2; w_1, w_2, w_3\}$, where w_1, w_2, and w_3 are affine transformations. Deduce the orbits of the points marked b and c in the figure.

4.4. Figure IV.115 shows the start of an orbit of a point under the shift dynamical system associated with a certain hyperbolic IFS. The IFS is of the form $\{\mathbb{R}; w_1, w_2, w_3\}$, where the transformations $w_n : \mathbb{R} \to \mathbb{R}$ are affine and the attractor is $[0, 1]$. Sketch part of the orbit of the point labelled b in the figure. (Notice that this IFS is actually just-touching: nonetheless it is straightforward to define uniquely the associated shift dynamics on $\mathcal{O} \cap A$ where $\mathcal{O}$ is the open set referred to in Definition 2.2.)

We can sharpen up the definition of the overlapping IFS with the aid of the mixing properties discussed in section 3. Let $\{\mathbf{X}; w_1, \ldots, w_N\}$ be a hyperbolic IFS, and define the set

$$M = \bigcup_{i \neq j} (w_i(A) \cap w_j(A))$$

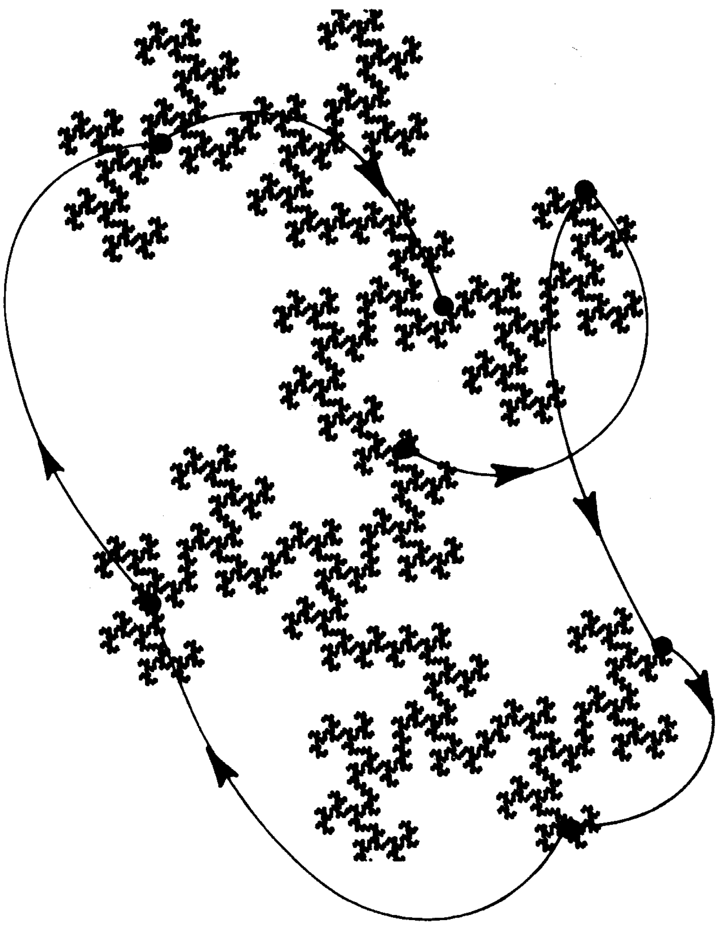


Figure IV.113. A chaotic orbit getting started. The shift dynamics are often wild. Why?

of points in various intersections of the maps of the IFS. Then the following properties hold:

open interior: If there is a set $\mathcal{O}$, open with respect to A, such that $\mathcal{O} \subset M$, then the IFS is overlapping. This allows the IFS to be declared overlapping easily in some cases. The proof is not too difficult: Suppose this to be the case, namely that M contains an open set $\mathcal{O}$. Suppose that $\mathcal{O}_1$ were an open set that we thought might satisfy the open set condition for just-touching IFS. Then $W^n(\mathcal{O}_1) \cap \mathcal{O} = \emptyset$ for all n, since $\mathcal{O}_1$ can't contain points in the overlap, and maps inside itself. Using the continuous map $\phi : \Sigma \to A$, we know that we would then have $\phi^{-1}(\mathcal{O}_1)$ and $\phi^{-1}(\mathcal{O})$ both open sets in code space. But $T^n(\phi^{-1}(\mathcal{O}))$ must intersect $\phi^{-1}(\mathcal{O}_1)$ in code space for some n, due to mixing, and an address in the intersection corresponds to a point a on the attractor such that $W^n(\{a\}$ thus intersects $\mathcal{O}$. Hence $\mathcal{O}_1$ cannot exist, and the IFS is overlapping.

dense address: Notice that in order to prevent the IFS from being just-touching in the proof just given, the orbit of $\phi^{-1}(M)$ only needs to be dense in Σ to end up with points in the image of any open set in A. Consequently,

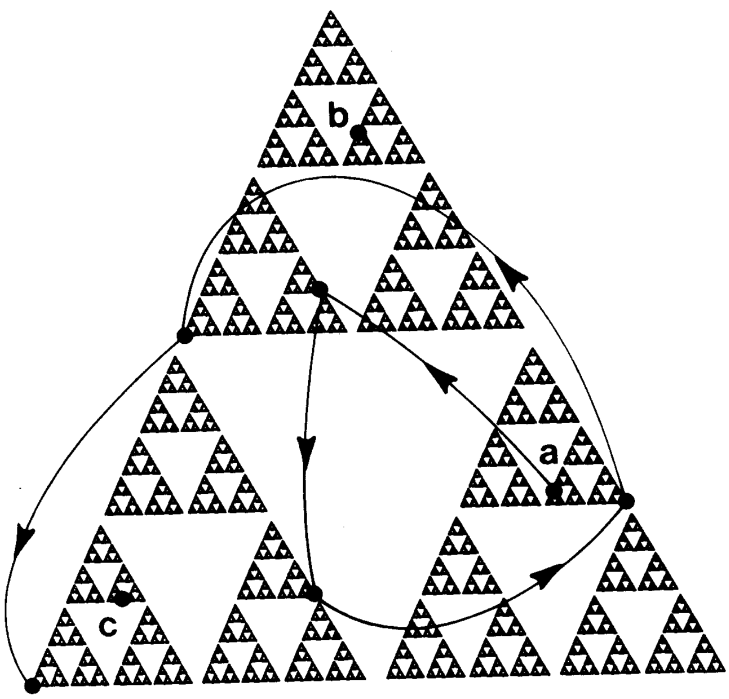


Figure IV.114. The orbit of the point a is shown. Can you plot the first few points of the orbits of b and c? Warning! The IFS here is not the usual one. See how the knowledge of some dynamics can imply some more!

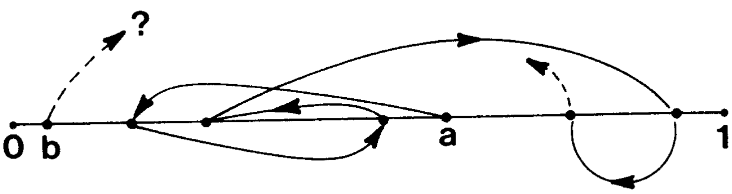


Figure IV.115. This figure shows a sketch of part of an orbit of an IFS $\{[0, 1]; w_1, w_2, w_3\}$ on its attractor $[0, 1]$. The transformation $w_i : [0, 1] \to [0, 1]$ is affine for $i = 1, 2, 3$. Sketch part of the orbit of b.

an IFS is overlapping if the orbit of $\phi^{-1}(M)$ in code space under the shift transformation is dense.

empty interiors: If the IFS is made up of affine maps, and it "looks" just-touching, that is, M does not have an interior, then it is just-touching. The important property of affine maps used here is that they map boundary points to boundary points, and boundary points come from boundary points.

4.5. The empty interiors property is not general, but is useful when it applies. To see why it is restricted, consider the following IFS, made of six translated copies of the following map:

On the interval $[-1, 1]$, let $\theta(x) = \text{Arccos}(x)$, that is, map x to the point directly above it on the unit circle. Then take $\theta(x)$ to the point $\theta(x) - \alpha \sin\theta(x)$, where $\alpha \in [0, 1/2)$. Then map the new point on the circle back to $[-1, 1]$ by taking $\theta(x)$ to $x' = \cos\theta(x)$. Now fold the interval over at 0 with the map x'^2, and ensure that it is contractive by dividing by 3. The interval has now been mapped to [0, 1/3]. Call this map $v(x)$. Explicitly, we have

$$v(x) = \frac{1}{3}\cos^2(\text{Arccos}(x) - \sin(\text{Arccos}(x)).$$

We now form 6 maps $w_1, \ldots, w_6$ by translations and inversions of $v(x)$:

$$\begin{aligned} w_1(x) &= v(x) - 1 & w_2(x) &= -v(x) - \frac{1}{3} \\ w_3(x) &= v(x) - \frac{1}{3} & w_4(x) &= -v(x) + \frac{1}{3} \\ w_5(x) &= v(x) + \frac{1}{3} & w_6(x) &= -v(x) + 1. \end{aligned}$$

The reader should be able to verify that the attractor of the IFS

$$\{[-1, 1]; w_1, w_2, w_3, w_4, w_5, w_6\}$$

is the interval $[-1, 1]$, and that each of these maps touches any neighbor at a single point. There is a point $x_0(\alpha)$ such that

$$\text{Arccos}(x_0) - \alpha \sin \text{Arccos}(x_0) = \pi/2,$$

whose image is the points

$$\{-1, -1/3, 1/3, 1\},$$

and whose endpoints map to $\{-2/3, 0, 2/3\}$. By choosing α at different values we can move x_0 around the interval $[-1/3, 0]$. If we pick $\sigma \in \Sigma$ to be a dense orbit in Σ under the shift transformation, we can successively approximate this address for x_0 such that the address of x_0 is 3σ, and we can do this for each such $\sigma \in \Sigma$. We can also do this for a variety of periodic orbits, which are not dense.

It turns out that for *most* values of $x_0 \in [-1/3, 0]$ (the probability of a value in this interval being one of these is 100%), this IFS is overlapping, although between every two values for which it is overlapping, there is a value for which it is just-touching. The attractors of this family of IFS are identical, as are the intersection points. This is thus both an example of an IFS that has a finite set of intersection points and is (sometimes) overlapping, and an example of one that does not go smoothly through the succession from totally disconnected to just-touching to overlapping. Small wonder these properties are defined for the IFS and not the attractor; they are really properties governing the behavior of addresses in code space.

5 Equivalent Dynamical Systems

Definition 5.1 *Two metric spaces* $(\mathbf{X}_1, d_1)$ *and* $(\mathbf{X}_2, d_2)$ *are said to be* topologically equivalent *if there is a homeomorphism* $f\colon \mathbf{X}_1 \to \mathbf{X}_2$. *Two subsets* $S_1 \subset \mathbf{X}_1$ *and* $S_2 \subset$

$\mathbf{X}_2$ *are topologically equivalent, or* homeomorphic, *if the metric spaces* (S_1, d_1) *and* (S_2, d_2) *are topologically equivalent.* S_1 *and* S_2 *are* metrically equivalent *if* (S_1, d_1) *and* (S_2, d_2) *are equivalent metric spaces.*

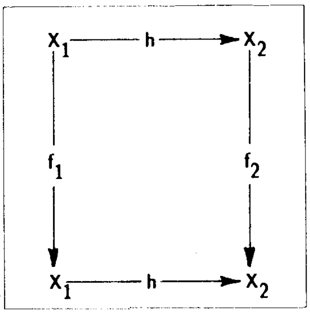
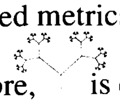
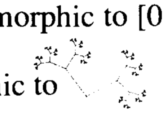


Figure IV.116. A commutative diagram that establishes the equivalence between the two dynamical systems $\{\mathbf{X}_1; f_1\}$ and $\{\mathbf{X}_2; f_2\}$. The function $h : \mathbf{X}_1 \to \mathbf{X}_2$ is a homeomorphism.

The Cantor set and code space, discussed following Theorem 2.3 in Chapter IV, are metrically equivalent. Theorem 8.5 in Chapter II tells us that if $f : \mathbf{X}_1 \to \mathbf{X}_2$ is a continuous one-to-one mapping from a compact metric $(\mathbf{X}_1, d_1)$ onto a compact metric space $(\mathbf{X}_2, d_2)$ then f is a homeomorphism. So by means of the code space mapping $\phi : \Sigma \to A$ (Theorem 2.1) one readily establishes that the attractor of a totally disconnected hyperbolic IFS is topologically equivalent to a classical Cantor set.

Topological equivalence permits a great deal more "stretching" and "compression" to take place than is permitted by metric equivalence. Later we will define a quantity called the fractal dimension. The fractal dimension of a subset of a metric space such as ($\mathbb{R}^2$, Euclidean) provides a measure of the geometrical complexity of the set; it measures the wildness of the set, and it may be used to predict your excitement and wonder when you look at a picture of the set. We will show that two metrically equivalent sets have the same fractal dimension. If they are merely topologically equivalent, their fractal dimensions may be different.

With the naturally implied metrics, $[0, 1]$ is homeomorphic to $[0, 2]$. ■ is homeomorphic to ●. What is more,  is even homeomorphic to , and ——— is homeomorphic to .

In fractal geometry we are especially interested in the *geometry* of sets, and in the way they *look,* when they are represented by pictures. Thus we use the restrictive condition of metric equivalence to start to define mathematically what we mean when we say that two sets are alike. However, in dynamical systems theory we are interested in *motion* itself, in the dynamics, in the way points move, in the existence of periodic orbits, in the asymptotic behavior of orbits, and so on. These structures are not damaged by homeomorphisms, as we will see, and hence we say that two dynamical systems are alike if they are related via a homeomorphism.

Definition 5.2 *Two dynamical systems* $\{\mathbf{X}_1; f_1\}$ *and* $\{\mathbf{X}_2; f_2\}$ *are said to be* equivalent, *or* topologically conjugate, *if there is a homeomorphism* $\theta : \mathbf{X}_1 \to \mathbf{X}_2$ *such that*

$$f_1(x_1) = \theta^{-1} \circ f_2 \circ \theta(x_1) \text{ for all } x_1 \in \mathbf{X}_1,$$

$$f_2(x_2) = \theta \circ f_1 \circ \theta^{-1}(x_2) \text{ for all } x_2 \in \mathbf{X}_2.$$

In other words, the two dynamical systems are related by the commutative diagram shown in Figure IV.116.

The following theorem expresses formally what should already be clear intuitively from our experience with shift dynamics on fractals.

Theorem 5.1 *Let $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ be a totally disconnected hyperbolic IFS and let $\{A; S\}$ be the associated shift dynamical system. Let Σ be the associated code space of N symbols and let $T : \Sigma \to \Sigma$ be defined by*

$$T(\sigma_1\sigma_2\sigma_3\ldots) = \sigma_2\sigma_3\sigma_4\ldots \qquad \text{for all } \sigma = \sigma_1\sigma_2\sigma_3\ldots \in \Sigma.$$

Then the two dynamical systems $\{A; S\}$ and $\{\Sigma; T\}$ are equivalent. The homeomorphism that provides this equivalence is $\phi : \Sigma \to A$, as defined in Theorem 4.2.1. Moreover, $\{a_1, a_2, \ldots, a_p\}$ is a repulsive cycle of period p for S if, and only if, $\{\phi(a_1), \phi(a_2), \ldots, \phi(a_p)\}$ is a repulsive cycle of period p for T.

Examples & Exercises

5.1. Let $\{\mathbf{X}_1; f_1\}$ and $\{\mathbf{X}_2; f_2\}$ be equivalent dynamical systems. Let a homeomorphism that provides this equivalence be denoted by $\theta : \mathbf{X}_1 \to \mathbf{X}_2$. Show that

$$\{x_1, x_2, \ldots, x_p\}$$

is a cycle of period p for $\{\mathbf{X}_1; f_1\}$ if and only if

$$\{\theta(x_1), \theta(x_2), \ldots, \theta(x_p)\}$$

is a cycle of period p for $\{\mathbf{X}_2; f_2\}$. Suppose that $\{x_1, x_2, \ldots, x_p\}$ is an attractive cycle for f_1. Show that this does not imply that $\{\theta(x_1), \ldots, \theta(x_p)\}$ is an attractive cycle for f_2.

5.2. Let $\{\mathbf{X}_1; f_1\}$ and $\{\mathbf{X}_2; f_2\}$ be equivalent dynamical systems. Let a homeomorphism that provides this equivalence be denoted by $\theta : \mathbf{X}_1 \to \mathbf{X}_2$. Let $\{f_1^{\circ n}(x)\}_{n=0}^{\infty}$ be an eventually periodic orbit of f_1. Show that $\{f_2^{\circ n}(\theta(x))\}_{n=0}^{\infty}$ is an eventually periodic orbit of f_2.

5.3. Let $\{\mathbf{X}_1; f_1\}$ and $\{\mathbf{X}_2; f_2\}$ be equivalent dynamical systems. Let a homeomorphism that provides this equivalence be denoted by $\theta : \mathbf{X}_1 \to \mathbf{X}_2$. Let this homeomorphism be such as to make the two spaces $(\mathbf{X}_1, d_1)$ and $(\mathbf{X}_2, d_2)$ metrically equivalent. Construct an example where $x_f \in \mathbf{X}_1$ is a repulsive fixed point of the dynamical system $\{\mathbf{X}_1, f_1\}$ yet $\theta(x_f)$ is not a repulsive fixed point of $\{\mathbf{X}_2, d_2\}$.

5.4. Let $\{\mathbf{X}_1; f_1\}$ and $\{\mathbf{X}_2; f_2\}$ be equivalent metric spaces. Let a homeomorphism that provides their equivalence be denoted by $\theta : \mathbf{X}_1 \to \mathbf{X}_2$. Let $x_f \in \mathbf{X}_1$ be a fixed point of f_1 . Suppose there is an open set $\mathcal{O}$ that contains x_f and is such that $x \in \mathcal{O}$ implies $\lim_{n\to\infty} f_1^{\circ n}(x) = x_f$. Show that there is an open neighborhood of $\theta(x_f)$ in $\mathbf{X}_2$ with a similar property.

5.5. Our definition of *attractive* and *repulsive* fixed points and cycles, Definition 3.4, has the feature that it depends heavily on the metric. It is motivated by the situation of analytic dynamics where small disks are almost mapped into disks. Show how one can use exercise 5.4 to make a definition of an attractive cycle in such a way that attractiveness of cycles is preserved under topological conjugacy.

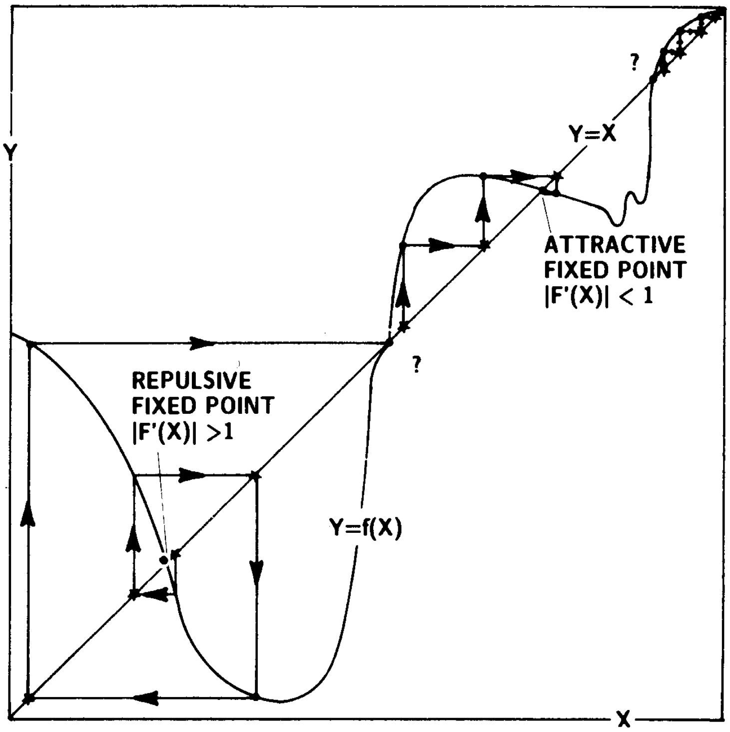


Figure IV.117. Attractive and repulsive fixed points in a web diagram for a differentiable dynamical system. Analyze the ? points.

5.6. Let $A \subset \mathbb{R}$. Then a function $f : A \to A$ is *differentiable* at a point $x_0 \in A$ if

$$\lim_{\substack{x \to x_0 \\ x \in A}} \left\{ \frac{f(x) - f(x_0)}{x - x_0} \right\}$$

exists. If this limit exists it is denoted by $f'(x_0)$. Let $\{\mathbb{R}; w_1, w_2, \ldots, w_N\}$ be a totally disconnected hyperbolic IFS acting on the metric space ($\mathbb{R}$, Euclidean). Suppose that, for each $n = 1, 2, \ldots, N$, $w_n(x)$ is differentiable, with $|w_n'(x)| > 0$ for all $x \in \mathbb{R}$. Show that the associated shift dynamical system $\{A; S\}$ is such that S is differentiable at each point $x_0 \in A$ and, moreover, $|S'(x_0)| > 1$ for all $x \in A$.

5.7. Let $\{\mathbb{R}; f\}$ and $\{\mathbb{R}; g\}$ be equivalent dynamical systems. Let a homeomorphism that provides their equivalence be denoted by $\theta : \mathbb{R} \to \mathbb{R}$. If $\theta(x)$ is differentiable for all $x \in \mathbb{R}$, then the dynamical systems are said to be *diffeomorphic*. Prove that a_1 is an attractive fixed point of f if and only if $\theta(a_1)$ is an attractive fixed point of g.

5.8. Let $\{\mathbb{R}; f\}$ be a dynamical system such that f is differentiable for all $x \in \mathbb{R}$. Consider the web diagrams associated with this system. Show that the fixed points of f are exactly the intersections of the line $y = x$ with the graph $y = f(x)$. Let a be a fixed point of f. Show that a is an attractive fixed point of f if and only if $|f'(a)| < 1$. Generalize this result to cycles. Note that if $\{a_1, a_2, \ldots, a_p\}$ is a cycle of period p, then $\frac{d}{dx}(f^{\circ p}(x)|_{x=a_1} = f'(a_1)f'(a_2)\ldots f'(a_p)$. Assure yourself that the situation is correctly summarized in the web diagram shown in Figure IV.117.

5.9. Consider the dynamical system $\{[0, 1]; f(x)\}$ where

$$f(x) = \begin{cases} 1 - 2x & \text{when } x \in [0, \frac{1}{2}], \\ 2x - 1 & \text{when } x \in [\frac{1}{2}, 1]. \end{cases}$$

Consider also the just-touching IFS $\{[0, 1], \frac{1}{2}x + \frac{1}{2}, -\frac{1}{2}x + \frac{1}{2}\}$. Show that it is possible to define a "shift transformation," S, on the attractor, A, of this IFS in such a way that $\{[0, 1]; S\}$ and $\{[0, 1]; f(x)\}$ are equivalent dynamical systems. To do this you should define $S : A \to A$ in the obvious manner for points with unique addresses; and you should make a suitable definition for the action of S on points with multiple addresses.

5.10. Let $\{\mathbb{R}^2; w_1, w_2, w_3\}$ denote a one-parameter family of IFS, where

$$w_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (\frac{1+p}{4}) & 0 \\ 0 & (\frac{1+p}{4}) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix};$$

$$w_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (\frac{1+p}{4}) & 0 \\ 0 & (\frac{1+p}{4}) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{3+p}{8} \\ \frac{p}{2} \end{pmatrix},$$

$$w_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1+p}{4}) & 0 \\ 0 & (\frac{1+p}{4}) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{3-p}{4} \\ 0 \end{pmatrix} \quad \text{for } p \in [0, 1].$$

Let the attractor of this IFS be denoted by $A(p)$. Show that $A(0)$ is a Cantor set and $A(1)$ is a Sierpinski triangle. Consider the associated family of code space maps $\phi(p) : \Sigma \to A(p)$. Show that $\phi(p)(\sigma)$ is continuous in p for fixed $\sigma \in \Sigma$; that is $\phi(p)(\sigma) : [0, 1] \to \mathbb{R}^2$ is a continuous path. Draw some of these paths, including ones that meet at $p = 1$. Interpret these observations in terms of the Cantor set becoming "joined to itself" at various points to make a Sierpinski triangle, as suggested in Figure IV.118.

Since the IFS is totally disconnected when $p = 0$, $\phi(p = 0) : \Sigma \to A(0)$ is invertible. Hence we can define a continuous transformation $\theta : A(0) \to A(1)$ by $\theta(x) = \phi(p = 1)(\phi^{-1}(p = 0)(x))$. Show that if we define a set $J(x) = \{y \in A(0) : \theta(y) = x\}$ for each $x \in A(1)$, then $J(x)$ is the set of points in $A(0)$ whose associated paths meet at $x \in A(1)$ when $p = 1$. Invent shift dynamics on paths.

6 The Shadow of Deterministic Dynamics

Our goal in this section is to extend the definition of the shift dynamical system associated with a totally disconnected hyperbolic IFS to cover the just-touching and overlapping cases. This will lead us to the idea of a random shift dynamical system and to the discovery of a beautiful theorem. This theorem will be called the Shadow Theorem.

Let $\{X; w_1, w_2, \ldots, w_N\}$ denote a hyperbolic IFS, and let A denote its attractor. Assume that $w_n : A \to A$ is invertible for each $n = 1, 2, \ldots, N$, but that the IFS is

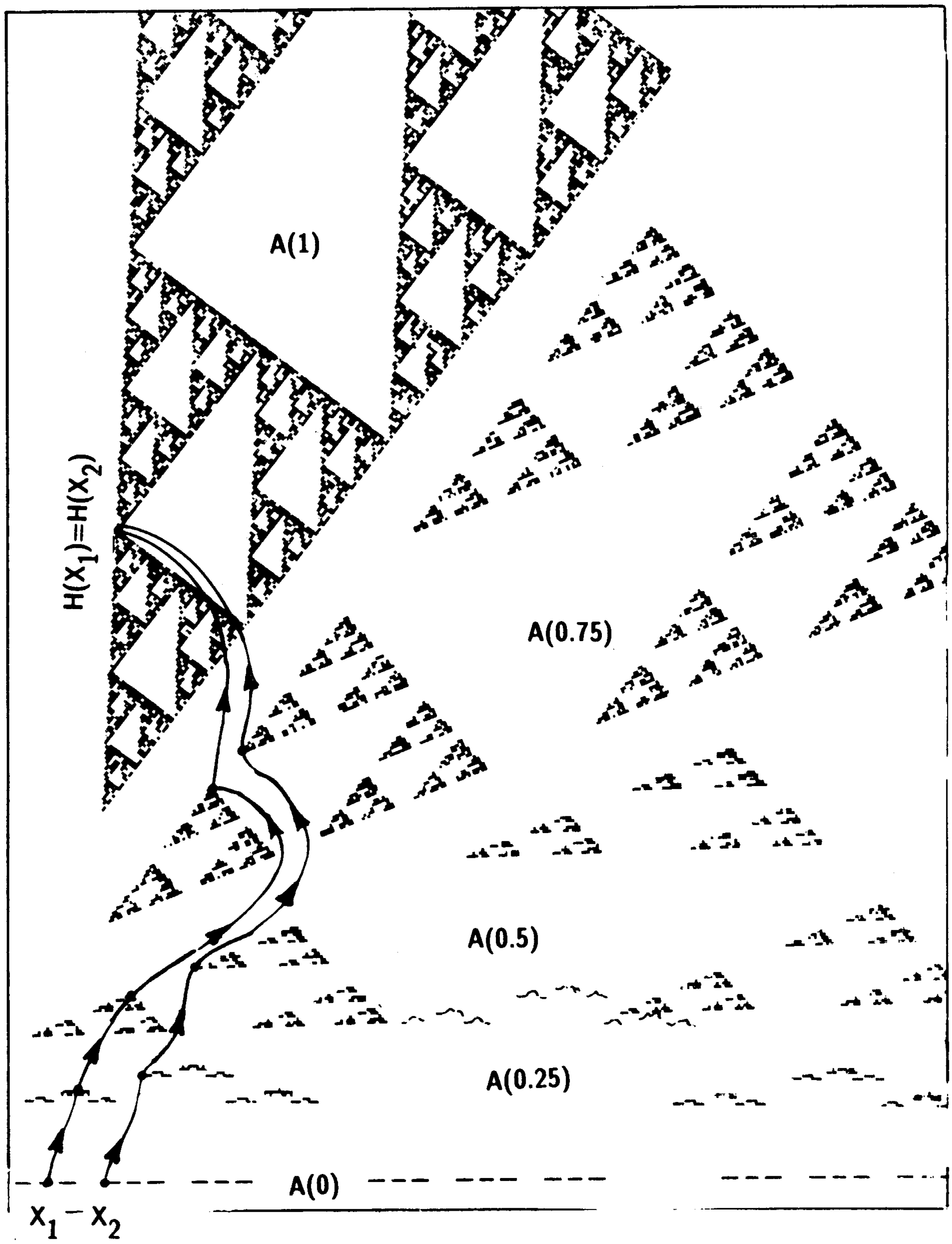


Figure IV.118. Continuous transformation of a Cantor set into a Sierpinski triangle. The inverse transformation would involve some ripping.

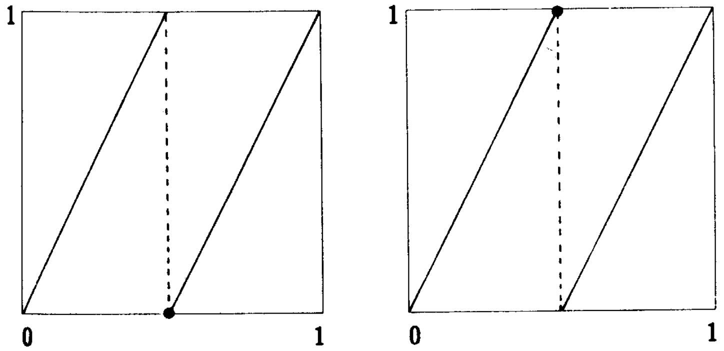


Figure IV.119. The two possible shift dynamical systems associated with the just-touching IFS $\{[0,1]; \frac{1}{2}x, \frac{1}{2}x + \frac{1}{2}\}$ are represented by the two possible graphs of $S(x)$. "Most" orbits are unaffected by the difference between the two systems.

not totally disconnected. We want to define a dynamical system $\{A; S\}$ analogous to the shift dynamical system defined earlier. Clearly, we should define

$$S(x) = w_n^{-1}(x) \qquad \text{when } x \in w_n(A), \text{ but } x \notin w_m(A) \qquad \text{for } m \neq n,$$

for each $n = 1, 2, \ldots, N$.

However, at least one of the intersections $w_m(A) \cap w_n(A)$ is nonempty for some $m \neq n$. One idea is simply to make an assignment of which inverse map is to be applied in the overlapping region. For the case $N = 2$ we might define, for example,

$$S(x) = \begin{cases} w_1^{-1}(x) & \text{when } x \in w_1(A), \\ w_2^{-1}(x) & \text{when } x \in A \setminus w_1(A). \end{cases}$$

In the just-touching case the assignment of where S takes points that lie in the overlapping regions does not play a very important role: only a relatively small proportion of points will have somewhat arbitrarily specified orbits. We look at some examples, just to get the flavor.

Examples & Exercises

6.1. Consider the shift dynamical systems associated with the IFS

$$\{[0,1]; \frac{1}{2}x, \frac{1}{2}x + \frac{1}{2}\}.$$

We have $S(x) = 2x$ for $x \in [0, \frac{1}{2})$ and $S(x) = 2x - 1$ for $x \in (\frac{1}{2}, 1]$. We can define the value of $S(\frac{1}{2})$ to be either 1 or 0. The two possible graphs for $S(x)$ are shown in Figure IV.119. The only points $x \in [0,1] = A$ whose orbits are affected by the definition are those rational numbers whose binary expansions end $\ldots 01\overline{1}$ or $\ldots 10\overline{0}$, the dyadic rationals.

6.2. Show that if we follow the ideas introduced above, there is only one dynamical system $\{A; S\}$ that can be associated with the just-touching IFS $\{[0,1]; -\frac{1}{2}x + \frac{1}{2}, \frac{1}{2}x\}$. The key here is that $w_1^{-1}(x) = w_2^{-1}(x)$ for all $x \in w_1(A) \cap w_2(A)$.

6.3. Consider some possible "shift" dynamical systems $\{A; S\}$ that can be associated with the IFS

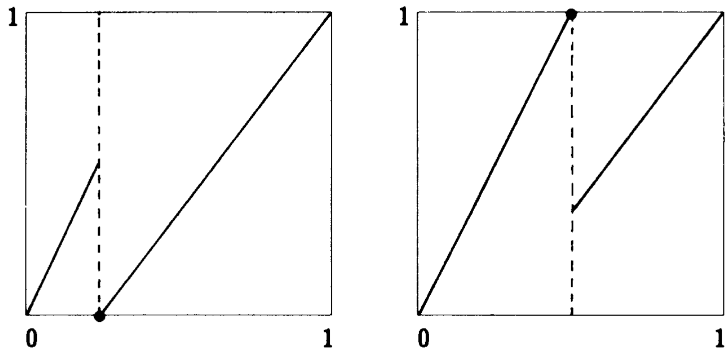


Figure IV.120. Two possible shift dynamical systems that can be associated with the overlapping IFS $\{[0,1]; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}$. In what ways are they alike?

$$\{C; \frac{1}{2}z, \frac{1}{2}z + \frac{1}{2}, \frac{1}{2}z + \frac{i}{2}\}.$$

The attractor, $\mathbb{A}$, is overlapping at the three points $a = w_1(\mathbb{A}) \cap w_2(\mathbb{A})$, $b = w_2(\mathbb{A}) \cap w_3(\mathbb{A})$, and $c = w_3(\mathbb{A}) \cap w_1(\mathbb{A})$. We might define $S(a) = w_1^{-1}(a)$ or $w_2^{-1}(a)$, $S(b) = w_2^{-1}(b)$ or $w_3^{-1}(b)$, and $S(c) = w_3^{-1}(c)$ or $w_1^{-1}(c)$. Show that regardless of which definition is made, the orbits of a, b, and c are eventually periodic.

6.4. Consider a just-touching IFS of the form $\{\mathbb{R}^2; w_1, w_2, w_3\}$ whose attractor is an equilateral Sierpinski triangle $\mathbb{A}$. Assume that each of the maps is a similitude of scaling factor 0.5. Consider the possibility that each map involves a rotation through 0°, 120°, or 240°. The attractor, $\mathbb{A}$, is overlapping at the three points $a = w_1(\mathbb{A}) \cap w_2(\mathbb{A})$, $b = w_2(\mathbb{A}) \cap w_3(\mathbb{A})$, and $c = w_3(\mathbb{A}) \cap w_1(\mathbb{A})$. Show that it is possible to choose the maps so that $w_1^{-1}(a) = w_2^{-1}(a)$, $w_2^{-1}(b) = w_3^{-1}(b)$, and $w_3^{-1}(c) = w_1^{-1}(c)$.

6.5. Is code space on two symbols topologically equivalent to code space on three symbols? Yes! Construct a homeomorphism that establishes this equivalence.

6.6. Consider the hyperbolic IFS $\{\Sigma; t_1, t_2, \ldots, t_N\}$, where Σ is code space on N symbols $\{1, 2, \ldots, N\}$ and

$$t_n\sigma = n\sigma \qquad \text{for all } \sigma \in \Sigma.$$

Show that the associated shift dynamical system is exactly $\{\Sigma; T\}$ defined in Theorem 4.5.1. Can two such shift dynamical systems be equivalent for different values of N? To answer this question consider how many fixed points the dynamical system $\{\Sigma; T\}$ possesses for different values of N.

6.7. Consider the overlapping hyperbolic IFS $\{[0,1]; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}$. Compare the two associated shift dynamical systems whose graphs are shown in Figure IV.120. What features do they share in common?

6.8. Demonstrate that code space on two symbols is not metrically equivalent to code space on three symbols.

In considering exercises such as 6.7, where two different dynamical systems are

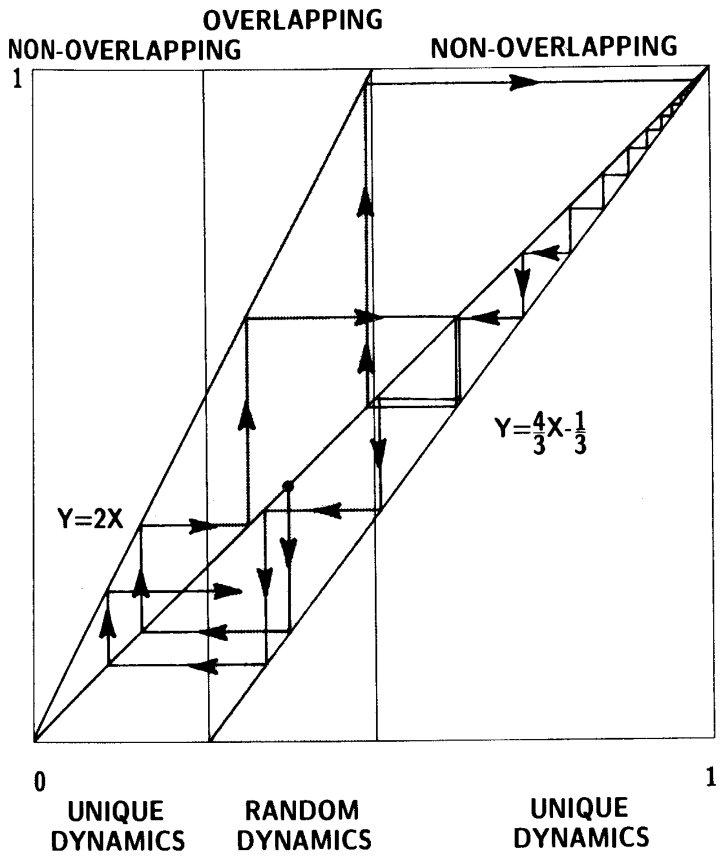


Figure IV.121. A partially random and partially deterministic shift dynamical system associated with the IFS $\{[0, 1]; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}$.

associated with an IFS in the overlapping case, we are tempted to entertain the idea that no particular definition of the shift dynamics in the overlapping regions is to be preferred. This suggests that we define the dynamics in overlapping regions in a somewhat random manner. Whenever a point on an orbit lands in an overlapping region we should allow the possibility that the next point on the orbit is obtained by applying any one of the available inverse transformations. This idea is illustrated in Figure IV.121, which should be compared with Figure IV.120.

Definition 6.1 *Let $\{X; w_1, w_2\}$ be a hyperbolic IFS. Let A denote the attractor of the IFS. Assume that both $w_1 : A \to A$ and $w_2 : A \to A$ are invertible. A sequence of points $\{x_n\}_{n=0}^{\infty}$ in A is called an orbit of the* random shift dynamical system *associated with the IFS if*

$$x_{n+1} = \begin{cases} w_1^{-1}(x_n) & \text{when } x_n \in w_1(A) \text{ and } x_n \notin w_1(A) \cap w_2(A), \\ w_2^{-1}(x_n) & \text{when } x_n \in w_2(A) \text{ and } x_n \notin w_1(A) \cap w_2(A), \\ \text{one of } \{w_1^{-1}(x_n), w_2^{-1}(x_n)\} & \text{when } x_n \in w_1(A) \cap w_2(A), \end{cases}$$

for each $n \in \{0, 1, 2, \ldots\}$. We will use the notation $x_{n+1} = S(x_n)$ although there may be no well-defined transformation $S : A \to A$ that makes this true. Also we will write $\{A; S\}$ to denote the collection of possible orbits defined here, and we will call $\{A; S\}$ the random shift dynamical system associated with the IFS.

Notice that if $w_1(A) \cap w_2(A) = \emptyset$ then the IFS is totally disconnected and the orbits defined here are simply those of the shift dynamical system $\{A; S\}$ defined earlier.

We now show that there is a completely deterministic dynamical system acting on a higher-dimensional space, whose projection into the original space X yields the "random dynamics" we have just described. Our random dynamics are seen as the shadow of deterministic dynamics. To achieve this we turn the IFS into a totally disconnected system by introducing an additional variable. To keep the notation succinct we restrict the following discussion to IFS's of two maps.

Definition 6.2 *The* lifted *IFS associated with a hyperbolic IFS* $\{X; w_1, w_2\}$ *is the hyperbolic IFS* $\{X \times \Sigma; \tilde{w}_1, \tilde{w}_2\}$, *where* Σ *is the code space on two symbols* $\{1, 2\}$, *and*

$$\tilde{w}_1(x, \sigma) = (w_1(x), 1\sigma) \qquad \text{for all } (x, \sigma) \in X \times \Sigma;$$

$$\tilde{w}_2(x, \sigma) = (w_2(x), 2\sigma) \qquad \text{for all } (x, \sigma) \in X \times \Sigma.$$

What is the nature of the attractor $\tilde{A} \subset X \times \Sigma$ of the lifted IFS? It should be clear that

$$A = \{x \in A : (x, \sigma) \in \tilde{A}\} \text{ and } \Sigma = \{\sigma \in \Sigma : (x, \sigma) \in \tilde{A}\}.$$

In other words, the projection of the attractor of the lifted IFS into the original space X is simply the attractor A of the original IFS. The projection of $\tilde{A}$ into Σ is Σ. Recall that Σ is equivalent to a classical Cantor set. This tells us that the attractor of the lifted IFS is totally disconnected.

Lemma 6.1 *Let* $\{X; w_1, w_2\}$ *be a hyperbolic IFS with attractor* A. *Let the two transformations* $w_1 : A \to A$ *and* $w_2 : A \to A$ *be invertible. Then the associated lifted IFS is hyperbolic and totally disconnected.*

Definition 6.3 *Let* $\{X; w_1, w_2\}$ *be a hyperbolic IFS. Let the two transformations* $w_1 : A \to A$ *and* $w_2 : A \to A$ *be invertible. Let* $\tilde{A}$ *denote the attractor of the associated lifted IFS. Then the shift dynamical system* $\{\tilde{A}; \tilde{S}\}$ *associated with the lifted IFS is called the* lifted shift dynamical system *associated with the IFS.*

Notice that

$$\tilde{S}(x, \sigma) = (w_{\sigma_1}^{-1}(x), T(\sigma)) \qquad \text{for all} (X, \sigma) \in \tilde{A},$$

where

$$T(\sigma_1\sigma_2\sigma_3\sigma_4 \ldots) = \sigma_2\sigma_3\sigma_4\sigma_5 \ldots \qquad \text{for all } \sigma = \sigma_1\sigma_2\sigma_3\sigma_4 \ldots . \in \Sigma.$$

Theorem 6.1 *[(The Shadow Theorem).] Let* $\{X; w_1, w_2\}$ *be a hyperbolic IFS of invertible transformations* w_1 *and* w_2 *and attractor* A. *Let* $\{x_n\}_{n=0}^{\infty}$ *be any orbit of the associated random shift dynamical system* $\{A; S\}$. *Then there is an orbit* $\{\tilde{x}_n\}_{n=0}^{\infty}$ *of the lifted dynamical system* $\{\tilde{A}; \tilde{S}\}$ *such that the first component of* $\tilde{x}_n$ *is* x_n *for all* n.

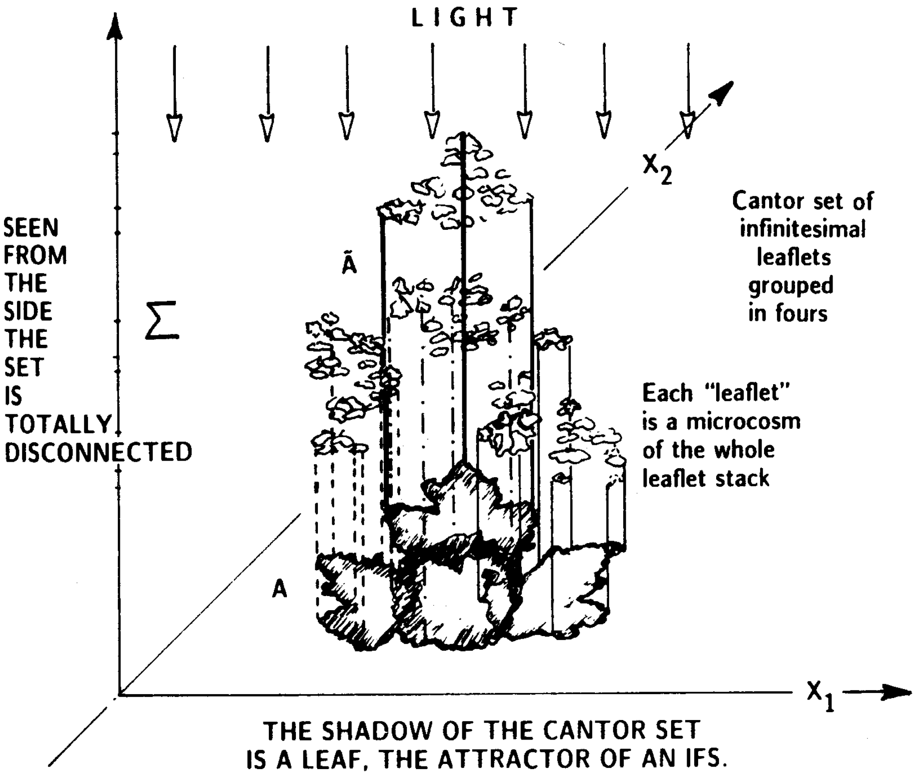


Figure IV.122. The lift of the overlapping leaf attractor is totally disconnected. Deterministic shift dynamics become possible. See also Figure IV.123.

We leave the proofs of Lemma 6.1 and Theorem 6.1 as exercises. It is fun, however, and instructive to look in a couple of different geometrical ways at what is going on here.

Examples & Exercises

6.9. Consider the IFS $\{C; w_1(z), w_2(z), w_3(z), w_4(z)\}$ where, in complex notation,

$$w_1(z) = (0.5)(\cos 45^\circ - \sqrt{-1}\sin 45^\circ)z + (0.4 - 0.2\sqrt{-1}),$$
$$w_2(z) = (0.5)(\cos 45^\circ + \sqrt{-1}\sin 45^\circ)z - (0.4 + 0.2\sqrt{-1}),$$
$$w_3(z) = (0.5)z + \sqrt{-1}(0.3),$$
$$w_4(z) = (0.5)z - \sqrt{-1}(0.3).$$

A sketch of its attractor is included in Figure IV.122. It looks like a maple leaf. The leaf is made of four overlapping leaflets, which we think of as separate entities, at different heights "above" the attractor. In turn, we think of each leaflet as consisting of four smaller leaflets, again at different heights. One quickly gets the idea: one ends up with a set of heights distributed on a Cantor set in such a way that the shadow of the whole collection of infinitesimal leaflets is the leaf attractor in the C plane. The Cantor set is essentially Σ. The lifted attractor is totally disconnected; it supports deterministic shift dynamics, as illustrated in Figure IV.123.

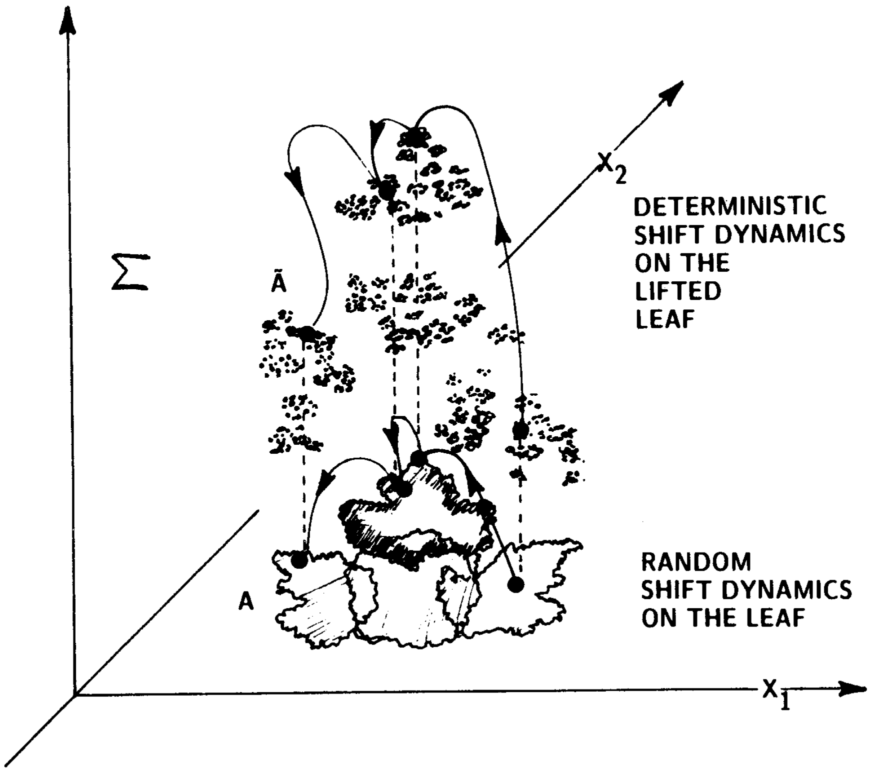


Figure IV.123. A picture of the Shadow Theorem. Deterministic dynamics on a totally disconnected dust has a shadow that is dancing random shift dynamics on a leaf attractor.

6.10. Consider the overlapping hyperbolic IFS $\{\mathbb{R}; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}$. We can lift this to the hyperbolic IFS $\{\mathbb{R}^2; w_1(x), w_2(x)\}$, where

$$w_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix};$$

$$w_2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ \frac{2}{3} \end{pmatrix}.$$

The attractor $\tilde{A}$ of this lifted system is shown in Figure IV.124, which also shows an orbit of the associated shift dynamical system. The shadow of this orbit is an apparently random orbit of the original system. The Shadow Theorem asserts that *any* orbit $\{x_n\}_{n=0}^{\infty}$ of a random shift dynamical system associated with the IFS $\{\mathbb{R}; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}$ is the projection, or shadow, of some orbit for the shift dynamical system associated with the lifted IFS.

6.11. As a compelling illustration of the Shadow Theorem, consider the IFS

$$\{\mathbb{R}; \frac{1}{2}x, \frac{3}{4}x + \frac{1}{4}\}.$$

Let us look at the orbits $\{x_n\}_{n=0}^{\infty}$ of the shift dynamical system specified in the left-hand graph of Figure IV.120. In this case we *always* choose $S(x) = w_2^{-1}(x)$ in the overlapping region. What orbits $\{\tilde{x}_n\}_{n=0}^{\infty}$ of the lifted system, described in exercise 6.7, are these orbits the shadows of? Look again at Figure IV.124! Define

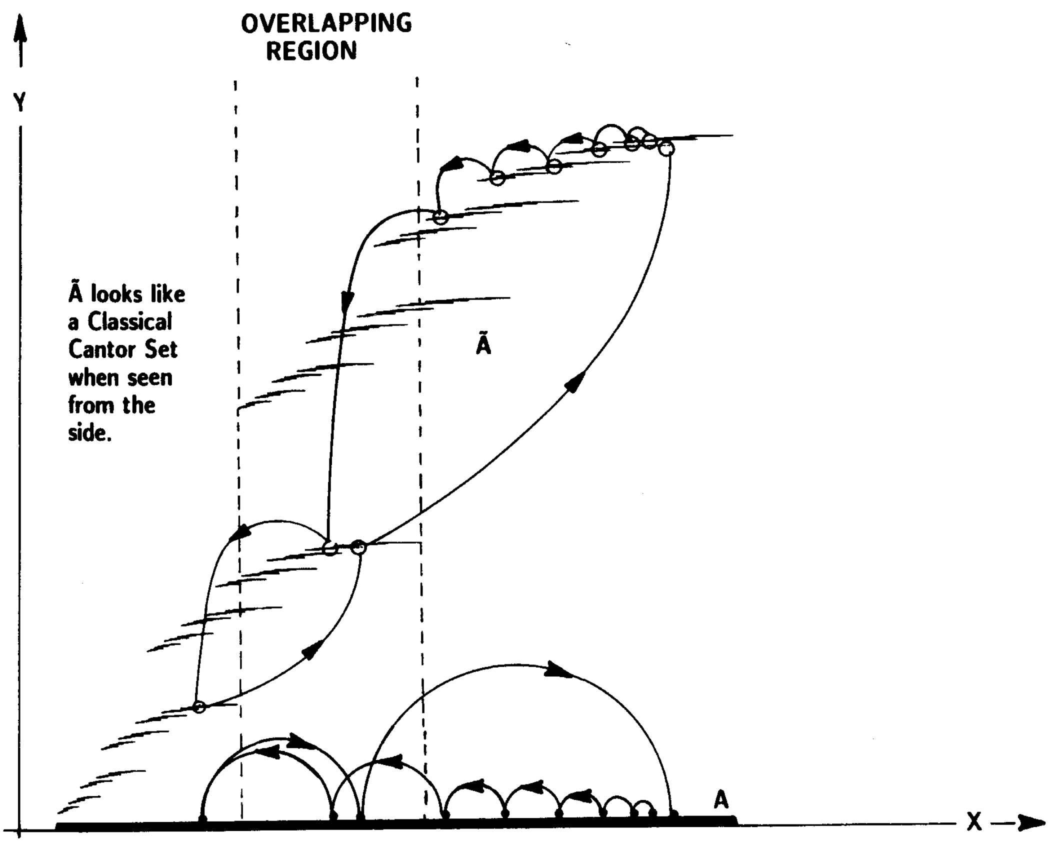


Figure IV.124. The Shadow Theorem asserts that the random shift dynamical system orbit on the overlapping attractor A is the shadow of a deterministic orbit on $\tilde{A}$.

the top of $\tilde{A}$ as

$$\tilde{A}_{top} = \{(x, y) \in \tilde{A} : (z, y) \in \tilde{A} \Rightarrow z \leq x, \qquad \text{and } y \in [0, 1]\}.$$

Notice that $\tilde{S} : \tilde{A}_{top} \to \tilde{A}_{top}$. It is easy to see that there is a one-to-one correspondence between orbits of the lifted system $\{\tilde{A}_{top}; \tilde{S}\}$ and orbits of the original system specified through the left-hand graph of Figure IV.120. Indeed,

$$\{(x_n, y_n)\}_{n=0}^{\infty} \text{ is an orbit of the lifted system and } (x_0, y_0) \in \tilde{A}_{\text{top}}$$

$$\Updownarrow$$

$$\{x_n\}_{n=0}^{\infty} \text{ is an orbit of the left-hand graph of Figure IV.120}$$

6.12. Draw some pictures to illustrate the Shadow Theorem in the case of the just-touching IFS $\{[0, 1]; \frac{1}{2}x, \frac{1}{2}x + \frac{1}{2}\}$.

6.13. Illustrate the Shadow Theorem using the overlapping IFS $\{[0, 1]; -\frac{3}{4}x + \frac{3}{4}, \frac{3}{4}x + \frac{1}{4}\}$. Find an orbit of period 2 whose lift has minimal period 4. Do there exist periodic orbits whose lifts are not periodic?

6.14. Prove Lemma 6.1.

6.15. Prove Theorem 6.1.

6.16. The IFS $\{\Sigma; w_1(\sigma), \ldots, w_N(\sigma)\}$ given by

$$w_n(\sigma) = n\sigma$$

for each $n = 1, 2, \ldots, N$, has an interesting lift. Show that the lift of this IFS, with a suitably defined inverse, is the shift automorphism on the space of shifts and therefore equivalent to the baker's transformation.

6.17. In section 5 it was shown that the associated shift dynamical system of any totally disconnected IFS is equivalent to the shift transformation on code space. Then we may replace the second map in the lift for the Shadow Theorem with such a totally disconnected IFS. That is, we could take a map like the leaf shown in Figures IV.122 and IV.123, and define the map

$$\{\mathbb{R}^2 \times \mathbb{A}; \tilde{w}_1(x, y), \ldots \tilde{w}_4(x, y)\},$$

where $\tilde{w}_i = (w_i^{-1}(x, y), v_i(x, y))$, where v_i are the maps of the totally disconnected IFS

$$v_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix},$$

$$v_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$v_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

$$v_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Since this IFS produces an attractor that is totally disconnected, and therefore a copy of code space, the resulting lift is totally disconnected. What would a rendition of the lifted system look like if the maple leaf were lifted using a totally disconnected tree?

7 The Meaningfulness of Inaccurately Computed Orbits Is Established by Means of a Shadowing Theorem

Let $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ be a hyperbolic IFS of contractivity $0 < s < 1$. Let A denote the attractor of the IFS, and assume that $w_n : A \to A$ is invertible for each $n = 1, 2, \ldots, N$. If the IFS is totally disconnected, let $\{A; S\}$ denote the associated shift dynamical system; otherwise let $\{A; S\}$ denote the associated random shift dynamical system. Consider the following model for the inaccurate calculation of an orbit of a point $x_0 \in A$. This model will surely describe the reader's experiences in computing shift dynamics directly on pictures of fractals. Moreover, it is a reasonable model for the occurrence of numerical errors when machine computation is used to compute an orbit.

Let an exact orbit of the point $x_0 \in A$ be denoted by $\{x_n\}_{n=0}^{\infty}$, where $x_n = S^{\circ n}(x_0)$ for each n. Let an approximate orbit of the point $x_0 \in A$ be denoted by $\{\tilde{x}_n\}_{n=0}^{\infty}$ where $\tilde{x}_0 = x_0$. Then we suppose that at each step there is made an error of at most θ for some $0 \leq \theta < \infty$; that is,

$$d(\tilde{x}_{n+1}, S(\tilde{x}_n)) \leq \theta \qquad \text{for} n = 0, 1, 2, \ldots.$$

We proceed to analyze this model. It is clear that the inaccurate orbit $\{\tilde{x}_n\}_{n=0}^{\infty}$ will usually start out by diverging from the exact orbit $\{x_n\}_{n=0}^{\infty}$ at an exponential rate. It may well occur "accidentally" that $d(x_n, \tilde{x}_n)$ is small for various large values of n, due to the compactness of A. But typically, if $d(x_n, \tilde{x}_n)$ is small enough, then $d(x_{n+j}, \tilde{x}_{n+j})$ will again grow exponentially with increasing j. To be precise, suppose $d(\tilde{x}_1, S(\tilde{x}_0)) = \theta$ and that we make no further errors. Suppose also that for some integer M, and some integers $\sigma_1, \sigma_2, \ldots, \sigma_M \in \{1, 2, \ldots, N\}$, we have

$$\tilde{x}_n \text{ and } x_n \in w_{\sigma_n}(A), \text{ for } n = 0, 1, 2, \ldots, M.$$

Moreover, suppose that

$$x_{n+1} = w_{\sigma_n}^{-1}(x_n) \text{ and } \tilde{x}_{n+1} = w_{\sigma_n}^{-1}(\tilde{x}_n), \text{ for } n = 0, 1, 2, \ldots, M.$$

Then we have

$$d(x_{n+1}, \tilde{x}_{n+1}) \geq s^{-n}\theta, \text{ for } n = 0, 1, 2, \ldots, M.$$

For some integer $J > M$ it is likely to be the case that

$$x_{J+1} = w_{\sigma_J}^{-1}(x_n) \text{ and } \tilde{x}_{n+1} = w_{\tilde{\sigma}_J}^{-1}(\tilde{x}_n), \text{ for some } \sigma_J \neq \tilde{\sigma}_J.$$

Then, without further assumptions, we cannot say anything more about the correlation between the exact orbit and the approximate orbit. Of course, we always have the error bound

$$d(x_n, \tilde{x}_n) \leq \text{diam}(A) = \max\{d(x, y) : x \in A, y \in A\}, \text{ for all } n = 1, 2, 3, \ldots.$$

Do the above comments make the situation hopeless? Are all of the calculations of shift dynamics we have done in this chapter without point because they are riddled with errors? No! The following wonderful theorem tells us that however many errors we make, there is an exact orbit that lies at every step within a small distance of our errorful one. This orbit shadows the errorful orbit. This type of theorem is extremely important in dynamics, and in any class of dynamical systems that has one (such as IFS) behavior that can be accurately analyzed using graphics on computers. Here we are use the word "shadows" in the sense of a secret agent who shadows a spy. The agent is always just out of sight, not too far away, usually not too close, but forever he follows the spy.

Theorem 7.1 The Shadowing Theorem. *Let $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ be a hyperbolic IFS of contractivity s, where $0 < s < 1$. Let A denote the attractor of the IFS and suppose that each of the transformations $w_n : A \to A$ is invertible. Let $\{A; S\}$ denote the associated shift dynamical system in the case that the IFS is totally* *disconnected; otherwise let* $\{A; S\}$ *denote the associated random shift dynamical system. Let* $\{\tilde{x}_n\}_{n=0}^{\infty} \subset A$ *be an approximate orbit of* S*, such that*

$$d(\tilde{x}_{n+1}, S(\tilde{x}_n)) \leq \theta \qquad \text{for all } n = 0, 1, 2, 3, \ldots,$$

for some fixed constant θ *with* $0 \leq \theta \leq \text{diam}(A)$. *Then there is an exact orbit* $\{x_n = S^{\circ n}(x_0)\}_{n=0}^{\infty}$ *for some* $x_0 \in A$*, such that*

$$d(\tilde{x}_{n+1}, x_{n+1}) \leq \frac{s\theta}{(1-s)} \qquad \text{for all } n = 0, 1, 2, \ldots.$$

Proof As usual we exploit code space! For $n = 1, 2, 3, \ldots,$ let $\sigma_n \in \{1, 2, \ldots, N\}$ be chosen so that $w_{\sigma_1}^{-1}, w_{\sigma_2}^{-1}, w_{\sigma_3}^{-1}, \ldots,$ is the actual sequence of inverse maps used to compute $S(\tilde{x}_0), S(\tilde{x}_1), S(x_2), \ldots$. Let $\phi : \Sigma \to A$ denote the code space map associated with the IFS. Then define

$$x_0 = \phi(\sigma_1\sigma_2\sigma_3\ldots).$$

Then we compare the exact orbit of the point x_0,

$$\{x_n = S^{\circ n}(x_0) = \phi(\sigma_{n+1}\sigma_{n+2}\ldots)\}_{n=0}^{\infty}$$

with the errorful orbit $\{\tilde{x}_n\}_{n=0}^{\infty}$.

Let M be a large positive integer. Then, since x_M and $S(\tilde{x}_{M-1})$ both belong to A, we have

$$d(S(x_{M-1}), S(\tilde{x}_{M-1}) \leq \text{diam}(A) < \infty.$$

Since $S(x_{M-1})$ and $S(\tilde{x}_{M-1})$ are both computed with the same inverse map $w_{\sigma_M}^{-1}$ it follows that

$$d(x_{M-1}, \tilde{x}_{M-1}) \leq s\,\text{diam}(A).$$

Hence

$$\begin{aligned} d(S(x_{M-2}), S(\tilde{x}_{M-2})) &= d(x_{M-1}, S(\tilde{x}_{M-2})) \\ &\leq d(x_{M-1}, \tilde{x}_{M-1}) + d(\tilde{x}_{M-1}, S(\tilde{x}_{M-2})) \\ &\leq \theta + s\,\text{diam}(A); \end{aligned}$$

and repeating the argument used above we now find

$$d(x_{M-2}, \tilde{x}_{M-2})) \leq s(\theta + s\,\text{diam}(A)).$$

Repeating the same argument k times we arrive at

$$d(x_{M-k}, \tilde{x}_{M-k}) \leq s\theta + s^2\theta + \cdots + s^{k-1}\theta + s^k\,\text{diam}(A).$$

Hence for any positive integer M and any integer n such that $0 < n < M$, we have

$$d(x_n, \tilde{x}_n) \leq s\theta + s^2\theta + \cdots + s^{M-n-1}\theta + s^{M-n}\,\text{diam}(A).$$

Now take the limit of both sides of this equation as $M \to \infty$ to obtain

$$d(x_n, \tilde{x}_n) \leq s\theta(1 + s + s^2 + \cdots) = \frac{s\theta}{(1-s)}, \qquad \text{for all } n = 1, 2, \ldots.$$

This completes the proof.

Examples & Exercises

7.1. Let us apply the Shadowing Theorem to an orbit on the Sierpinski triangle, using the random shift dynamical system associated with the IFS

$$\{C; \frac{1}{2}z, \frac{1}{2}z + \frac{1}{2}, \frac{1}{2}z + \frac{i}{2}\}.$$

Since the system is just-touching we must assign values to the shift transformation applied to the just-touching points. We do this by defining

$$S(x_1 + ix_2) = 2x_1 \bmod 1 + i(2x_2 \bmod 1).$$

We consider the orbit of the point $\tilde{x}_0 = (0.2147, 0.0353)$. We compute the first 11 points on the exact orbit of this point, and compare it to the results obtained when a deliberate error $\theta = 0.0001$ is introduced at each step. We obtain:

Errorful	*Exact*
$\tilde{x}_0 = (0.2147, 0.0353)$	$S^{\circ 0}(\tilde{x}_0) = (0.2147, 0.0353)$
$\tilde{x}_1 = (0.4295, 0.0705)$	$S^{\circ 1}(\tilde{x}_0) = (0.4294, 0.0706)$
$\tilde{x}_2 = (0.8591, 0.1409)$	$S^{\circ 2}(\tilde{x}_0) = (0.8588, 0.1412)$
$\tilde{x}_3 = (0.7183, 0.2817)$	$S^{\circ 3}(\tilde{x}_0) = (0.7176, 0.2824)$
$\tilde{x}_4 = (0.4365, 0.5635)$	$S^{\circ 4}(\tilde{x}_0) = (0.4352, 0.5648)$
$\tilde{x}_5 = (0.8731, 0.1269)$	$S^{\circ 5}(\tilde{x}_0) = (0.8704, 0.1296)$
$\tilde{x}_6 = (0.7463, 0.2537)$	$S^{\circ 6}(\tilde{x}_0) = (0.7408, 0.2592)$
$\tilde{x}_7 = (0.4927, 0.5073)$	$S^{\circ 7}(\tilde{x}_0) = (0.4816, 0.5184)$
$\tilde{x}_8 = (0.9855, 0.0145)$	$S^{\circ 8}(\tilde{x}_0) = (0.9632, 0.0368)$
$\tilde{x}_9 = (0.9711, 0.0289)$	$S^{\circ 9}(\tilde{x}_0) = (0.9264, 0.0736)$
$\tilde{x}_{10} = (0.9423, 0.0577)$	$S^{\circ 10}(\tilde{x}_0) = (0.8528, 0.1472)$

Notice how the orbit with errors diverges from the exact orbit of $\tilde{x}_0$. Nonetheless, the shadowing theorem asserts that there is an *exact* orbit $\{x_n\}$ such that

$$d(x_n, \tilde{x}_n) \leq \frac{\frac{1}{2}}{1 - \frac{1}{2}}(0.0001) = 0.0001,$$

where $d(\cdot, \cdot)$ denotes the Manhattan metric. This really *seems* unlikely; but it must be true! Here's an example of such a shadowing orbit, also computed exactly.

Exact Shadowing Orbit $x_n = S^{\circ n}(x_0)$	$d(x_n, \tilde{x}_n) \leq 0.0001$
$x_0 = (0.21478740234375, 0.03521259765625)$	0.00009
$x_1 = (0.4295748046875, 0.0704251953125)$	0.00008
$x_2 = (0.8591496093750, 0.1408503906250)$	0.00005
$x_3 = (0.7182992187500, 0.2817007812500)$	0.000001
$x_4 = (0.4365984375000, 0.5634015625000)$	0.0001
$x_5 = (0.8731968750000, 0.1268031250000)$	0.0001
$x_6 = (0.7463937500000, 0.2536062500000)$	0.0001
$x_7 = (0.4927875000000, 0.5072125000000)$	0.00009
$x_8 = (0.9855750000000, 0.0144250000000)$	0.00008

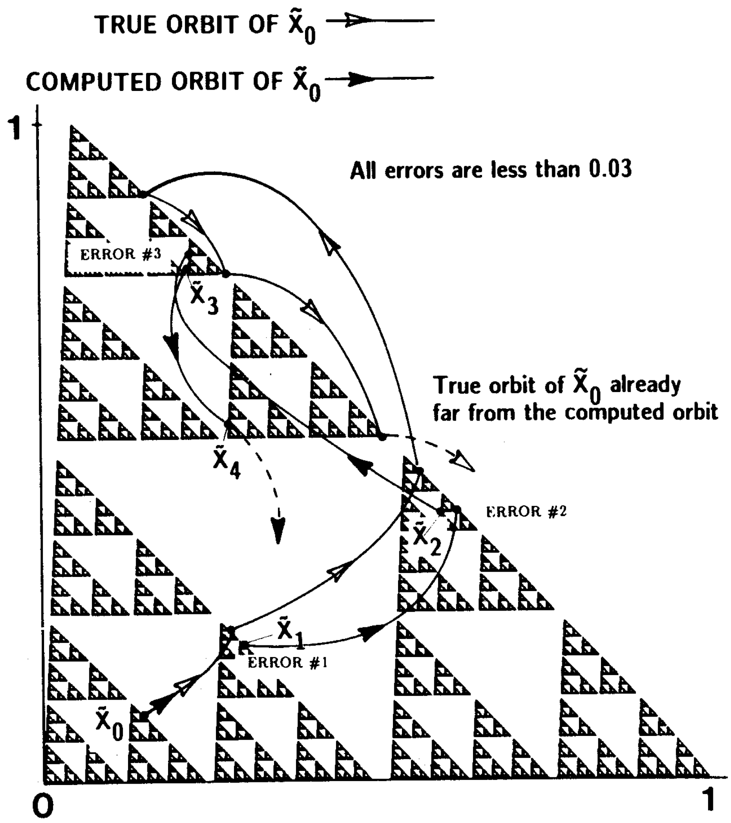


Figure IV.125. The Shadowing Theorem tells us there is an exact orbit closer to $\{\tilde{x}_n\}$ than 0.03 for all n.

$x_9 = (0.9711500000000, 0.0288500000000)$	0.00005
$x_{10} = (0.9423000000000, 0.0577000000000)$	0.000000

Figure IV.125 illustrates the idea.

7.2. Consider the shift dynamical system $\{\Sigma; T\}$ on the code space of two symbols $\{1, 2\}$. Show that the sequence of points $\{\tilde{x}_n\}$ given by

$$\tilde{x}_0 = 21\overline{2}, \text{ and } \tilde{x}_n = 1\overline{2} \qquad \text{for all } n = 1, 2, 3, \ldots$$

is an errorful orbit for the system. Illustrate the divergence of $T^{\circ n}\tilde{x}_0$ from $\tilde{x}_n$. Find a shadowing orbit $\{x_n\}_{n=0}^{\infty}$ and verify the error estimate provided by the Shadowing Theorem.

7.3. Illustrate the Shadowing Theorem by constructing an erroneous orbit, and an orbit that shadows it, for the shift dynamical system $\{[0, 1]; \frac{1}{3}x, \frac{1}{2}x + \frac{1}{2}\}$.

7.4. Compute an orbit for a random shift dynamical system associated with the overlapping IFS $\{[0, 1]; \frac{3}{4}x, \frac{1}{2}x + \frac{1}{2}\}$.

7.5. An orbit of the shift dynamical system associated with the IFS

$$\left\{\mathbb{R}^2; \frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}, \frac{3}{4}\begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \frac{1}{8}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 7 \end{pmatrix}\right\},$$

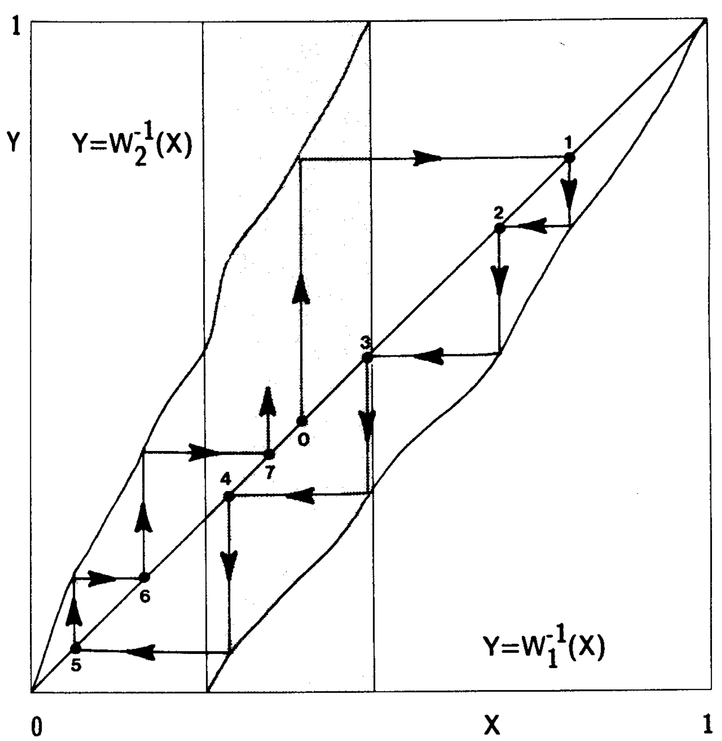


Figure IV.126. An exact orbit shadows the orbit "computed" by "drawing" in this web diagram for a random shift dynamical system.

is computed to accuracy 0.0005. How close a shadowing orbit does there exist? Use the Manhattan metric.

7.6. In Figure IV.126 an orbit of the random shift dynamical system associated with the overlapping IFS $\{[0, 1], w_1(x), w_2(x)\}$ is computed by drawing a web diagram. The computer in this case consists of a pencil and a drafting table. Estimate the errors in the drawing and then deduce how closely an exact orbit shadows the plotted one. You will need to estimate the contractivity of the IFS. Also draw a tube around the plotted orbit, within which an exact orbit lies.

7.7. Figure IV.127 shows an orbit $\{x_n\}$ of the random shift dynamical system associated with the IFS $\{[0, 1]; w_1(x), w_2(x)\}$. It was obtained by defining $S(x) = w_2^{-1}(x)$ for $x \in w_1(A) \cap w_2(A)$. A contractivity factor for the IFS is readily estimated from the drawing to be $\frac{3}{5}$. Hence if the web diagram is accurate to within 1 mm at each iteration, that is

$$d(\tilde{x}_{n+1}, S(\tilde{x}_n)) \leq 1\,\text{mm},$$

then there is an exact orbit $\{x_n = S^{\circ n}(x_0)\}_{n=0}^{\infty}$ such that

$$d(x_n, \tilde{x}_n) \leq \frac{(\frac{3}{5})}{(\frac{2}{5})} = 1.5\,\text{mm}.$$

Thus there is an actual orbit that remains within the "orbit tube" shown in Figure IV.127.

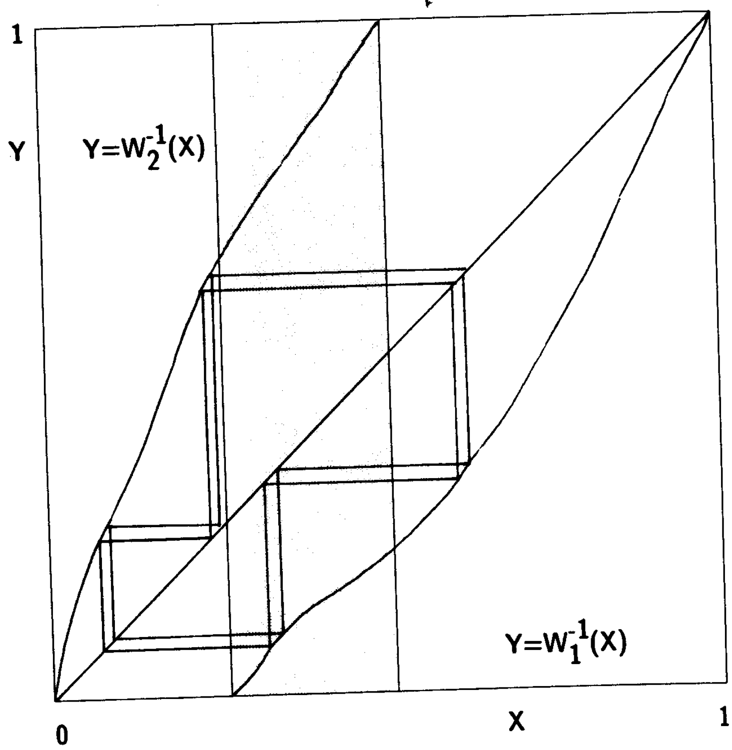


Figure IV.127. Only the *Shadow* knows. Inside the "orbit tube" there is an *exact* orbit $\{x_n\}_{n=0}^{\infty}$ of the random shift dynamical system associated with the IFS.

8 Chaotic Dynamics on Fractals

The shift dynamical system $\{A; S\}$ associated with a totally disconnected hyperbolic IFS is equivalent to the shift dynamical system $\{\Sigma, T\}$, where Σ is the code space associated with the IFS. As we have seen, this equivalence means that the two systems have a number of properties in common; for example, the two systems have the same number of cycles of minimal period 7. A particularly important property that they share is that they are both "chaotic" dynamical systems, a concept that we explain in this section. First, however, we want to underline that the two systems are deeply different from the point of view of the interplay of their dynamics with the geometry of the underlying spaces.

Consider the case of an IFS of three transformations. Let Σ denote the code space of the three symbols $\{1, 2, 3\}$, and look at the orbit of the point $\sigma \in \Sigma$ given by

$\sigma =$

1231112132122233132331111
121131211221231311321332112122 13
22122222323123223331131231332132
23233313323331111112111311211 12
21123113111321133121112121213122
11222122312311232123313
1113121212............**FOREVER.**

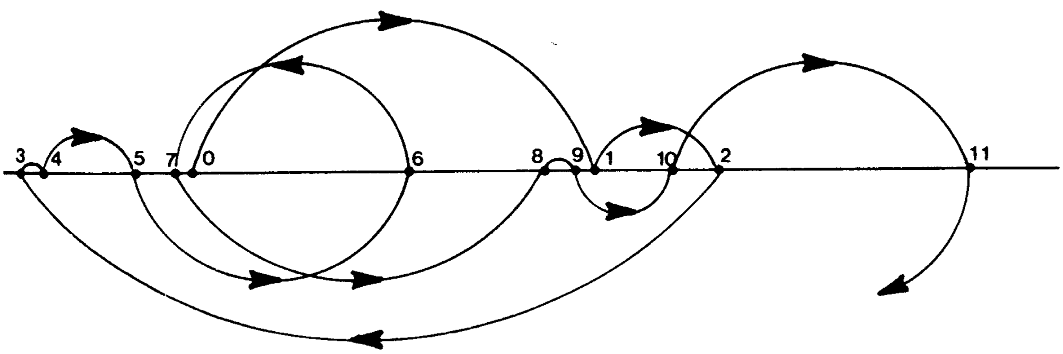


Figure IV.128. The start of a chaotic orbit on a Ternary Cantor set.

This orbit $\{T^{\circ n}\sigma\}_{n=0}^{\infty}$ may be plotted on a Cantor set of three symbols, as sketched in Figure IV.128. This can be compared with the orbit $\{S^{\circ n}(\phi(\sigma))\}_{n=0}^{\infty}$ of the shift dynamical system $\{A, S\}$ associated with an IFS of three maps, as plotted in Figure IV.129. Figure IV.130 shows an equivalent orbit, but this time for the just-touching IFS $\{[0, 1]; \frac{1}{3}x, \frac{1}{3}x + \frac{1}{3}, \frac{1}{3}x + \frac{2}{3}\}$, displayed using a web diagram.

In each case the "same" dynamics look entirely different. The qualities of beauty and harmony present in the observed orbits are different. This is not suprising: the equivalence of the dynamical systems is a topological equivalence. It does not provide much information about the interplay of the dynamics with the geometries of the spaces on which they act. This interplay is an open area for research. For example, what are the special conserved properties of two metrically equivalent dynamical systems? Can you quantify the grace and delicacy of a dancing orbit on a fractal?

This said, we turn our attention back to an important collection of properties shared by all shift dynamical systems. For simplicity we formalize the discussion for the case of the shift dynamical system $\{A, S\}$ associated with a totally disconnected hyperbolic IFS.

Definition 8.1 *Let $(\mathbf{X}, d)$ be a metric space. A subset $B \subset \mathbf{X}$ is said to be* dense *in $\mathbf{X}$ if the closure of B equals $\mathbf{X}$. A sequence $\{x_n\}_{n=0}^{\infty}$ of points in $\mathbf{X}$ is said to be dense in $\mathbf{X}$ if, for each point $a \in \mathbf{X}$, there is an subsequence $\{x_{\sigma_n}\}_{n=0}^{\infty}$ that converges to a. In particular an orbit $\{x_n\}_{n=0}^{\infty}$ of a dynamical system $\{\mathbf{X}, f\}$ is said to be dense in $\mathbf{X}$ if the sequence $\{x_n\}_{n=0}^{\infty}$ is dense in $\mathbf{X}$.*

By now you will have had some experience with using the random iteration algorithm, Program 2 of Chapter III, for computing images of the attractor A of IFS in $\mathbb{R}^2$. If you run the algorithm starting from a point $x_0 \in A$, then all of the computed points lie on A. Apparently, the sequences of points we plot are examples of sequences that are dense in the metric space (A, d).

The property of being dense is invariant under homeomorphism : if B is dense in a metric space $(\mathbf{X}, d)$ and if $\theta : \mathbf{X} \to \mathbf{Y}$ is a homeomorphism, then $\theta(B)$ is dense in $\mathbf{Y}$. If $\{\mathbf{X}; f\}$ and $\{\mathbf{Y}, g\}$ are equivalent dynamical systems under θ; and if $\{x_n\}$ is an orbit of f dense in $\mathbf{X}$, then $\{\theta(x_n)\}$ is an orbit of g dense in $\mathbf{Y}$.

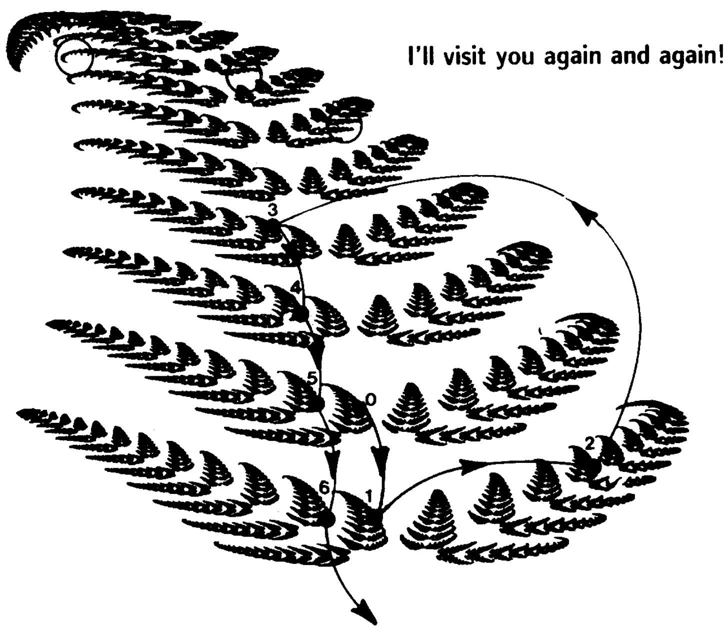


Figure IV.129. The start of an orbit of a deterministic shift dynamical system. This orbit is chaotic. It will visit the part of the attractor inside each of these little circles infinitely many times.

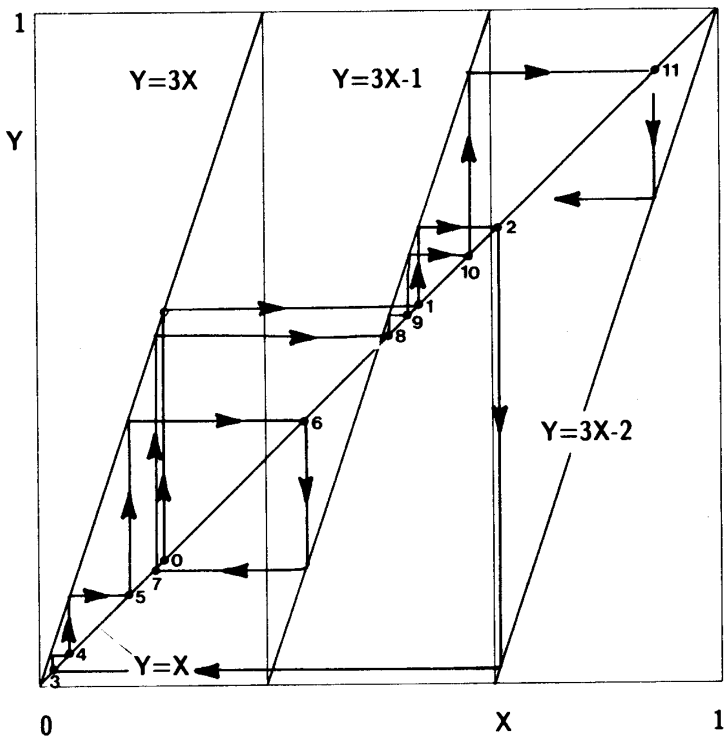


Figure IV.130. Equivalent orbit to the one in Figures IV.128 and IV.129, this time ploted using a web diagram. The starting point has address 12311121321222331.... This manifestation of an orbit, which goes arbitrarily close to any point, takes place on a just-touching attractor.

Definition 8.2 *A dynamical system* $\{\mathbf{X}, f\}$ *is* transitive *if, whenever* $\mathcal{U}$ *and* $\mathcal{V}$ *are open subsets of the metric space* $(\mathbf{X}, d)$, *there exists a finite integer n such that*

$$\mathcal{U} \cap f^{\circ n}(\mathcal{V}) \neq \emptyset.$$

The dynamical system $\{[0, 1]; f(x) = \min\{2x, 2 - 2x\}\}$ is topologically transitive. To verify this just let $\mathcal{U}$ and $\mathcal{V}$ be any pair of open intervals in the metric space ([0, 1], Euclidean). Clearly, each application of the transformation increases the length of the interval $\mathcal{U}$ in such a way that it eventually overlaps $\mathcal{V}$.

Definition 8.3 *The dynamical system* $\{\mathbf{X}; f\}$ *is* sensitive to initial conditions *if there exists* $\delta > 0$ *such that, for any* $x \in \mathbf{X}$ *and any ball* $B(x, \epsilon)$ *with radius* $\epsilon > 0$, *there is* $y \in B(x, \epsilon)$ *and an integer* $n \geq 0$ *such that* $d(f^{\circ n}(x), f^{\circ n}(y)) > \delta$.

Roughly, orbits that begin close together get pushed apart by the action of the dynamical system. For example, the dynamical system $\{[0, 1]; 2x \bmod 1\}$ is sensitive to initial conditions.

Examples & Exercises

8.1. Show that the rational numbers are dense in the metric space ($\mathbb{R}$, Euclidean).

8.2. Let $C(n)$ be a counting function that counts all of the rational numbers that lie in the interval [0, 1]. Let $r_{C(n)}$ denote the nth rational number in [0, 1]. Prove that the sequence of real numbers $\{r_{C(n)} \in [0, 1] : n = 1, 2, 3, ...\}$ is dense in the metric space ([0, 1], Euclidean).

8.3. Consider the dynamical system $\{[0, 1]; f(x) = 2x \bmod 1\}$. Find a point $x_0 \in [0, 1]$ whose orbit is dense in [0, 1].

8.4. Show that the dynamical system $\{[0, \infty) : f(x) = 2x\}$ is sensitive to initial conditions, but that the dynamical system $\{[0, \infty) : f(x) = (0.5)x\}$ is not.

8.5. Show that the shift dynamical system $\{\Sigma; T\}$, where Σ is the code space of two symbols, is transitive and sensitive to initial conditions.

8.6. Let $\{\mathbf{X}, f\}$ and $\{\mathbf{Y}, g\}$ be equivalent dynamical systems. Show that $\{\mathbf{X}, f\}$ is transitive if and only if $\{\mathbf{Y}, g\}$ is transitive. In other words, the property of being transitive is preserved between equivalent dynamical systems.

Definition 8.4 *A dynamical system* $\{\mathbf{X}, f\}$ *is* chaotic *if*

(1) it is transitive;
(2) it is sensitive to initial conditions;
(3) the set of periodic orbits of f is dense in $\mathbf{X}$.

Theorem 8.1 *The shift dynamical system associated with a totally disconnected hyperbolic IFS of two or more transformations is chaotic.*

Sketch of Proof: First one establishes that the shift dynamical system $\{\Sigma; T\}$ is chaotic where Σ is the code space of N symbols, with $N \geq 2$. One then uses the code space map $\phi : \Sigma \to A$ to carry the results over to the equivalent dynamical system $\{A; S\}$.

Theorem 1 applies to the lifted IFS associated with a hyperbolic IFS. Hence the lifted shift dynamical system associated with an IFS of two or more transformations is chaotic. In turn this implies certain characteristics to the behavior of the projection of a lifted shift dynamical system, namely a random shift dynamical system.

Let us consider now why the random iteration algorithm works, from an intuitive point of view. Consider the hyperbolic IFS $\{\mathbb{R}^2; w_1, w_2\}$. Let $a \in A$; suppose that the address of a is $\sigma \in \Sigma$, the associated code space. That is

$$a = \phi(\sigma).$$

With the aid of a random-number generator, a sequence of one million ones and twos is selected. For example, suppose that the the actual sequence produced is the following one, *which has been written from right to left*,

$$21\ldots 1212112112121112111211111211211121111211212122211$$

By this we mean that the first number chosen is a 1, then a 1, then three 2's, and so on. Then the following sequence of points on the attractor is computed:

$$a = \phi(\sigma)$$

$$w_1(a) = \phi(1\sigma)$$

$$w_1 \circ w_1(a) = \phi(11\sigma)$$

$$w_2 \circ w_1 \circ w_1(a) = \phi(211\sigma)$$

$$w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(2211\sigma)$$

$$w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(22211\sigma)$$

$$w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(122211\sigma)$$

$$w_2 \circ w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(2122211\sigma)$$

$$w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(12122211\sigma)$$

$$w_2 \circ w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(212122211\sigma)$$

$$w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(1212122211\sigma)$$

$$w_1 \circ w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(11212122211\sigma)$$

$$\vdots$$

$$w_2 \circ w_1 \circ \ldots w_1 \circ w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_1 \circ w_2 \circ w_2 \circ w_2 \circ w_1 \circ w_1(a) = \phi(21\ldots 11212122211\sigma)$$

We imagine that instead of plotting the points as they are computed, we keep a list of the one million computed points. This done, we plot the points in the reverse order from the order in which they were computed. That is, we begin by plotting the point $\phi(21\ldots 11212122211\sigma)$ and we finish by plotting the point $\phi(\sigma)$. What will we see? We will see a million points on the orbit of the shift dynamical system $\{A; S\}$; namely, $\{S^{\circ n}(\phi(21\ldots 11212122211\sigma))\}_{n=0}^{1,000,000}$.

Now from our experience with shift dynamics and from our theoretical knowledge and intuitions what do we expect of such an orbit? We expect it to be *chaotic* and to visit a widely distributed collection of points on the attractor. We are looking at part of a "randomly chosen" orbit of the shift dynamical system; we expect it to be dense in the attractor.

For example, suppose that you are doing shift dynamics on a picture of a totally disconnected fractal, or a fern. You should be convinced that by making sly adjustments in the orbit at each step, as in the Shadowing Theorem, you can most easily coerce an orbit into visiting, to within a distance $\epsilon > 0$, each point in the image. But then the Shadowing Theorem ensures that there is an actual orbit close to our artificial one, and it too goes close to every point on the fractal, say to within a distance of 2ϵ of each point on the image. This suggests that "most" orbits of the shift dynamical system are dense in the attractor.

Examples & Exercises

8.7. Make experiments on a picture of the attractor of a totally disconnected hyperbolic IFS to verify the assertion in the last paragraph that "by making sly adjustments in an orbit ... you can most easily coerce the orbit into visiting to within a distance $\epsilon > 0$ of each point in the image." Can you make some experimental estimates of how many orbits go to within a distance $\epsilon > 0$, for several values of ϵ, of every point in the picture? One way to do this might be to work with a discretized image and to try to count the number of available orbits.

8.8. Run the Random Iteration Algorithm, Program 2 in Chapter III, to produce an image of a fractal, for example a fern without a stem as used in Figure IV.129. As the points are calculated and plotted, keep a list of them. Then plot the points over again in reverse order, this time making them flash on and off on the picture of the attractor on the screen, so that you can see where they land. This way you will see the interplay of the geometry with the shift dynamics on the attractor. See if the orbit is beautiful. If you think that it is, try to make your impression objective.

We want to begin to formulate the idea that "most" orbits of the shift dynamical system associated with a totally disconnected IFS are dense in the attractor. The following lemma counts the number of cycles of minimal period p.

Lemma 8.1 *Let $\{A; S\}$ be the shift dynamical system associated with a totally disconnected hyperbolic IFS $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$. Let $\mathcal{N}(p)$ denote the number of distinct cycles of minimal period p, for $p \in \{1, 2, 3, \ldots\}$. Then*

$$\mathcal{N}(p) = \left(N^p - \sum_{\substack{k=1 \\ k \text{ divides } p}}^{p-1} k\mathcal{N}(k) \right) / p \qquad \text{for } p = 1, 2, 3, \ldots.$$

Proof It suffices to restrict attention to code space, and to give the main idea, consider only the case $N = 2$. For $p = 1$, the cycles of period 1 are the fixed points of T. The equation

$$T\sigma = \sigma \sigma \in \Sigma$$

implies $\sigma = \overline{1111}$ or $\sigma = \overline{2222}$. Thus $\mathcal{N}(1) = 2$. For $p = 2$, any point that lies on a cycle of period 2 must be a fixed point of $T^{\circ 2}$, namely

$$T^{\circ 2}\sigma = \sigma,$$

where $\sigma = \overline{11}, \overline{12}, \overline{21}$, or $\overline{22}$. The only cycles here that are not of minimal period 2 must have minimal period 1. Furthermore, there are two distinct points on a cycle of minimal period 2, so

$$\mathcal{N}(2) = (2^2 - \mathcal{N}(1))/2 = 2/2 = 1.$$

One quickly gets the idea. Mathematical induction on p completes the proof for $N = 2$.

For $N = 2$, we find, for example, $\mathcal{N}(2) = 1$, $\mathcal{N}(3) = 2$, $\mathcal{N}(4) = 3$, $\mathcal{N}(5) = 6$, $\mathcal{N}(6) = 9$, $\mathcal{N}(7) = 18$, $\mathcal{N}(8) = 30$, $\mathcal{N}(9) = 56$, $\mathcal{N}(10) = 99$, $\mathcal{N}(11) = 186$, $\mathcal{N}(12) = 335$, $\mathcal{N}(13) = 630$, $\mathcal{N}(14) = 1161$, $\mathcal{N}(15) = 2182$, $\mathcal{N}(16) = 4080$, $\mathcal{N}(17) = 7710$, $\mathcal{N}(18) = 14532$, $\mathcal{N}(19) = 27594$, $\mathcal{N}(20) = 52377$. In particular, 99.9% of all points lying on cycles of period 20 lie on cycles of minimal period 20.

Here is the idea we are getting at. We know that the set of periodic cycles are dense in the attractor of a hyperbolic IFS. It follows that we may approximate the attractor by the set of all cycles of some finite period, say period 12 billion. Thus we replace the attractor A by such an approximation $\tilde{A}$, which consists of $2^{12{,}000{,}000{,}000}$ points. Suppose we pick one of these points at random. Then this point is extremely likely to lie on a cycle of *minimal* period 12 billion. Hence the orbit of a point chosen "at random" on the approximate attractor $\tilde{A}$ is extremely likely to consist of 12 billion *distinct* points on A.

In fact one can show that a statistically random sequence of symbols contains every possible finite subsequence. So we expect that the set of 12 billion distinct points on A is likely to contain at least one representative from each part of the attractor!