3.4 THE HÉNON MAP

About 15 years ago the French astronomer-mathematician Michel Hénon was searching for a simple two-dimensional function possessing special properties of more complicated systems. The result was a family of functions denoted by H_{ab} and given by

$$H_{ab}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix}, \text{ where } a \text{ and } b \text{ are real numbers} \qquad (1)$$

The maps defined in (1) are called **Hénon maps**. Many authors refer to the Hénon maps as H, and just call them the Hénon map.

Notice that if $b = 1$, $x = t$ and $y = 0$, then (1) becomes

$$H_{ab}\begin{pmatrix} t \\ 0 \end{pmatrix} = \begin{pmatrix} 1 - at^2 \\ t \end{pmatrix}$$

Thus the image of the real line is the parabola given parametrically by $x = 1 - at^2$ and $y = t$ (Figure 3.9). As a result, the Hénon maps constitute a 2-dimensional generalization of the family $\{F_C\}$ mentioned in Section 2.3, where $F_C(x) = 1 - Cx^2$. Next we will find the Jacobian of H_{ab}.

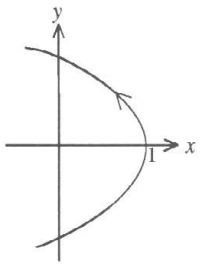


Figure 3.9

THEOREM 3.19. Let a and b be any fixed real numbers. Then $\det DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix} = -b$ for all x, y in R^2. If $a^2x^2 + b \geq 0$, then the eigenvalues of $DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix}$ are the real numbers $-ax \pm \sqrt{a^2x^2 + b}$.

Proof. Since the coordinate functions of H_{ab} are given by

$$f\begin{pmatrix} x \\ y \end{pmatrix} = 1 - ax^2 + y \quad \text{and} \quad g\begin{pmatrix} x \\ y \end{pmatrix} = bx$$

we find that

$$DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2ax & 1 \\ b & 0 \end{pmatrix}$$

so that

$$\det DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix} = \det\begin{pmatrix} -2ax & 1 \\ b & 0 \end{pmatrix} = -b$$

To determine the eigenvalues of $DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix}$ we observe that

$$\det\left(DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix} - \lambda I\right) = \det\begin{pmatrix} -2ax - \lambda & 1 \\ b & -\lambda \end{pmatrix} = \lambda^2 + 2ax\lambda - b$$

Therefore λ is an eigenvalue of $DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix}$ if $\lambda^2 + 2ax\lambda - b = 0$. This means that

$$\lambda = \frac{-2ax \pm \sqrt{4a^2x^2 + 4b}}{2} = -ax \pm \sqrt{a^2x^2 + b}$$

Thus the eigenvalues are real if $a^2x^2 + b \geq 0$. ■

The map H_{ab} has a constant Jacobian $\det DH_{ab}$. Hénon noted in his original paper (1976) that H_{ab} is the "most general quadratic mapping [on R^2] with constant Jacobian." Next, recall that the Jacobian of H_{ab} determines whether H_{ab} is area-expanding or area-contracting (or neither). By Theorem 3.19, the map H_{ab} is area-contracting if $0 \leq b < 1$; it is genuinely a two-dimensional map if $b \neq 0$. Thus we will henceforth assume $0 < b < 1$. For such values of b, $H_{ab}(\mathbf{v})$ has distinct real eigenvalues for every value of the parameter a, and all $\mathbf{v}$.

It is straightforward to show that H_{ab} is one-to-one.

THEOREM 3.20. H_{ab} is one-to-one.

Proof. Let x, y, z, and w be real numbers. Then

$$H_{ab}\begin{pmatrix} x \\ y \end{pmatrix} = H_{ab}\begin{pmatrix} z \\ w \end{pmatrix} \text{ if and only if } \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix} = \begin{pmatrix} 1 - az^2 + w \\ bz \end{pmatrix}$$

that is,

$$1 - ax^2 + y = 1 - az^2 + w \quad \text{and} \quad bx = bz$$

Since $b \neq 0$, it follows that $x = z$. Therefore $y = w$ as well, so that $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} z \\ w \end{pmatrix}$

Consequently H_{ab} is one-to-one. ■

The map H_{ab} is composed of the three functions H_1, H_2, and H_3, where

$$H_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 1 - ax^2 + y \end{pmatrix}, \quad H_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} bx \\ y \end{pmatrix}, \quad \text{and} \quad H_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$$

More precisely, $H_{ab} = H_3 \circ H_2 \circ H_1$ (where for convenience we suppress the subscript ab on H_1, H_2, and H_3). To interpret H_1, H_2, and H_3 geometrically, suppose that $a > 1$. Then H_1 begins the folding process. The effect of H_1 on the ellipse in Figure 3.10(a) is shown in Figure 3.10(b). Next, H_2 contracts curves in the x direction, since $0 < b < 1$ by hypothesis (Figure 3.10(c)). The folding started by H_1 is enhanced by H_2. Finally, H_3 flips shapes across the line $y = x$. The total effect of H_1, H_2, and H_3 (that is, of H_{ab}) on an ellipse is shown in Figure 3.10(d).

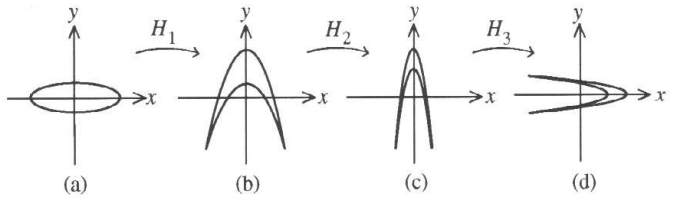


Figure 3.10

Now we will prove that H_{ab} is invertible.

THEOREM 3.21. H_{ab} is invertible, and $H_{ab}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{b}\, y \\ -1 + \frac{a}{b^2}\, y^2 + x \end{pmatrix}$.

Proof. We could show that H_1, H_2, and H_3 are invertible, and then that

$$H_{ab}^{-1} = (H_3 \circ H_2 \circ H_1)^{-1} = H_1^{-1} \circ H_2^{-1} \circ H_3^{-1}$$

It is easier to show that by the formula for H_{ab}^{-1}, we have $H_{ab} \circ H_{ab}^{-1} = I$:

$$(H_{ab} \circ H_{ab}^{-1})\begin{pmatrix} x \\ y \end{pmatrix} = H_{ab}\begin{pmatrix} \frac{1}{b}\, y \\ -1 + \frac{a}{b^2}\, y^2 + x \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \text{ for all } x \text{ and } y \quad ■$$

Next we determine the values of a and b for which H_{ab} has fixed points.

THEOREM 3.22. Let $a \neq 0$. Then H_{ab} has a fixed point if $a \geq -\frac{1}{4}(1 - b)^2$.

Proof. The point $\begin{pmatrix} x \\ y \end{pmatrix}$ is a fixed point of H_{ab} provided that

$$\begin{pmatrix} x \\ y \end{pmatrix} = H_{ab}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix}$$

The left-hand and right-hand vectors are equal if $y = bx$ and $x = 1 - ax^2 + y$, which implies that $x = 1 - ax^2 + bx$. This is equivalent to $ax^2 + (1 - b)x - 1 = 0$, which by the quadratic formula yields

$$x = \frac{1}{2a}\left(b - 1 \pm \sqrt{(1 - b)^2 + 4a}\right)$$

Such an x exists if $(1 - b)^2 + 4a \geq 0$, that is, if $a \geq -\frac{1}{4}(1 - b)^2$. ■

In the event that H_{ab} has two fixed points $\mathbf{p}$ and $\mathbf{q}$, they are given by

$$\mathbf{p} = \begin{pmatrix} \frac{1}{2a}\left(b - 1 + \sqrt{(1-b)^2 + 4a}\right) \\ \frac{b}{2a}\left(b - 1 + \sqrt{(1-b)^2 + 4a}\right) \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} \frac{1}{2a}\left(b - 1 - \sqrt{(1-b)^2 + 4a}\right) \\ \frac{b}{2a}\left(b - 1 - \sqrt{(1-b)^2 + 4a}\right) \end{pmatrix} \tag{2}$$

Since we know the fixed points of H_{ab} and the eigenvalues of $DH_{ab}\begin{pmatrix} x \\ y \end{pmatrix}$ for all x and y, we can determine conditions under which the fixed point $\mathbf{p}$ is attracting.

THEOREM 3.23. The fixed point $\mathbf{p}$ is attracting provided that a is a nonzero number lying in the interval

$$J = \left(-\frac{1}{4}(1-b)^2, \frac{3}{4}(1-b)^2 \right)$$

Proof. Theorem 3.18 tells us that $\mathbf{p}$ is attracting if the eigenvalues of $DH_{ab}(\mathbf{p})$ are less than 1 in absolute value. Letting $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$, we know from (2) that

$$p_1 = \frac{1}{2a}\left(b - 1 + \sqrt{(1-b)^2 + 4a} \right) \tag{3}$$

so that $2ap_1 = b - 1 + \sqrt{(1-b)^2 + 4a}$. Therefore

$$2ap_1 > b - 1, \text{ or equivalently, } 2ap_1 + 1 > b \tag{4}$$

By Theorem 3.19 the eigenvalues of $DH_{ab}(\mathbf{p})$ are less than 1 in absolute value if $\left| -ap_1 \pm \sqrt{a^2p_1^2 + b} \right| < 1$. We will show that if a is in the interval J, then

$$0 \le -ap_1 + \sqrt{a^2p_1^2 + b} < 1 \tag{5}$$

On the one hand, because $b > 0$ we have

$$-ap_1 + \sqrt{a^2p_1^2 + b} \ge -ap_1 + \sqrt{a^2p_1^2} = -ap_1 + |ap_1| \ge 0$$

On the other hand, $a > -(1-b)^2/4$ by hypothesis, so that $(1-b)^2 + 4a > 0$. Consequently p_1 is a real number by (3), and by (4),

$$(ap_1 + 1)^2 = a^2p_1^2 + 2ap_1 + 1 > a^2p_1^2 + b > 0$$

It follows that $ap_1 + 1 > \sqrt{a^2p_1^2 + b}$, so that $-ap_1 + \sqrt{a^2p_1^2 + b} < 1$. Therefore (5) is proved. An analogous argument proves that

$$-1 < -ap_1 - \sqrt{a^2p_1^2 + b} < 0$$

(see Exercise 9). Consequently the eigenvalues of $DH_{ab}(\mathbf{p})$ are less than 1 in absolute value, so that $\mathbf{p}$ is an attracting fixed point. ■

From Theorem 3.23, the fixed point $\mathbf{p}$ is attracting for certain values of a. By contrast the fixed point $\mathbf{q}$ given in (2) is a saddle point (Exercise 8). Thus we have the following situation for a given value of b in $(0, 1)$:

If $a < -\frac{1}{4}(1-b)^2$, then H_{ab} has no fixed points.

If $-\frac{1}{4}(1-b)^2 < a < \frac{3}{4}(1-b)^2$ and $a \neq 0$, then H_{ab} has two fixed points, $\mathbf{p}$ and $\mathbf{q}$, of which $\mathbf{p}$ is attracting and $\mathbf{q}$ is a saddle point.

For the present, let b be fixed in the interval $(0, 1)$, and let the parameter a increase. In addition to the bifurcation at $-(1-b)^2/4$, H_{ab} has a bifurcation at $a = 3(1-b)^2/4$, because one of the two eigenvalues of $H_{ab}(\mathbf{p})$ descends through -1 (Exercise 10), so that $\mathbf{p}$ is transformed from an attracting fixed point to a saddle point. Recollect that an attracting 2-cycle for the quadratic family $\{Q_\mu\}$ emerges as μ increases and passes through 3. Thus we might suspect that as a passes through $3(1-b)^2/4$, an attracting 2-cycle for H_{ab} would be born. This is the case. In order to prove it, one would have to solve the equation

$$\begin{pmatrix} x \\ y \end{pmatrix} = H_{ab}^{[2]}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 - a(1 - ax^2 + y)^2 + bx \\ b(1 - ax^2 + y) \end{pmatrix}$$

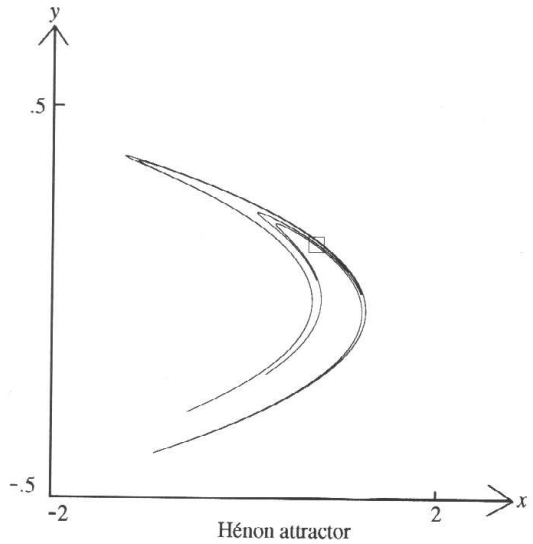
for x and y. Of course this entails solving a fourth-degree equation in x, which is possible because two roots are known from the two fixed points of H_{ab}. The result is that H_{ab} has a period-doubling bifurcation at $a = 3(1-b)^2/4$.

As a increases further, H_{ab} undergoes a period-doubling cascade. For certain special values of b the bifurcation values of a, as well as the Feigenbaum constant, are known. In particular, Derrida, Gervois, and Pomeau (1979) have calculated the following bifurcation values of a for $b = 0.3$:

bifurcation point	period-n cycle appears
-0.1225	1
0.3675	2
0.9125	4
1.0260⋯	8
1.0510⋯	16
1.0565⋯	32

The cascade terminates at approximately 1.0580459, which is the "Feigenbaum constant" for the Hénon map.

How does $H_{a(.3)}$ behave when $a > 1.06$? One might imagine that for an arbitrary $a > 1.06$, the iterates of virtually any initial point would be sprinkled unpredictably throughout a region in the plane. However, that does not happen. For example, let $a = 1.4$, and designate $H_{(1.4)(.3)}$ by H. If we neglect the first few iterates of $\mathbf{0}$ and plot the next 10,000 iterates, then we obtain the shape A_H appearing in Figure 3.11, on the front cover, and in Color Plate 1. The set A_H is called the **Hénon attractor** of the map, because the iterates of every point in a certain quadrilateral Q surrounding A_H approach the attractor.



Hénon attractor

Figure 3.11

Although A_H may appear to consist of a few fairly simple curves, when we zoom in on a small rectangle containing the fixed point $\mathbf{p}$, we see that there are several strands (Figure 3.12(a)). No matter how much we magnify the region, nearly identical new sets of strands appear (Figures 3.12(b) and (c)). It turns out that there are in reality an infinite number of such strands that make the region near to $\mathbf{p}$ look like a product of a line and a Cantor set.

The iterates of nearly all points in a rectangular region Q containing A_H not only converge to A_H but seem to trace out a dense subset of A_H. Thus whether the initial point is $\mathbf{0}$ or another point, after a few initial iterates the next several thousand iterates yield a virtually identical shape.

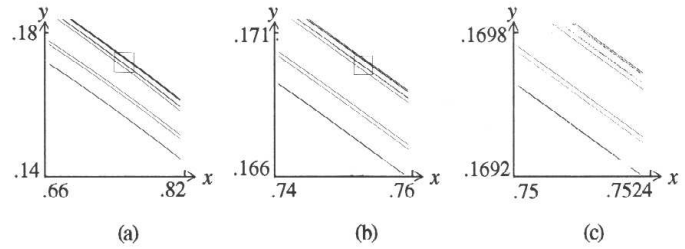


Figure 3.12

Two significant features of the Hénon attractor are evident from Figure 3.13, which displays iterates 34 through 40 of the origin (denoted by an isolated asterisk). First, the figure shows that iterates bounce around the attractor erratically. Second, the figure suggests that the Hénon attractor has sensitive dependence on initial conditions. Iterates of the origin identified by asterisks were computed by a CRAY supercomputer with single precision accuracy, whereas iterates identified by little squares were computed by the same CRAY supercomputer, but with double precision. The difference in accuracy between single and double precision is approximately 10^{-14} units. However, after 40 iterates the results are completely unrelated. Because of its sensitive dependence, A_H is called a **chaotic attractor**.

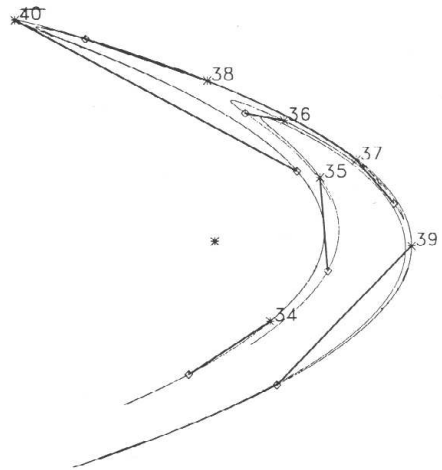


Figure 3.13

In his famous article of 1976, Hénon featured the attractor A_H with $a = 1.4$ and $b = 0.3$. Why did he select these parameters? On the one hand, he noticed that if a lies somewhere near 1.4, then as b increases from 0 to 0.3 the attractor A_H grows from nothing to a robust set in R^2; moreover, if b is much larger than 0.3, then the iterates of various points near $\mathbf{0}$ are unbounded. On the other hand, he found that if $b = 0.3$, and if $a <$ the Feigenbaum constant ≈ 1.06, then H_{ab} has an attracting cycle, so the attractor is a finite point set; however, if $a > 1.55$ then iterates of all points are unbounded. These observations led Hénon to select $a = 1.4$ and $b = 0.3$ as parameters that provide an attractor that is full and has interesting characteristics. Nevertheless, other parameters near 1.4 and 0.3 yield interesting attractors. In fact, some colleagues use $a = 1.42$ instead of $a = 1.4$. You might be interested to see what differences there are between the attractors corresponding to these two values of a.

EXERCISES 3.4

1. Using the program HENON, determine whether it takes 100, 500, 1000, 2000, or 5000 iterates of $\mathbf{0}$ to fill out A_H so that it appears as a collection of (relatively) complete curves.

2. Using the program HENON, show that after the first few iterates, the iterates of the points $\mathbf{0}$ and $\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$ yield virtually the same shape for A_H.

3. Recall that H has parameters $a = 1.4$ and $b = 0.3$.
   a. Find the (approximate) coordinates of the fixed point $\mathbf{p}$ for H.
   b. Find the eigenvalues λ and μ of $DH(\mathbf{p})$, with $\lambda < \mu$.
   c. Find eigenvectors $\mathbf{v}_\lambda$ and $\mathbf{v}_\mu$ of $DH(\mathbf{p})$. Confirm that H stretches distances along (that is, in the direction of) the attractor at $\mathbf{p}$, and contracts distances in a direction oblique to the direction of the attractor at $\mathbf{p}$.

4. Let $a = 1.4$. Use the program HENON to discuss what happens to the attractor of H_{ab} when
   a. b increases from 0.3 to 0.5
   b. b decreases from 0.3 to 0.1

5. Let $b = 0.3$. Use the program HENON to discuss what happens to the attractor of H_{ab} when
   a. a increases from 1.4
   b. a decreases from 1.4

6. Let $b = 0.3$. Use the program HENON to find a value of a such that H_{ab} has an attracting n-cycle, and determine an n-cycle.
   a. $n = 4$ b. $n = 8$ c. $n = 7$ d. $n = 3$

7. Let $a = 0$.
   a. Show that for each b in the interval $(0, 1)$, H_{ab} has a unique fixed point $\mathbf{p}$, and find $\mathbf{p}$.
   b. Determine whether $\mathbf{p}$ is attracting, repelling, or a saddle point.
   c. Find an eigenvector corresponding to each eigenvalue of H_{ab}.

8. a. Let $a = -(1-b)^2/4$ and $0 < b < 1$. Find the eigenvalues of $H_{ab}(\mathbf{q})$.
   b. Let $a > -(1-b)^2/4$ and $a \neq 0$. Show that the fixed point $\mathbf{q}$ in (2) is a saddle point of H_{ab} for each b in the interval $(0, 1)$.
   c. Use the results of (a) and (b) to discuss the type of bifurcation of H_{ab} that occurs at $a = -0.1225$.

9. Assume that $0 < b < 1$ and $-(1-b)^2/4 < a < 3(1-b)^2/4$. Let

$$p_1 = \frac{1}{2a}\left(b - 1 + \sqrt{(1-b)^2 + 4a}\right)$$

   Show that $-1 < -ap_1 - \sqrt{a^2 p_1^2 + b} < 0$.

10. Let b be in $(0, 1)$ and $a = 3(1-b)^2/4$. Find the eigenvalues of H_{ab}, and convince yourself that H_{ab} has a bifurcation at this value of a.

11. Let $H^*\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a - x^2 + by \\ x \end{pmatrix}$.
   a. Show that $H_{ab} \underset{E}{\approx} H^*$, where E is an appropriate linear function.
   b. What does the result of (a) tell you about the attractor of H^*? (H^* is often used as an alternative to H_{ab}.)

12. Use the bifurcation table for $H_{a(.3)}$ that appears in this section to compute an approximate Hénon version of the Feigenbaum number (see Section 1.5). Compare your answer with the Feigenbaum number for the quadratic family.

3.5 THE HORSESHOE MAP

One of the earliest examples of a function defined on R^2 that exhibits interesting dynamics is the **horseshoe map** described in the 1960's by the American mathematician Stephen Smale (1967).

The horseshoe map will be denoted by M. Its domain is the set S in R^2 composed of the unit square $T = [0, 1] \times [0, 1]$, bounded on the left and right by semicircles B and E (Figure 3.14(a)). We assume that S contains its boundary. The function M shrinks S vertically by a factor of $a < 1/3$, and expands S horizontally by a factor of $b = 3$, with the semicircles B and E altered so as to continue to be semicircular. The resulting figure is folded by M so that it fits again inside S, with only the semicircles protruding to the left of T (Figure 3.14(b)). Thus the range of M looks like a horseshoe. When S is partitioned as in Figure 3.14(c), we can see the effect of M on each member of the partition (Figure 3.14(d)). Specifically, M sends semicircles B and E into B, and sends the square T into two strips inside T plus a curved strip inside E.

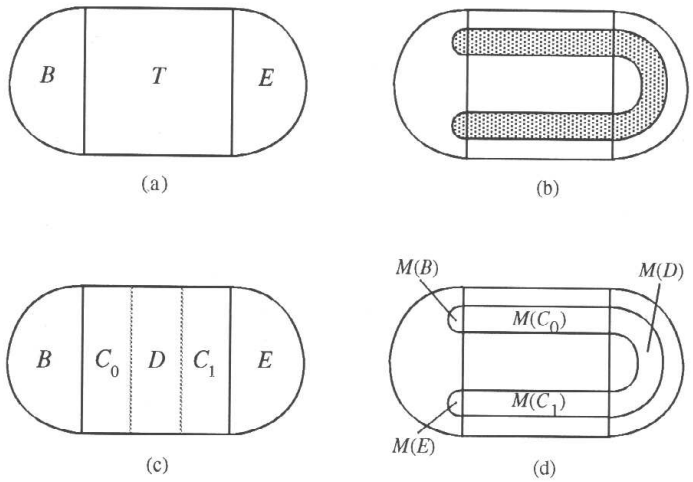


Figure 3.14

Even though we have not defined M by a formula or a series of formulas, we are able to systematically analyze it. To start with, we note that M is well-defined and the range is contained in S. Next we will show that M is a homeomorphism.

THEOREM 3.24. $M: S \to M(S)$ is a homeomorphism.

Proof. By definition, M maps onto $M(S)$. That M is one-to-one follows from the fact that stretching and contracting are one-to-one operations, and M folds in a non-overlapping manner. To prove that M is continuous, we compare subregions in Figure 3.14(c) with their corresponding images in Figure 3.14(d). The map M expands distances in C_0 and C_1 by a factor of 3, and shrinks distances in B and E. The largest expansion of distances for points in D occurs at the top boundary, which maps onto the exterior boundary γ of $M(D)$. Since the length of the top boundary of D is 1/3 and the length of γ is less than $\pi/2$, it follows that M expands distances in D by no more than a factor of 6. Consequently if $\mathbf{v}$ and $\mathbf{w}$ are in S and $\|\mathbf{v} - \mathbf{w}\| < \varepsilon$, then $\|M(\mathbf{v}) - M(\mathbf{w})\| < 6\varepsilon$. Therefore M is continuous. The inverse map M^{-1} is continuous by the same kind of argument. Therefore M is a homeomorphism. ■

Assuming that B in Figure 3.14(c) contains its boundary, we show next that M has a fixed point that lies in B and on that boundary.

THEOREM 3.25. M has a unique fixed point in B, to which the iterates of all points in B and E converge.

Proof. Since M shrinks the domain vertically by a factor of $a < 1/3$, this means that for each n, $M^{[n]}(B)$ is closed and semicircular with diameter a^n. In addition,

$$B \supseteq M(B) \supseteq M^{[2]}(B) \supseteq \cdots \supseteq M^{[n]}(B)$$

The two-dimensional version of the Heine-Borel Theorem (Theorem 2.7) implies that a nested sequence of closed, bounded sets in R^2 has a common point. Thus the intersection B_∞ of the sets $M^{[n]}(B)$ for $n \geq 1$ has at least one element. Notice that the diameter of $M^{[n]}(B)$ is a^n, and $\lim_{n\to\infty} a^n = 0$. Thus B_∞ contains exactly one element, which we denote by $\mathbf{p}$. Now $\mathbf{p}$ lies on the boundary of B and C_0. Since $\mathbf{p}$ and $M(\mathbf{p})$ are in $M^{[n]}(B)$ for all n, it follows that $\|\mathbf{p} - M(\mathbf{p})\| \leq a^n \to 0$ as n grows without bound. Consequently $\mathbf{p} = M(\mathbf{p})$, which means that $\mathbf{p}$ is a fixed point. Finally, because $B \supseteq M(E)$ and all elements of B converge to $\mathbf{p}$, we conclude that all elements of E also converge to $\mathbf{p}$. ■

Although $\mathbf{p}$ attracts all points in B and E, $\mathbf{p}$ is not an attracting fixed point because the iterates of points in the interior of C_0 are drawn away from $\mathbf{p}$.

Figure 3.15 shows the image of $M^{[2]}$, which is composed of two connected horseshoes. Similarly, for each positive integer n, the image of $M^{[n]}$ contains 2^{n-1} connected horseshoes whose width is approximately a^n.

From the definition of M, if $\mathbf{v}$ is in T, then $M(\mathbf{v})$ is also in T only if $\mathbf{v}$ is in $C_0 \cup C_1$. Let C_+ denote the collection of points in $C_0 \cup C_1$ *all* of whose iterates lie in $C_0 \cup C_1$. Thus

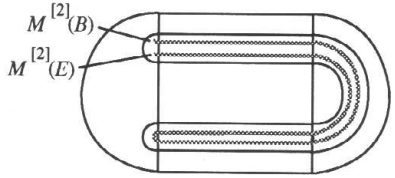


Figure 3.15

$$C_+ = \{\mathbf{v} \text{ in } C_0 \cup C_1 : M^{[n]}(\mathbf{v}) \text{ is in } C_0 \cup C_1 \text{ for } n = 0, 1, 2, 3, \ldots\}$$

All points in the domain of M either migrate toward $\mathbf{p}$, which is the most illustrious member of C_+, or start out in C_+ and stay there. Consequently C_+ serves as the attractor A_M of M. What does C_+ look like geometrically?

The set C_+ is an intersection of a nested sequence of ever thinner vertical strips. To show this we notice first that the collection of all points $\mathbf{v}$ such that $M(\mathbf{v})$ is in $C_0 \cup C_1$ consists of the four strips in Figure 3.16(a), where C_{ik} is the set of all $\mathbf{v}$ such that $\mathbf{v}$ is in C_i and $M(\mathbf{v})$ is in C_k, for $i, k = 0, 1$. (For example, $\mathbf{v}$ is in C_{01} if and only if $\mathbf{v}$ is in C_0 and $M(\mathbf{v})$ is in C_1.) Similarly, the collection of all $\mathbf{v}$ such that $M^{[2]}(\mathbf{v})$ is in $C_0 \cup C_1$ consists of the eight strips in Figure 3.16(b), and so forth. Therefore C_+, and hence the attractor A_M, is a collection of vertical lines in the square $T = [0, 1] \times [0, 1]$ whose intersection with the x axis is a Cantor-like set.

If $\mathbf{v}$ is in C_+, then each iterate of $\mathbf{v}$ lies either in C_0 or in C_1, so we can associate with $\mathbf{v}$ the **forward sequence** $z = z_0 z_1 z_2 \cdots$, where

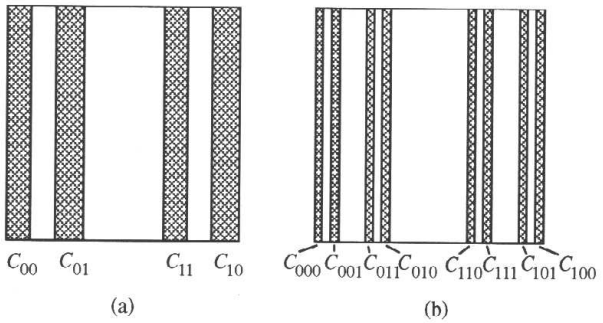


Figure 3.16

$$z_n = \begin{cases} 0 \text{ if } M^{[n]}(\mathbf{v}) \text{ is in } C_0 \\ 1 \text{ if } M^{[n]}(\mathbf{v}) \text{ is in } C_1 \end{cases}$$

Notice that the sequence $z_0 z_1 z_2 \cdots$ identifies the forward iterates of $\mathbf{v}$. This is reminiscent of the sequence identified with the numbers in $[0, 1]$ for the tent function described in Section 2.2. However, since C_0 and C_1 are separated by a rectangle of width 1/3, *every* sequence of 0's and 1's is the image of an element of C_+. (The corresponding result is false for T!)

In contrast to the tent function, each sequence of 0's and 1's corresponds to a whole vertical line, not just an individual point. To identify sequences of 0's and 1's with unique points in S, we need to examine the pre-images of points in S.

Figure 3.14(d) indicates that $M(C_0)$ and $M(C_1)$ are horizontal strips in T that we denote by V_0 and V_1, respectively (Figure 3.17(a)). Similarly, $M^{[2]}(C_{00})$, $M^{[2]}(C_{01})$, $M^{[2]}(C_{11})$, and $M^{[2]}(C_{10})$ are horizontal strips in Figure 3.15, and we designate these strips as V_{00}, V_{01}, V_{11}, and V_{10} respectively (Figure 3.17(b)). Letting $M^{-[m]}(P)$ denote the set Q such that $M^{[m]}(Q) = P$, we find that

$$M^{-[1]}(V_i) = C_i \quad \text{and} \quad M^{-[2]}(V_{ik}) = C_{ik} \quad \text{for } i = 0, 1 \text{ and } k = 0, 1$$

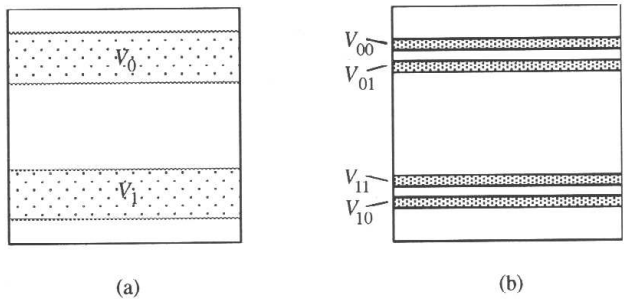


Figure 3.17

Continuing in this fashion, we obtain a nested sequence of ever thinner horizontal strips whose intersection we denote by C_-. Thus

$$C_- = \{\mathbf{v} \text{ in } S: M^{-[n]}(\mathbf{v}) \text{ is in } C_0 \cup C_1 \text{ for } n = 1, 2, 3, \ldots\}$$

In the same way as we identified a sequence of 0's and 1's for each element of C_+, we now assign the **backward sequence** $\cdots z_{-3} z_{-2} z_{-1}$ to each $\mathbf{v}$ in C_-, where

$$z_{-n} = \begin{cases} 0 \text{ if } M^{-[n]}(\mathbf{v}) \text{ is in } C_0 \\ 1 \text{ if } M^{-[n]}(\mathbf{v}) \text{ is in } C_1 \end{cases}$$

Each such sequence corresponds to a horizontal line in T, and C_- is a collection of horizontal lines in T whose intersection with the y axis is a Cantor-like set.

Combining the forward sequence $z_0z_1z_2\cdots$ and the backward sequence $\cdots z_{-3}z_{-2}z_{-1}$, we obtain the **two-sided sequence** (or **bi-infinite sequence**) $\cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots$, where the decimal point separates the backward part from the forward part of the sequence. Finally, we define the set C^* by

$$C^* = C_+ \cap C_-$$

The set C^* consists of the points in T all of whose forward and backward iterates lie in $C_0 \cup C_1$.

Let Z denote the collection of all two-sided sequences of 0's and 1's, and define h: $C^* \to Z$ by

$$h(\mathbf{v}) = \text{the two-sided sequence corresponding to } \mathbf{v}, \text{ for } \mathbf{v} \text{ in } C^* \qquad (1)$$

Then h is well-defined, and identifies C^* with Z, as Lemma 1 asserts.

LEMMA 1. h: $C^* \to Z$ is one-to-one and onto.

Proof. If $\mathbf{v}$ and $\mathbf{w}$ are in C^* and $h(\mathbf{v}) = h(\mathbf{w})$, then because $h(\mathbf{v})$ and $h(\mathbf{w})$ have the same forward (backward) sequence, they lie on the same vertical (horizontal) line in T. Therefore $\mathbf{v} = \mathbf{w}$, so that h is one-to-one. To show that h is onto, assume that $x = \cdots x_{-3}x_{-2}x_{-1}.x_0x_1x_2\cdots$ is in Z. For $n \geq 0$, let

$$J_n = \{\mathbf{v} \text{ in } C_0 \cup C_1 : h(\mathbf{v}) = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots \text{ and } z_0z_1z_2\cdots z_n = x_0x_1x_2\cdots x_n\}$$

and

$$J_{-n} = \{\mathbf{v} \text{ in } C_0 \cup C_1 : h(\mathbf{v}) = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots \text{ and } z_{-n}\cdots z_{-3}z_{-2}z_{-1} = x_{-n}\cdots x_{-3}x_{-2}x_{-1}\}$$

Then J_n and J_{-n} are closed for all n. Because $\bigcap_{n\geq 0} J_n$ is a single vertical line and $\bigcap_{n<0} J_n$ is a single horizontal line in T, it follows that $\bigcap_{\text{all } n} J_n$ is a unique point $\mathbf{v}^*$. By construction, $h(\mathbf{v}^*) = x$, so that h is onto. ■

The function h has the property that

$$\text{if } h(\mathbf{v}) = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots, \text{ then } h(M(\mathbf{v})) = \cdots z_{-2}z_{-1}z_0.z_1z_2z_3\cdots$$

This means that the sequence associated with $M(\mathbf{v})$ is the sequence associated with $\mathbf{v}$, shifted to the left one place with respect to the decimal point. This fact, along with the association between points of C^* and two-sided sequences of 0's and 1's, bears immediate fruit. In particular, the two doubly-repeated sequences

$$\cdots \overline{0}.\overline{0}\cdots \quad \text{and} \quad \cdots \overline{1}.\overline{1}\cdots$$

correspond to fixed points of M. You can check that the fixed point $\mathbf{p}$ on the border between B and C_0 corresponds to $\cdots \overline{0}.\overline{0}\cdots$. Can you locate the other fixed point in T? Next, the sequences $\cdots\overline{10}.\overline{10}\cdots$ and $\cdots \overline{01}.\overline{01}\cdots$ comprise a 2-cycle for M. Using these sequences as models, one can exhibit two-sided sequences corresponding to n-cycles for any positive integer n. From the definition of $\cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots$, one can even indicate where in T members of such a cycle lie.

Next we will introduce a distance on the set Z of two-sided sequences. This distance will make it possible for us to show that C^* and Z are homeomorphic. Let $x = \cdots x_{-3}x_{-2}x_{-1}.x_0x_1x_2\cdots$ and $z = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots$. Then we define the **distance** $\|x - z\|$ between x and z by the formula

$$\|x - z\| = \sum_{k=-\infty}^{\infty} \frac{|x_k - z_k|}{2^{|k|}} \qquad (2)$$

The distance is a metric on the space of two-sided sequences (Exercise 6). If $x_k = z_k$ for $|k| \leq n$, then $\|x - z\| \leq 1/2^{n-1}$. Moreover, if $\|x - z\| \leq 1/2^n$, then $x_k = z_k$ for $|k| \leq n + 1$. Thus the distance between x and z is small provided that the central blocks of x and z are identical.

THEOREM 3.26. h: $C^* \to Z$ is a homeomorphism.

Proof. By Lemma 1, h is one-to-one and onto. Therefore we need only show that h and h^{-1} are continuous. Let $\varepsilon > 0$, and choose n so large that $1/2^{n-1} < \varepsilon$. Next, let $\mathbf{v}$ and $\mathbf{w}$ be in C^*, with

$$h(\mathbf{v}) = x = \cdots x_{-3}x_{-2}x_{-1}.x_0x_1x_2\cdots \quad \text{and} \quad h(\mathbf{w}) = z = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots$$

If $\|\mathbf{v} - \mathbf{w}\| < 1/3^{n+1}$, then $\mathbf{v}$ and $\mathbf{w}$ lie in the same vertical strip of width $1/3^{n+1}$, so that $x_k = z_k$ for $k = 0, 1, 2, \ldots, n$. Similarly, there is a $\delta_1 > 0$ such that if $\|\mathbf{v} - \mathbf{w}\| < \delta_1$ then $\mathbf{v}$ and $\mathbf{w}$ lie in the same horizontal strip at the nth stage, which means that $x_k = z_k$ for $k = -1, -2, \ldots, -n$. Now choose $\delta > 0$ such that $\delta < 1/3^{n+1}$ and $\delta \leq \delta_1$. It follows that if $\|\mathbf{v} - \mathbf{w}\| < \delta$, then $x_k = z_k$ for $|k| \leq n$, and thus

$$\|x - z\| = \sum_{|k|=n+1}^{\infty} \frac{|x_k - z_k|}{2^{|k|}} \leq \frac{1}{2^{n-1}} < \varepsilon$$

Consequently h is continuous. The proof that h^{-1} is continuous follows by a similar argument (but with the roles of vertical contraction and horizontal expansion interchanged). ■

The **left shift map** $\sigma: Z \to Z$ is defined as the name suggests:

$$\text{if } z = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots, \text{ then } \sigma(z) = \cdots z_{-3}z_{-2}z_{-1}z_0.z_1z_2\cdots$$

Thus σ shifts the entries to the left one place with respect to the decimal point, or equivalently, shifts the decimal point one place to the right. That σ is a homeomorphism on Z can be proved in a straightforward manner (Exercise 7). Moreover, we can show that σ is **strongly chaotic**, which by definition means that

i. its domain has a dense set of periodic points
ii. it has sensitive dependence on initial conditions
iii. it is transitive (that is, it has an element with dense orbit)

THEOREM 3.27. σ is strongly chaotic on Z.

Proof. First we will show that the set of periodic points of σ is dense in Z. To that end, let $z = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots$ be an arbitrary element of Z, and let n be an arbitrary positive integer. If x is the doubly-repeating two-sided sequence $z_{-n}\cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots z_n$, then it follows that x is periodic (with period $2n+1$). Moreover, $z_k = x_k$ for $|k| \le n$, so that $\|x - z\| \le 1/2^{n-1}$. Thus the periodic points are dense in Z. To show that σ has sensitive dependence on initial conditions, let z be in Z, $\varepsilon = 1/2$, $\delta > 0$, and n so large that $1/2^{n-1} < \delta$. If x is chosen with $x_k = z_k$ for all k and such that $|k| \le n$ but $x_{n+1} \ne z_{n+1}$, then $\|x - z\| \le 1/2^{n-1} < \delta$. However,

$$\sigma^{[n+1]}(x) = \cdots x_n \cdot x_{n+1}\cdots \quad \text{and} \quad \sigma^{[n+1]}(z) = \cdots z_n \cdot z_{n+1}\cdots$$

so that $\|\sigma^{[n+1]}(x) - \sigma^{[n+1]}(x)\| \ge 1 > \varepsilon$. Therefore σ has sensitive dependence on initial conditions. Finally, we will show that σ is transitive. To see this, let the forward portion of the two-sided sequence z^* have the form

$$0\ 1\ 000\ 001\ 010\ 011\ 100\ 101\ 110\ 111\ 00000\ 00001\cdots$$

(where for each positive odd integer n, all possible n-tuples appear in order), and let the backward portion of z^* have the form

$$\cdots\ 0100\ 0011\ 0010\ 0001\ 0000\ 11\ 10\ 01\ 00$$

(where for each positive even integer n, all possible n-tuples appear in backward order). Then it is possible to show that the orbit of z^* is dense in Z (Exercise 4). With this we have completed the proof that σ is strongly chaotic on Z. ■

The functions $M: C^* \to C^*$ and $\sigma: Z \to Z$ are **conjugate**, in the sense that there is a homeomorphism $h: C^* \to Z$ such that $h \circ M = \sigma \circ h$. The homeomorphism we have in mind, of course, is the one defined in (1). You should check that indeed $h(M(\mathbf{v})) = \sigma(h(\mathbf{v}))$ for all $\mathbf{v}$ in C^* (Exercise 8). The strong chaotic nature of one function is inherited by another function conjugate to it (as we saw in Section 2.3). Thus we obtain the following consequence of Theorem 3.27.

COROLLARY 3.28. M is strongly chaotic on C^*.

The Smale horseshoe map is justifiably famous, for several reasons. First, it has a very simple geometric definition and is easy to visualize. Second, it has all the characteristics of a strongly chaotic map: sensitive dependence, plenty of periodic points, and a dense orbit. It also has, in the most obvious way, the telltale signs of a chaotic map: stretching and folding. The stretching yields sensitive dependence, and the folding allows the map to be bounded. It is conjectured that in some sense every map that is strongly chaotic has a subset of its domain on which the map acts like the horseshoe map M.

Homoclinic Points

Let f be a function, and p a fixed point of f. It can happen that there is a point q in the domain of f whose forward iterates converge to p and such that a sequence of backward iterates of q also converges to p. The following example illustrates such behavior.

EXAMPLE 1. Let T be the tent function, defined by

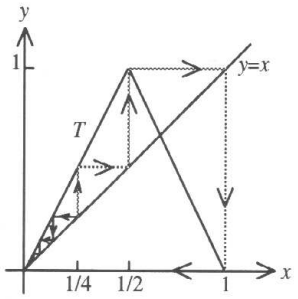
$$T(x) = \begin{cases} 2x & \text{for } 0 \le x \le 1/2 \\ 2(1-x) & \text{for } 1/2 < x \le 1 \end{cases}$$

Show that the forward and backward iterates of $1/4$ converge to the fixed point 0.

Solution. First we notice that

$$T(\frac{1}{4}) = \frac{1}{2}, \quad T^{[2]}(\frac{1}{4}) = 1, \text{ and } T^{[3]}(\frac{1}{4}) = 0$$

so that the forward iterates of $1/4$ are eventually 0, and hence converge to 0. By



Forward and backward iterates of 1/4

Figure 3.18

contrast, $T^{-1}(1/4)$ contains 1/8, $T^{-[2]}(1/4)$ contains 1/16, and in general, $T^{-[n]}(1/4)$ contains $1/2^{n+2}$. Thus there are backward iterates of 1/4 that converge to 0. Figure 3.18 shows both forward and backward iterates of 1/4. ❑

Since the forward iterates of 1/4 converge to 0, 1/4 is in the local stable manifold $W^s_{\text{loc}}(0)$ of 0. In the same manner, 1/4 is in the local unstable manifold $W^u_{\text{loc}}(0)$ of 0, because there are backward iterates of 1/4 that converge to 0. This means that 1/4 is in both $W^s_{\text{loc}}(0)$ and $W^u_{\text{loc}}(0)$. Points with this property are called homoclinic points.

DEFINITION 3.29. Let F be defined on a subset of R^n, where $n = 1$ or 2, and let $\mathbf{p}$ be a fixed point of F. A point $\mathbf{q}$ is **homoclinic to** $\mathbf{p}$, or is a **homoclinic point**, if $\mathbf{q} \neq \mathbf{p}$ but $\mathbf{q}$ is in both $W^s_{\text{loc}}(\mathbf{p})$ and $W^u_{\text{loc}}(\mathbf{p})$.

From Example 1 we know that 1/4 is homoclinic to 0 for T. In fact, there are infinitely many numbers in (0, 1) that are homoclinic to 0 for T (Exercise 11). Analogously, there are numbers in (0, 1) that are homoclinic to 0 for the quadratic function Q_4 (Exercise 12). By contrast, there are no numbers in (0, 1) that are homoclinic to 0 for the quadratic function Q_μ if $0 < \mu < 4$ (also Exercise 12).

The horseshoe map M has points homoclinic to the fixed point $\mathbf{p}$ in B. Since M is conjugate to the left shift σ, we can verify this by showing, equivalently, that the left shift σ has points homoclinic to the fixed point $z^{**} = \cdots \bar{0}.\bar{0}\cdots$ of σ. We will indicate how one can determine such homoclinic points of σ, but will leave the details to be completed in Exercise 10. On the one hand, $W^s_{\text{loc}}(z^{**})$ consists of all two-sided sequences all of whose entries to the right of some entry are identical to those of z^{**}. On the other hand, $W^u_{\text{loc}}(z^{**})$ consists of those two-sided sequences all of whose entries to the left of some entry are identical to those of z^{**}. It follows from these facts that σ has points homoclinic to z^{**}. What is more, there are infinitely many points homo-clinic to z^{**}.

Homoclinic points play a significant role in the study of the dynamics of higher-dimensional functions.

Conclusion

Chapter 3 has introduced dynamics of functions defined in R^2. Besides the linear functions, which exhibit regularity, we have focused on three prominent functions: the baker's function, the Hénon map, and the Smale horseshoe map. For each of these three functions there is an attractor to which the iterates of all points in the domain converge, and which contains in a natural way a two-dimensional version of a Cantor set. In each there is a stretching and some form of folding, which are characteristic of chaotic behavior. Finally, each represents a two-dimensional analogue of a one-dimensional chaotic function studied in Chapters 1–2.

EXERCISES 3.5

1. Find the approximate location of the point in S corresponding to the sequence $\cdots\bar{1}.\bar{1}\cdots$ in Z.

2. Determine the number of n-cycles in C^*.

3. Let $x = \cdots x_{-3}x_{-2}x_{-1}.x_0x_1x_2\cdots$ and $z = \cdots z_{-3}z_{-2}z_{-1}.z_0z_1z_2\cdots$.

   a. Show that if $x_k = z_k$ for $|k| \leq n$, then $\|x - z\| \leq 1/2^{n-1}$.

   b. Find x and z such that $x_k = z_k$ for $|k| \leq n$ and $\|x - z\| = 1/2^{n-1}$.

   c. Find x and z in Z such that $\|x - z\| = 3$.

4. Show that the sequence z^* in the proof of Theorem 3.27 has a dense orbit.

5. Show that there is an element of C^* that neither is periodic nor has a dense orbit under M.

6. Show that the distance defined in (2) has the properties of a metric as defined in Section 2.4.

7. Show that the left shift $\sigma: Z \to Z$ is a homeomorphism.

8. Show that $h(M(\mathbf{v})) = \sigma(h(\mathbf{v}))$ for all $\mathbf{v}$ in C^*.

9. Show that there is a one-to-one function whose domain is C^* and whose range is the Cantor ternary set C.

10. Let $z^{**} = \cdots\overline{0}.\overline{0}\cdots$, one of the two fixed points of σ.
    a. Show that $W^s_{\text{loc}}(z^{**})$ consists of all two-sided sequences all of whose entries to the right of some entry are identical to those of z^{**}.
    b. Show that $W^u_{\text{loc}}(z^{**})$ consists of those two-sided sequences all of whose entries to the left of some entry are identical to those of z^{**}.
    c. Show that there are infinitely many points homoclinic to z^{**}.

11. Show that there are infinitely many numbers in the interval $(0, 1)$ that are homoclinic to 0 for the tent function T.

12. Let $Q_\mu(x) = \mu x(1 - x)$ for $0 \le x \le 1$.
    a. Show that there are no points homoclinic to the fixed point 0 whenever $0 < \mu < 4$.
    b. Show that there are infinitely many points homoclinic to 0 if $\mu = 4$.
    c. Show that there are infinitely many points homoclinic to the fixed point $p_\mu = 1 - 1/\mu$ if $\mu = 4$.