Chapter III

Transformations on Metric Spaces; Contraction Mappings; and the Construction of Fractals

1 Transformations on the Real Line

Fractal geometry studies "complicated" subsets of geometrically "simple" spaces such as $\mathbb{R}^2$, $\mathbb{C}$, $\mathbb{R}$, and $\hat{\mathbb{C}}$. In deterministic fractal geometry the focus is on those subsets of a space that are generated by, or possess invariance properties under, simple geometrical transformations of the space into itself. A simple geometrical transformation is one that is easily conveyed or explained to someone else. Usually it can be completely specified by a small set of parameters. Examples include affine transformations in $\mathbb{R}^2$, which are expressed using 2×2 matrices and 2-vectors, and rational transformations on the Riemann Sphere, which require the specification of the coefficients in a pair of polynomials.

Definition 1.1 *Let* $(\mathbf{X}, d)$ *be a metric space. A* transformation *on* $\mathbf{X}$ *is a function* $f : \mathbf{X} \to \mathbf{X}$, *which assigns exactly one point* $f(x) \in \mathbf{X}$ *to each point* $x \in \mathbf{X}$. *If* $S \subset \mathbf{X}$ *then* $f(S) = \{f(x) : x \in S\}$. f *is* one-to-one *if* $x, y \in \mathbf{X}$ *with* $f(x) = f(y)$ *implies* $x = y$. f *is* onto *if* $f(\mathbf{X}) = \mathbf{X}$. f *is called* invertible *if it is one-to-one and onto: in* *this case it is possible to define a transformation $f^{-1} : \mathbf{X} \to \mathbf{X}$, called the inverse of f, by $f^{-1}(y) = x$, where $x \in \mathbf{X}$ is the unique point such that $y = f(x)$.*

Definition 1.2 *Let $f: \mathbf{X} \to \mathbf{X}$ be a transformation on a metric space. The* forward iterates *of f are transformations $f^{\circ n} : \mathbf{X} \to \mathbf{X}$ defined by $f^{\circ 0}(x) = x$, $f^{\circ 1}(x) = f(x)$, $f^{\circ(n+1)}(x) = f \circ f^{(n)}(x) = f(f^{(n)}(x))$ for $n = 0, 1, 2, \ldots$. If f is invertible then the* backward iterates *of f are transformations $f^{\circ(-m)}(x) : \mathbf{X} \to \mathbf{X}$ defined by $f^{\circ(-1)}(x) = f^{-1}(x)$, $f^{\circ(-m)}(x) = (f^{\circ m})^{-1}(x)$ for $m = 1, 2, 3, \ldots$.*

In order to work in fractal geometry one needs to be familiar with the basic families of transformations in $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{C}$, and $\hat{\mathbb{C}}$. One needs to know well the relationship between "formulas" for transformations and the geometric changes, stretchings, twistings, foldings, and skewings of the underlying fabric, the metric space upon which they act. It is more important to understand what the transformations do to sets than how they act on individual points. So, for example, it is more useful to know how an affine transformation in $\mathbb{R}^2$ acts on a straight line, a circle, or a triangle, than to know to where it takes the origin.

Examples & Exercises

1.1. Let f: $\mathbf{X} \to \mathbf{X}$ be an invertible transformation. Show that

$$f^{\circ m} \circ f^{\circ n} = f^{\circ(m+n)} \qquad \text{for all integers } m \text{ and } n.$$

1.2. A transformation $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 2x$ for all $x \in \mathbb{R}$. Is f invertible? Find a formula for $f^{\circ n}(x)$ that applies for all integers n.

1.3. A transformation $f : [0, 1] \to [0, 1]$ is defined by $f(x) = \frac{1}{2}x$. Is this transformation one-to-one? Onto? Invertible?

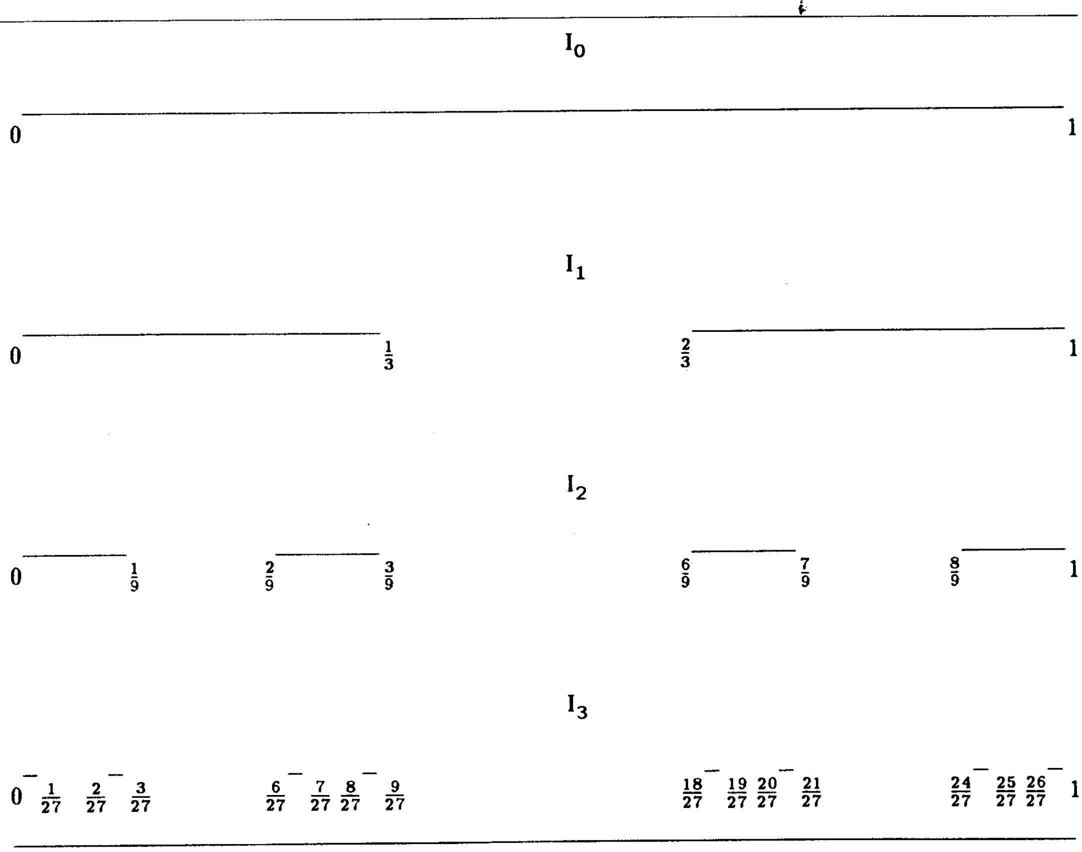
1.4. The mapping $f : [0, 1] \to [0, 1]$ is defined by $f(x) = 4x \cdot (1 - x)$. Is this transformation one-to-one? Onto? Is it invertible?

1.5. Let C denote the *Classical Cantor Set*. This subset of the metric space $[0, 1]$ is obtained by successive deletion of middle-third open subintervals as follows. We construct a nested sequence of closed intervals

$$I_0 \supset I_1 \supset I_2 \supset I_3 \supset I_4 \supset I_5 \supset I_6 \supset I_7 \ldots \supset I_n \supset \ldots,$$

where

Figure III.28. Construction of the Classical Cantor Set $\mathcal{C}$.



$$
\begin{aligned}
I_0 &= [0, 1],\\
I_1 &= [0, \tfrac{1}{3}] \cup [\tfrac{1}{3}, \tfrac{2}{3}],\\
I_2 &= [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{3}{9}] \cup [\tfrac{6}{9}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, \tfrac{9}{9}],\\
I_3 &= [0, \tfrac{1}{27}] \cup [\tfrac{2}{27}, \tfrac{3}{27}] \cup [\tfrac{6}{27}, \tfrac{7}{27}] \cup [\tfrac{8}{27}, \tfrac{9}{27}]\\
&\quad \cup [\tfrac{18}{27}, \tfrac{19}{27}] \cup [\tfrac{20}{27}, \tfrac{21}{27}] \cup [\tfrac{24}{27}, \tfrac{25}{27}] \cup [\tfrac{26}{27}, \tfrac{27}{27}],\\
I_4 &= I_3 \text{ take away the middle open third of each interval in } I_3,\\
&\vdots\\
I_n &= I_{N-1} \text{ take away the middle open third of each interval in } I_{N-1}.
\end{aligned}
$$

This construction is illustrated in Figure III.28. We define

$$\mathcal{C} = \cap_{n=0}^{\infty} I_n.$$

$\mathcal{C}$ contains the point $x = 0$, so it is nonempty. In fact $\mathcal{C}$ is a perfect set that contains uncountably many points, as discussed in Chapter IV. $\mathcal{C}$ is an official fractal and we will often refer to it.

We are now able to work in the metric space ($\mathcal{C}$, Euclidean). A transformation $f : \mathcal{C} \to \mathcal{C}$ is defined by $f(x) = \frac{1}{3}x$. Show that this transformation is one-to-one but

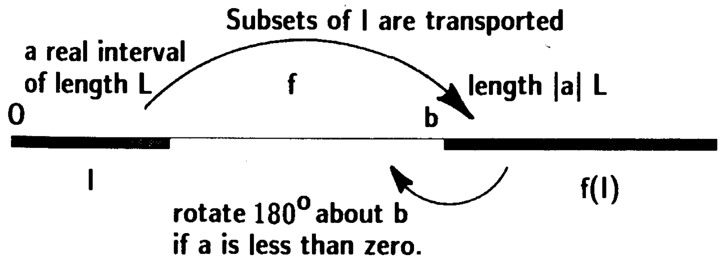


Figure III.29. The action of the affine transformation $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = ax + b$.

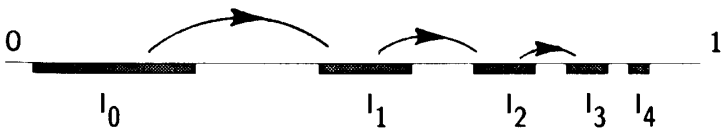


Figure III.30. This figure suggests a sequence of intervals $\{I_n\}_{n=0}^{\infty}$. Find an affine transformation $f : \mathbb{R} \to \mathbb{R}$ so that $f^{\circ n}(I_0) = I_n$ for $n = 0, 1, 2, 3, \ldots$. Use a straight-edge and dividers to help you.

not onto. Also, find another affine transformation (see example 1.7, which maps $\mathcal{C}$ one-to-one into $\mathcal{C}$).

1.6. $f : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $f(x_1, x_2) = (2x_1, x_2^2 + x_1)$ for all $(x_1, x_2) \in \mathbb{R}^2$. Show that f is not invertible. Give a formula for $f^{\circ 2}(x)$.

1.7. *Affine transformations* in $\mathbb{R}^1$ are transformations of the form $f(x) = a \cdot x + b$, where a and b are real constants. Given the interval $I = [0, 1]$, $f(I)$ is a new interval of length $|a|$, and f rescales by a. The left endpoint 0 of the interval is moved to b, and $f(I)$ lies to the left or right of b according to whether a is positive or negative, respectively (see Figure III.29).

We think of the action of an affine transformation on all of $\mathbb{R}$ as follows: the whole line is stretched away from the origin if $|a| > 1$, or contracted toward it if $|a| < 1$; flipped through 180° about $\mathcal{O}$ if $a < 0$; and then *translated* (shifted as a whole) by an amount b (shift to the left if $b < 0$, and to the right if $b > 0$).

1.8. Describe the set of affine transformations that takes the real interval $\mathbf{X} = [1, 2]$ into itself. Show that if f and g are two such transformations then $f \circ g$ and $g \circ f$ are also affine transformations on $[1, 2]$. Under what conditions does $f \circ g(\mathbf{X}) \cup g \circ f(\mathbf{X}) = \mathbf{X}$?

1.9. A sequence of intervals $\{I_n\}_{n=0}^{\infty}$ is indicated in Figure III.30. Find an affine transformation $f : \mathbb{R} \to \mathbb{R}$ so that $f^{\circ n}(I_0) = I_n$ for $n = 0, 1, 2, 3, \ldots$. Use a straight-edge and dividers to help you. Also show that $\{I_n\}_{n=1}^{\infty}$ is a Cauchy sequence in $(\mathcal{H}(\mathbb{R}), h)$, where h is the Hausdorff distance on $\mathcal{H}(\mathbb{R})$ induced by the Euclidean metric on $\mathbb{R}$. Evaluate $I = \lim_{n\to\infty} I_n$.

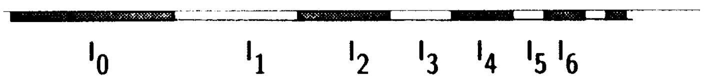


Figure III.31. Picture of a convergent geometric series in $\mathbb{R}^1$ (see exercise 1.10).

1.10. Consider the geometric series $\sum_{n=0}^{\infty} b \cdot a^n = b + a \cdot b + a^2 b + a^3 b + a^4 b + \cdots > 0$, $0 < b < 1$. This is associated with a sequence of intervals $I_0 = [0, b]$, $I_n = f^{\circ n}(I_0)$, where $f(x) = ax + b$, $n = 1, 2, 3, \ldots$, as illustrated in Figure III.31.

Let $I = \cup_{n=0}^{\infty} I_n$ and let l denote the total length of I. Show that $f(I) = I \setminus I_0$, and hence deduce that $al = l - b$ so that $l = b/(1-a)$. Deduce at once that

$$\sum_{n=0}^{\infty} b \cdot a^n = b/(1-a).$$

Thus we see from a geometrical point of view a well-known result about *geometric series*. Make a similar geometrical argument to cover the case $-1 < a < 0$.

Definition 1.3 *A transformation $f : \mathbb{R} \to \mathbb{R}$ of the form*

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n,$$

where the coefficients a_i $(i = 0, 1, 2, \ldots, N)$ are real numbers, $a_n \neq 0$, and N is a nonnegative integer, is called a polynomial transformation. *N is called the* degree *of the transformation.*

Examples & Exercises

1.11. Show that if $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are polynomial transformations, then so is $f \circ g$. If f is of degree N, calculate the degee of $f^{\circ m}(x)$ for $m = 1, 2, 3, \ldots$.

1.12. Show that for $n > 1$ a polynomial transformation $f : \mathbb{R} \to \mathbb{R}$ of degree n is not generally invertible.

1.13. Show that far enough out (i.e., for large enough $|x|$), a polynomial transformation $f : \mathbb{R} \to \mathbb{R}$ always stretches intervals. That is, view f as a transformation from ($\mathbb{R}$, Euclidean) into itself. Show that if I is an interval of the form $I = \{x : |x - a| \leq b\}$ for fixed $a, b \in \mathbb{R}$, then for any number $M > 0$ there is a number $\beta > 0$ such that if $b > \beta$, then the ratio (length of $f(I)$)/(length of I) is larger than M. This idea is illustrated in Figure III.32.

1.14. A polynomial transformation $f : \mathbb{R} \to \mathbb{R}$ of degree n can produce at most $(n-1)$ *folds*. For example $f(x) = x^3 - 3x + 1$ behaves as shown in Figure III.33.

1.15. Find a family of polynomial transformations of degree 2 which map the interval $[0, 2]$ into itself, such that, with one exception, if $y \in f([0, 2])$ then there exist two distinct points x_1 and x_2 in $[0, 2]$ with $f(x_1) = f(x_2) = y$.

1.16. Show that the one-parameter family of polynomial transformations f_λ :

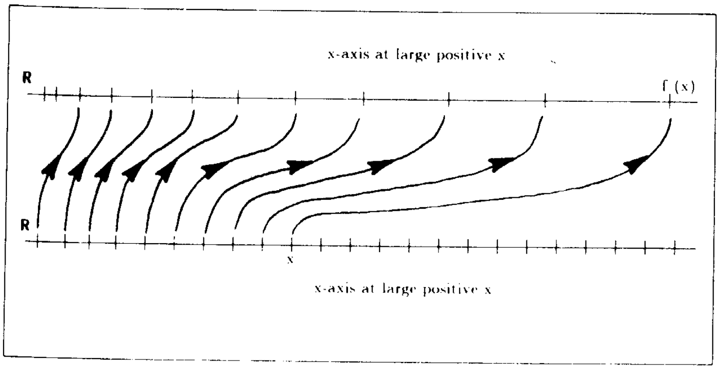


Figure III.32. A polynomial transformation $f : \mathbb{R} \to \mathbb{R}$ of degree > 1 stretches $\mathbb{R}$ more and more the farther out one goes.

$[0, 2] \to [0, 2]$, where

$$f_\lambda(x) = \lambda \cdot x \cdot (2 - x),$$

and the parameter λ belongs to $[0, 1]$, indeed takes the interval $[0, 2]$ into itself. Locate the value of x at which the fold occurs. Sketch the behavior of the family, in the spirit of Figure III.33.

1.17. Let $f : \mathbb{R} \to \mathbb{R}$ be a polynomial transformation of degree n. Show that values of x that are transformed into fold points are solutions of

$$\frac{df}{dx}(x) = 0, x \in \mathbb{R}.$$

Solutions of this equation are called (real) *critical points* of the function f. If c is a critical point then $f(c)$ is a *critical value*. Show that a critical value need not be a fold point.

1.18. Find a polynomial transformation such that Figure III.34 is true.

1.19. Recall that a polynomial transformation of an interval $f : I \subset \mathbb{R} \to I$ is normally represented as in Figure III.35. This will be useful when we study iterates $f^{\circ n}(x)_{n=1}^{\infty}$. However, the folding point of view helps us to understand the idea of the deformation of space.

1.20. Polynomial transformations can be lifted to act on subsets of $\mathbb{R}^2$ in a simple

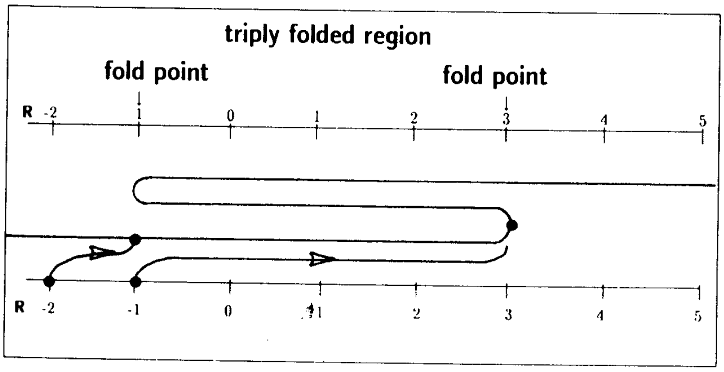


Figure III.33. The polynomial transformation $f(x) = x^3 - 3x + 1$.

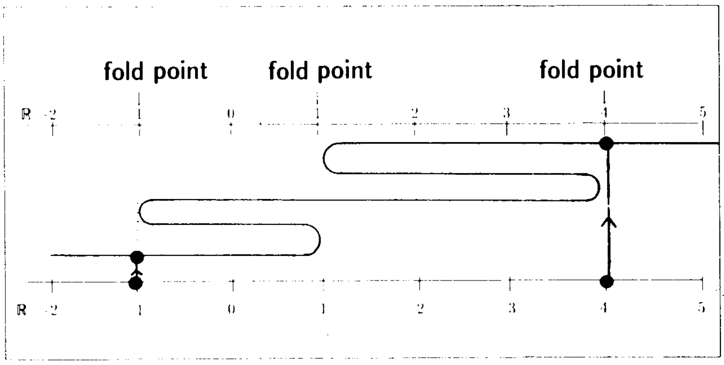


Figure III.34. Find a polynomial transformation $f : \mathbb{R} \to \mathbb{R}$, so that this figure correctly represents the way it folds on the real line.

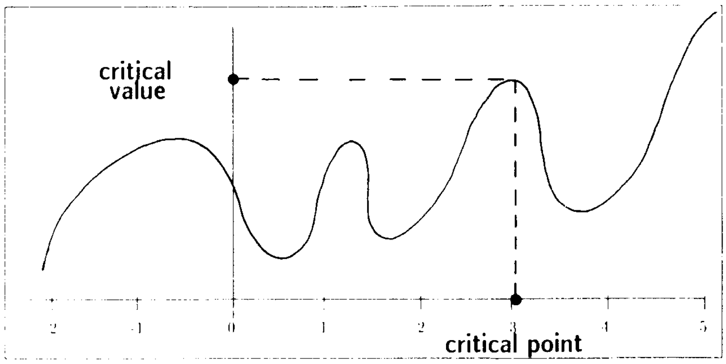


Figure III.35. The usual way of picturing a polynomial transformation.

way: we can define, for example, $F(x) = (f_1(x_1), f_2(x_2))$, where f_1 and f_2 are polynomial transformations in $\mathbb{R}$, so that $F : \mathbb{R}^2 \to \mathbb{R}^2$. Desired foldings in two orthogonal directions can be produced; or shrinking in one direction and folding in another. Show that the transformation $F(x_1, x_2) = (\frac{8}{5}x_1^3 - \frac{36}{5}x_1^2 + \frac{48}{5}x_1, x_2)$ acts on the triangular set S in Figure III.36 as shown.

The real line can be extended to a space which is topologically a circle by including the point at infinity. One way to do this is to think of $\mathbb{R}$ as a subset of $\hat{\mathbb{C}}$, and

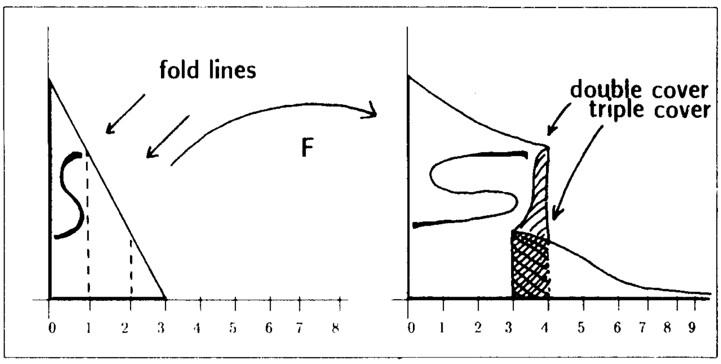


Figure III.36. A polynomial transformation acting on a set S in the plane.

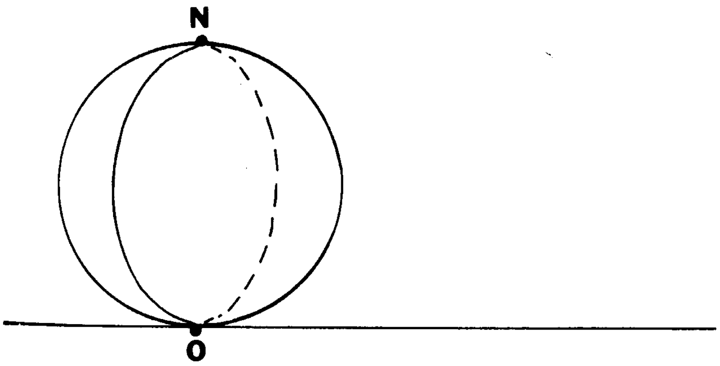


Figure III.37. $\mathbb{R} \cup \{\infty\}$ becomes a circle on a sphere.

then include the North Pole on $\hat{\mathbb{C}}$. We define this space to be $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ and will usually give it the spherical metric.

Definition 1.4 *A transformation $f : \hat{\mathbb{R}} \to \hat{\mathbb{R}}$ defined in the form*

$$f(x) = \frac{ax+b}{cx+d}, \qquad a, b, c, d \in \mathbb{R}, ad \neq bc,$$

is called a linear fractional transformation *or a* Möbius transformation. *If $c \neq 0$, then $f(-d/c) = \infty$, and $f(\infty) = a/c$. If $c = 0$, then $f(\infty) = \infty$.*

Examples & Exercises

1.21. Show that a Möbius transformation is invertible.

1.22. Show that if f_1 and f_2 are both Möbius transformations then so is $f_1 \circ f_2$.

1.23. What does $f(x) = 1/x$ do to $\hat{\mathbb{R}}$ on the sphere?

1.24. Show that the set of Möbius transformations f such that $f(\infty) = \infty$ is the set of affine transformations.

1.25. Find a Möbius transformation $f : \hat{\mathbb{R}} \to \hat{\mathbb{R}}$ so that $f(1) = 2$, $f(2) = 0$, $f(0) = \infty$. Evaluate $f(\infty)$.

1.26. Figure III.38 shows a Sierpinski triangle before and after the polynomial transformation $x \mapsto ax(x-b)$ has been applied to the x-axis. Evaluate the real constants a and b. Notice how well fractals can be used to illustrate how a transformation acts.

2 Affine Transformations in the Euclidean Plane

Definition 2.1 *A transformation $w : \mathbb{R}^2 \to \mathbb{R}^2$ of the form*

$$w(x_1, x_2) = (ax_1 + bx_2 + e, cx_1 + dx_2 + f),$$

where $a, b, c, d, e,$ and f are real numbers, is called a (two-dimensional) affine *transformation.*

We will often use the following equivalent notations

$$w(x) = w\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} = Ax + t.$$

Here $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a two-dimensional, 2×2 real matrix and t is the column vector $\begin{pmatrix} e \\ f \end{pmatrix}$, which we do not distinguish from the coordinate pair $(e, f) \in \mathbb{R}^2$. Such transformations have important geometrical and algebraic properties. From this point on, we shall assume that the reader is familiar with matrix multiplication.

The matrix A can always be written in the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} r_1 \cos\theta_1 & -r_2 \sin\theta_2 \\ r_1 \sin\theta_1 & r_2 \cos\theta_2 \end{pmatrix},$$

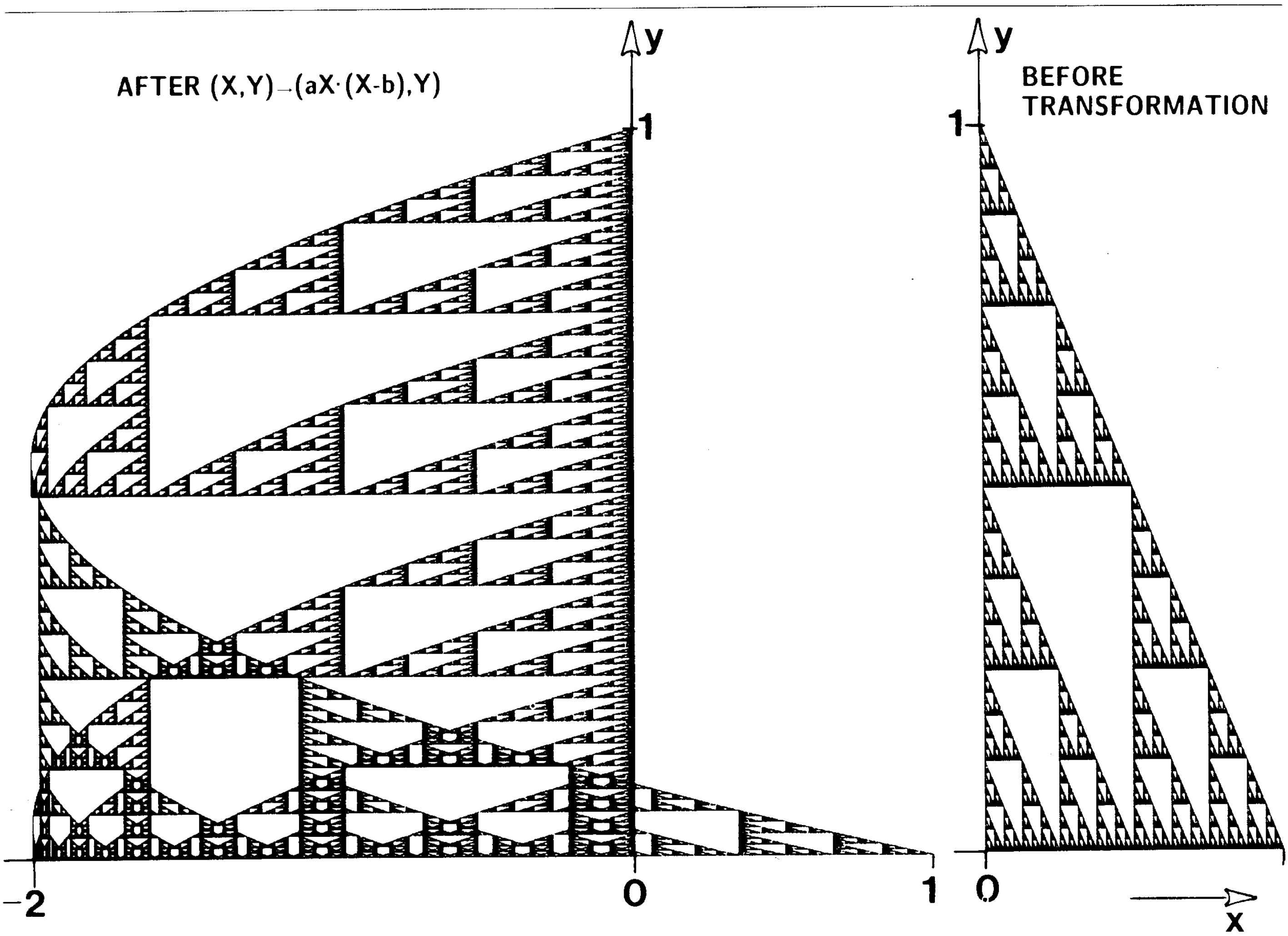


Figure III.38. A Sierpinski triangle before and after the polynomial transformation $x \mapsto ax(x - b)$ is applied to the x-axis. Evaluate the real constants a and b.

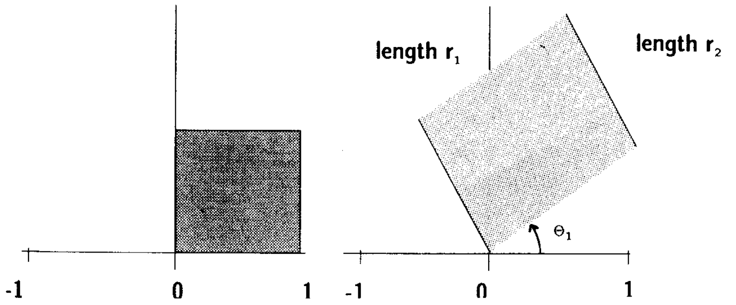


Figure III.39. An affine transformation takes parallelograms into parallelograms.

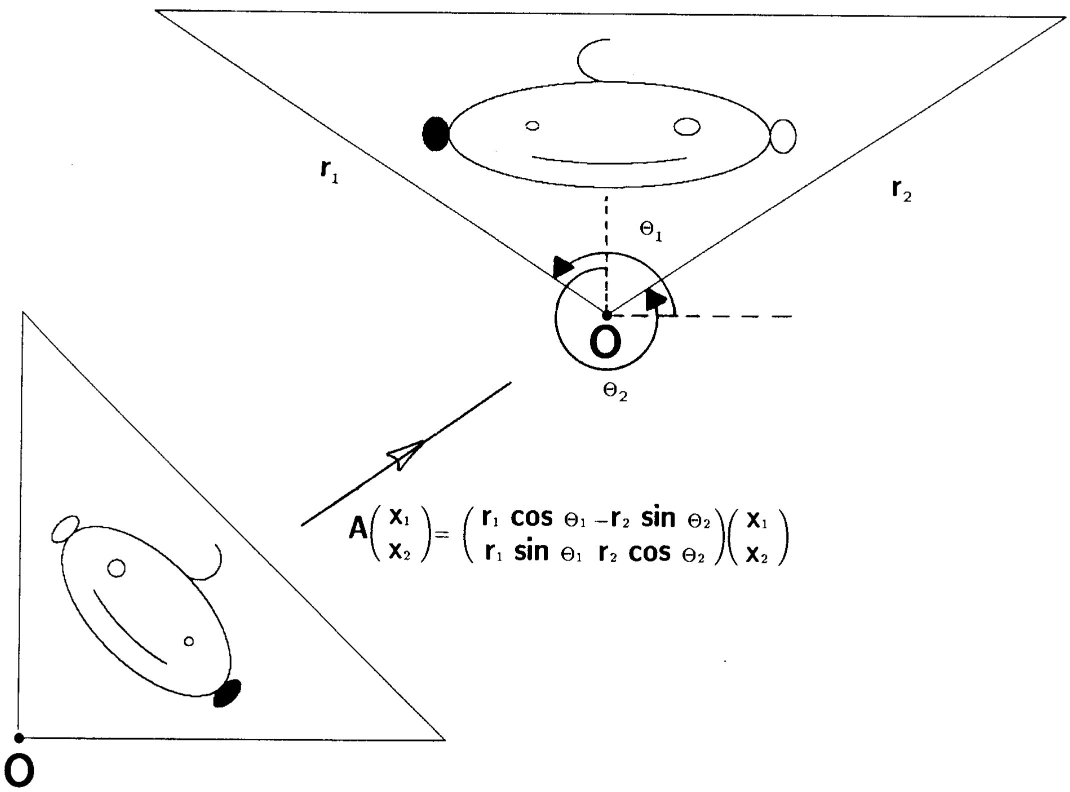


Figure III.40. A linear transformation can turn pictures over.

where (r_1, θ_1) are the polar coordinates of the point (a, c) and $(r_2, (\theta_2 + \pi/2))$ are the polar coordinates of the point (b, d). The *linear* transformation

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

in $\mathbb{R}^2$ maps any parallelogram with a vertex at the origin to another parallelogram with a vertex at the origin, as illustrated in Figure III.39. Notice that the parallelogram may be "turned over" by the transformation, as illustrated in Figure III.40.

The general affine transformation $w(x) = Ax + t$ in $\mathbb{R}^2$ consists of a linear transformation, A which deforms space relative to the origin, as described above, followed by a *translation* or *shift* specified by the vector t (see Figure III.41).

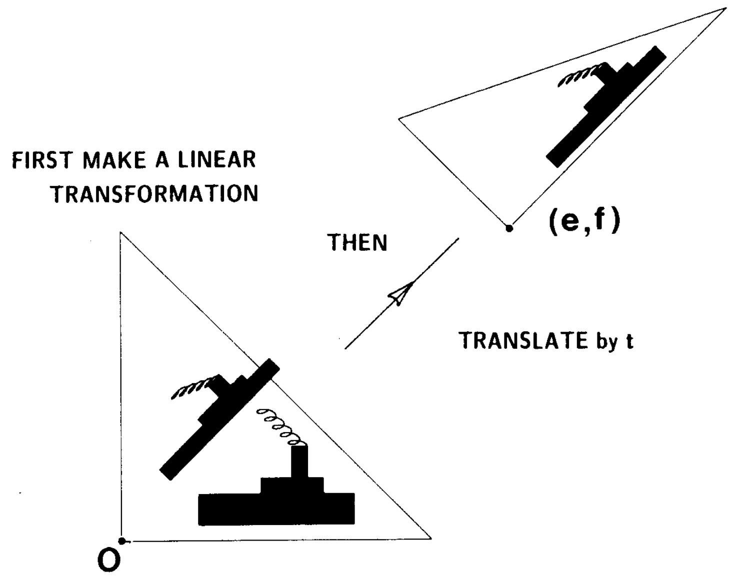


Figure III.41. An affine transformation consists of a linear transformation followed by a translation.

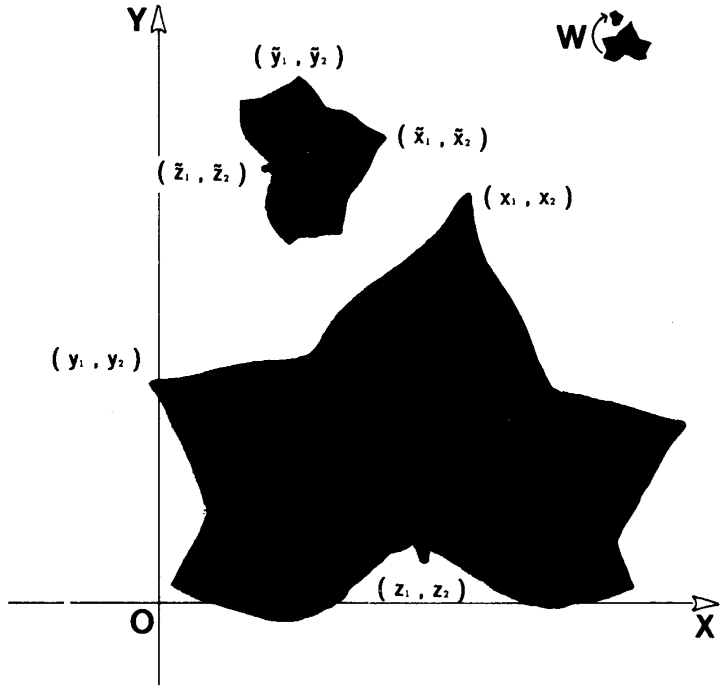


Figure III.42. Two ivy leaves lying on the Euclidean Plane determine an affine transformation.

How can one find an affine transformation that approximately transforms one given set into another given set in $\mathbb{R}^2$? Let's show how to find the affine transformation that almost takes the big leaf to the little leaf in Figure III.42. This figure actually shows a photocopy of two real ivy leaves. We wish to find the numbers a, b, c, d, e, and f defined above, so that

$$w(\text{BIG LEAF}) \text{ approximately equals LITTLE LEAF.}$$

Begin by introducing x and y coordinate axes, as already shown in Figure III.42. Mark three points on the big leaf (we've chosen the leaf tip, a side spike, and the point where the stem joins the leaf) and determine their coordinates (x_1, x_2), (y_1, y_2), and (z_1, z_2). Mark the corresponding points on the little leaf, assuming that a caterpillar hasn't eaten them, and determine their coordinates; say $(\tilde{x}_1, \tilde{x}_2)$, $(\tilde{y}_1, \tilde{y}_2)$, and $(\tilde{z}_1, \tilde{z}_2)$, respectively.

Then a, b, and e are obtained by solving the three linear equations

$$x_1 a + x_2 b + e = \tilde{x}_1,$$

$$y_1 a + y_2 b + e = \tilde{y}_1,$$

$$z_1 a + z_2 b + e = \tilde{z}_1;$$

while c, d, and f satisfy

$$x_1 c + x_2 d + f = \tilde{x}_2,$$

$$y_1 c + y_2 d + f = \tilde{y}_2,$$

$$z_1 c + z_2 d + f = \tilde{z}_2.$$

Examples & Exercises

2.1. Find an affine transformation in $\mathbb{R}^2$ that takes the triangle with vertices at $(0, 0)$, $(0, 1)$, $(1, 0)$ to the triangle with vertices at $(4, 5)$, $(-1, 2)$, and $(3, 0)$. Show what this transformation does to a circle inscribed in the first triangle.

2.2. Show that a necessary and sufficient condition for the affine transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} = Ax + t$$

to be invertible is $\det A \neq 0$, where $\det A = (ad - bc)$ is the determinant of the 2×2 matrix A.

2.3. Show that if $f_1 : \mathbb{R}^2 \to \mathbb{R}^2$ and $f_2 : \mathbb{R}^2 \to \mathbb{R}^2$ are both affine transformations, then so is

$$f_3 = f_1 \circ f_2.$$

If $f_i(x) = A_i x + t_i$, $i = 1, 2, 3$, where A_i is a 2×2 real matrix, express A_3 in terms of A_1 and A_2.

2.4. Let A and B be 2×2 matrices, with determinants $\det A$ and $\det B$, respectively. Show that the determinant of the product is the product of the determinants, i.e.,

$$\det(AB) = \det A \cdot \det B.$$

Definition 2.2 *A transformation $w : \mathbb{R}^2 \to \mathbb{R}^2$ is called a* similitude *if it is an affine transformation having one of the special forms*

$$w\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

$$w\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta & r\sin\theta \\ r\sin\theta & -r\cos\theta \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

for some translation $(e, f) \in \mathbb{R}^2$, some real number $r \neq 0$, and some angle $\theta, 0 \leq \theta < 2\pi$. θ is called the rotation angle while r is called the scale factor *or* scaling. *The linear transformation*

$$R_\theta\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is a rotation. *The linear transformation*

$$R\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is a reflection.

Figure III.43 shows some of the things a similitude can do. Notice that a similitude preserves angles.

Examples & Exercises

2.5. Find the scaling ratios r_1, r_2 and the rotation angles θ_1, θ_2 for the affine transformation that takes the triangle $(0,0)$, $(0,1)$, $(1,0)$ onto the straight-line segment from $(1,1)$ to $(2,2)$ in $\mathbb{R}^2$ in such a way that both $(0,1)$ and $(1,0)$ go to $(1,1)$.

2.6. Let S be a region in $\mathbb{R}^2$ bounded by a polygon or other "nice" boundary. Let $w : \mathbb{R}^2 \to \mathbb{R}^2$ be an affine transformation, $w(x) = Ax + t$. Show that

$$(\text{area of } w(S)) = |\det A| \cdot (\text{area of } S);$$

see Figure III.44. Show that $\det A < 0$ has the interpretation that S is "flipped over" by the transformation. (Hint: suppose first that S is a triangle.)

2.7. Show that if $w : \mathbb{R}^2 \to \mathbb{R}^2$ is a similitude, $w(x) = Ax + t$, where t is the translation and A is a 2×2 matrix, then A can always be written either $A = rR_\theta$ or $A = rRR_\theta$.

2.8. View the railway tracks image in Figure III.45 as a subset S of $\mathbb{R}^2$. Find a similitude $w : \mathbb{R}^2 \to \mathbb{R}^2$ such that $w(S) \subset S$, $w(S) \neq S$.

2.9. We use the notation introduced in Definition 2.2. Find a nonzero real number r, an angle θ, and a translation vector t such that the similitude $wx = rR_\theta x + t$ on $\mathbb{R}^2$ obeys

$$w(\blacktriangle) \subset \blacktriangle, \qquad \text{with } w(\blacktriangle) \neq \blacktriangle,$$

where $\blacktriangle$ denotes a Sierpinski triangle with vertices at $(0,0)$, $(1,0)$, and $(\frac{1}{2}, 1)$.

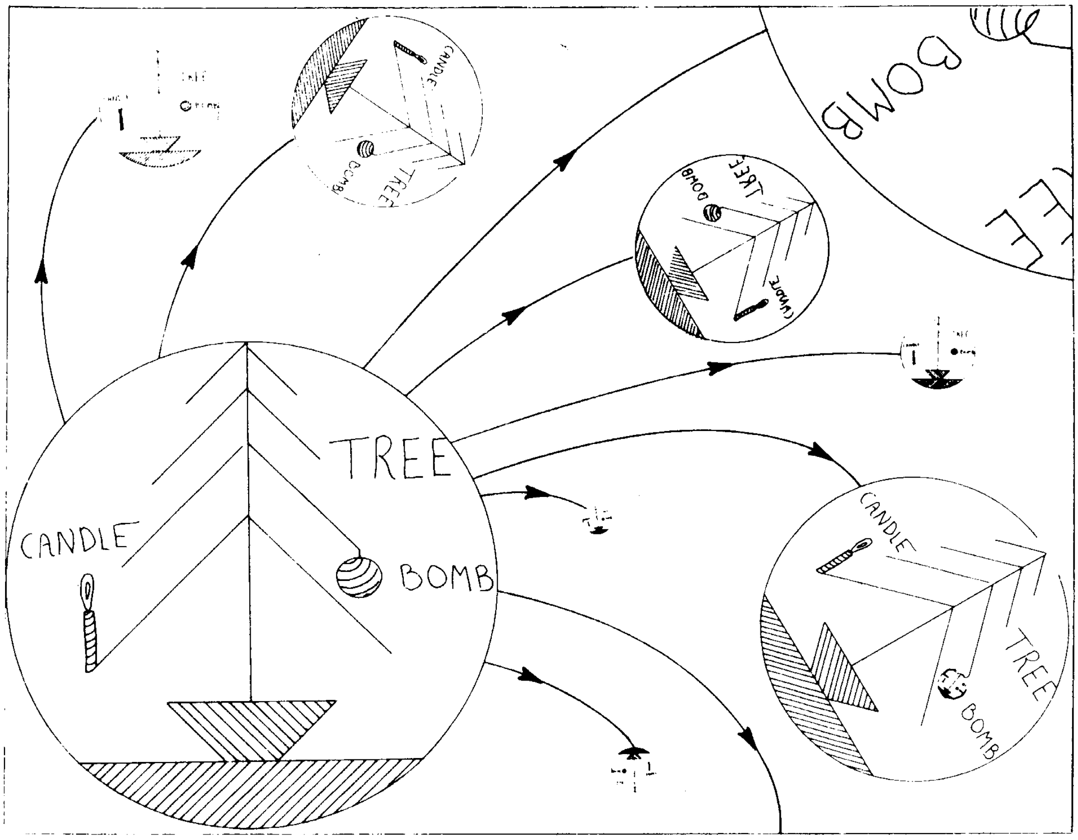


Figure III.43. Some of the things that a similitude can do.

2.10. Show that if $w : \mathbb{R}^2 \to \mathbb{R}^2$ is affine, $w(x) = Ax + t$, then it can be reexpressed

$$w(x) = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} R_\theta \begin{pmatrix} r_3 & 0 \\ 0 & r_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + t,$$

where $r_i \in \mathbb{R}$ and $0 \le \theta < 2\pi$. We call a transformation of the form

$$w\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

a coordinate rescaling.

2.11. Let S denote the two-dimensional orchard subset of $\mathbb{R}^2$ shown in Figure III.46. Find two fundamentally different affine transformations that map S into S but not onto S. Define the transformations by specifying how they act on three points.

2.12. Show that if A is a 2×2 matrix such that $\det A \neq 0$, with

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then the inverse of A, denoted A^{-1}, is given by

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \frac{d}{\det A} & \frac{-b}{\det A} \\ \frac{-c}{\det A} & \frac{a}{\det A} \end{pmatrix}.$$

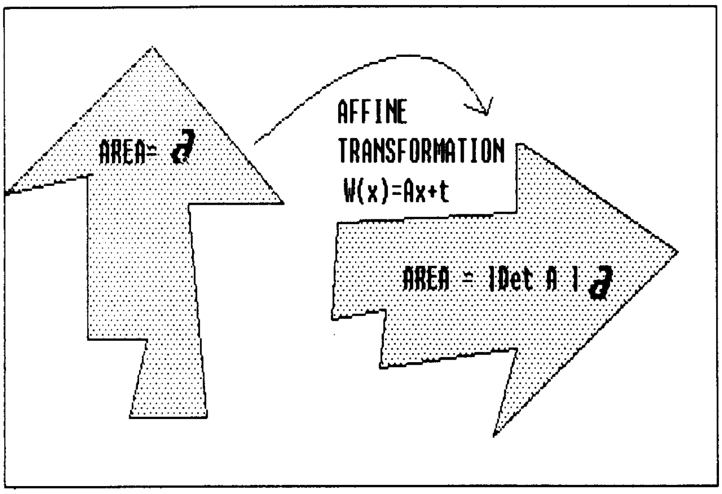
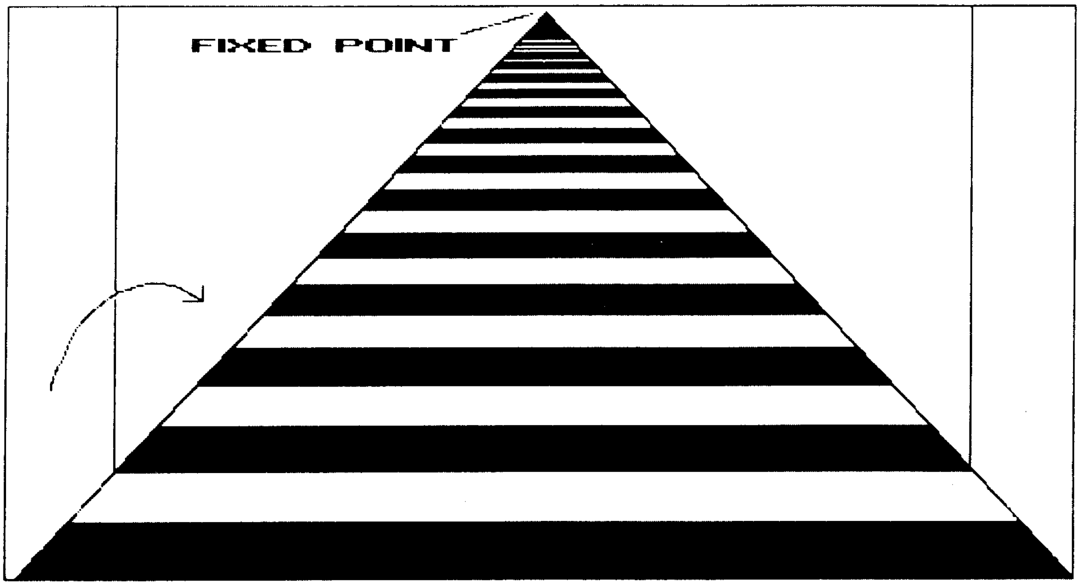


Figure III.44. The scaling factor by which an affine transformation changes area is determined by the determinant of its linear part.



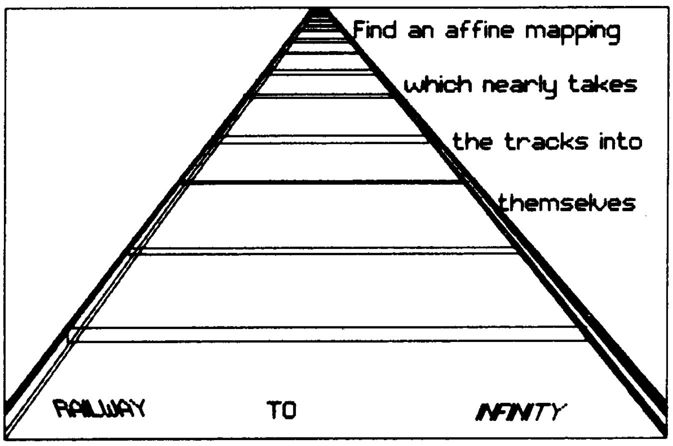


Figure III.45. Railway to infinity. Can you find an affine transformation that nearly maps the track ties into themselves?

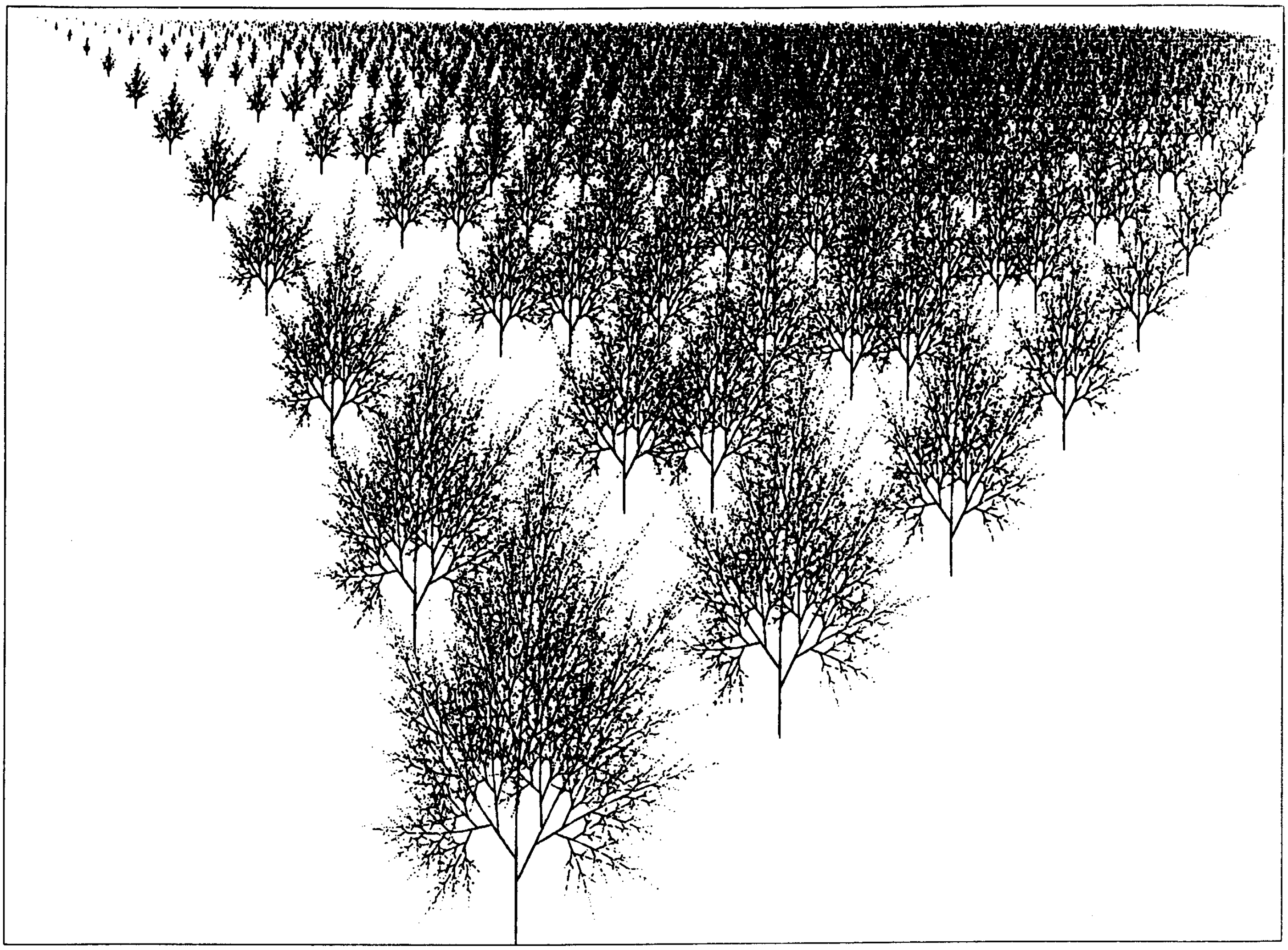


Figure III.46. Orchard subset of $\mathbb{R}^2$. Can you find some interesting affine transformations that map this set into itself?

2.13. The *trace* of a matrix A is the sum of the elements along the diagonal, that is

$$\operatorname{tr} A = \sum a_{ii}.$$

Let A be a 2×2 matrix, and let B be another 2×2 matrix such that $\det B \neq 0$. Show that

$$\operatorname{tr}(BAB^{-1}) = \operatorname{tr} A$$

and

$$\det(BAB^{-1}) = \det A.$$

2.14. Let $w(x) = Ax$ denote a linear transformation in the metric space $(\mathbb{R}^2, D)$ where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Define the *norm* of a point $x \in \mathbb{R}^2$ to be $|x| = D(x, O)$, where O denotes the origin. Define the norm of the linear transformation A by

$$|A| = \max\left\{\frac{|Ax|}{|x|} : x \in \mathbb{R}^2, x \neq 0\right\}$$

when this maximum exists. Show that $|A|$ is defined when D is the Euclidean metric and when it is the Manhattan metric. Find an expression for $|A|$ in terms of a, b, c, and d in each case. Make a geometrical interpretation of $|A|$. Show that when $|A|$ exists we have

$$|Ax| \leq |A| \cdot |x| \qquad \text{for all } x \in \mathbb{R}^2.$$

3 Möbius Transformations on the Riemann Sphere

Definition 3.1 *A transformation* $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ *defined by*

$$f(z) = \frac{(az+b)}{(cz+d)},$$

where $a, b, c,$ *and* $d \in \mathbb{C}$, $ad - bc \neq 0$, *is called a* Möbius transformation *on* $\hat{\mathbb{C}}$. *If* $c \neq 0$ *then* $f(-d/c) = \infty$, *and* $f(\infty) = a/c$. *If* $c = 0$, *then* $f(\infty) = \infty$.

As shown by the following exercises and examples, one can think of a Möbius transformation as follows. Map the whole plane $\mathbb{C}$, together with the point at infinity, onto the sphere $\hat{\mathbb{C}}$, as described in Chapter II. A sequence of operations is then applied to the sphere. Each operation is elementary and has the property that it takes circles to circles. The possible operations are rotation about an axis, rescaling (uniformly expand or contract the sphere), and translation (the whole sphere is picked up and moved to a new place on the plane, without rotation). Finally, the sphere is mapped back onto the plane in the usual way. Since the mappings back and forth from the plane to the sphere take straight lines and circles in the plane to circles on the sphere, we see that a Möbius transformation transforms the set of straight lines and circles in the plane onto itself. We also see that a Möbius transformation is invertible. It is wonderful how the quite complicated geometry of Möbius transformations is handled by straightforward complex algebra, where we simply manipulate expressions of the form $(az+b)/(cz+d)$.

Examples & Exercises

3.1. Show that the most general Möbius transformation, which maps ∞ to ∞, is of the form $f(z) = az + b$, $a, b \in \mathbb{C}$, $a \neq 0$, and that this is a similitude. Show that any

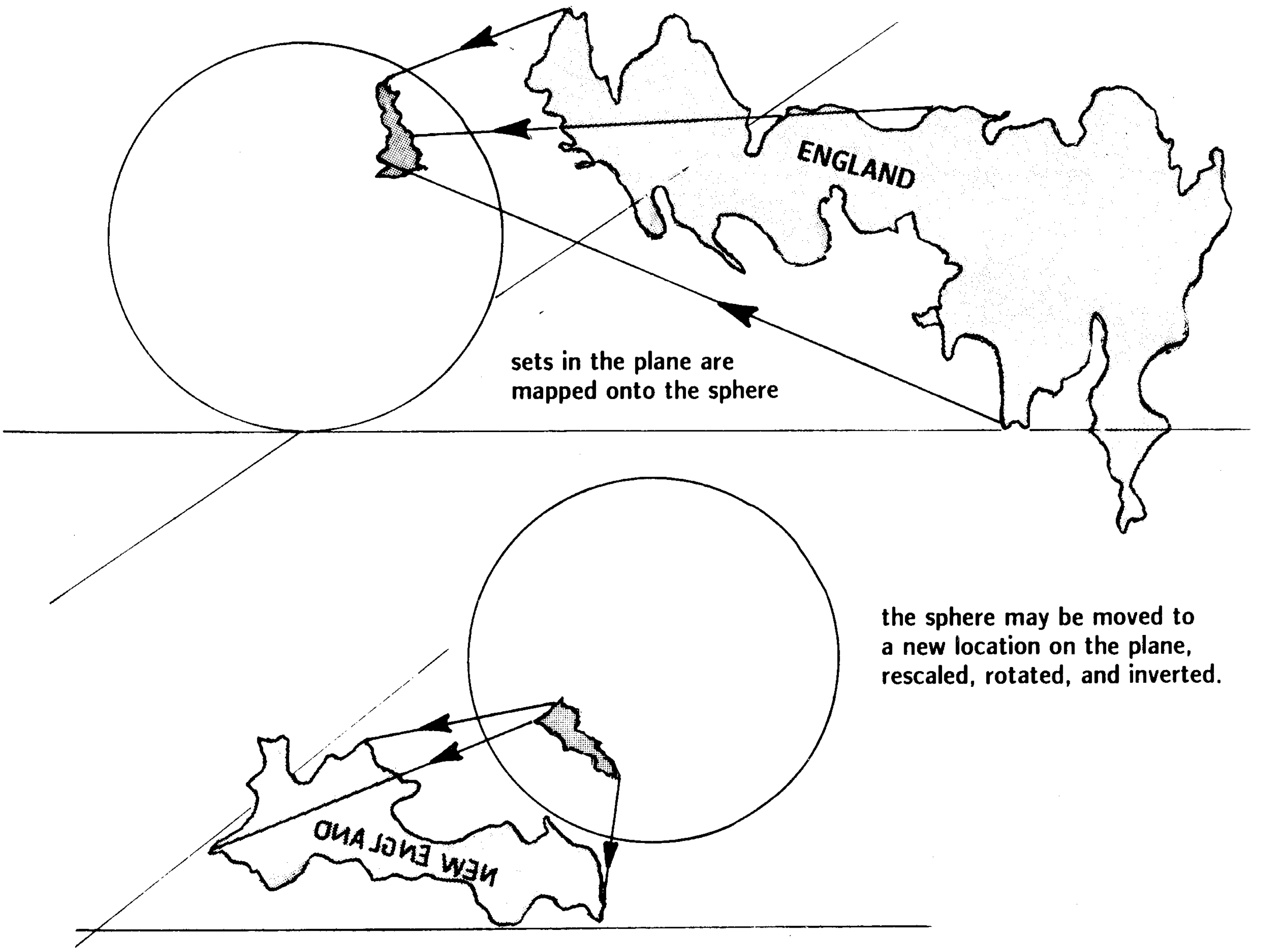


Figure III.47. A Möbius transformation acting on England to produce a new country.

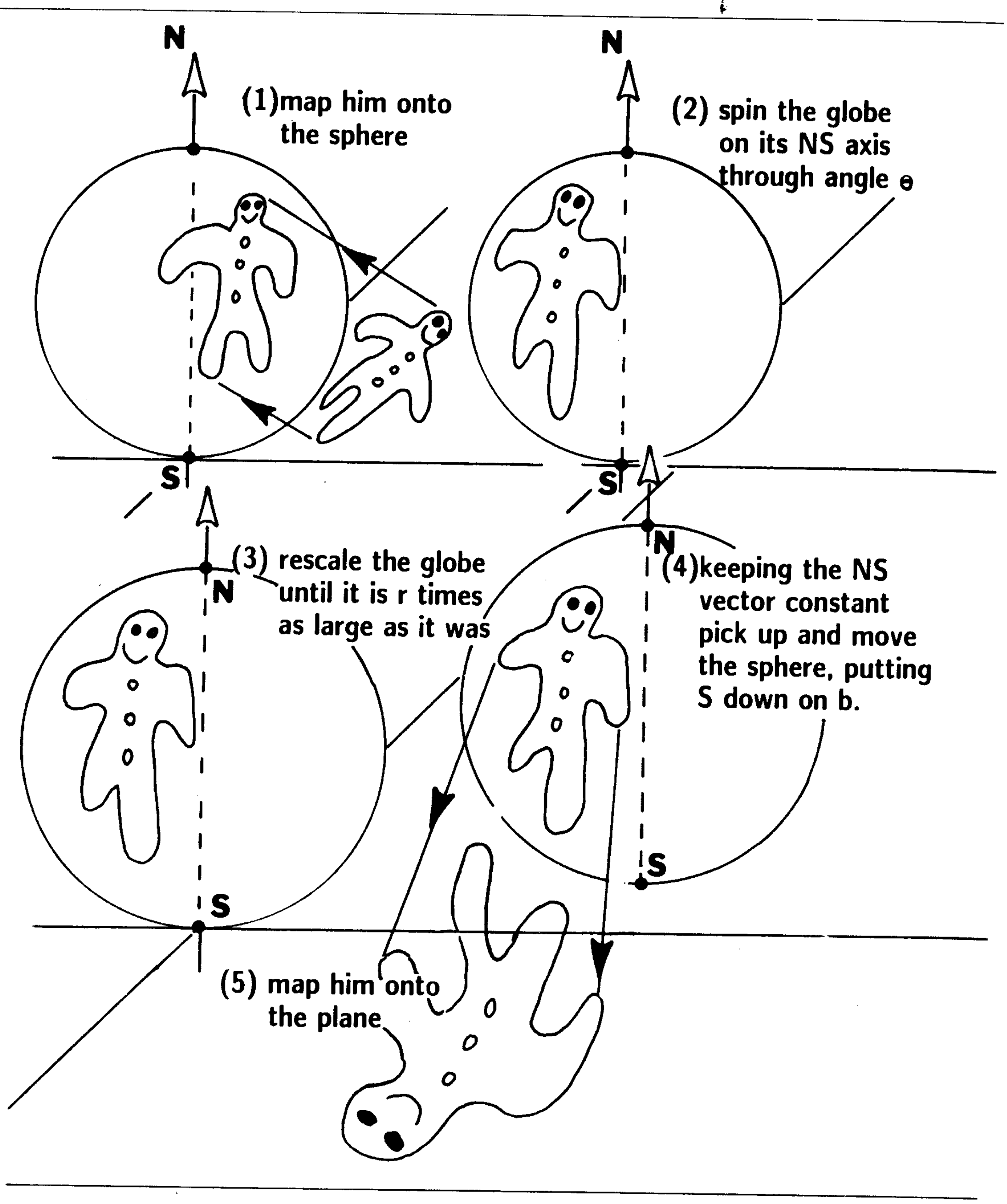
two-dimensional similitude that does not involve a reflection can be written in this form. That is, disregarding changes in notation,

$$\begin{aligned} f(z) &= f(x_1 + ix_2) = (a_1 + ia_2)(x_1 + ix_2) + (b_1 + ib_2) \\ &= re^{i\theta}(x_1 + ix_2) + (b_1 + ib_2), \qquad (i = \sqrt{-1}) \\ &= \begin{pmatrix} r\cos\theta & -r\sin\theta \\ r\sin\theta & r\cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \end{aligned}$$

Find r and θ in terms of a_1 and a_2. Show that the transformation can be achieved as illustrated in Figure III.48.

3.2. Show that the Möbius transformation $f(z) = 1/z$ corresponds to first mapping the plane to the sphere in such a way that the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ goes to the equator, followed by an inversion of the sphere (turn it upside down by rotating

Figure III.48. The mechanism of the similitude $f(z) = re^{i\theta}z + b$ in terms of the sphere.



about an axis through $+1$ and -1 on the equator), and finally mapping back to the plane.

3.3. Show that any Möbius transformation that is not a similitude may be written

$$f(z) = e + \frac{f}{z+g} \qquad \text{for some } e, f, g \in \mathbb{C}, f \neq 0.$$

3.4. Sketch what happens to the picture in Figure III.49 under the Möbius transformation $f(z) = \frac{1}{z}$.

3.5. What happens to Figure III.49 under the Möbius transformation $f(z) = 1 + iz$?

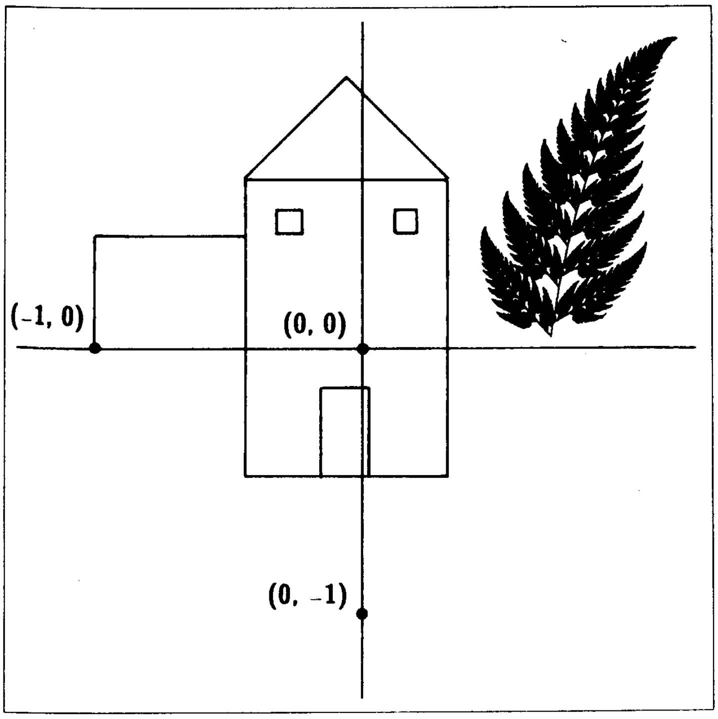


Figure III.49. Up the Garden Path. What does the Möbius transformation $z \mapsto 1 + iz$ do to this picture?

3.6. Show algebraically that a Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is always invertible.

3.7. Show that if f_1 and f_2 are Möbius transformations, then $f_1 \circ f_2$ is a Möbius transformation.

3.8. Find a Möbius transformation that takes the real line to the unit circle centered at the origin.

3.9. Evaluate $f^{\circ n}(z)$ if $f(z) = 1/(1+z)$, $n \in \{-2, -1, 0, 1, 2, 3, \ldots\}$.

3.10. Interpret the Möbius transformation $f(z) = i + 1/(z - i)$ in terms of operations on the sphere.

4 Analytic Transformations

In this section we continue the discussion of transformations on the metric spaces ($\mathbb{C}$, Euclidean) and ($\hat{\mathbb{C}}$, Spherical). We introduce a generalization of the Möbius transformations, called analytic transformations. We concentrate on the behavior of quadratic transformations. It is recommended that during a first reading or first course the reader obtains a good mental picture of how the quadratic transformation acts on the sphere. The reader may then want to study this section more closely after reading about Julia sets in Chapter VII.

The similitude $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by the formula $f(z) = 3z + 1$ is an example of an analytic transformation. It maps circles to circles magnified by a factor of three. A disk with center at z_0 is taken to a disk with center at $f(z_0) = 3z_0 + 1$. The tranformation is continuous, and it maps open sets to open sets. Nowhere does it "fold back along the dotted line."

The similitude $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by $f(z) = (3 + 3i)z + (1 - 2i)$ is similarly described. The circles and disks are now rotated by 45° in addition to being magnified and translated.

Loosely a transformation on $\hat{\mathbb{C}}$ is analytic if it is continuous and it locally "behaves like" a similitude. If you take a very small region indeed (How small? Small enough! There is a smallness such that what is about to be said is true!) and you watch what the transformation does to that tiny region, you will typically find that it is magnified or shrunk, rotated, and translated, in almost exactly the same manner that some similitude would do the job. The similitude will always be of the special type discussed in exercise 3.1 above.

We make this description more precise. Let us decide to look at what our transformation does in the vicinity of a point $z_0 \in \hat{\mathbb{C}}$. Assume that z_0 is not a critical point, defined below. Let T denote a tiny region, a disk for example, which contains the point z_0. Let $f(T)$ be its image under the transformation. Then one can rescale T by a factor that makes it roughly the size of the unit square, and one can rescale $f(T)$ by the same factor. The assertion of the previous paragraph is that the action of the transformation, viewed as taking T, rescaled, onto $f(T)$, rescaled, can be described more accurately by a similitude. If you like, one could consider a picture P drawn in T and examine the transformed image $f(P)$: if P and $f(P)$ are rescaled by the same factor so that P is the size of the unit square, then $f(P)$ looks more and more like a similitude applied to P. This description becomes more and more precise the tinier the region under discussion.

Consider the quadratic transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by

$$f(z) = z^2 = (x_1 + ix_2)^2 = (x_1^2 - x_2^2) + 2x_1x_2 i = f_1(x_1, x_2) + f_2(x_1, x_2)i,$$

where $f_1(x_1, x_2) = (x_1^2 - x_2^2)$ is called the real part of $f(z)$, and $f_2(x_1, x_2) = 2x_1x_2$ is called the imaginary part of f. Pictures of what this transformation does to some Sierpinski triangles in $\mathbb{C}$ are illustrated in Figure III.50.

Two features are to be noticed. (I) Provided that we stay away from the origin, the transformation behaves locally like a similitude: for points z close to z_0, $f(z)$ is approximated by the similitude

$$w(z) = az + b \qquad \text{where } a = 2z_0 \text{ and } b = -z_0^2.$$

This fact shows up in Figure III.50: upon close examination (we suggest the use of a magnifying glass here) of the transformed Sierpinski triangles, one sees that they are built up out of small triangles whose shapes are only slightly different from that of their preimages. The only place where this is not true is at the forward image of the

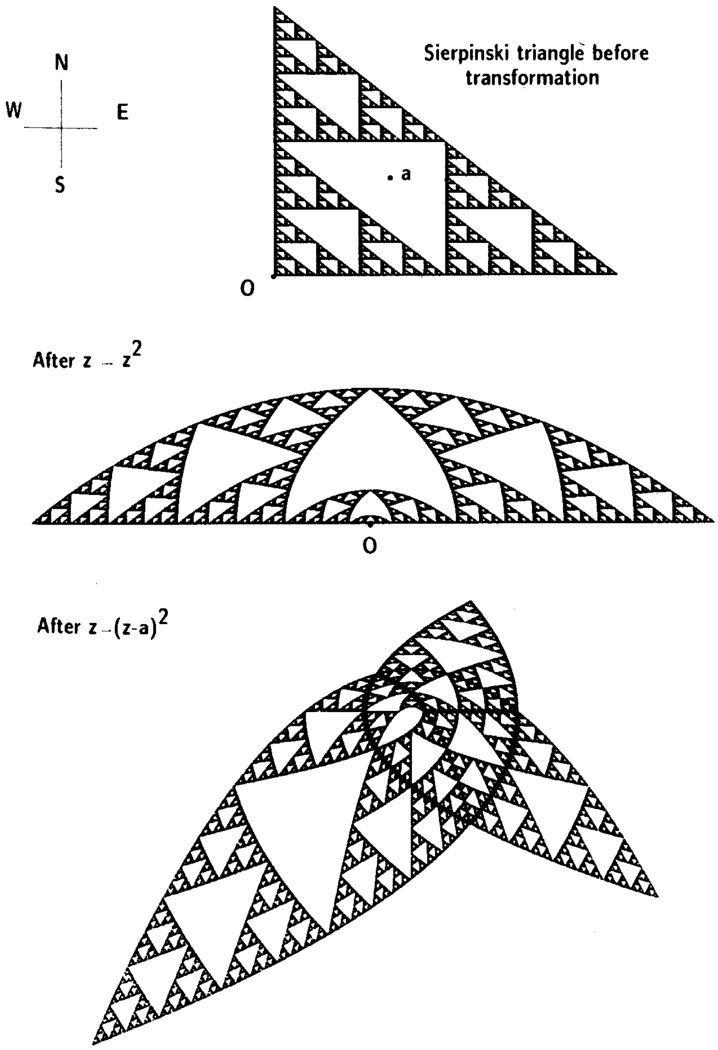


Figure III.50. Quadratic transformations are described by showing how they act on a Sierpinski triangle. Use a magnifying glass to check that the transformations behave locally like similitudes.

origin, which is a critical point. (II) The transformation maps the space twice around the origin.

One can track analytically what happens to the point

$$z = R\cos t + iR\sin t,$$

where $R > 0$. As the time parameter t goes from zero to 2π, z moves anticlockwise once around the circle of radius R. The transformed point $f(z)$ is given by

$$f(z) = R^2\cos 2t + iR^2\sin 2t.$$

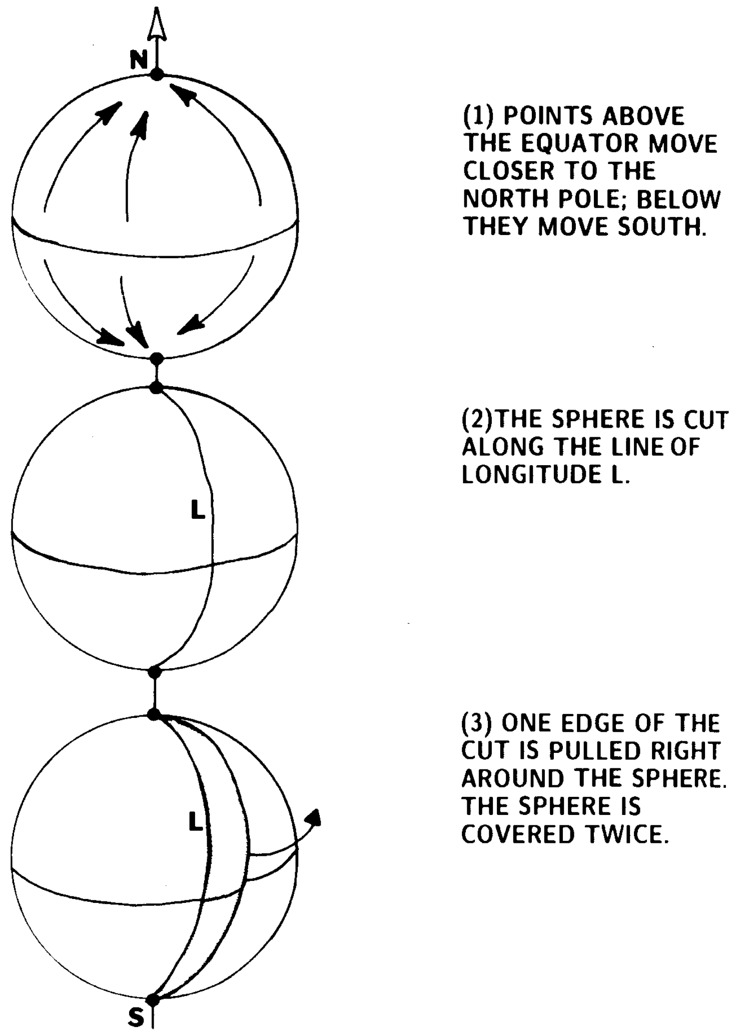
As the time parameter t goes from 0 to 2π, $f(z)$ moves twice around the circle of radius R^2.

On the Riemann sphere the transformation $z \mapsto z^2$ can be described as follows. Let us say that the Equator corresponds to the circle of unit radius in the plane, that the South Pole corresponds to the Origin, and that the North Pole corresponds to the

Point at Infinity. Then the transformation leaves both Poles fixed. The Line of Longitude L connecting the Poles, which corresponds to the positive real axis, is mapped into itself, and the Equator is mapped into itself. Here is what we must picture. First, points that lie above the Equator are moved closer to the North Pole; points that lie below the Equator are moved closer to the South Pole; and the Equator is not shifted. Second, the skin of the sphere is cut along the Line of Longitude L. One side of the cut is held fixed while the other side is pulled around the sphere (following the terminator when the Sun is high above the Equator), uniformly stretching the space, until the edge of the cut is back over L. The two lips of the the cut are rejoined. The sphere has been mapped twice over itself. The Poles are the critical points of the transformation; they are the points about which wrapping occurs. This description is illustrated in Figure III.51.

The most general quadratic transformation on the sphere is expressible by a formula of the form $f(z) = Az^2 + Bz + C$, where A, B, and C are complex numbers.

Figure III.51. The action of the quadratic transformation $z \mapsto z^2$ in terms of the sphere.



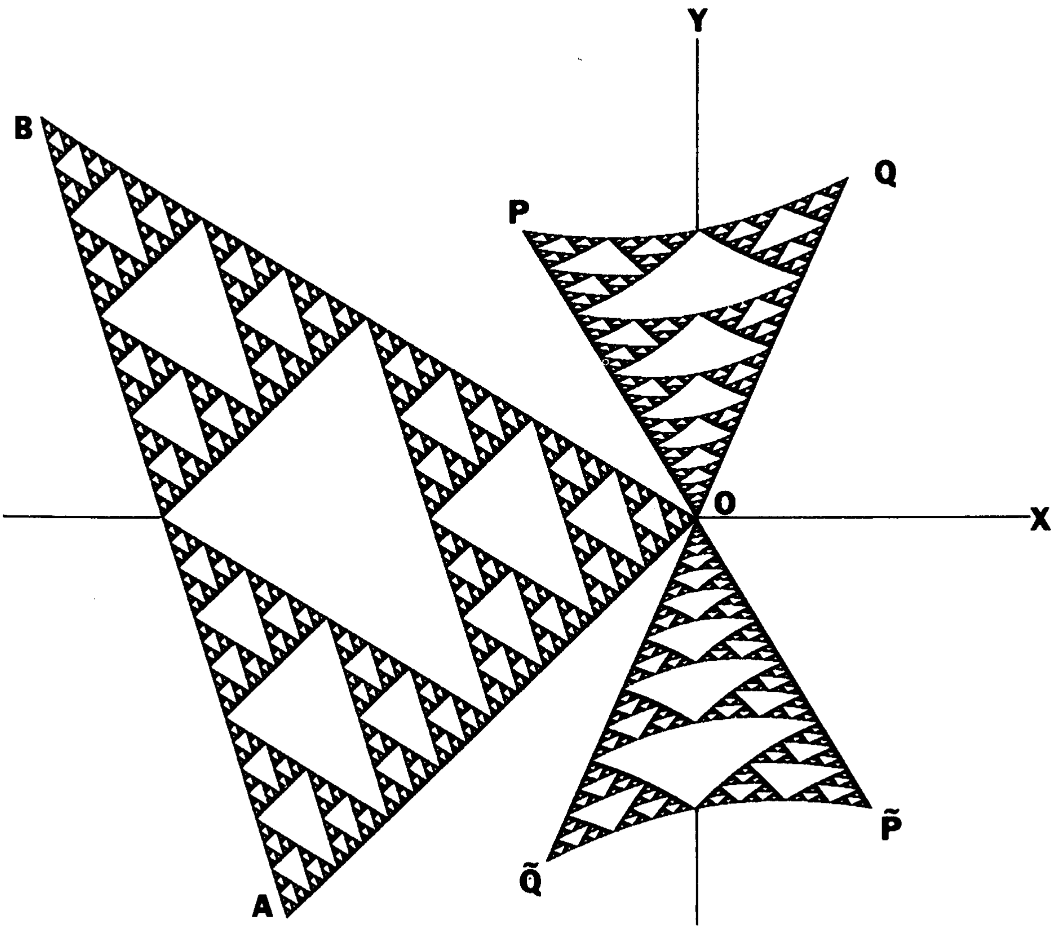


Figure III.52. The set valued inverse, f^1, of the quadratic transformation $f(z) = z^2$, maps the Sierpinski triangle AOB into $POQ \cup \tilde{P}O\tilde{Q}$. More generally, f^{-1} maps the Space of Fractals into itself. Look carefully at this image! Several important features of analytic transformations are illustrated here.

One can show there is a change of coordinates, $z \mapsto \theta(z)$, where θ is a similitude, such that $f(z)$ becomes expressible in the special form $f(z) = z^2 + \tilde{C}$ for some complex number $\tilde{C}$; see Exercise 5.20 in the following section. Hence the description of the most general quadratic transformation on the sphere can be made in the same terms as above, except that at the end there is a translation by some constant amount $\tilde{C}$. This translation leaves the Point at Infinity fixed.

The quadratic transformation $f(z) = z^2$ maps the punctured plane $\mathbb{C}$ onto itself twice. Each point on $z \in \mathbb{C}\{0\}$ has two preimages. Hence $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is not an invertible transformation. In such situations we can define a set-valued inverse function.

Definition 4.1 *Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be an analytic transformation such that $f(\hat{\mathbb{C}}) = \hat{\mathbb{C}}$. Then the* set-valued inverse *of f is the mapping $f^{-1} : \mathcal{H}(\hat{\mathbb{C}}) \to \mathcal{H}(\hat{\mathbb{C}})$ defined by*

$$f^{-1}(A) = \{w \in \hat{\mathbb{C}} : f(w) \in A\} \qquad \text{for all } A \in \mathcal{H}(X).$$

In Figure III.52 we illustrate the transformation f^{-1} acting on the Space of Fractals, in the case of the quadratic transformation $f(z) = z^2$.

One can obtain explicit formulas for $f^{-1}(z)$ when f is a quadratic transformation. For example for $f(z) = z^2$, $f^{-1}(O) = O$, $f^{-1}(\infty) = \infty$, and $f^{-1}(z) = \{w_1(z), w_2(z)\}$ for $z \in \hat{\mathbb{C}} \setminus \{0, \infty\}$. Here $w_1(x_1 + ix_2) = a(x_1, x_2) + ib(x_1, x_2)$, and $w_2(x_1, x_2) = -a(x_1, x_2) - ib(x_1, x_2)$, where

$$a(x_1, x_2) = \sqrt{\frac{\sqrt{x_1^2 + x_2^2} + x_1}{2}} \qquad \text{when } x_2 \geq 0,$$

$$a(x_1, x_2) = -\sqrt{\frac{\sqrt{x_1^2 + x_2^2} + x_1}{2}} \qquad \text{when } x_2 < 0,$$

$$b(x_1, x_2) = \sqrt{\frac{\sqrt{x_1^2 + x_2^2} - x_1}{2}}.$$

Each of the two functions $w_1(z)$ and $w_2(z)$ is itself analytic on $\mathbb{C} \setminus \{0, \infty\}$.

The following definition formalizes what is meant by an analytic transformation on the complex plane. We recommend further reading, for example [Rudin, 1966].

Definition 4.2 *Let* $(\mathbb{C}, d)$ *denote the complex plane with the Euclidean metric. A transformation* $f : \mathbb{C} \to \mathbb{C}$ *is called* analytic *if for each* $z_0 \in \mathbb{C}$ *there is a similitude of the form*

$$w(z) = az + b, \qquad \text{for some pair of numbers } a, b \in \mathbb{C},$$

such that $d(f(z), w(z))/d(z, z_0) \to 0$ *as* $z \to z_0$. *The numbers a and b depend on* z_0. *If, corresponding to a certain point* $z_0 = c$, *we have* $a = 0$, *then c is called a* critical point *of the transformation, and* $f(c)$ *is called a* critical value.

If the analytic transformation $f(z)$ is a rational transformation, which means that it is expressible as a ratio of two polynomials in z, such as

$$\text{(i)} f(z) = 1 + 2i + 27z^2 - 9z^3,$$

$$\text{(ii)} f(z) = \frac{1+z}{1-z},$$

$$\text{(iii)} f(z) = \frac{1+z+z^2}{1-z+z^2};$$

then the numbers a and b in the similitude $w(z)$ in Definition 4.2 are given by the formulas

$$a = f'(z_0) \text{ and } b = f(z_0) - az_0.$$

The derivative $f'(z)$ of the rational function $f(z)$ can be calculated by treating z as though it were the real variable x and applying the standard differentiation rules of calculus. The critical points $c \in \mathbb{C}$ are the solutions of the equation $f'(c) = 0$.

For example, close enough to any point $z_0 \in \mathbb{C}$ such that $f'(z_0) \neq 0$, the cubic transformation (i) is well described by the similitude

$$w(z) = (54z_0 - 27z_0^2)z + (1 + 2i - 27z_0^2 + 18z_0^3).$$

The finite critical points associated with (i) may be obtained by solving

$$54c - 27c^2 = 0$$

and are accordingly $c = 0 + i0$ and $c = 2 + i0$. By making the change of coordinates $z' = 1/z$ (see section 5), one can also analyze the behavior near the point at infinity. It turns out that $c = \infty$ is always a critical point for a polynomial transformation $f(z)$ on $\hat{\mathbb{C}}$. The space is "wrapped" an integral number of times about the image of critical point. For example, the cubic transformation (i) wraps space twice about each of the points $f(0 + i0) = 1 + 2i$, and $f(2 + i0) = 37 + 2i$, and it wraps it three times about $f(\infty) = \infty$.

Examples & Exercises

4.1. Sketch a globe representing $\hat{\mathbb{C}}$, including a subset that looks like Africa, and show what happens to the subset under the quadratic transformation $f(z) = z^2$.

4.2. Verify the following explicit formulas for $f^{-1}(z)$, corresponding to $f(z) = z^2 - 1 : f^{-1}(-1) = O$; $f^{-1}(\infty) = \infty$; and $f^{-1}(z) = \{w_1(z), w_2(z)\}$ for $z \in \hat{\mathbb{C}} \setminus \{-1, \infty\}$, where $w_1(x_1 + ix_2) = a(x_1, x_2) + ib(x_1, x_2)$, and $w_2(x_1, x_2) = -a(x_1, x_2) - ib(x_1, x_2)$. Here

$$a(x_1, x_2) = \sqrt{\frac{\sqrt{(1 + x_1)^2 + x_2^2} + 1 + x_1}{2}} \qquad \text{when } x_2 \geq 0,$$

$$a(x_1, x_2) = -\sqrt{\frac{\sqrt{(1 + x_1)^2 + x_2^2} + 1 + x_1}{2}} \qquad \text{when } x_2 < 0,$$

and

$$b(x_1, x_2) = \sqrt{\frac{\sqrt{(1 + x_1)^2 + x_2^2} - 1 - x_1}{2}}.$$

Both $w_1(z)$ and $w_2(z)$ are analytic on $\mathbb{C} \setminus \{-1\}$.

4.3. Locate the critical points and critical values of the quadratic transformation $f(z) = z^2 + 1$.

4.4. Draw a side view of a man with an arm stretched out in front of him, holding a knife. The blade should point down. Choose the origin of coordinates to be his navel. Draw another picture to explain how *hara-kiri* can be achieved by applying the inverse of the quadratic transformation $f(z) = z^2$ to your image.

4.5. Find a similitude that approximates the behavior of the given analytic transformation in the vicinity of the given point: (a) $f(z) = z^2$ near $z_0 = 1$; (b) $f(z) = 1/z$ near $z_0 = 1 + i$; (c) $f(z) = (z - 1)^3$ near $z_0 = 1 - i$.

5 How to Change Coordinates

In describing transformations on spaces we usually make use of an underlying coordinate system. Most spaces have a coordinate system by means of which the points in the space are located. This underlying coordinate system is implied by the specification of the space: for example, $\mathbf{X} = [1, 2]$ provides a collection of points together with the natural coordinate x restricted by $l \le x \le 2$. We can think of either the space, made of points $x \in \mathbf{X}$, or equivalently the system of coordinates. If the space $\mathbf{X}$ is $\mathbb{R}^2$ or $\mathbb{C}$, then the underlying coordinate system may be Cartesian coordinates. If $\mathbf{X} = \hat{\mathbb{C}}$, then the coordinate system may be angular coordinates on the sphere.

In each case the underlying coordinate system is itself a subset of a metric space. We denote this metric space by $\mathbf{X}_C$. Usually we do not consciously distinguish between a point $x \in \mathbf{X}$ and its coordinate $x \in \mathbf{X}_C$. Notice, however, that the space $\mathbf{X}_C$ may contain points (coordinates) that do not correspond to any point in the space $\mathbf{X}$. For example, in the case of the space $\mathbf{X} = \blacksquare$ it is natural to take $\mathbf{X}_C = \mathbb{R}^2$; then points $x \in \mathbf{X}$ in the space correspond to coordinates $x = (x_1, x_2) \in \mathbf{X}_C$ restricted by $0 \le x_1 \le 1$ and $0 \le x_2 \le 1$. However, the coordinates $(3, 5) \in \mathbf{X}_C$ do not correspond to a point in $\mathbf{X}$. We would like the reader to think of the space itself as "lying above" its coordinate system, as suggested in Figure III.53.

A change of coordinate system may be described by a transformation $\theta : \mathbf{X}_C \to \mathbf{X}_C$. We can think of a change of coordinates being effected by physically moving each point $x \in \mathbf{X}$ so that it no longer lies above $x \in \mathbf{X}_C$ but instead above the coordinate $x' = \theta(x) \in \mathbf{X}_C$. Thus we must now distinguish between a point x lying in the space, $\mathbf{X}$, from its coordinate $x \in \mathbf{X}_C$. Then we want to think of the change of coordinates $\theta : \mathbf{X}_C \to \mathbf{X}_C$ as moving $\mathbf{X}$ relative to the underlying coordinate space $\mathbf{X}_C$, as illustrated in Figure III.54.

Example

5.1. Let $\mathbf{X} = [1, 2]$ and take $\mathbf{X}_C$ to be $\mathbb{R}$. Let $\theta : \mathbb{R} \to \mathbb{R}$ be defined by $\theta(x) = 2x + 1$. Then the coordinate of the point $x = 1.5$ becomes changed to 4. We want to think of the space $\mathbf{X}$ as being moved relative to the coordinate space $\mathbf{X}_C$, which is held fixed , as illustrated in Figure III.55.

Let $\theta : \mathbf{X}_C \to \mathbf{X}_C$ denote a change of coordinates. In order that the new coordinate system be useful, it is usually necessary that θ, treated as a transformation from $\mathbf{X}$ to $\theta(\mathbf{X})$, be one-to-one and onto, and hence invertible. Let $f : \mathbf{X} \to \mathbf{X}$ be a transformation on a metric space $\mathbf{X}$. We want to consider how the transformation f should

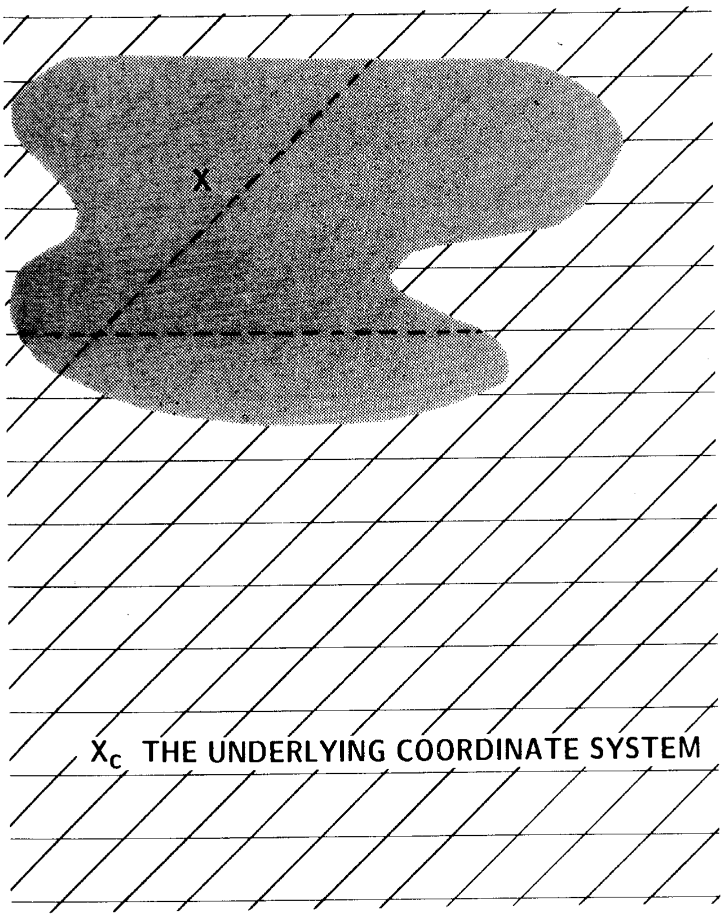


Figure III.53. The underlying coordinate system $\mathbf{X}_C$ for the space $\mathbf{X}$.

be expressed after the change of coordinates. Let x denote simultaneously a point in $\mathbf{X}$ and the coordinates of that point. Let $f(x)$ denote simultaneously the point to which x is transformed by f, and the coordinates of that point. Let x' denote the point $x \in \mathbf{X}$ in the new coordinate system. That is, $x' = \theta(x) \in \mathbf{X}_C$ denotes the new coordinates of the point x. Let $f'(x')$ denote the same transformation $f : \mathbf{X} \to \mathbf{X}$, but expressed in the new coordinate system. Then the relation between the two coordinate systems is expressed by the commutative diagram in Figure III.57, and is illustrated in Figure III.56.

Theorem 5.1 *Let $\mathbf{X}$ be a space and let $\mathbf{X}_C \supset \mathbf{X}$ be a coordinate space for $\mathbf{X}$. Let a change of coordinates be provided by a transformation $\theta : \mathbf{X}_C \to \mathbf{X}_C$. Let θ be invertible when treated as a transformation from $\mathbf{X}$ to $\theta(\mathbf{X})$. Let the coordinates of a point $x \in \mathbf{X}$ be denoted by x before the change of coordinates, and by x' after the change of coordinates, so that*

$$x' = \theta(x).$$

Let $f : \mathbf{X} \to \mathbf{X}$ be a transformation on the space $\mathbf{X}$. Let $x \mapsto f(x)$ be the formula for f expressed in the original coordinates. Let $x' \mapsto f'(x')$ be the formula for f expressed in the new coordinates. Then

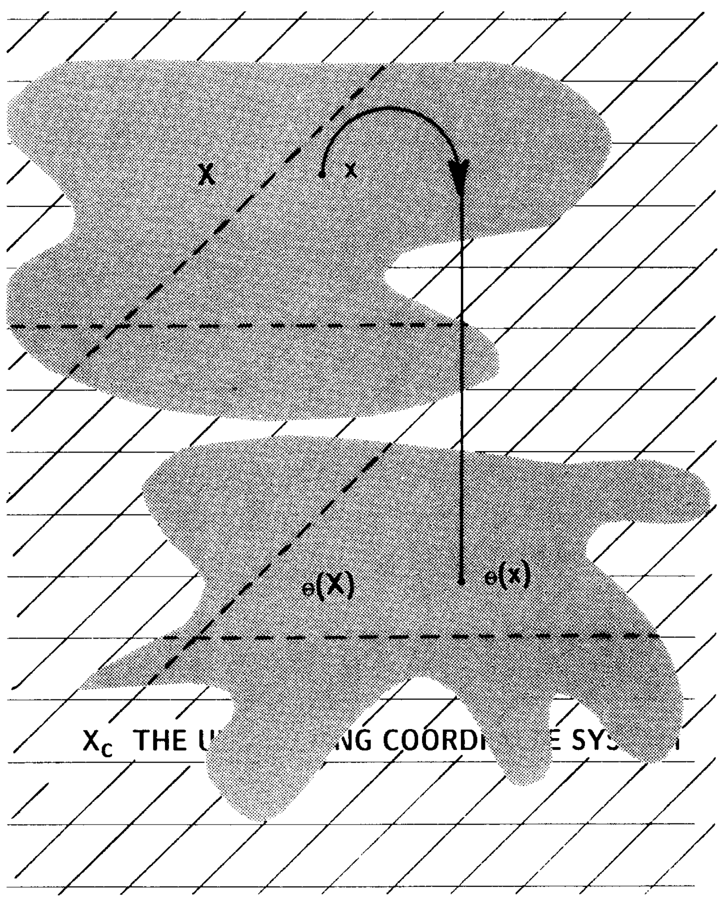


Figure III.54. A change of coordinates in terms of $\mathbf{X}$ and $\mathbf{X}_C$. We think of $\mathbf{X}$ as being removed relative to the underlying coordinate space $\mathbf{X}_C$.

$$f(x) = (\theta^{-1} \circ f' \circ \theta)(x),$$
$$f'(x') = (\theta \circ f \circ \theta^{-1})(x').$$

Examples & Exercises

5.2. Consider an affine transformation $f(x) = ax + b$, $a \neq 0$, $a \neq 1$, $a, b \in \mathbb{R}$. This has a fixed point $x_f \in \mathbb{R}$ defined by $f(x_f) = x_f$. We find $x_f = b/(1 - a)$. x_f is clearly the interesting point in the action of an affine transformation on $\mathbb{R}$. Accordingly let us change coordinates to move x_f to the origin: that is $x' = \theta(x) = x - x_f$. What does f look like in this new coordinate system?

$$f'(x') = (\theta \circ f \circ \theta^{-1})(x') = \theta \circ (x' + x_f) = a(x' + x_f) + b - x_f;$$

$f'(x') = ax'$, which is simply a rescaling! Now using the first formula we get

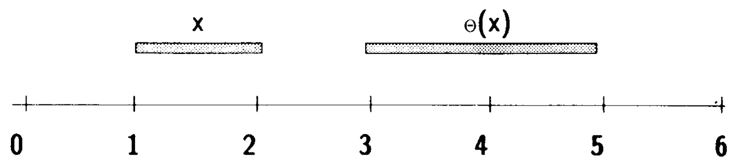


Figure III.55. A change of coordinates for the space $[1, 2]$ given by the transformation $x' = \theta(x) = 2x + 1$.

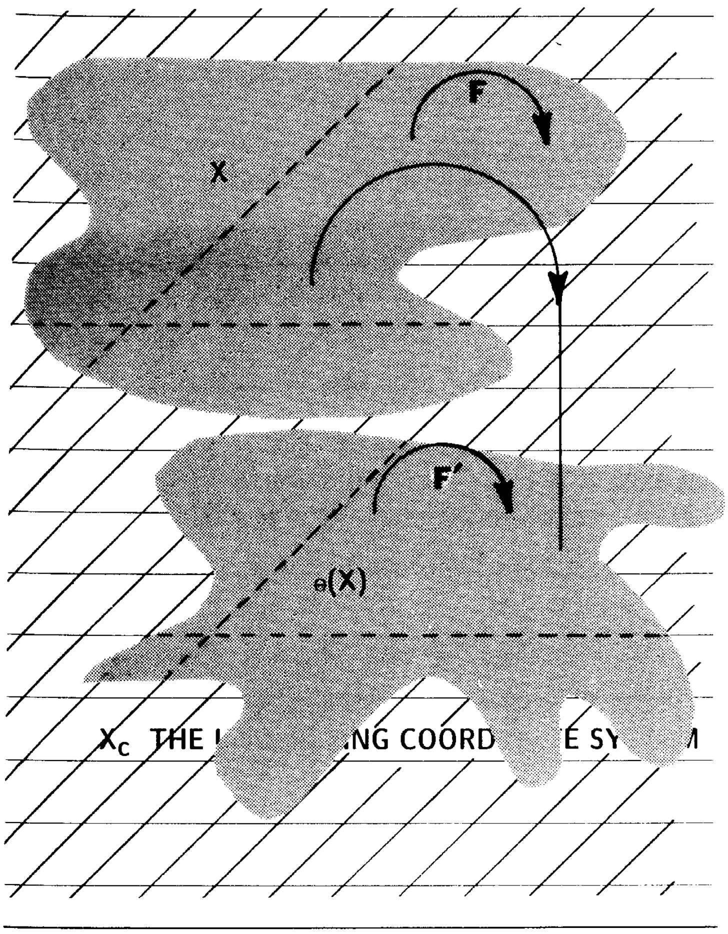


Figure III.56. The transformation F acting on $\mathbf{X}$ is equivalent to F' acting on $\theta(\mathbf{X})$.

$$f(x) = a(x - x_f) + x_f$$

and

$$f^{\circ n}(x) = a^n(x - x_f) + x_f \qquad \text{for all } n \in \{0, 1, 2, 3, \ldots\}.$$

We now see a new way of visualizing an affine transformation on $\mathbb{R}$: for example, if $a > 1$, we see the image in Figure III.58.

5.3. Show that for any affine transformation $f(x) : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x) = Ax + t$, with fixed point x_f, that the coordinate transformation $\theta(x) = x - x_f$ transforms the function $f'(x') = Ax'$.

5.4. Let $\mathbf{X} = [1, 2]$ and let a change of coordinates be defined by $x' = 2x - 1$. Let a transformation $f : \mathbf{X} \to \mathbf{X}$ be defined by $f(x) = (x - 1)^2 + 1$. Express f in the new coordinate system.

Definition 5.1 *Let $f : \mathbf{X} \to \mathbf{X}$ be a transformation on a metric space. A point $x_f \in X$ such that $f(x_f) = x_f$ is called a* fixed point *of the transformation.*

The fixed points of a transformation are very important. They tell us which parts of the space are pinned in place, not moved, by the transformation. The fixed points of a transformation restrict the motion of the space under nonviolent, nonripping transformations of bounded deformation.

Examples & Exercises

5.5. Find the fixed points x_1 and x_2 of the Möbius transformation

$$f(z) = \frac{(z+2)}{(4-z)}$$

on $\hat{\mathbb{C}}$. Make a change of coordinates so that x_1 becomes the origin and x_2 becomes the point at infinity. Hence interpret the action of $f(z)$ on the sphere in geometrical terms.

5.6. Let $W(x) = Ax + t$ where $\det A \neq 0$ is a two-dimensional affine transformation acting on the space $\mathbf{X} = \mathbb{R}^2$. Find the fixed point x_f. Change coordinates so that x_f becomes the origin of coordinates. Hence describe the action geometrically of a two-dimensional, nondegenerate affine transformation. What can happen if $\det A = 0$?

5.7. Suppose we can find a coordinate transformation $BAB^{-1} = D$, where D is a diagonal matrix we denote by

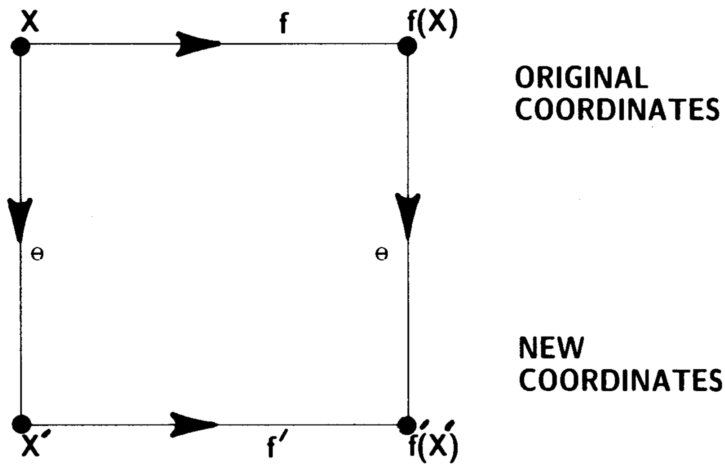


Figure III.57. Commutative diagram for the coordinate change $\theta : \mathbf{X}_C \to \mathbf{X}_C$.

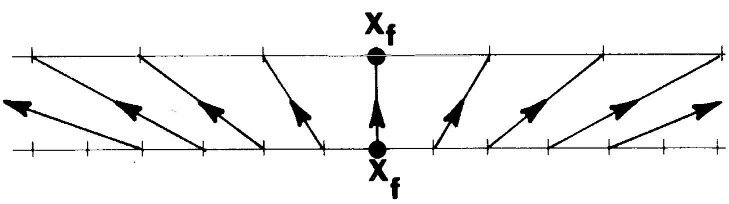


Figure III.58. An affine transformation on $\mathbb{R}$. We see rescaling (magnification or diminution) centered at the fixed point, together with a flip of 180° if $a < 0$.

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Show that λ_1 and λ_2 satisfy the equation

$$\det \begin{pmatrix} e-\lambda & f \\ g & h-\lambda \end{pmatrix} = \begin{vmatrix} e-\lambda & f \\ g & h-\lambda \end{vmatrix} = \lambda^2 - \operatorname{tr} A\lambda + \det A = 0.$$

5.8. Analyze the behavior of the affine transformation $w(z) = 7z + 1$ on $\hat{\mathbb{C}}$ near the point at infinity by making the change of coordinates $h(z) = 1/z$.

5.9. Two one-parameter families of transformations on $\mathbb{R}$ are $f_\mu(x) = x^2 - \mu$ and $g_\lambda(x) = \lambda x(1-x)$, where μ and λ are real parameters. Find a change of coordinates and a function $\mu = \mu(\lambda)$ so that $f'_{\mu(\lambda)}(x') = g_\lambda(x')$ is valid for an appropriate interval on the λ-axis.

5.10. Find the real fixed points of $g(x) = x^2 - \frac{1}{2}$. Analyze the behavior of g near each of its fixed points by changing coordinates so as to move first one then the other to the origin. Another method for looking at the behavior of g near a fixed point is to approximate $g(x)$ by the first two terms of its Taylor series expansion about the fixed point. Compare these methods.

5.11. Suppose that the 2×2 matrix

$$A = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

satisfies the condition $(\operatorname{tr} A)^2 - 4 \det A > 0$. Show that there is a B such that

$$BAB^{-1} = D,$$

where D is a diagonal matrix. Furthermore, show that one choice for B is given by

$$B = \begin{pmatrix} 1 & \frac{f}{\lambda_1 - h} \\ 1 & \frac{f}{\lambda_2 - h} \end{pmatrix}.$$

What do you think happens if $(\operatorname{tr} A)^2 - 4 \det A < 0$?

5.12. Let $w : \mathbb{R}^2 \to \mathbb{R}^2$ denote the affine transformation

$$w\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Make a change of coordinates so that the transformation is simply a coordinate rescaling. What are the rescaling factors?

Definition 5.2 *Let F denote a set of transformations on a metric space $\mathbf{X}$. F is called a* semigroup *if $f, g \in F$ implies $f \circ g \in F$. F is called a* group *if it is a semigroup of invertible transformations, and $f \in F$ implies $f^{-1} \in F$.*

We introduce this definition because we will use semigroups (and groups) of transformations both to characterize and to compute fractal subsets of $\mathbf{X}$. However, we do not use any deep theorems from group theory.

Examples & Exercises

5.13. Let $f : \mathbf{X} \to \mathbf{X}$ be a transformation on a metric space. Show that the set of transformations $\{f^{\circ n} : n{=}0, 1, 2, 3, \ldots\}$ forms a semigroup.

5.14. A transformation $T : \Sigma \to \Sigma$ on code space is defined by

$$T(x_1x_2x_3x_4x_5\ldots) = x_2x_3x_4x_5x_6\ldots$$

and is called a *shift operator*. Describe the semigroup of transformations $\{T^{\circ n} : n = 0, 1, 2, 3, \ldots\}$. What are the *fixed points* of $T^{\circ 3}$ if the code space is built up from the two symbols $\{0, 1\}$?

5.15. Show that the set of Möbius transformations on $\hat{\mathbb{R}}$ forms a group.

5.16. Show that the set of Möbius transformations on $\hat{\mathbb{C}}$ forms a group.

5.17. Show that the set of invertible affine transformations on $\mathbb{R}^2$ forms a group.

5.18. Show that the set of transformations $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that $f(\blacktriangle) \subset \blacktriangle$ forms a semigroup.

5.19. Show that a group of transformations is provided by the set of affine transformations of the form $w(x) = Ax + t$, where $A = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$ for $a, b, c \in \mathbb{R}$, with $ac \neq 0$, and the translation vector t is arbitrary.

5.20. The most general analytic quadratic transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ can be expressed by a formula of the form $f(z) = Az^2 + Bz + C$, where A, B, and C are complex numbers, and $A \neq 0$. Show that by means of a suitable change of coordinates, $z' = \theta(z)$, where θ is a similitude, show that $f(z)$ can be reexpressed as a quadratic transformation of the special form $f'(z) = (z')^2 + \tilde{C}$ for some complex number $\tilde{C}$.

6 The Contraction Mapping Theorem

Definition 6.1 *A transformation $f : \mathbf{X} \to \mathbf{X}$ on a metric space $(\mathbf{X}, d)$ is called* contractive *or* a contraction mapping *if there is a constant $0 \le s < 1$ such that*

$$d(f(x), f(y)) \le s \cdot d(x, y) \forall x, \qquad y \in \mathbf{X}.$$

Any such number s is called a contractivity factor *for f.*

It would be convenient to be able to talk about the largest number and the smallest number in a set of real numbers. However, a set such as $S = (-\infty, 3)$ does not possess either. This difficulty is overcome by the following definition.

Definition 6.2 *Let S denote a set of real numbers. Then the* infimum *of S is equal to $-\infty$ if S contains negative numbers of arbitrarily large magnitude. Otherwise the infimum of $S = \max\{x \in \mathbb{R} : x \le s$ for all $s \in S\}$. The infinum of S always* *exists because of the nature of the real number system, and it is denoted by* $\inf S$. *The* supremum *of S is similarly defined. It is equal to* $+\infty$ *if S contains arbitrarily large numbers; otherwise it is the minimum of the set of numbers that are greater than or equal to all of the numbers in S. The supremum of S always exists, and it is denoted by* $\sup S$.

Examples & Exercises

6.1. Find the supremum and the infimum of the following sets of real numbers: (a) $(-\infty, 3)$; (b) C, the Classical Cantor Set; (c) $\{1, 2, 3, 4, \ldots\}$; (d) the positive real numbers.

6.2. Let $f : \mathbf{X} \to \mathbf{X}$ be a contraction mapping on a compact metric space $(\mathbf{X}, d)$. Show that $\inf\{s \in \mathbb{R} : s \text{ is a contractivity factor for } f\}$ is a contractivity factor for f.

6.3. Show that if $f : \mathbf{X} \to \mathbf{X}$ and $g : \mathbf{X} \to \mathbf{X}$ are contraction mappings on a space $(\mathbf{X}, d)$, with contractivity factors s and t, respectively, then $f \circ g$ is a contraction mapping with contractivity factor st.

Theorem 6.1 *[(The Contraction Mapping Theorem).] Let* $f : \mathbf{X} \to \mathbf{X}$ *be a contraction mapping on a complete metric space* $(\mathbf{X}, d)$. *Then f possesses exactly one fixed point* $x_f \in \mathbf{X}$ *and moreover for any point* $x \in \mathbf{X}$, *the sequence* $\{f^{\circ n}(x) : n = 0, 1, 2, \ldots\}$ *converges to* x_f. *That is,*

$$\lim_{n\to\infty} f^{\circ n}(x) = x_f, \qquad \text{for each } x \in \mathbf{X}.$$

Figure III.59 illustrates the idea of a contractive transformation on a compact metric space.

Proof Let $x \in \mathbf{X}$. Let $0 \le s < 1$ be a contractivity factor for f. Then

$$d(f^{\circ n}(x), f^{\circ m}(x)) \le s^{m\wedge n} d(x, f^{\circ|n-m|})(x) \tag{1}$$

for all $m, n = 0, 1, 2, \ldots$, where we have fixed $x \in \mathbf{X}$. The notation $u \wedge v$ denotes the minimum of the pair of real numbers u and v. In particular, for $k = 0, 1, 2, \ldots$, we have

$$\begin{aligned} d(x, f^{\circ k}(x)) &\le d(x, f(x)) + (f(x), f^{\circ 2}(x)) + \cdots + d(f^{\circ(k-1)}(x), f^{\circ k}(x)) \\ &\le (1 + s + s^2 + \cdots + s^{k-1}) d(x, f(x)) \\ &\le (1 - s)^{-1} d(x, f(x)), \end{aligned}$$

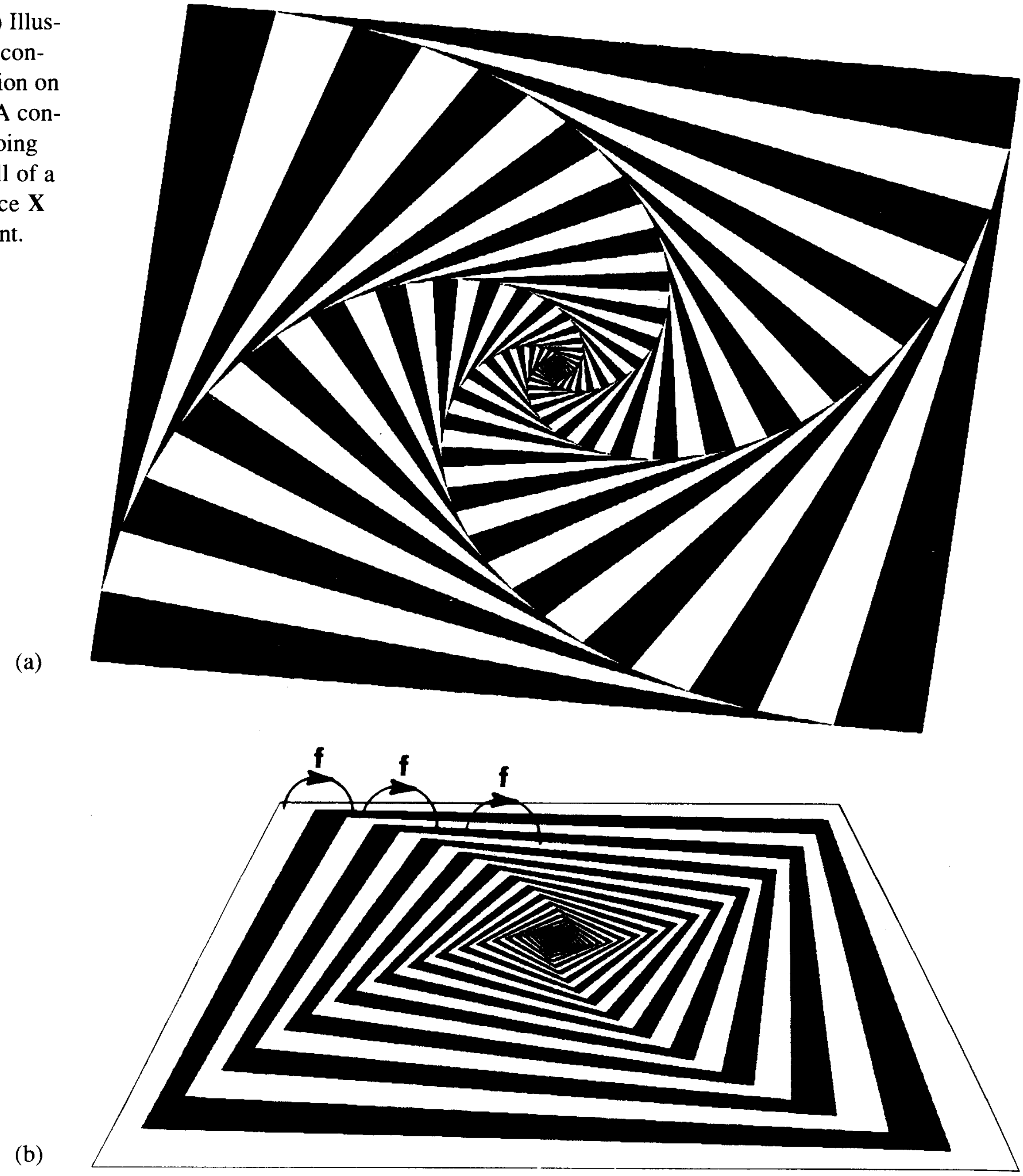
so substituting into equation (1) we now obtain

$$d(f^{\circ n}(x), f^{\circ m}(x)) \le s^{m\wedge n} \cdot (1 - s)^{-1} \cdot (d(x, f(x)),$$

from which it immediately follows that $\{f^{\circ n}(x)\}_{n=0}^{\infty}$ is a Cauchy sequence. Since $\mathbf{X}$ is complete this Cauchy sequence possesses a limit $x_f \in \mathbf{X}$, and we have

$$\lim_{n\to\infty} f^{\circ n}(x) = x_f.$$

Figure III.59. (a) Illustrates the idea of a contractive transformation on a metric space. (b) A contraction mapping doing its work, drawing all of a compact metric space **X** toward the fixed point.



Now we shall show that x_f is a fixed point of f. Since f is contractive it is continuous and hence

$$f(x_f) = f(\lim_{n\to\infty} f^{\circ n}(x)) = \lim_{n\to\infty} f^{\circ(n+1)}(x) = x_f.$$

Finally, can there be more than one fixed point? Suppose there are. Let x_f and y_f be two fixed points of f. Then $x_f = f(x_f)$, $y_f = f(y_f)$, and

$$d(x_f, y_f) = d(f(x_f), f(x_f)) \le s d(x_f, y_f),$$

where $(1-s)d(x_f, y_f) \le 0$, which implies $d(x_f, y_f) = 0$ and hence $x_f = y_f$. This completes the proof.

Examples & Exercises

6.4. Let $w(x) = Ax + t$ be an affine transformation in two dimensions. Make the change of coordinates $h(x) = x' = x - x_f$, under the assumption that $\det(I - A) \neq 0$, and show that $w'(x') = h \circ w \circ h^{-1}(x') = Ax'$, that $w(x) = (h^{-1} \circ w' \circ h)(x) = A(x - x_f) + x_f$, and hence that

$$w^{\circ n}(x) = A^n(x - x_f) + x_f \qquad \text{for } n = 0, 1, 2, 3, \ldots. \tag{2}$$

Give conditions on A such that it is contractive (a) in the Euclidean metric, and (b) in the Manhattan metric. Show that if $|A| < 1$, where $|A|$ denotes any appropriate norm of A viewed as a linear operator on a two-dimensional vector space, then $\{w^{\circ n}(x)\}$ is a Cauchy sequence that converges to x_f, for each $x \in \mathbb{R}^2$.

6.5. Let $f : \blacksquare \to \blacksquare$ be a contraction mapping on ($\blacksquare$, Euclidean). Show that Figure III.59 gives the right idea.

6.6. Let $f : \mathbb{R} \to \mathbb{R}$ be the affine transformation $f(x) = \frac{1}{2}x + \frac{1}{2}$. Verify f is a contraction mapping and deduce

$$\lim_{n\to\infty} f^{\circ n}(x) = x_f \qquad \text{for each } x \in \mathbb{R}.$$

Use this formula with $x = 1$ to obtain a geometrical series for the fixed point $x_f \in \mathbb{R}$. Observe, however, $f(\mathbb{R}) = \mathbb{R}$; indeed f is invertible.

6.7. Let $(\mathbf{X}, d)$ be a compact metric space that contains more than one point. Show that the situation in exercise 6.6 cannot occur for any contraction mapping $f : \mathbf{X} \to \mathbf{X}$. That is, show that $f(\mathbf{X}) \subset \mathbf{X}$ but $f(\mathbf{X}) \neq \mathbf{X}$. That is, show that a contraction mapping on a nontrivial compact metric space is not invertible. Hint: use the compactness of the space to show that there is a point in the space that is farthest away from the fixed point. Then show that there is a point that is not in $f(\mathbf{X})$.

6.8. Show that the set of contraction mappings on a metric space forms a semigroup.

6.9. Show that the affine transformation $w : \blacktriangle \to \blacktriangle$ defined by $w(x) = Ax + t$ is a contraction, where

$$A = \begin{pmatrix} \frac{1}{2}\cos 120^\circ & -\frac{1}{2}\sin 120^\circ \\ \frac{1}{2}\sin 120^\circ & \frac{1}{2}\cos 120^\circ \end{pmatrix} \text{ and } t = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}.$$

Here $\blacktriangle$ is an equilateral Sierpinski triangle with a vertex at the origin and one at $(1, 0)$. You need to begin by verifying that w does indeed map $\blacktriangle$ into itself! Locate the fixed point x_f. Make a picture of this contraction mapping "doing its work, mapping all of the compact metric space $\blacktriangle$ toward the fixed point." Use different colors to denote the successive regions $f^{\circ(n)}(\blacktriangle) \setminus f^{\circ(n+1)}(\blacktriangle)$ for $n = 0, 1, 2, 3, \ldots$.

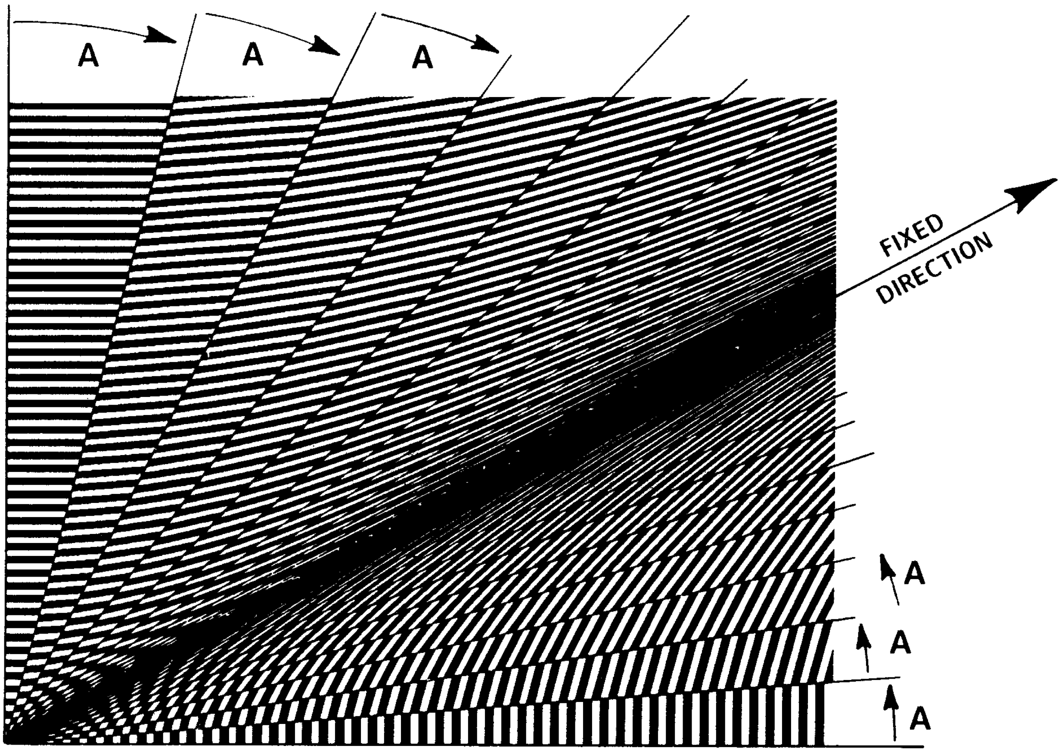


Figure III.60. The existence of a positive eigenvalue of an "angle-squeezing" linear transformation.

6.10. Define a mapping on the code space of two symbols $\{0, 1\}$ by $f(x_1x_2x_3x_4\ldots) = 1x_1x_2x_3x_4\ldots$. (Recall that the metric is $d(x, y) = \sum_{i=1}^{\infty} \frac{|x_i - y_i|}{3^i}$, or equivalent.) Show that f is a contraction mapping. Locate the fixed point of f.

6.11. Let $(\mathbf{X}, d)$ be a compact metric space, and let $f : \mathbf{X} \to \mathbf{X}$ be a contraction mapping. Show that $\{f^{\circ n}(\mathbf{X})\}_{n=0}^{\infty}$ is a Cauchy sequence of points in $(\mathcal{H}(\mathbf{X}), h)$ and $\lim_{n\to\infty} f^{\circ n}(\mathbf{X}) = \{x_f\}$, where x_f is the fixed point of f.

6.12. Let $(\mathbf{X}, d)$ be a compact metric space. Let $f : \mathbf{X} \to \mathbf{X}$ have the property $\lim_{n\to\infty} f^{\circ n}(\mathbf{X}) = x_f$. Find a metric $\tilde{d}$ on $\mathbf{X}$ such that f is a contraction mapping, and the identity is a homeomorphism from $(\overline{\mathbf{X}}, d) \to (\overline{\mathbf{X}}, \tilde{d})$.

6.13. Let $Ax = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ with $a, b, c, d \in \mathbb{R}$, all strictly positive, be a linear transformation on $\mathbb{R}^2$. Show that A maps the positive quadrant $\{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\}$ into itself. Let a mapping $f : [0, 90°] \to [0, 90°]$ be defined by

$$A\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = (\text{some positive number})\begin{pmatrix} \cos f(\theta) \\ \sin f(\theta) \end{pmatrix}.$$

Show that $\{f^{\circ n}(\theta)\}$ converges to the unique fixed point of f. Deduce that there exists a unique positive number λ, and an angle $0 < \theta < 90°$ such that $A\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \lambda\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$. See Figure III.60.

7 Contraction Mappings on the Space of Fractals

Let $(\mathbf{X}, d)$ be a metric space and let $(\mathcal{H}(\mathbf{X}), h(d))$ denote the corresponding space of nonempty compact subsets, with the Hausdorff metric $h(d)$. We introduce the notation $h(d)$ to show that d is the underlying metric for the Hausdorff metric h. For example, we may discuss $(\mathcal{H}(\hat{\mathbb{C}}), h(\text{spherical}))$ or $(\mathcal{H}(\mathbb{R}^2), h(\text{Manhattan}))$. We will drop this additional notation when we evaluate Hausdorff distances.

We have repeatedly refused to define fractals: we have agreed that they are subsets of simple geometrical spaces, such as $(\mathbb{R}^2$, Euclidean) and $(\hat{\mathbb{C}}$, Spherical). If we were to define a *deterministic fractal*, we *might* say that it is a fixed point of a contractive transformation on $(\mathcal{H}(\mathbf{X}), h(d))$. We would require that the underlying metric space $(\mathbf{X}, d)$ be "geometrically simple." We would require also that the contraction mapping be constructed from simple, easily specified, contraction mappings on $(\mathbf{X}, d)$, as described below.

Lemma 7.1 *Let $w : \mathbf{X} \to \mathbf{X}$ be a contraction mapping on the metric space $(\mathbf{X}, d)$. Then w is continuous.*

Proof Let $\epsilon > 0$ be given. Let $s > 0$ be a contractivity factor for w. Then

$$d(w(x), w(y)) \leq sd(x, y) < \epsilon$$

whenever $d(x, y) < \delta$, where $\delta = \epsilon/s$. This completes the proof.

Lemma 7.2 *Let $w : \mathbf{X} \to \mathbf{X}$ be a continuous mapping on the metric space $(\mathbf{X}, d)$. Then w maps $\mathcal{H}(\mathbf{X})$ into itself.*

Proof Let S be a nonempty compact subset of $\mathbf{X}$. Then clearly $w(S) = \{w(x) : x \in S\}$ is nonempty. We want to show that $w(S)$ is compact. Let $\{y_n = w(x_n)\}$ be an infinite sequence of points in S. Then $\{x_n\}$ is an infinite sequence of points in S. Since S is compact there is a subsequence $\{x_{N_n}\}$ that converges to a point $\hat{x} \in S$. But then the continuity of w implies that $\{y_{N_n} = f(x_{N_n})\}$ is a subsequence of $\{y_n\}$ that converges to $\hat{y} = f(\hat{x}) \in w(S)$. This completes the proof.

The following lemma tells us how to make a contraction mapping on $(\mathcal{H}(\mathbf{X}), h)$ out of a contraction mapping on $(\mathbf{X}, d)$.

Lemma 7.3 *Let $w : \mathbf{X} \to \mathbf{X}$ be a contraction mapping on the metric space $(\mathbf{X}, d)$ with contractivity factor s. Then $w : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ defined by*

$$w(B) = \{w(x) : x \in B\} \forall B \in \mathcal{H}(\mathbf{X})$$

is a contraction mapping on $(\mathcal{H}(\mathbf{X}), h(d))$ with contractivity factor s.

Proof From Lemma 7.1 it follows that $w : \mathbf{X} \to \mathbf{X}$ is continuous. Hence by Lemma 7.2 w maps $\mathcal{H}(\mathbf{X})$ into itself. Now let $B, C \in \mathcal{H}(\mathbf{X})$. Then

$$\begin{aligned} d(w(B), w(C)) &= \max\{\min\{d(w(x, y), w(y)) : y \in C\} : x \in B\} \\ &\leq \max\{\min\{s \cdot d(x, y) : y \in C\} : x \in B\} = s \cdot d(B, C). \end{aligned}$$

Similarly, $d(w(C), w(B)) \leq s \cdot d(C, B)$. Hence

$$h(w(B), w(C)) = d(w(B), w(C)) \vee d(w(C), w(B)) \leq s \cdot d(B, C) \vee d(C, B)$$
$$\leq s \cdot d(B, C).$$

This completes the proof.

The following lemma gives a characteristic property of the Hausdorff metric which we will shortly need. The proof follows at once from exercise 6.13 of Chapter II.

Lemma 7.4 *For all B, C, D, and E, in $\mathcal{H}(\mathbf{X})$*

$$h(B \cup C, D \cup E) \leq h(B, D) \vee h(C, E),$$

where as usual h is the Hausdorff metric.

The next lemma provides an important method for combining contraction mappings on $(\mathcal{H}(\mathbf{X}), h)$ to produce new contraction mappings on $(\mathcal{H}(\mathbf{X}), h)$. This method is distinct from the obvious one of composition.

Lemma 7.5 *Let $(\mathbf{X}, d)$ be a metric space. Let $\{w_n : n = 1, 2, \ldots, N\}$ be contraction mappings on $(\mathcal{H}(\mathbf{X}), h)$. Let the contractivity factor for w_n be denoted by s_n for each n. Define $W : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ by*

$$W(B) = w_1(B) \cup w_2(B) \cup \ldots \cup w_n(B)$$
$$= \cup_{n=1}^{n} w_n(B), \qquad \text{for each } B \in \mathcal{H}(\mathbf{X}).$$

Then W is a contraction mapping with contractivity factor $s = \max\{s_n : n = 1, 2, \ldots, N\}$.

Proof We demonstrate the claim for $N = 2$. An inductive argument then completes the proof. Let $B, C \in \mathcal{H}(\mathbf{X})$. We have

$$h(W(B), W(C)) = h(w_1(B) \cup w_2(B), w_1(C) \cup w_2(C))$$
$$\leq h(w_1(B), w_1(C)) \vee h(w_2(B), w_2(C)) \text{ (by Lemma 7.2)}$$
$$\leq s_1 h(B, C) \vee s_2 h(B, C) \leq s h(B, C).$$

This completes the proof.

Definition 7.1 *A (hyperbolic)* iterated function system *consists of a complete metric space $(\mathbf{X}, d)$ together with a finite set of contraction mappings $w_n : \mathbf{X} \to \mathbf{X}$, with respective contractivity factors s_n, for $n = 1, 2, \ldots, N$. The abbreviation "IFS" is used for "iterated function system." The notation for the IFS just announced is $\{\mathbf{X}; w_n, n = 1, 2, \ldots, N\}$ and its contractivity factor is $s = \max\{s_n : n = 1, 2, \ldots, N\}$.*

We put the word "hyperbolic" in parentheses in this definition because it is sometimes dropped in practice. Moreover, we will sometimes use the nomenclature "IFS" to mean simply a finite set of maps acting on a metric space, with no particular conditions imposed upon the maps.

The following theorem summarizes the main facts so far about a hyperbolic IFS.

Theorem 7.1 *Let $\{\mathbf{X}; w_n, n = 1, 2, \ldots, N\}$ be a hyperbolic iterated function system with contractivity factor s. Then the transformation $W : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ defined by*

$$W(B) = \cup_{n=1}^{n} w_n(B)$$

for all $B \in \mathcal{H}(\mathbf{X})$, is a contraction mapping on the complete metric space $(\mathcal{H}(\mathbf{X}), h(d))$ with contractivity factor s. That is

$$h(W(B), W(C)) \leq s \cdot h(B, C)$$

for all $B, C \in \mathcal{H}(\mathbf{X})$. Its unique fixed point, $A \in \mathcal{H}(\mathbf{X})$, obeys

$$A = W(A) = \cup_{n=1}^{n} w_n(A)$$

and is given by $A = \lim_{n\to\infty} W^{\circ n}(B)$ for any $B \in \mathcal{H}(\mathbf{X})$.

Definition 7.2 *The fixed point $A \in \mathcal{H}(\mathbf{X})$ described in the theorem is called the* attractor *of the IFS.*

Sometimes we will use the name "attractor" in connection with an IFS that is simply a finite set of maps acting on a complete metric space $\mathbf{X}$. By this we mean that one can make an assertion analagous to the last sentence of Theorem 7.1.

We wanted to use the words "deterministic fractal" in place of "attractor" in Definition 7.2. We were tempted, but resisted. The nomenclature "iterated function system" is meant to remind one of the name "dynamical system." We will introduce dynamical systems in Chapter 4. Dynamical systems often possess attractors, and when these are interesting to look at they are called *strange attractors*.

Examples & Exercises

7.1. This exercise takes place in the metric spaces ($\mathbb{R}$, Euclidean) and $(\mathcal{H}(R), h(\text{Euclidean}))$. Consider the IFS $\{\mathbb{R}; w_1, w_2\}$, where $w_1(x) = \frac{1}{3}x$ and $w_2(x) = \frac{1}{3}x + \frac{2}{3}$. Show that this is indeed an IFS with contractivity factor $s = \frac{1}{3}$. Let $B_0 = [0, 1]$. Calculate $B_n = W^{\circ n}(B_0)$, $n = 1, 2, 3, \ldots$. Deduce that $A = \lim_{n\to\infty} B_n$ is the classical Cantor set. Verify directly that $A = \frac{1}{3}A \cup \{\frac{1}{3}A + \frac{2}{3}\}$. Here we use the following notation: for a subset A of $\mathbb{R}$, $xA = \{xy : y \in A\}$ and $A + x = \{y + x : y \in A\}$.

7.2. With reference to example 7.1, show that if $w_1(x) = s_1 x$ and $w_2(x) = (1 - s_1)x + s_1$, where s_1 is a number such that $0 < s_1 < 1$, then $B_1 = B_2 = B_3 = \ldots$. Find the attractor.

7.3. Repeat example 7.1 with $w_1(x) = \frac{1}{3}x$ and $w_2(x) = \frac{1}{2}x + \frac{1}{2}$. In this case $A = \lim_{n\to\infty} B_n$ will not be the classical Cantor set, but it will be something like it. Describe A. Show that A contains no intervals. How many points does A contain?

7.4. Consider the IFS $\{\mathbb{R}; \frac{1}{4}x + \frac{3}{4}, \frac{1}{2}x, \frac{1}{4}x + \frac{1}{4}\}$. Verify that the attractor looks like the image in Figure III.61. Show, precisely, how the set in Figure III.61 is a union of three "shrunken copies of itself." This attractor is interesting: it contains countably many holes and countably many intervals.

Figure III.61. Attractor for three affine maps on the real line. Can you find the maps?

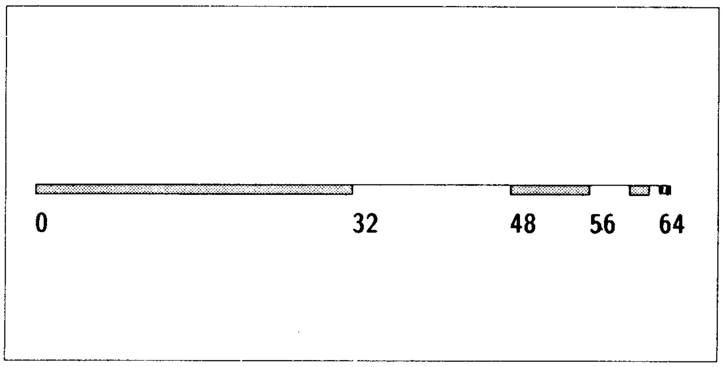
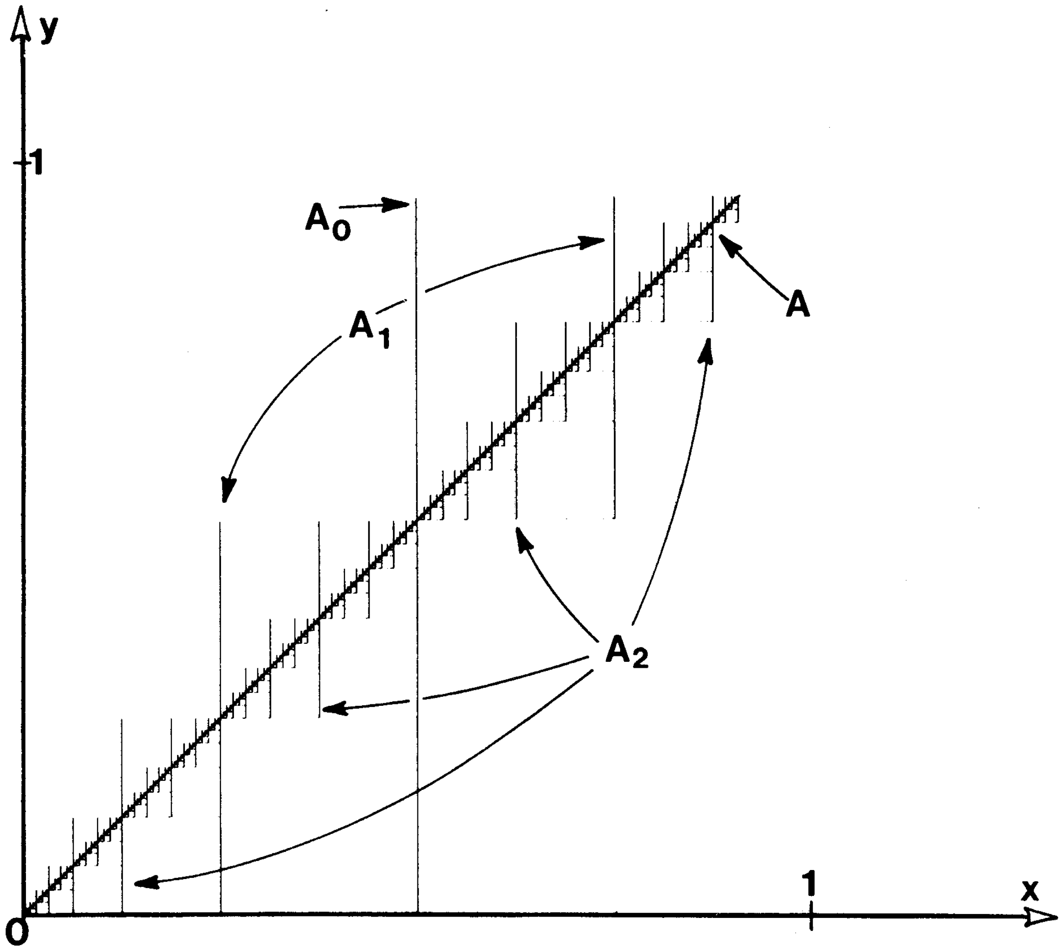


Figure III.62. A sequence of sets converging to a line segment.



7.5. Show that the attractor of an IFS having the form $\{\mathbb{R}; w_1(x) = ax + b, w_2(x) = cx + d\}$, where a, b, c, and $d \in \mathbb{R}$, is either connected or totally disconnected.

7.6. Does there exist an IFS of three affine maps in $\mathbb{R}^2$ whose attractor is the union of two disjoint closed intervals?

7.7. Consider the IFS

$$\left\{\mathbb{R}^2; \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right\}.$$

Let $A_0 = \{(\frac{1}{2}, y) : 0 \le y \le 1\}$, and let $W^{\circ n}(A_0) = A_n$, where W is defined on $\mathcal{H}(\mathbb{R}^2)$ in the usual way. Show that the attractor is $A = \{(x, y) : x = y, 0 \le x \le 1\}$ and that Figure III.62 is correct. Draw a sequence of pictures to show what happens if $A_0 = \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1, 0 \le y \le 1\}$.

7.8. Consider the attractor for the IFS $\{\mathbb{R}; w_1(x) = 0, w_2(x) = \frac{2}{3}x + \frac{1}{3}\}$. Show that it consists of a countable increasing sequence of real points $\{x_n : n = 0, 1, 2, \ldots\}$ together with $\{1\}$. Show that x_n can be expressed as the nth partial sum of an infinite geometric series. Give a succinct formula for x_n.

7.9. Describe the attractor A for the IFS $\{[0, 2]; w_1(x) = \frac{1}{9}x^2, w_2(x) = \frac{3}{4}x + \frac{1}{2}\}$ by describing a sequence of sets which converges to it. Show that A is totally disconnected. Show that A is perfect. Find the contractivity factor for the IFS.

7.10. Let $(r, \theta), 0 \le r \le \infty, 0 \le \theta < 2\pi$ denote the polar coordinates of a point in the plane, $\mathbb{R}^2$. Define $w_1(r, \theta) = (\frac{1}{2}r + \frac{1}{2}, \frac{1}{2}\theta)$, and $w_2(r, \theta) = (\frac{2}{3}r + \frac{1}{3}, \frac{2}{3}\theta + \frac{2\pi}{3})$. Show that $\{\mathbb{R}^2; w_1, w_2\}$ is not a hyperbolic IFS because both maps w_1 and w_2 are discontinuous on the whole plane. Show that $\{\mathbb{R}^2; w_1, w_2\}$ nevertheless has an attractor; find it (just consider r and θ separately).

7.11. Show that the sequence of sets illustrated in Figure III.63 can be written in the form $A_n = W^{\circ n}(A_0)$ for $n = 1, 2, \ldots$, and find $W : \mathcal{H}(\mathbb{R}^2) \to \mathcal{H}(\mathbb{R}^2)$.

7.12. Describe the collection of functions that constitutes the attractor A for the IFS

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

Find the contractivity factor for the IFS.

7.13. Let $C^0[0, 1] = \{f \in C[0, 1] : f(0) = f(1) = 0\}$, and define $d(f, g) = \max\{|f(x) - g(x)| : x \in [0, 1]\}$. Define $w_1 : C^0[0, 1] \to C^0[0, 1]$ by $(w_1(f))(x) = \frac{1}{2}f(2x \bmod 1) + 2x(1 - x)$ and $(w_2(f))(x) = \frac{1}{2}f(x)$. Show that $\{C^0[0, 1]; w_1, w_2\}$ is an IFS, find its contractivity factor, and find its attractor. Draw a picture of the attractor.

7.14. Find conditions such that the Möbius transformation $w(x) = (ax + b)/(cz + d), a, b, c, d \in \mathbb{C}, ad - bc \ne 0$, provides a contraction mapping on the unit disk

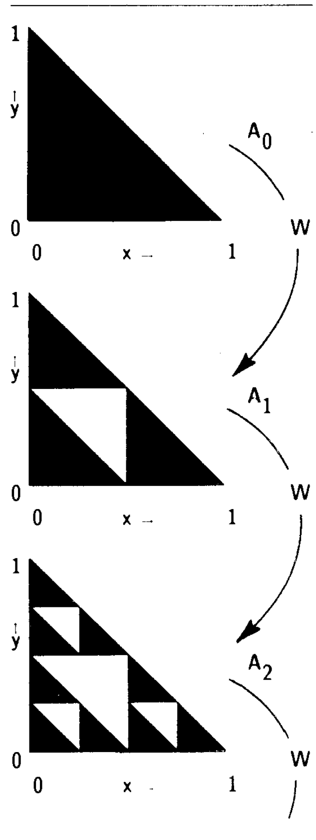


Figure III.63. The first three sets A_0, A_1, and A_2 in a convergent sequence of sets in $\mathcal{H}(\mathbb{R}^2)$. Can you find a transformation $W : \mathcal{H}(\mathbb{R}^2) \to \mathcal{H}(\mathbb{R}^2)$ such that $A_{n+1} = W(A_n)$?

$\mathbf{X} = \{z \in \mathbb{C} : |z| \leq 1\}$. Find an upper bound for the contractivity factor. Construct an IFS using two Möbius transformations on $\mathbf{X}$, and describe its attractor.

7.15. Show that a Möbius transformation on $\hat{\mathbb{C}}$ is never a contraction in the spherical metric.

7.16. Let (Σ, d) be the code space of three symbols $\{0, 1, 2\}$, with metric

$$d(x, y) = \sum_{n=1}^{\infty} \frac{|x_n - y_n|}{4^n}.$$

Define $w_1 : \Sigma \to \Sigma$ by $w_1(x) = 0x_1x_2x_3\ldots$ and $w_2(x) = 2x_1x_2x_3\ldots$. Show that w_1 and w_2 are both contraction mappings and find their contractivity factors. Describe the attractor of the IFS $\{\Sigma; w_1, w_2\}$. What happens if we include in the IFS a third transformation defined by $w_3x = 1x_1x_2x_3\ldots$?

7.17. Let $\blacktriangle \subset \mathbb{R}^2$ denote the compact metric space constisting of an equilateral Sierpinski triangle with vertices at $(0, 0)$, $(1, 0)$, and $(\frac{1}{2}, \frac{\sqrt{3}}{2})$, and consider the IFS $\{\blacktriangle, \frac{1}{2}z + \frac{1}{2}, \frac{1}{2}e^{2\pi i/3}z + \frac{1}{2}\}$ where we use complex number notation. Let $A_0 = \blacktriangle$, and $A_n = W^{\circ n}(A_0)$ for $n = 1, 2, 3, \ldots$. Describe A_1, A_2, and the attractor A. What happens if the third transformation $w_3(z) = \frac{1}{2}z + \frac{1}{4} + (\sqrt{3}/4)i$ is included in the IFS?

8 Two Algorithms for Computing Fractals from Iterated Function Systems

In this section we take time out from the mathematical development to provide two algorithms for rendering pictures of attractors of an IFS on the graphics display device of a microcomputer or workstation. The reader should establish a computer-graphical environment that includes one or both of the software tools suggested in this section.

The algorithms presented are (1) the Deterministic Algorithm and (2) the Random Iteration Algorithm. The Deterministic Algorithm is based on the idea of directly computing a sequence of sets $\{A_n = W^{\circ n}(A)\}$ starting from an initial set A_0. The Random Iteration Algorithm is founded in ergodic theory; its mathematical basis will be presented in Chapter IX. An intuitive explanation of why it works is presented in Chapter IV. We defer important questions concerning discretization and accuracy. Such questions are considered to some extent in later chapters.

For simplicity we restrict attention to hyperbolic IFS of the form $\{\mathbb{R}^2; w_n : n = 1, 2, \ldots, N\}$, where each mapping is an affine transformation. We illustrate the algorithms for an IFS whose attractor is a Sierpinski triangle. Here's an example of such an IFS:

$$w_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$w_2\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}+\begin{bmatrix}1\\50\end{bmatrix},$$

$$w_3\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}0.5 & 0\\0 & 0.5\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}+\begin{bmatrix}25\\50\end{bmatrix}.$$

This notation for an IFS of affine maps is cumbersome. Let us agree to write

$$w_i(x)=w_i\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}a_i & b_i\\c_i & d_i\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}+\begin{bmatrix}e_i\\f_i\end{bmatrix}=A_ix+t_i.$$

Then Table III.1 is a tidier way of conveying the same iterated function system.

Table III.1 also provides a number p_i associated with w_i for $i=1,2,3$. These numbers are in fact probabilities. In the more general case of the IFS $\{\mathbf{X}; w_n : n = 1, 2, \ldots, N\}$, there would be N such numbers $\{p_i : i = 1, 2, \ldots, N\}$ that obey

$$p_1+p_2+p_3+\cdots+p_n=1 \text{ and } p_i>0 \qquad \text{for } i=1,2,\ldots,N.$$

These probabilities play an important role in the computation of images of the attractor of an IFS using the Random Iteration Algorithm. They play no role in the Deterministic Algorithm. Their mathematical significance is discussed in later chapters. For the moment we will use them only as a computational aid, in connection with the Random Iteration Algorithm. To this end we take their values to be given approximately by

$$p_i\approx\frac{|\det A_i|}{\sum_{i=1}^{N}|A_i|}=\frac{|a_id_i-b_ic_i|}{\sum_{i=1}^{N}|a_id_i-b_ic_i|} \qquad \text{for } i=1,2,\ldots,N.$$

Here the symbol $\approx$ means "approximately equal to." If, for some i, $\det A_i=0$, then p_i should be assigned a small positive number, such as 0.001. Other situations should be treated empirically. We refer to the data in Table III.1 as an IFS *code*. Other IFS codes are given in Tables III.2, III.3, and III.4.

Algorithm 8.1 The Deterministic Algorithm. *Let* $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ *be a hyperbolic IFS. Choose a compact set* $A_0 \subset \mathbb{R}^2$. *Then compute successively* $A_n = W^{\circ n}(A)$ *according to*

$$A_{n+1}=\cup_{j=1}^{n}w_j(A_n) \qquad \text{for } n=1,2,\ldots.$$

Thus construct a sequence $\{A_n : n = 0, 1, 2, 3, \ldots\} \subset \mathcal{H}(\mathbf{X})$. Then by Theorem 7.1 the sequence $\{A_n\}$ converges to the attractor of the IFS in the Hausdorff metric.

Table III.1. IFS code for a Sierpinski triangle.

w	a	b	c	d	e	f	p
1	0.5	0	0	0.5	1	1	0.33
2	0.5	0	0	0.5	1	50	0.33
3	0.5	0	0	0.5	50	50	0.34

Table III.2. IFS code for a square.

w	a	b	c	d	e	f	p
1	0.5	0	0	0.5	1	1	0.25
2	0.5	0	0	0.5	50	1	0.25
3	0.5	0	0	0.5	1	50	0.25
4	0.5	0	0	0.5	50	50	0.25

Table III.3. IFS code for a fern.

w	a	b	c	d	e	f	p
1	0	0	0	0.16	0	0	0.01
2	0.85	0.04	-0.04	0.85	0	1.6	0.85
3	0.2	-0.26	0.23	0.22	0	1.6	0.07
4	-0.15	0.28	0.26	0.24	0	0.44	0.07

Table III.4. IFS code for a fractal tree.

w	a	b	c	d	e	f	p
1	0	0	0	0.5	0	0	0.05
2	0.42	-0.42	0.42	0.42	0	0.2	0.4
3	0.42	0.42	-0.42	0.42	0	0.2	0.4
4	0.1	0	0	0.1	0	0.2	0.15

We illustrate the implementation of the algorithm. The following program computes and plots successive sets A_{n+1} starting from an initial set A_0, in this case a square, using the IFS code in Table III.1. The program is written in BASIC. It should run without modification on an IBM PC with Color Graphics Adaptor or Enhanced Graphics Adaptor, and Turbobasic. It can be modified to run on any personal computer with graphics display capability. On any line the words preceded by a ' are comments and not part of the program.

Program 1. (Example of the Deterministic Algorithm)

```
screen 1 : cls 'initialize graphics
dim s(100,100) : dim t(100,100) 'allocate two arrays of pixels
a(1)=0.5:b(1)=0:c(1)=0:d(1)=0.5:e(1)=1:f(1)=1 'input the IFS code
a(2)=0.5:b(2)=0:c(2)=0:d(2)=0.5:e(2)=50:f(2)=1
a(3)=0.5:b(3)=0:c(3)=0:d(3)=0.5:e(3)=25:f(3)=50
for i=1 to 100 'input the initial set A(0), in this case
     a square, into the array t(i,j)
```

```
t(i,1)=1: pset(i,1) 'A(0) can be used as a condensation set
t(1,i)=1:pset(1,i)  'A(0) is plotted on the screen
t(100,i)=1:pset(100,i)
t(i,100)=1:pset(i,100)
next: do
for i=1 to 100 'apply W to set A(n) to make A(n+1) in the
      array s(i,j)
for j=1 to 100 : if t(i,j)=1 then
s(a(1)*i+b(1)*j+e(1),c(1)*i+d(1)*j+f(1))=1 'and apply W to A(n)
s(a(2)*i+b(2)*j+e(2),c(2)*i+d(2)*j+f(2))=1
s(a(3)*i+b(3)*j+e(3),c(3)*i+d(3)*j+f(3))=1
end if: next j: next i
cls 'clears the screen--omit to obtain sequence with a A(0) as
     condensation set (see section 9 in Chapter II)
for i=1 to 100 : for j=1 to 100
t(i,j)=s(i,j) 'put A(n+1) into the array t(i,j)
s(i,j)=0 'reset the array s(i,j) to zero
if t(i,j)=1 then
pset(i,j) 'plot A(n+1)
end if : next : next
loop until instat 'if a key has been pressed then stop,
     otherwise compute A(n+1)=W(A(n+1))
```

The result of running a higher-resolution version of this program on a Masscomp 5600 workstation and then printing the contents of the graphics screen is presented in Figure III.64. In this case we have kept each successive image produced by the program.

Notice that the program begins by drawing a box in the array $t(i, j)$. This box has no influence on the finally computed image of a Sierpinski triangle. One could just as well have started from any other (nonempty) set of points in the array $t(i, j)$, as illustrated in Figure III.65.

To adapt Program 1 so that it runs with other IFS codes will usually require changing coordinates to ensure that each of the transformations of the IFS maps the pixel array $s(i, j)$ into itself. Change of coordinates in an IFS is discussed in exercise 10.14. As it stands in Program 1, the array $s(i, j)$ is a discretized representation of the square in $\mathbb{R}^2$ with lower left corner at $(1, 1)$ and upper right corner at $(100, 100)$. Failure to adjust coordinates correctly will lead to unpredictable and exciting results!

Algorithm 8.2 The Random Iteration Algorithm. *Let* $\{\mathbf{X}; w_1, w_2, \ldots, w_N\}$ *be a hyperbolic IFS, where probability* $p_i > 0$ *has been assigned to to* w_i *for* $i = 1, 2, \ldots, N$, *where* $\sum_{i=1}^{n} p_i = 1$. *Choose* $x_0 \in \mathbf{X}$ *and then choose recursively, independently,*

$$x_n \in \{w_1(x_{n-1}), w_2(x_{n-1}), \ldots, w_N(x_{n-1})\} \qquad \text{for } n = 1, 2, 3, \ldots,$$

where the probability of the event $x_n = w_i(x_{n-1})$ is p_i. Thus, construct a sequence $\{x_n : n = 0, 1, 2, 3, \ldots\} \subset X$.

★ The reader should skip the rest of this paragraph and come back to it after reading Section 9. If $\{\mathbf{X}, w_0, w_1, w_2, \ldots, w_N\}$ is an IFS with condensation map w_0 and associated condensation set $C \subset \mathcal{H}(\mathbf{X})$, then the algorithm is modified by (a) attaching a probability $p_0 > 0$ to w_0, so now $\sum_{i=0}^{n} p_i = 1$; (b) whenever $w_0(x_{n-1})$ is selected for some n, choose x_n "at random" from C. Thus, in this case too, we construct a sequence $\{x_n : n = 0, 1, 2, \ldots\}$ of points in $\mathbf{X}$.

The sequence $\{x_n\}_{n=0}^{\infty}$ "converges to" the attractor of the IFS, under various conditions, in a manner that will be made precise in Chapter IX.

We illustrate the implementation of the algorithm. The following program computes and plots a thousand points on the attractor corresponding to the IFS code in Table III.1. The program is written in BASIC. It runs without modification on an

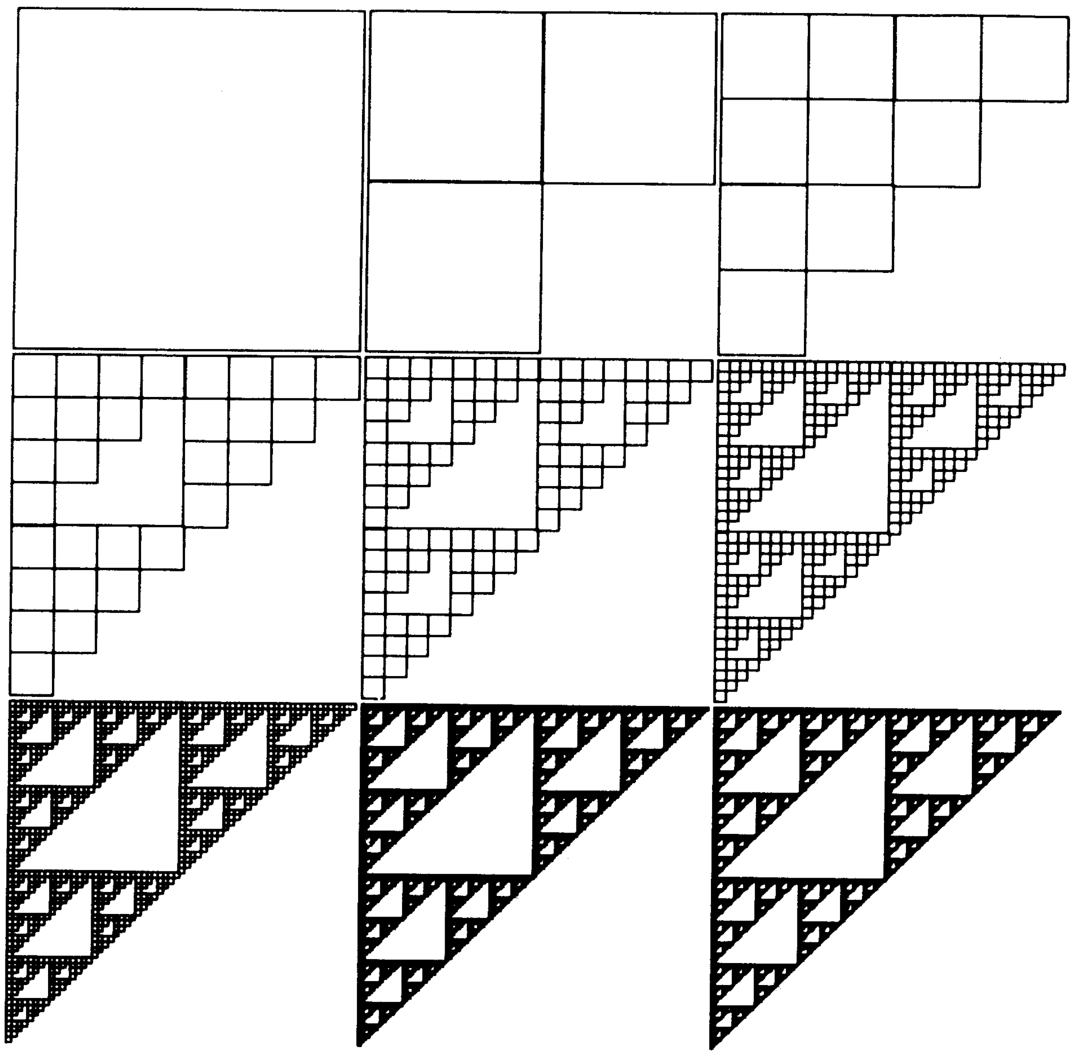


Figure III.64. The result of running the Deterministic Algorithm (Program 1) with various values of N, for the IFS code in Table III.1.

IBM PC with Enhanced Graphics Adaptor and Turbobasic. On any line the words preceded by a ' are comments: they are not part of the program.

Program 2. (Example of the Random Iteration Algorithm)

```
   'Iterated Function System Data
 a[1] = 0.5 : b[1] =0 : c[1] =0 : d[1] =.5 : e[1] =1 : f[1] =1
 a[2] = 0.5 : b[2] =0 : c[2] =0 : d[2] =.5 : e[2] =50 : f[2] =1
 a[3] = 0.5 : b[3] =0 : c[3] =0 : d[3] =.5 : e[3] =50 : f[3] =50

 screen 1 : cls 'initialize computer graphics
 window (0,0)-(100,100) 'set plotting window to 0<x<100, 0<y<100
 x =0  : y  =  0: numits =1000  'initialize (x,y) and define
      the number of iterations, numits
```

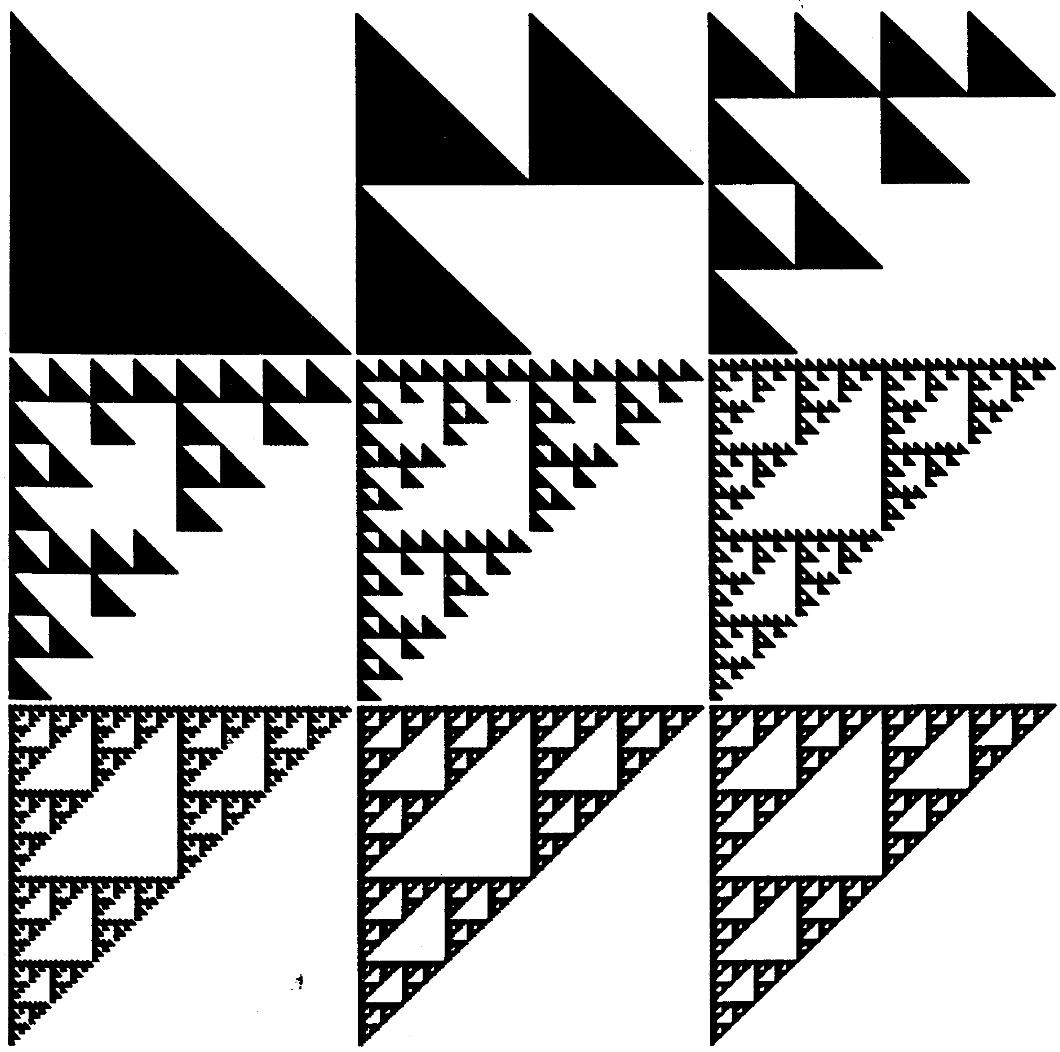


Figure III.65. The result of running the Deterministic Algorithm (Program 1), again for the IFS code in Table III.1, but starting from a different initial array. The final result is always the same!

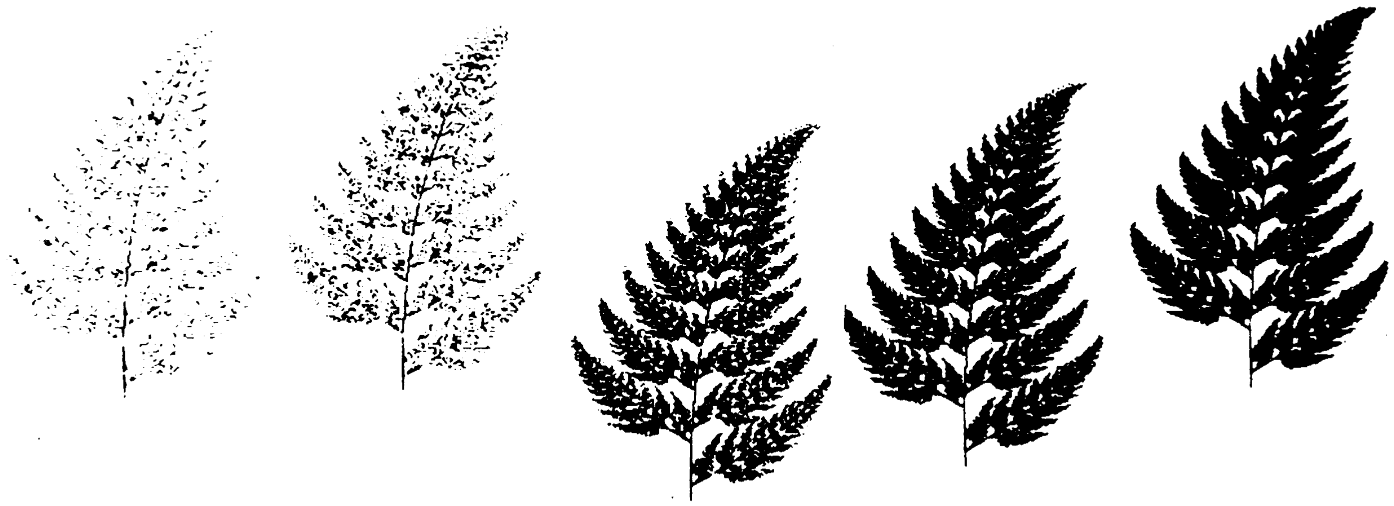


Figure III.66. The result of running the Random Iteration Algorithm for increasing numbers of iterations. The randomly dancing point starts to suggest the structure of the attractor of the IFS given in Table III.3.

```
for n =1  to numits 'Random Iteration begins!
k  =  int(3*rnd-0.00001)  +1  'choose one of the numbers 1, 2,
     and 3 with equal probability

'apply affine transformation number k to (x,y)
newx =a[k]*x+b[k]*y+e[k]  : newy =c[k]*x+d[k]*y+f[k]
x =newx  : y =newy  'set (x,y) to the point thus obtained
if n  >  10 then pset (x,y) 'plot (x,y) after the first 10
     iterations
next :   end
```

The result of running an adaptation of this program on a Masscomp workstation and then printing the contents of the graphics screen is presented in Figure III.66. Notice that if the size of the plotting window is decreased, for example by replacing the window call by WINDOW (0,0)–(50,50), then only a portion of the image is plotted, but at a higher resolution. Thus we have a simple means for "zooming in" on images of IFS attractors. The number of iterations may be increased to improve the quality of the computed image.

Examples & Exercises

8.1. Rewrite Programs 1 and 2 in a form suitable for your own computer environment, then run them and obtain hardcopy of the output. Compare their performance.

8.2. Modify Programs 1 and 2 so that they will compute images associated with the IFS code given in Table III.2.

8.3. Modify Program 2 so that it will compute images associated with the IFS codes given in Tables III.3 and III.4.

8.4. By changing the window size in Program 2, obtain images of "zooms" on the Sierpinski triangle. For example, use the following windows: (1, 1) – (50, 50); (1, 1) – (25, 25); (1, 1) – (12, 12); . . . ; (1, 1) – (N, N). How must the total number of iterations be adjusted as a function of N in order that (approximately) the number of points that land within the window remains constant? Make a graph of the total number of iterations against the window size.

8.5. What should happen, theoretically, to the sequence of images computed by Program 1 if the set A_0 is changed? What happens in practice? Make a computational experiment to see if there is any difference in say A_{10} corresponding to two different choices for A_0.

8.6. Rewrite Program 2 so that it applies the transformation w_i with probability p_i, where the probabilities are input by the user. Compare the number of iterations needed to produce a "good" rendering of the Sierpinski triangle, for the cases (a) $p_1 = 0.33$, $p_2 = 0.33$, $p_3 = 0.34$; (b) $p_1 = 0.2$, $p_2 = 0.46$, $p_3 = 0.34$; (c) $p_1 = 0.1$, $p_2 = 0.56$, $p_3 = 0.34$.

9 Condensation Sets

There is another important way of making contraction mappings on $\mathcal{H}(\mathbf{X})$.

Definition 9.1 *Let* $(\mathbf{X}, d)$ *be a metric space and let* $C \in \mathcal{H}(\mathbf{X})$. *Define a transformation* $w_0 : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ *by* $w_0(B) = C$ *for all* $B \in \mathcal{H}(\mathbf{X})$. *Then* w_0 *is called a* condensation transformation *and* C *is called the associated* condensation set.

Observe that a condensation transformation $w_0 : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ is a contraction mapping on the metric space $(\mathcal{H}(\mathbf{X}), h(d))$, with contractivity factor equal to zero, and that it possesses a unique fixed point, namely the condensation set.

Definition 9.2 *Let* $\{\mathbf{X}; w_1, w_2, \ldots, w_n\}$ *be a hyperbolic IFS with contractivity factor* $0 \le s < 1$. *Let* $w_0 : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ *be a condensation transformation. Then* $\{\mathbf{X}; w_0, w_1, \ldots, w_n\}$ *is called a* hyperbolic IFS with condensation, *with contractivity factor* s.

Theorem 7.1 can be modified to cover the case of an IFS with condensation.

Theorem 9.1 *Let* $\{\mathbf{X}; w_n : n = 0, 1, 2, \ldots, N\}$ *be a hyperbolic iterated function system with condensation, with contractivity factor* s. *Then the transformation* $W : \mathcal{H}(\mathbf{X}) \to \mathcal{H}(\mathbf{X})$ *defined by*

$$W(B) = \cup_{n=0}^{n} w_n(B) \forall B \in \mathcal{H}(\mathbf{X})$$

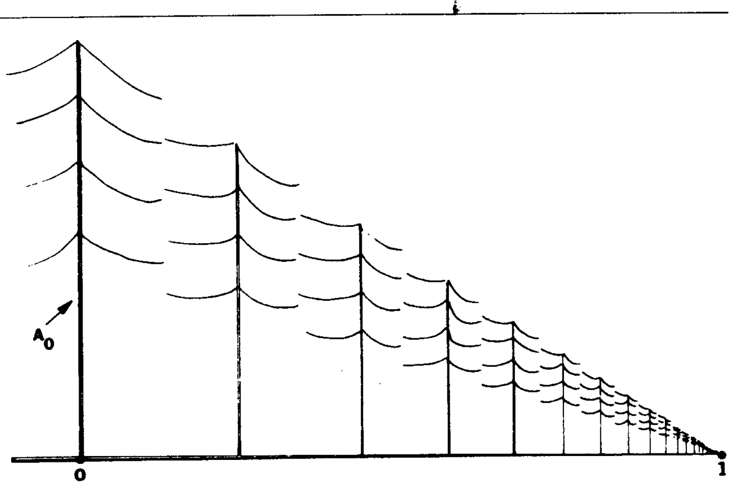


Figure III.67. A geometric series of pine trees, the attractor of an IFS with condensation.

is a contraction mapping on the complete metric space $(\mathcal{H}(\mathbf{X}), h(d))$ *with contractivity factor* s. *That is*

$$h(W(B), W(C)) \le s \cdot h(B, C) \forall B, C \in \mathcal{H}(\mathbf{X}).$$

Its unique fixed point, $A \in \mathcal{H}(\mathbf{X})$, *obeys*

$$A = W(A) = \cup_{n=0}^{n} w_n(A)$$

and is given by $A = \lim_{n\to\infty} W^{\circ n}(B)$ *for any* $B \in \mathcal{H}(\mathbf{X})$.

Examples & Exercises

9.1. A sequence of sets $\{A_n \subset \mathbf{X}\}_{n=0}^{\infty}$, where $(\mathbf{X}, d)$ is a metric space, is said to be *increasing* if $A_0 \subset A_1 \subset A_2 \subset \cdots$ and *decreasing* if $A_0 \supset A_1 \supset A_2 \supset \cdots$. The inclusions are not necessarily strict. A decreasing sequence of sets $\{A_n \subset \mathcal{H}(\mathbf{X})\}_{n=0}^{\infty}$ is a Cauchy sequence (prove it!). If $\mathbf{X}$ is compact then an increasing sequence of sets $\{A_n \subset \mathcal{H}(\mathbf{X})\}_{n=0}^{\infty}$ is a Cauchy sequence (prove it!). Let $\{\mathbf{X}; w_0, w_1, \ldots, w_n\}$ be a hyperbolic IFS with condensation set C, and let $\mathbf{X}$ be compact. Let $W_0(B) = \cup_{n=0}^{n} w_n(B) \forall B \in \mathcal{H}(\mathbf{X})$ and let $W(B) = \cup_{n=1}^{n} w_n(B)$. Define $\{C_n = W_0^{\circ n}(C)\}_{n=0}^{\infty}$. Then Theorem 9.1 tells us $\{C_n\}$ is a Cauchy sequence in $\mathcal{H}(\mathbf{X})$ that converges to the attractor of the IFS. Independently of the theorem observe that

$$C_n = C \cup W(C) \cup W^{\circ 2}(C) \cup \ldots \cup W^{\circ n}(C)$$

provides an increasing sequence of compact sets. It follows immediately that the limit set A obeys $W_0(A) = A$.

9.2. This example takes place in $(\mathbb{R}^2$, Euclidean). Let $C =$ $= A_0 \subset \mathbb{R}^2$

denote a set that looks like a scorched pine tree standing at the origin, with its trunk perpendicular to the x-axis. Let

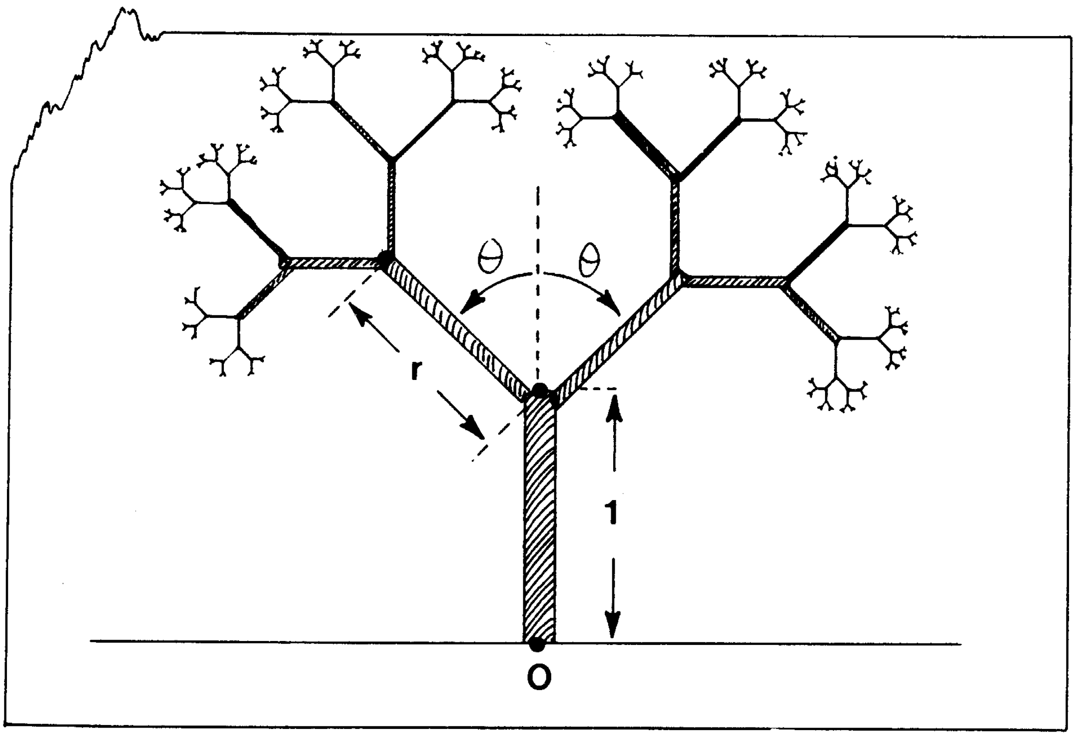


Figure III.68. Sketch of a fractal tree, the attractor of an IFS with condensation.

$$w_1\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0.75 & 0\\0 & 0.75\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}0.25\\0\end{pmatrix}.$$

Show that $\{\mathbb{R}^2; w_0, w_1\}$ is an IFS with condensation and find its contractivity factor. Let $A_n = W^{\circ n}(A_0)$ for $n = 1, 2, 3, \ldots$, where $W(B) = \cup_{n=0}^{n} w_n(B)$ for $B \in \mathcal{H}(\mathbb{R}^2)$. Show that A_n consists of the first $(n+1)$ pine trees reading from left to right in Figure III.67. If the first tree required 0.1% of the ink in the artist's pen to draw, and if the artist had been very meticulous in drawing the whole attractor correctly, find the total amount of ink used to draw the whole attractor.

9.3. What happens to the trees in Figure III.67 if $w_1\begin{pmatrix}x\\y\end{pmatrix}$ is replaced by

$$w_1\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0.5 & 0\\0 & 0.75\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}0.5\\0\end{pmatrix}$$

in exercise 9.2?

9.4. Find the attractor for the IFS with condensation $\{\mathbb{R}^2; w_0, w_1\}$, where the condensation set is the interval $[0, 1]$ and $w_1(x) = \frac{1}{2}x + 2$. What happens if $w_1(x) = \frac{1}{2}x$?

9.5. Find an IFS with condensation that generates the treelike set in Figure III.68. Give conditions on r and θ such that the tree is simply connected. Show that the tree is either simply connected or infinitely connected.

9.6. Find an IFS with condensation that generates Figure III.69.

9.7. You are given a condensation map $w_0(x)$ in $\mathbb{R}^2$ that provides the largest tree

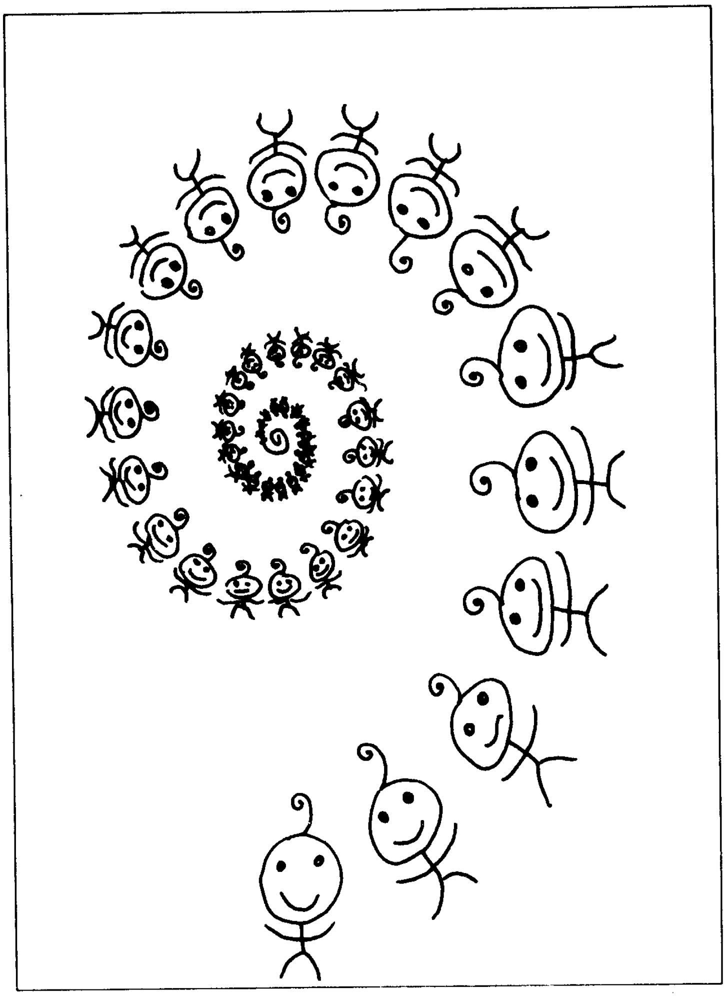


Figure III.69. An endless spiral of little men.

in Figure III.46. Find a hyperbolic IFS with condensation, of the form $\{\mathbb{R}^2; w_0, w_1, w_2\}$, which produces the whole orchard. What is the contractivity factor for this IFS? Find the attractor of the IFS $\{\mathbb{R}^2; w_1, w_2\}$.

9.8. Explain why removing the command that clears the screen ("cls") from Program 1 will result in the computation of an image associated with an IFS with condensation. Identify the condensation set. Run your version of Program 1 with the "cls" command removed.

10 How to Make Fractal Models with the Help of the Collage Theorem

The following theorem is central to the design of IFS's whose attractors are close to given sets.

Theorem 10.1 (The Collage Theorem, (Barnsley 1985b)). *Let* $(\mathbf{X}, d)$ *be a complete metric space. Let* $L \in \mathcal{H}(\mathbf{X})$ *be given, and let* $\epsilon \geq 0$ *be given. Choose an IFS (or IFS with condensation)* $\{\mathbf{X}; (w_0), w_1, w_2, \ldots, w_n\}$ *with contractivity factor*

$0 \leq s < 1$, *so that*

$$h(L, \cup^n_{\substack{n=1 \\ (n=0)}} w_n(L)) \leq \epsilon,$$

where $h(d)$ *is the Hausdorff metric. Then*

$$h(L, A) \leq \epsilon/(1-s),$$

where A *is the attractor of the IFS. Equivalently,*

$$h(L, A) \leq (1-s)^{-1} h(L, \cup^n_{\substack{n=1 \\ (n=0)}} w_n(L)) \qquad \text{for all } L \in \mathcal{H}(\mathbf{X}).$$

The proof of the Collage Theorem is given in the next section. The theorem tells us that to find an IFS whose attractor is "close to" or "looks like" a given set, one must endeavor to find a set of transformations—contraction mappings on a suitable space within which the given set lies—such that the union, or collage, of the images of the given set under the transformations is near to the given set. Nearness is measured using the Hausdorff metric.

Examples & Exercises

10.1. This example takes place in ($\mathbb{R}$, Euclidean). Observe that $[0, 1] = [0, \frac{1}{2}] \cup [\frac{1}{2}, 1]$. Hence the attractor is $[0, 1]$ for any pair of contraction mappings $w_1 : \mathbb{R} \to \mathbb{R}$ and $w_2 : \mathbb{R} \to \mathbb{R}$ such that $w_1([0, 1]) = [0, \frac{1}{2}]$ and $w_2([0, 1]) = [\frac{1}{2}, 1]$. For example, $w_1(x) = \frac{1}{2}x$ and $w_2(x) = \frac{1}{2}x + \frac{1}{2}$ does the trick. The unit interval is a collage of two smaller "copies" of itself.

10.2. Suppose we are using a trial-and-error procedure to adjust the coefficients in two affine transformations $w_1(x) = ax + b$, $w_2(x) = cx + d$, where $a, b, c, d \in \mathbb{R}$, to look for an IFS $\{\mathbb{R}; w_1, w_2\}$ whose attractor is $[0, 1]$. We might come up with $w_1(x) = 0.51x - 0.01$ and $w_2(x) = 0.47x + 0.53$. How far from $[0, 1]$ will the attractor for the IFS be? To find out compute

$$h\left([0, 1], \cup_{i=1}^2 w_i([0, 1])\right) = h([0, 1], [-0.0l, 0.5] \cup [0.53, 1]) = 0.015$$

and observe that the contractivity factor of the IFS is $s = 0.51$. So by the Collage Theorem, if A is the attractor,

$$h([0, 1], A) \leq 0.015/0.49 < 0.04.$$

10.3. Figure III.70 shows a target set $L \subset \mathbb{R}^2$, a leaf, represented by the polygonalized boundary of the leaf. Four affine transformations, contractive, have been applied to the boundary at lower left, producing the four smaller deformed leaf boundaries. The Hausdorff distance between the union of the four copies and the original is approximately 1.0 units, where the width of the whole frame is taken to be 10 units. The contractivity of the associated IFS $\{\mathbb{R}^2; w_1, w_2, w_3, w_4\}$ is approximately 0.6. Hence the Hausdorff distance h(Euclidean) between the original target leaf L and the attractor A of the IFS will be less than 2.5 units. (This is not promising much!) The actual attractor, translated to the right, is shown at lower right. Not surprisingly,

Figure III.70. The Collage Theorem applied to a region bounded by a polygonalized leaf boundary.

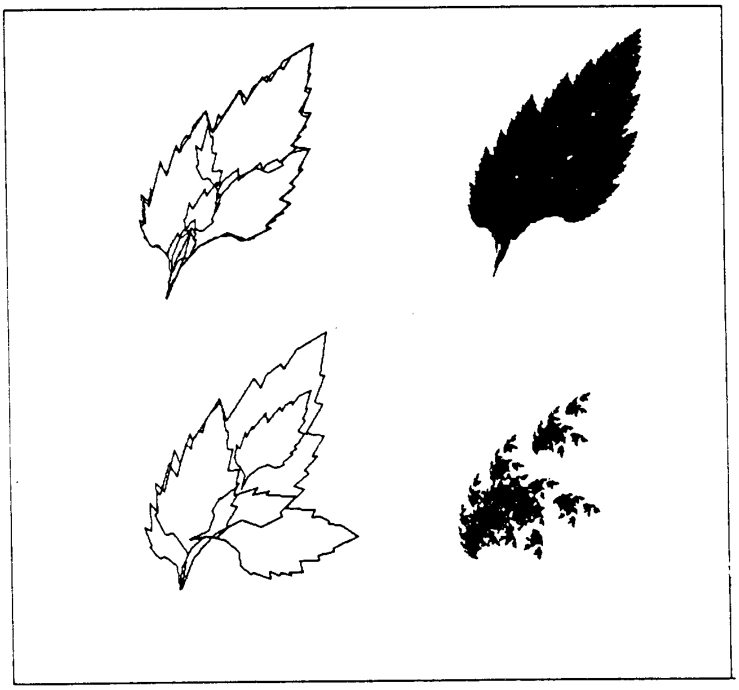
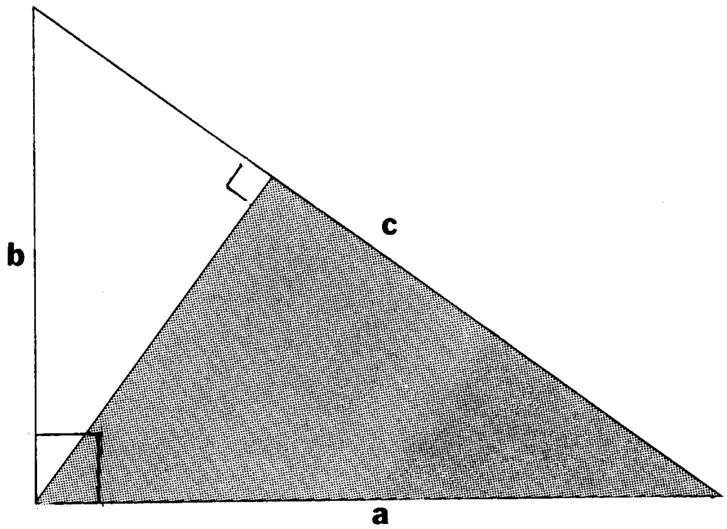


Figure III.71. The region bounded by a right-angle triangle is the union of the results of two similitudes applied to it.



it does not look much like the original leaf! An improved collage is shown at the upper left. The distance $h(L, \cup_{n=1}^{4} w_n(L))$ is now less than 0.02 units, while the contractivity of the IFS is still approximately 0.6. Hence $h(L, A)$ should now be less than 0.05 units, and we expect that the attractor should look quite like L at the resolution of the figure. A, translated to the right, is shown at the upper right.

10.4. To find an IFS whose attractor is a region bounded by a right-angle triangle, observe the collage in Figure III.71.

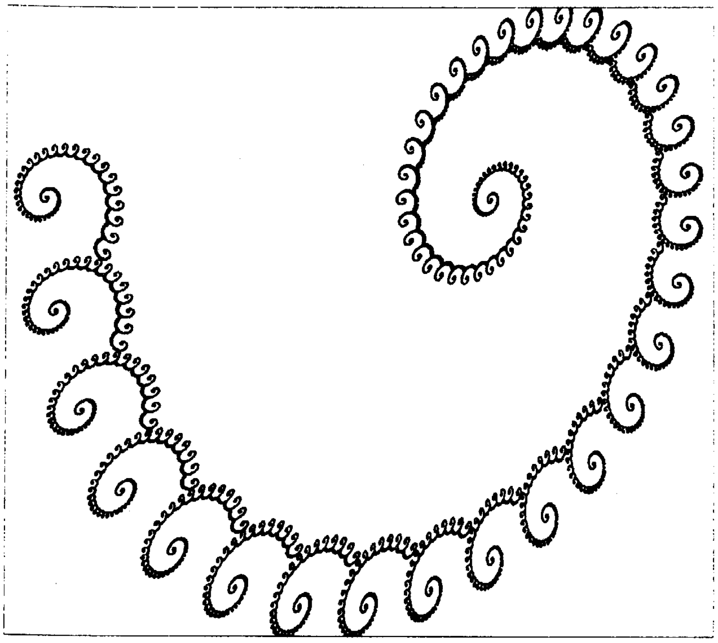
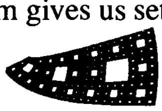


Figure III.72. Use the Collage Theorem to help you find an IFS consisting of two affine maps in $\mathbb{R}^2$ whose attractor is close to this set.

10.5. A nice proof of Pythagoras' Theorem is obtained from the collage in Figure III.71. Clearly both transformations involved are similitudes. The contractivity factors of these similitudes involved are (b/c) and (a/c). Hence the area $\mathcal{A}$ obeys $\mathcal{A} = (b/c)^2\mathcal{A} + (a/c)^2\mathcal{A}$. This implies $c^2 = a^2 + b^2$ since $\mathcal{A} > 0$.

10.6. Figures III.72–III.76 provide exercises in the application of the Collage Theorem. Condensation sets are not allowed in working these examples!

10.7. It is straightforward to see how the Collage Theorem gives us sets of maps for IFS's that generate ▲. A Menger Sponge looks like this: . Find an IFS for which it is the attractor.

10.8. The IFS that generates the *Black Spleenwort* fern, shown in Figure III.77, consists of four affine maps in the form

$$w_i\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta & -s\sin\theta \\ r\sin\theta & s\cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} h \\ k \end{pmatrix} \quad (i = 1, 2, 3, 4);$$

see Table III.5.

10.9. Find a collage of affine transformations in $\mathbb{R}^2$, corresponding to Figure III.78.

10.10. A collage of a leaf is shown in Figure III.79 (a). This collage implies the IFS $\{\mathbb{C}; w_1, w_2, w_3, w_4\}$ where, in complex notation,

$$w_i(z) = s_i z + (1 - s_i)a_i \qquad \text{for } i = 1, 2, 3, 4.$$

Verify that in this formula a_i is the fixed point of the transformation. The values found for s_i and a_i are listed in Table III.6. Check that these make sense in relation to the collage. The attractor for the IFS is shown in Figure III.79 (b).

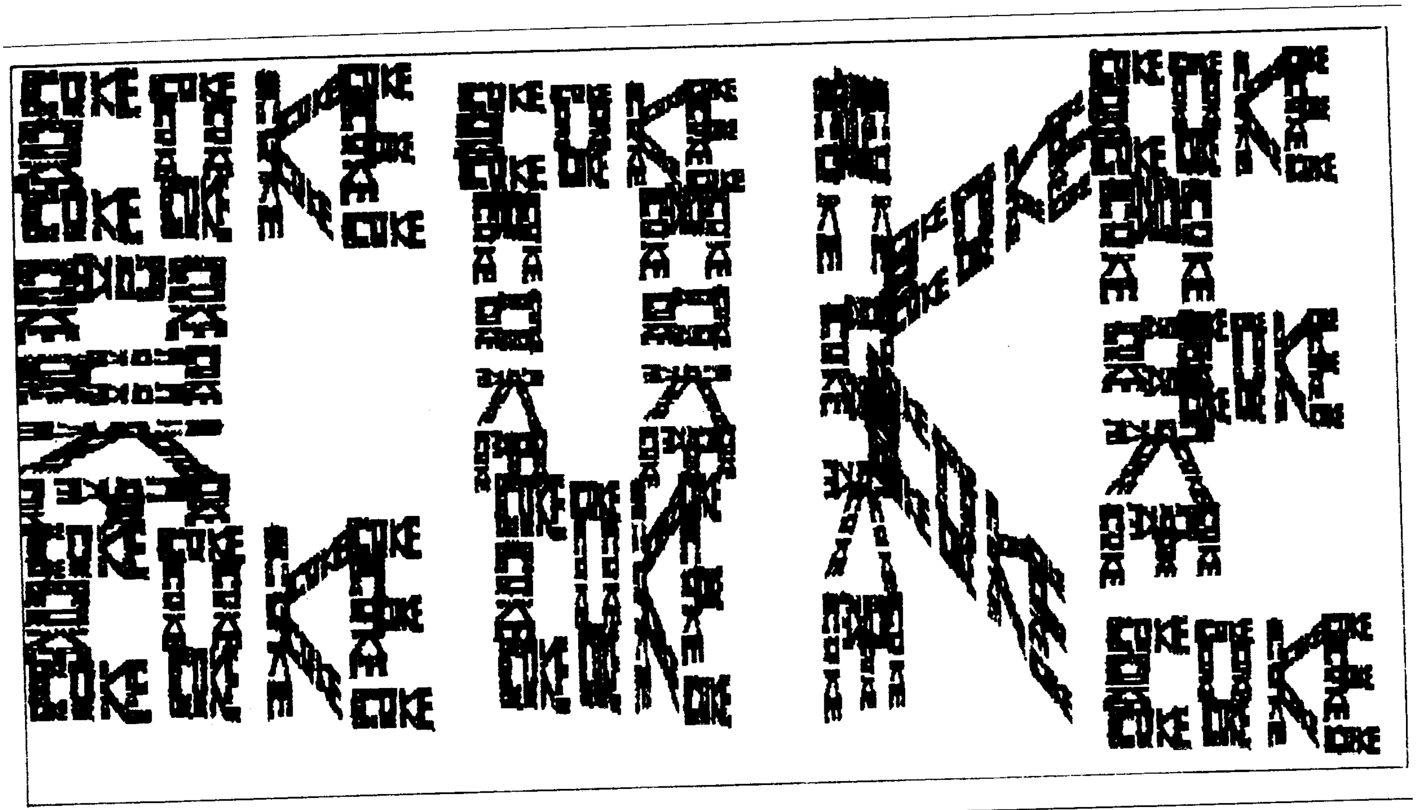
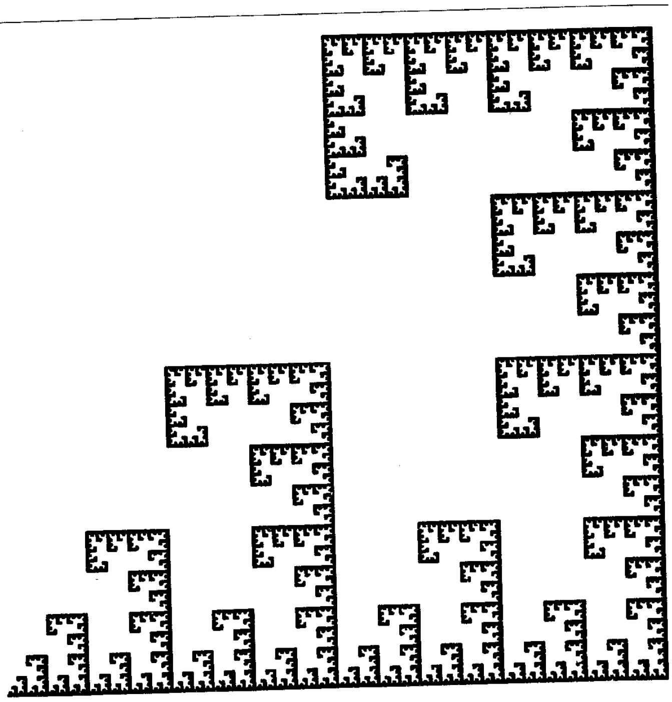


Figure III.73. This image represents the attractor of 14 affine transformations in $\mathbb{R}^2$. Use the Collage Theorem to help you find them.

Figure III.74. Use the Collage Theorem to help find a hyperbolic IFS of the form $\{\mathbb{R}^2; w_1, w_2, w_3\}$, where w_1, w_2, and w_3 are similitudes in $\mathbb{R}^2$, whose attractor is represented here. You choose the coordinate system.



10.11. The attractor in Figure III.80 is determined by two affine maps. Locate the fixed points of two such affine transformations on $\mathbb{R}^2$.

10.12. Figure III.81 shows the attractor for an IFS $\{\mathbb{R}^2; w_i, i = 1, 2, 3, 4\}$ where each w_i is a three-dimensional affine transformation. See also Color Plate 3. The attractor is contained in the region $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : -10 \leq x_1 \leq 10, 0 \leq x_2 \leq 10, -10 \leq x_3 \leq 10\}$.

$$w_1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.18 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$w_2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.85 & 0 & 0 \\ 0 & 0.85 & 0.1 \\ 0 & -0.1 & 0.85 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1.6 \\ 0 \end{bmatrix}$$

$$w_3 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.2 & -0.2 & 0 \\ 0.2 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.8 \\ 0 \end{bmatrix}$$

$$w_4 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -0.2 & 0.2 & 0 \\ 0.2 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.8 \\ 0 \end{bmatrix}$$

10.13. Find an IFS of similitudes in $\mathbb{R}^2$ such that the attractor is represented by the shaded region in Figure III.82. The collage should be "just-touching," by which we mean that the transforms of the region provide a tiling of the region: they should fit together like the pieces of a jigsaw puzzle.

10.14. This exercise suggests how to change the coordinates of an IFS. Let $\{\mathbf{X}_1, d_1\}$ and $\{\mathbf{X}_2, d_2\}$ be metric spaces. Let $\{\mathbf{X}_1; w_1, w_2, \ldots, w_N\}$ be a hyperbolic IFS with attractor A_1. Let $\theta : \mathbf{X}_1 \to \mathbf{X}_2$ be an invertible continuous transformation. Consider the IFS $\{\mathbf{X}_2; \theta \circ w_1 \circ \theta^{-1}, \theta \circ w_2 \circ \theta^{-1}, \ldots, \theta \circ w_N \circ \theta^{-1}\}$. Use θ to define a metric on $\mathbf{X}_2$ such that the new IFS is indeed a hyperbolic IFS. Prove that if $A_2 \in \mathcal{H}(\mathbf{X}_2)$ is

Table III.5. The IFS code for the Black Spleenwort, expressed in scale and angle formats.

	Translations		*Rotations*		*Scalings*	
Map	h	k	θ	ϕ	r	s
1	0.0	0.0	0	0	0.0	0.16
2	0.0	1.6	-2.5	-2.5	0.85	0.85
3	0.0	1.6	49	49	0.3	0.34
4	0.0	0.44	120	-50	0.3	0.37

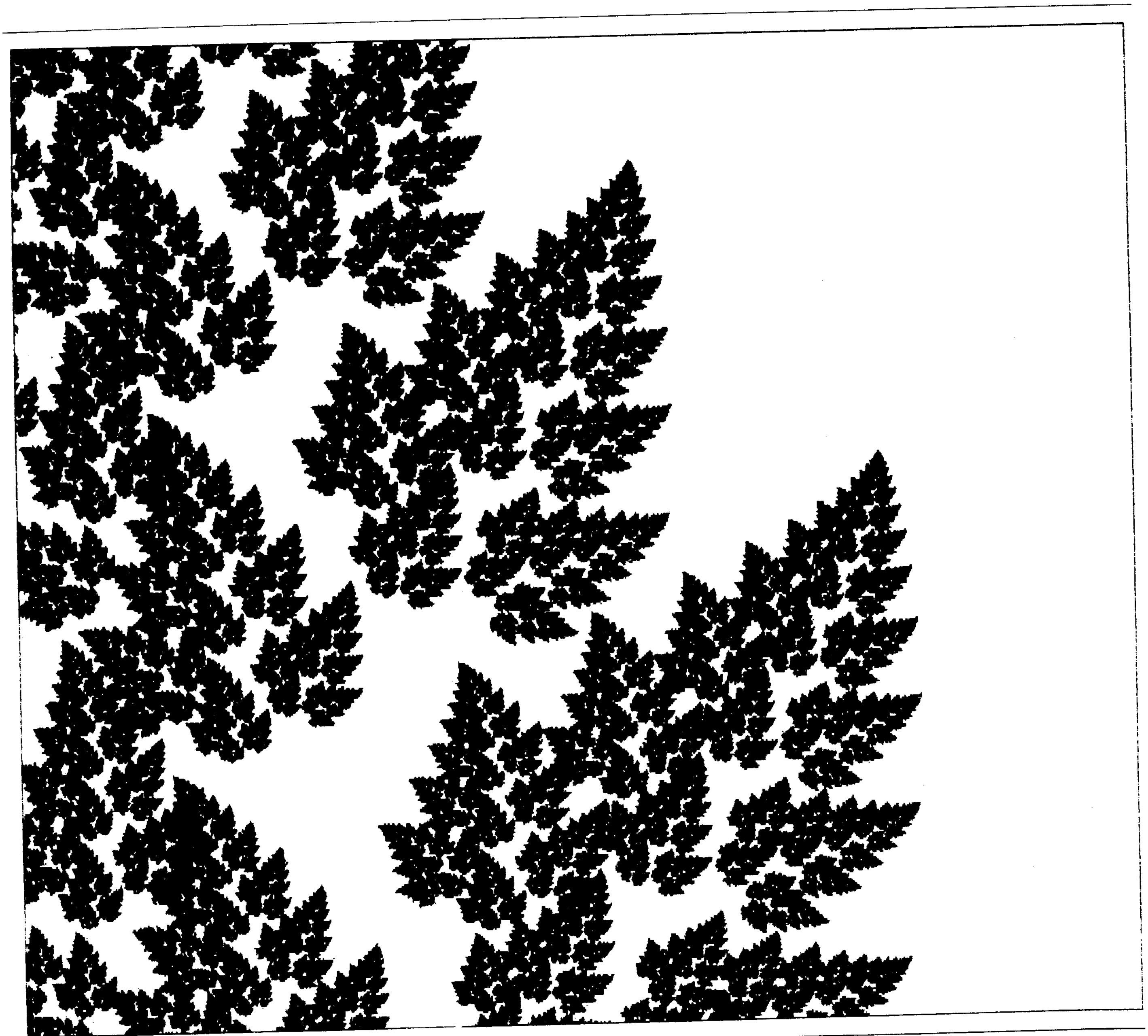


Figure III.75. Find an IFS of the form $\{\mathbb{R}^2; w_1, w_2, w_3, w_4\}$, where the w_i's are affine transformations on $\mathbb{R}^2$, whose attractor when rendered contains this image. Check your conclusion using Program 2.

the attractor of the new IFS, then $A_2 = \theta(A_1)$. Thus we can readily construct an IFS whose attractor is a transform of the attractor of another IFS.

10.15. Find some of the affine transformations used in the design of the fractal scene in Figure III.83.

10.16. Use the Collage Theorem to find an IFS whose attractor approximates the set in Figure III.84.

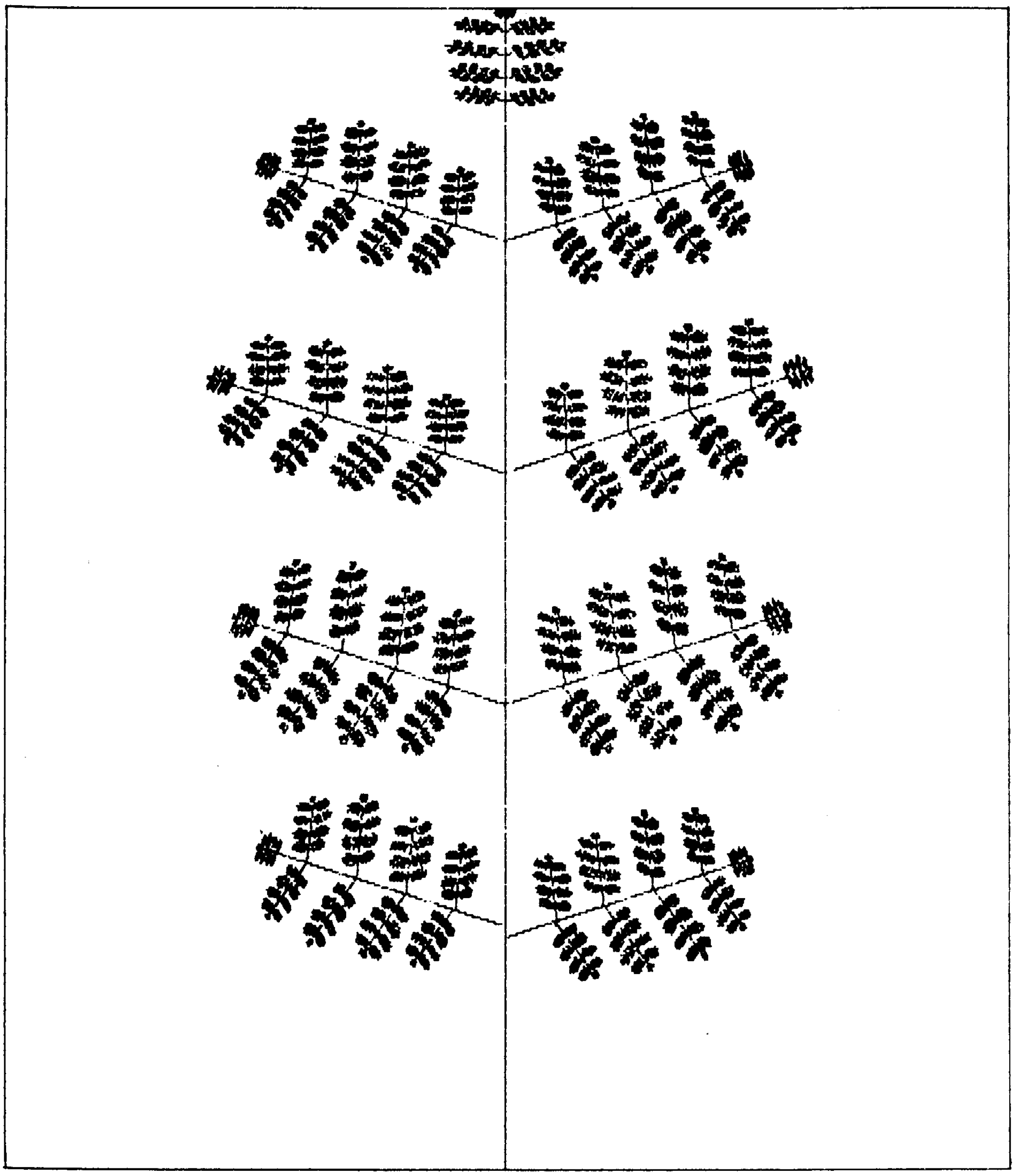
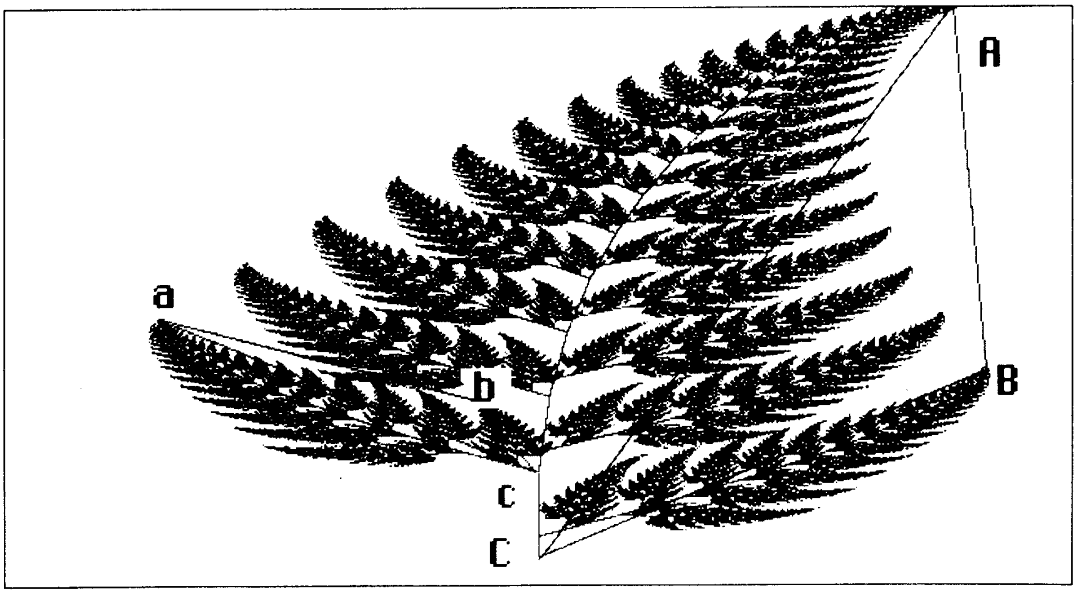


Figure III.76. How many affine transformations in $\mathbb{R}^2$ are needed to generate this attractor? You do not need to use a condensation set.

10.17. Solve the problems proposed in the captions of (a) Figure III.85, (b) Figure III.86, (c) Figure III.87.

11 Blowing in the Wind: The Continuous Dependence of Fractals on Parameters

The Collage Theorem provides a way of approaching the inverse problem: given a set L, find an IFS for which L is the attractor. The underlying mathematical principle



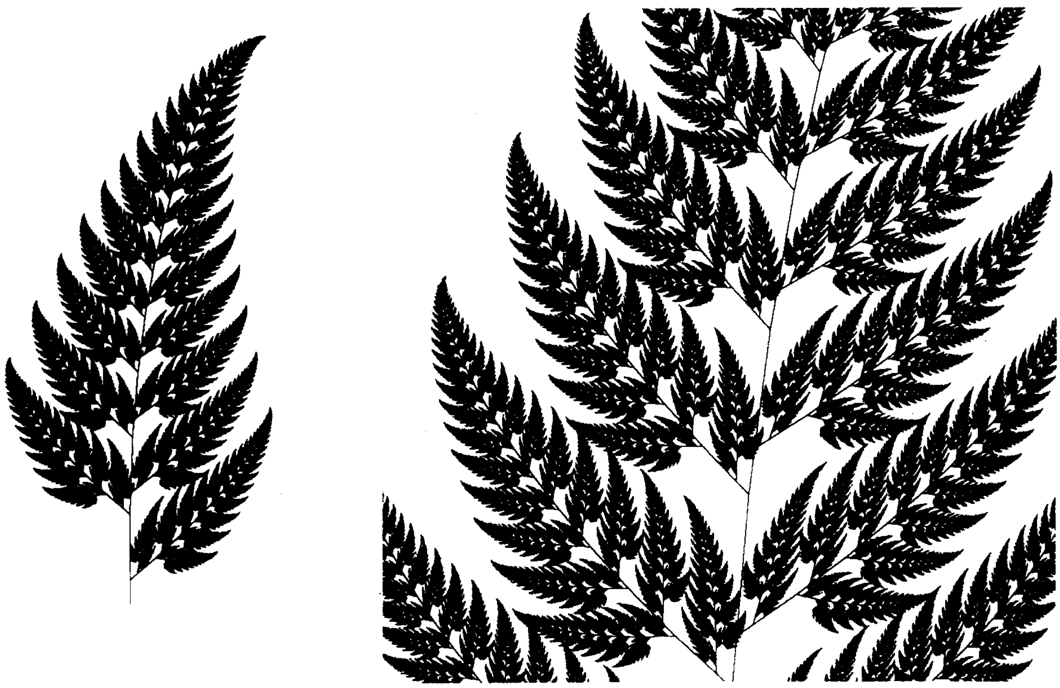


Figure III.77. The Black Spleenwort fern. The top image illustrates one of the four affine transformations in the IFS whose attractor was used to render the fern. The transformation takes the triangle ABC to triangle abc. The Collage Theorem provides the other three transformations. The IFS coded for this image is given in Table III.3. Observe that the stem is the image of the whole set under one of the transformations. Determine to which map number in Table III.3 the stem corresponds.The bottom image shows the Black Spleenwort fern and a close-up.

is very easy: the proof of the Collage Theorem is just the proof of the following lemma.

Lemma 11.1 *Let* $(\mathbf{X}, d)$ *be a complete metric space. Let* $f : \mathbf{X} \to \mathbf{X}$ *be a contraction mapping with contractivity factor* $0 \le s < 1$, *and let the fixed point of* f *be* $x_f \in \mathbf{X}$. *Then*

$$d(x, x_f) \le (1 - s)^{-1} \cdot d(x, f(x)) \text{ for all } x \in \mathbf{X}.$$

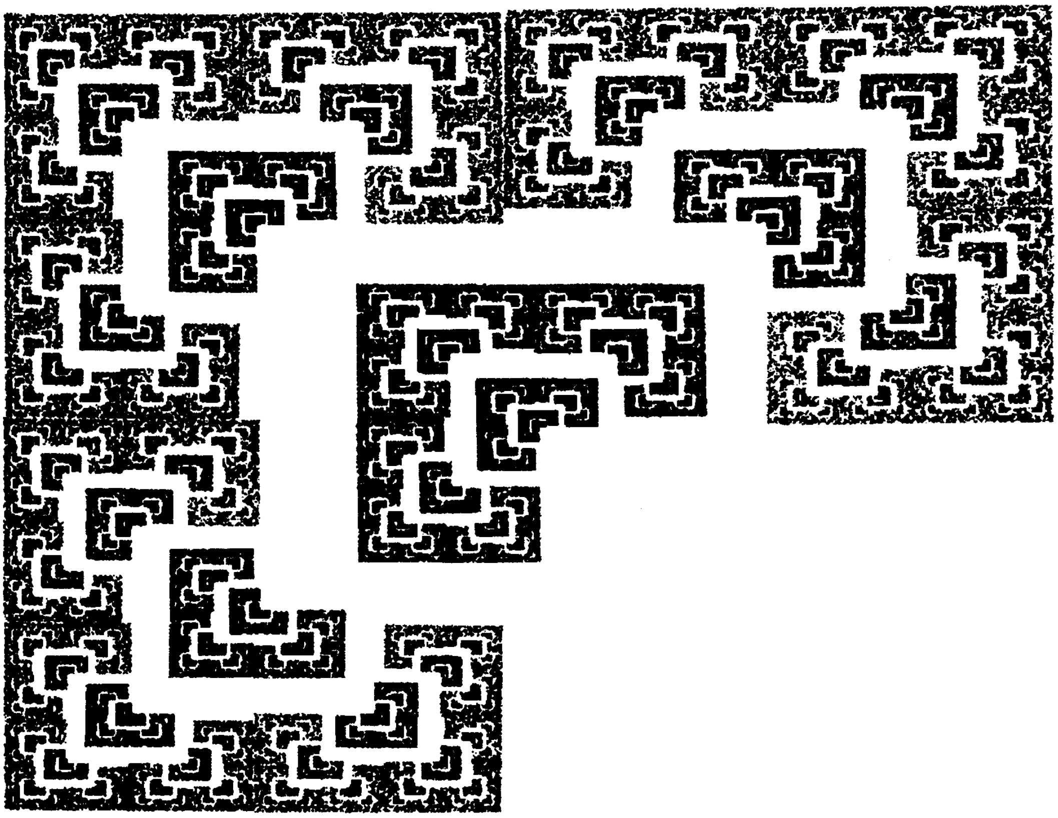
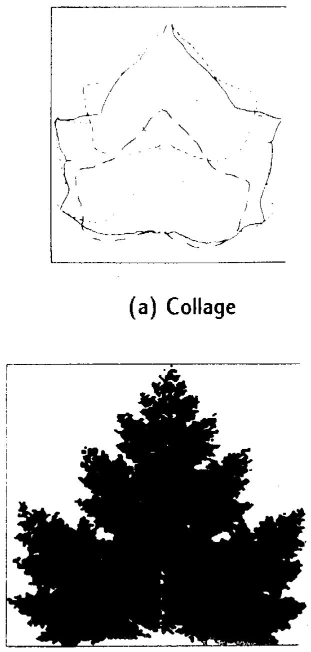


Figure III.78. Use the Collage Theorem to find the four affine transformations corresponding to this image. Can you find a transformation which will put in the "missing corner"?



(a) Collage

(b) Attractor

Figure III.79. A collage of a leaf is obtained using four similitudes, as illustrated in (a). The corresponding IFS is presented in complex notation in Table III.6. The attractor of the IFS is rendered in (b).

Proof The distance function $d(a, b)$, for fixed $a \in \mathbf{X}$, is continuous in $b \in \mathbf{X}$. Hence

$$\begin{aligned} d(x, x_f) &= d\left(x, \lim_{n\to\infty} f^{\circ n}(x)\right) = \lim_{n\to\infty} d(x, f^{\circ n}(x)) \\ &\le \lim_{n\to\infty} \sum_{m=1}^{n} d(f^{\circ(m-1)}(x), f^{\circ(m)}(x)) \\ &\le \lim_{n\to\infty} d(x, f(x))(1 + s + \cdots + s^{n-1}) \le (1-s)^{-1} d(x, f(x)). \end{aligned}$$

This completes the proof.

The following results are important and closely related to the above material. They establish the continuous dependence of the attractor of a hyperbolic IFS on parameters in the maps that constitute the IFS.

Table III.6. Scaling factors and fixed points for the collage in Figure III.79.

s	a
0.6	0.45 + 0.9i
0.6	0.45 + 0.3i
0.4 - 0.3i	0.60 + 0.3i
0.4 + 0.3i	0.30 + 0.3i

Lemma 11.2 *Let (P, d_p) and $(\mathbf{X}, d)$ be metric spaces, the latter being complete. Let $w : P \times \mathbf{X} \to \mathbf{X}$ be a family of contraction mappings on $\mathbf{X}$ with contractivity factor $0 \le s < 1$. That is, for each $p \in P$, $w(p, \cdot)$ is a contraction mapping on $\mathbf{X}$. For each fixed $x \in \mathbf{X}$ let w be continuous on P. Then the fixed point of w depends continuously on p. That is, $x_f : P \to \mathbf{X}$ is continuous.*

Proof Let $x_f(p)$ denote the fixed point of w for fixed $p \in P$. Let $p \in P$ and $\epsilon > 0$ be given. Then for all $q \in P$,

$$\begin{aligned} d(x_f(p), x_f(q)) &= d(w(p, x_f(p)), w(q, x_f(q))) \\ &\le d(w(p, x_f(p)), w(q, x_f(p))) \\ &\quad + d(w(q, x_f(p)), w(q, x_f(q))) \\ &\le d(w(p, x_f(p)), w(q, x_f(p))) + sd(x_f(p), x_f(q)), \end{aligned}$$

which implies

$$d(x_f(p), x_f(q)) \le (1-s)^{-1} d(w(p, x_f(p)), w(q, x_f(p))).$$

The right-hand side here can be made arbitrarily small by restricting q to be sufficiently close to p. (Notice that if there is a real constant C such that

$$d(w(p, x), w(q, x)) \le Cd(p, q) \qquad \text{for all } p, q \in P, \qquad \text{for all } x \in \mathbf{X},$$

then $d(x_f(p), x_f(q)) \le (1-s)^{-1} \cdot C \cdot d(p, q)$, which is a useful estimate.) This completes the proof.

Examples & Exercises

11.1. The fixed point of the contraction mapping $w : \mathbb{R} \to \mathbb{R}$ defined by $w(x) = \frac{1}{2}x + p$ depends continuously on the real parameter p. Indeed, $x_f = 2p$.

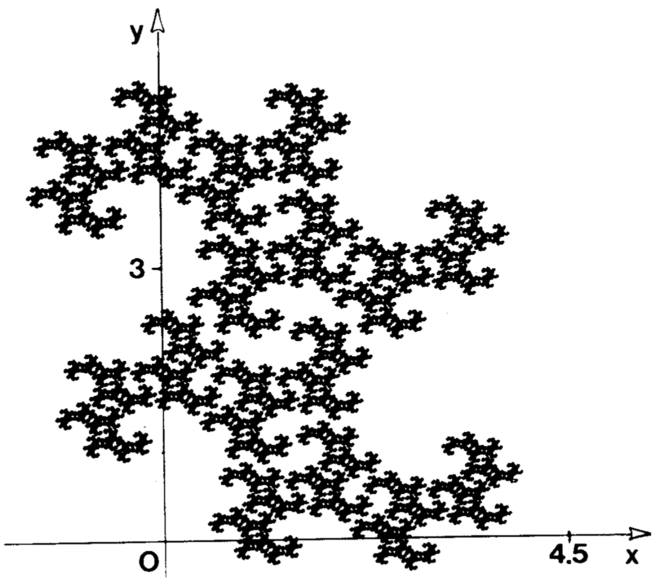


Figure III.80. Locate the fixed points of a pair of affine transformations in $\mathbb{R}^2$ whose attractor is rendered here.

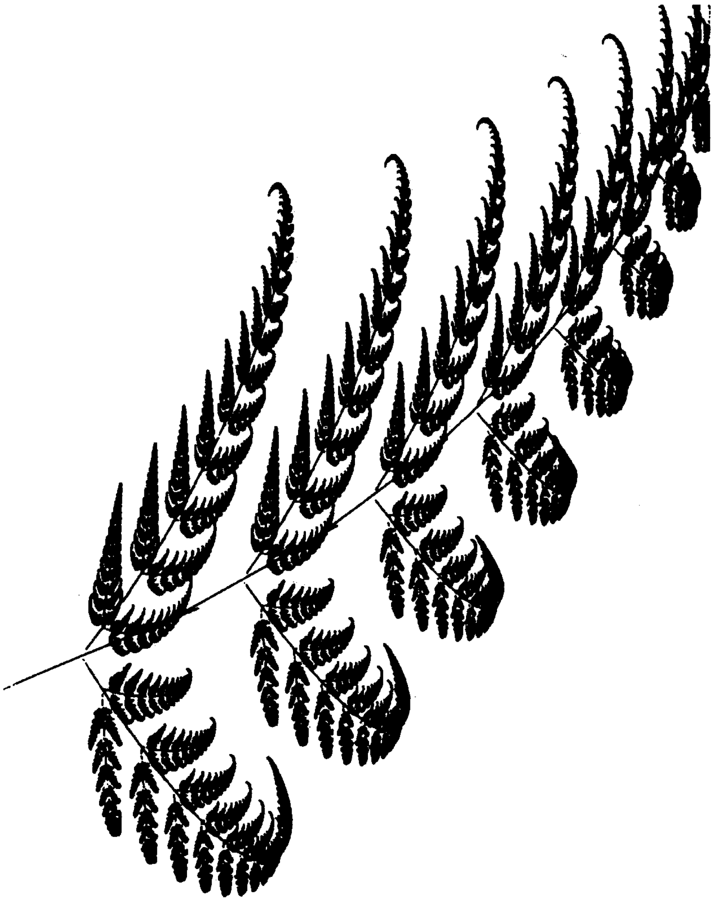


Figure III.81. Single three-dimensional fern. The attractor of an IFS of affine maps in $\mathbb{R}^3$.

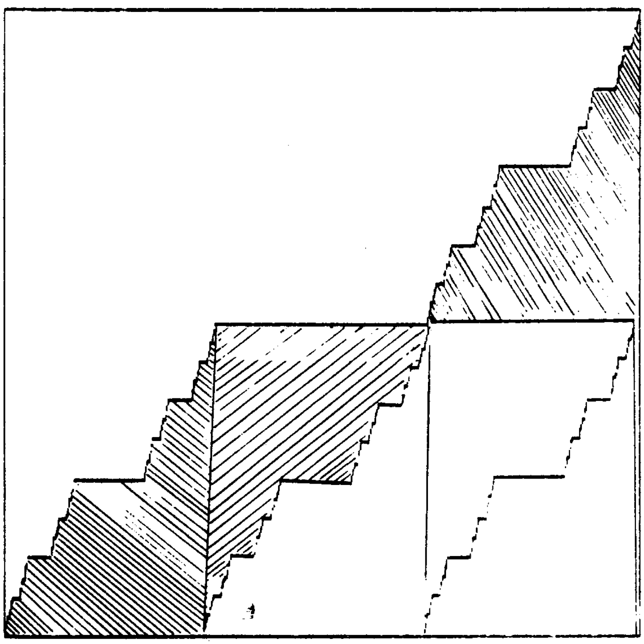


Figure III.82. Find a "just-touching" collage of the area under this Devil's Staircase.

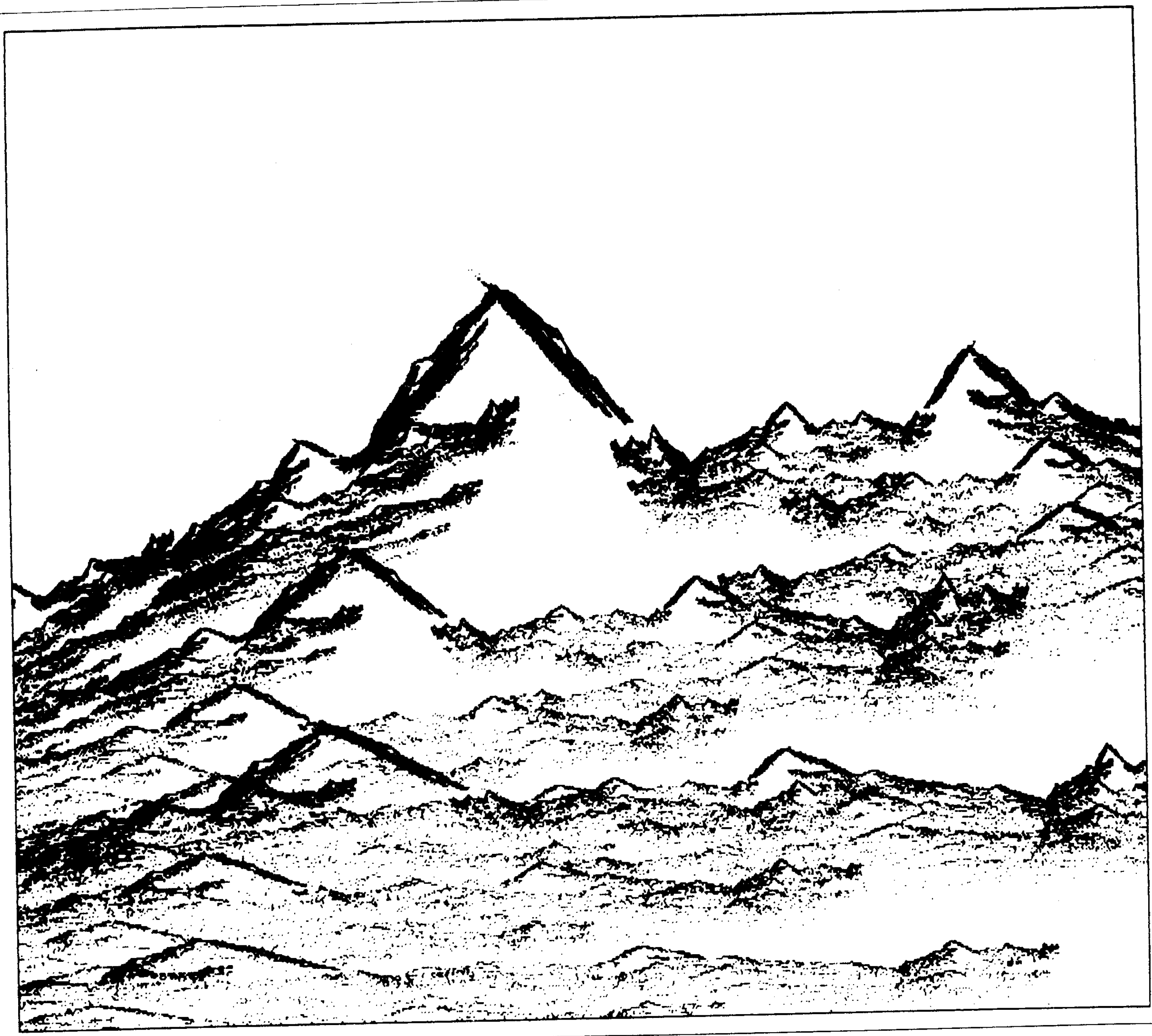


Figure III.83. Determine some of the affine transformations used in the design of this fractal scene. For example, where do the dark sides of the largest mountain come from?

11.2. Show that the fixed function for the transformation $w : C^0[0,1] \to C^0[0,1]$ defined by $w(f(x)) = pf(2x mod 1) + x(1-x)$ is continuous in p for $p \in (-1,1)$. Here, $C^0[0,1] = \{f \in C[0,1] : f(0) = f(1) = 0\}$ and the distance is $d(f,g) = \max\{|f(x) - g(x)| : x \in [0,1]\}$.

In order for this to be of use to us, we need some method of moving the continuous dependence on the parameter p to $\mathcal{H}(X)$. We cannot do this just because the image of a point in some set B depends continuously on p, since, although this gives

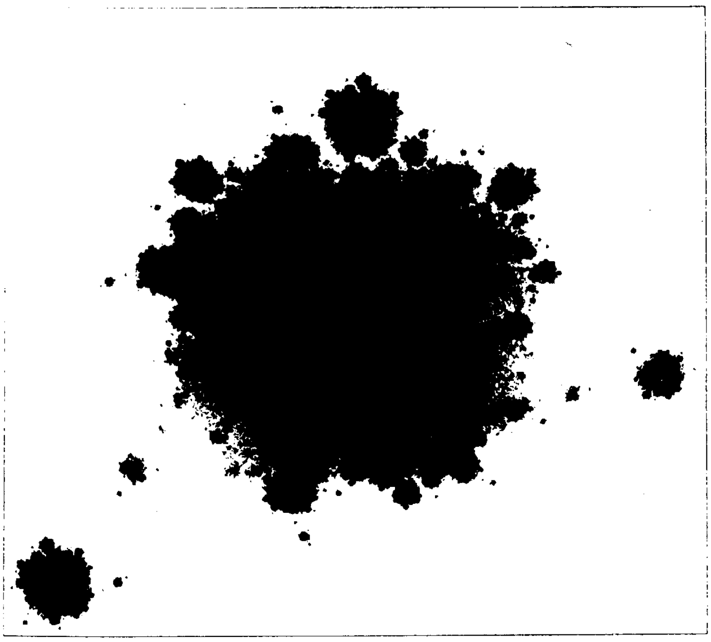


Figure III.84. "Typical" fractals are not pretty: use the Collage Theorem to find an IFS whose attractor approximates this set.

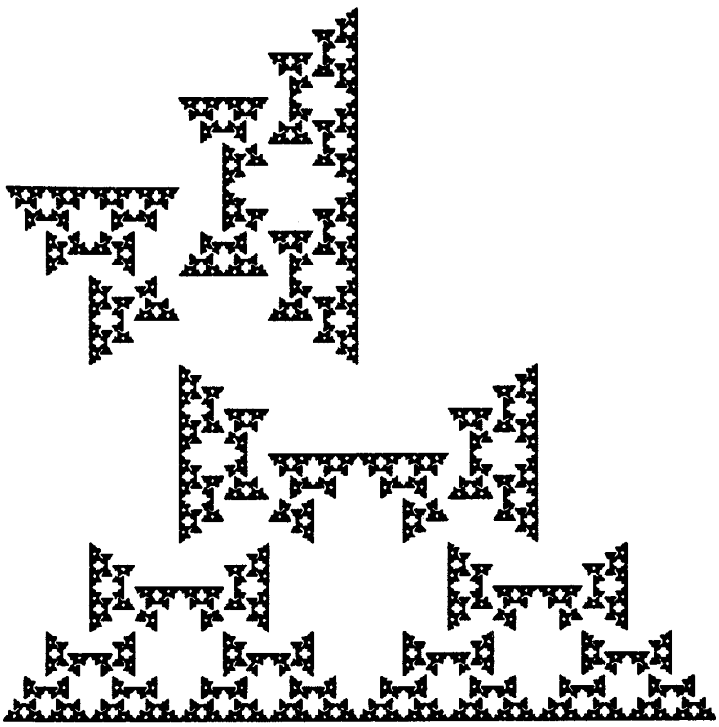


Figure III.85. Determine the affine transformations for an IFS corresponding to this fractal. Can you see, just by looking at the picture, if the linear part of any of the transformations has a negative determinant?

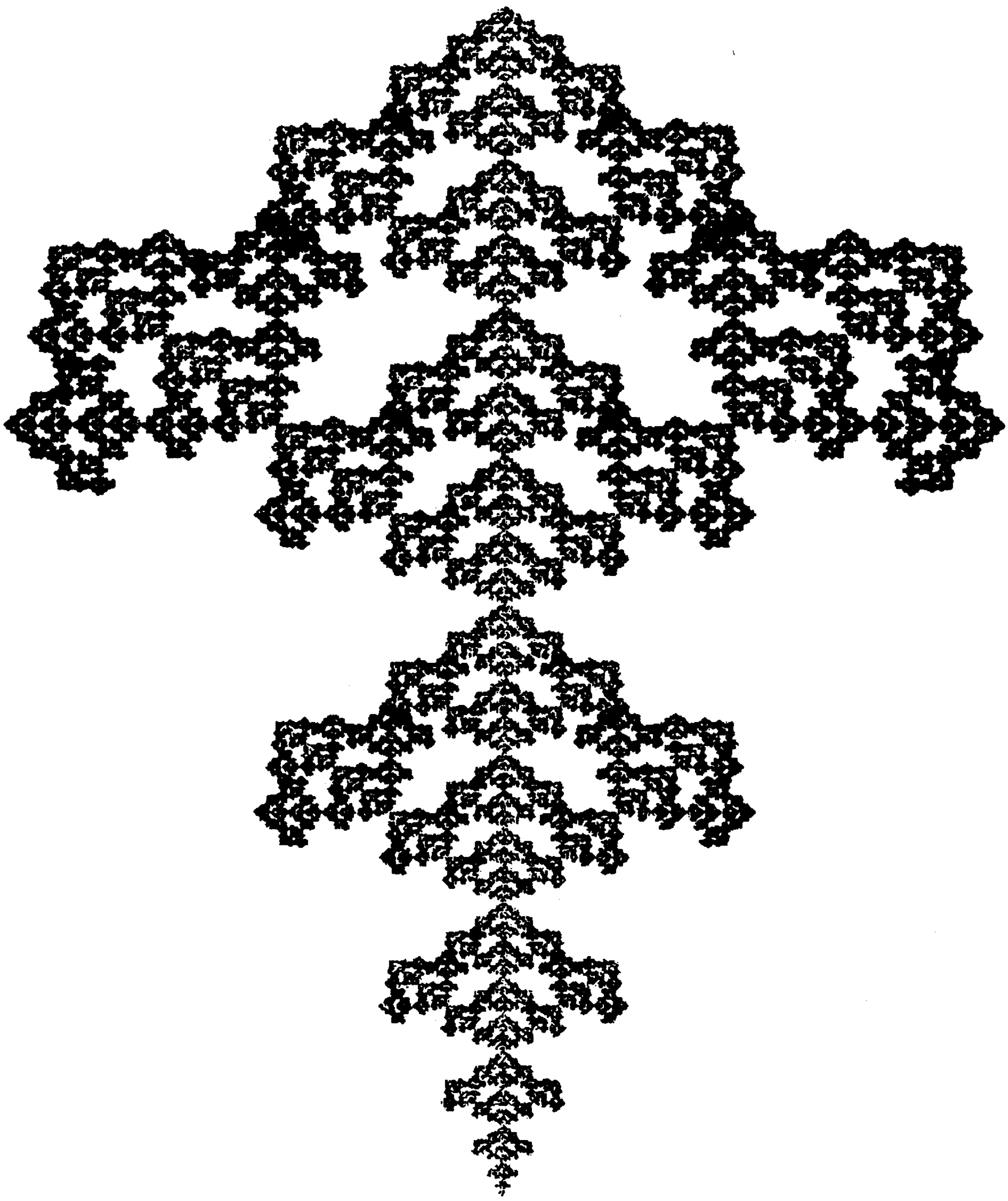


Figure III.86. Use the Collage Theorem to analyze this fractal. On how many different scales is the whole image apparently repeated here? How many times is the smallest clearly discernible copy repeated?

us a δ to constrain p with in order that $w(p, x)$ moves by less than ϵ, this relation is still dependent on the point (p, x). A set $B \in \mathcal{H}(X)$, which is interesting, contains an infinite number of such points, giving us no δ greater than 0 to constrain p with to limit the change in the whole set. We can get such a condition by further restricting $w(p, x)$. Many constraints will do this; we pick one that is simple to understand. For our IFS, parametrized by $p \in P$, that is $\{X: w_{1_p}, \ldots, w_{N_p}\}$, we want the conditions under which given $\epsilon > 0$, we can find a $\delta > 0$ such that

$$d_p(p, q) < \delta \Rightarrow h(w_p(B), w_q(B)) < \epsilon.$$

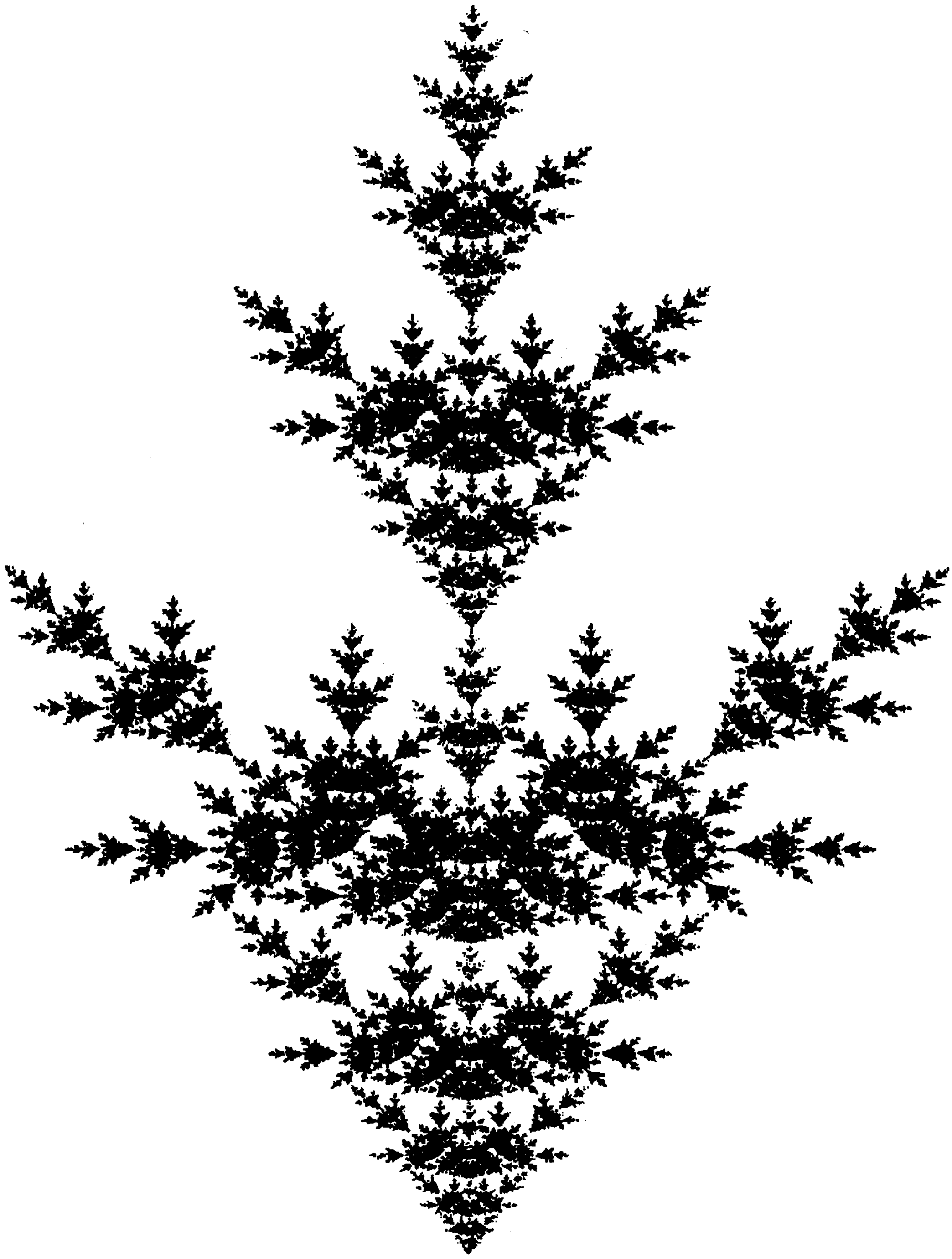


Figure III.87. Consider the white areas in this figure to represent a set S in $\mathbb{R}^2$. Locate the boundary of the largest pathwise-connected subset of S. It is recommended that you work with a photocopy of the image, a magnifying glass, and a fine red felt-tip pen.

Suppose that for every $p \in P$, $w_{i_p}(x)$ is a continuous function on X. Furthermore, we ask that there is a $k > 0$, independent of x and p such that for each fixed $x \in X$ and for each w_{i_p}, the condition

$$d(w_{i_p}(x), w_{i_q}(x)) \leq k \cdot d_p(p, q)$$

holds. This condition is called *Lipshitz* continuity. It is not the most general condition to prove what we need; we really only need some continuous function of $d(p, q)$ which is independent of x on the right-hand side. We choose Lipshitz continuity here because for the maps we are interested in, it is the easiest condition to check. If we can show that for any set $B \in \mathcal{H}(X)$ we have

$$h(w_{i_p}(B), w_{i_q}(B)) \leq k \cdot d_p(p, q),$$

then we can easily get the condition we want from the Collage Theorem. Proving this is simply a matter of writing down the definitions for the metric h.

$$h(w_p(B), w_q(B)) = d(w_p(B), w_q(B)) \vee d(w_q(B), w_p(B)),$$

where

$$d(w_p(B), w_q(B)) = \max_{x \in w_p(B)} (d(x, w_q(B)))$$
$$d(x, w_q(B)) = \min_{y \in w_q(B)} (d(x, y)).$$

Now, $x \in w_p(B)$ implies that there is an $\tilde{x} \in B$ such that $x = w_p(\tilde{x})$. Then there is a point $w_q(\tilde{x}) \in w_q(B)$, which is the image of $\tilde{x}$ under w_q. For this point, our condition holds, and

$$d(x, w_q(\tilde{x})) \leq k \cdot d_p(p, q) \Rightarrow \min_{y \in w_q(B)} (d(x, y)) \leq d(x, w_q(\tilde{x})) \leq k \cdot d_p(p, q)$$

Since this condition holds, for every $x \in w_p(B)$ the maximum over these points is at most $k \cdot d_p(p, q)$, and we have

$$d(w_p(B), w_q(B)) \leq k \cdot d_p(p, q).$$

The argument is nearly identical for $d(w_q(B), w_p(B))$, so we have

$$h(w_p(B), w_q(B)) \leq k \cdot d_p(p, q),$$

and a small change in the parameter on a particular map produces a small change in the image of any set $B \in \mathcal{H}(X)$. For a finite set of maps,$w_{1_p}, \ldots, w_{N_p}$, and their corresponding constants $k_1, \ldots, k_N$, it is then certainly the case that if $k = max_{i=1,\ldots,N}(k_i)$, we have

$$h(w_{i_p}(B), w_{i_q}(B)) \leq k \cdot d_p(p, q).$$

Now the union of such image sets cannot vary from parameter to parameter by more than the maximum Hausdorff distance above, consequently,

$$h(W_p(B), W_q(B)) \leq k \cdot d_p(p, q).$$

We now apply the results of Lemma 11.2 to the complete metric space $\mathcal{H}(X)$, yielding

$$h(A_p, A_q) \leq (1 - s)^{-1} h(A_p, W_q(A_p)) \leq (1 - s)^{-1} k \cdot d_p(p, q).$$

Theorem 11.1 *Let (X, d) be a complete metric space. Let $\{X; w_1, \ldots, w_N\}$ be a hyperbolic IFS with contractivity s. For $n = 1, 2, \ldots, N$, let w_n depend on the parameter $p \in (P, d_p)$ subject to the condition $d(w_{n_p}(x), w_{n_q}(x)) \leq k \cdot d_p(p, q)$ for all $x \in X$ with k independent of n, p, or x. Then the attractor $A(p) \in \mathcal{H}(X)$ depends continuously on the parameter $p \in P$ with respect to the Hausdorff metric $h(d)$.*

In other words, small changes in the parameters will lead to small changes in the attractor, provided that the system remains hyperbolic. This is very important because it tells us that we can continuously control the attractor of an IFS by adjusting parameters in the transformations, as is done in image compression applications. It also means we can smoothly interpolate between attractors: this is useful for image animation, for example.

Examples & Exercises

11.3. Construct a one-parameter family of IFS, of the form $\{\mathbb{R}^2; w_1, w_2, w_3\}$, where each w_i is affine and the parameter p lies in the interval $[0, 24]$. The attractor should tell the time, as illustrated in Figure III.88. $A(p)$ denotes the attractor at time p.

11.4. Imagine a slightly more complicated clockface, generated by using a one-parameter family of IFS of the form $\{\mathbb{R}^2; w_0, w_1, w_2, w_3\}$, $p \in [0, 24]$. w_0 creates the clockface, w_1 and w_2 are as in Exercise 11.3, and w_3 is a similitude that places a copy of the clockface at the end of the hour hand, as illustrated in Figure III.89. Then as p goes from 0 to 12 the hour hand sweeps through 360°, the hour hand on the smaller clockface sweeps through 720°, and the hour hand on the yet smaller clockface sweeps through 1080°, and so on. Thus as p advances, there exist lines on the attractor which are rotating at arbitrarily great speeds. Nonetheless we have continuous dependence of the image on p in the Hausdorff metric! At what times do all of the hour hands point in the same direction?

11.5. Find a one-parameter family of IFS in $\mathbb{R}^2$, whose attractors include the three trees in Figure III.90.

11.6. Run your version of Program 1 or Program 2, making small changes in the IFS code. Convince yourself that resulting rendered images "vary continuously" with respect to these changes.

11.7. Solve the following problems with regard to the images (a)–(f) in Figure III.91. Recall that a "just-touching" collage in $\mathbb{R}^2$ is one where the transforms of the target set do not overlap. They fit together like the pieces of a jigsaw puzzle.

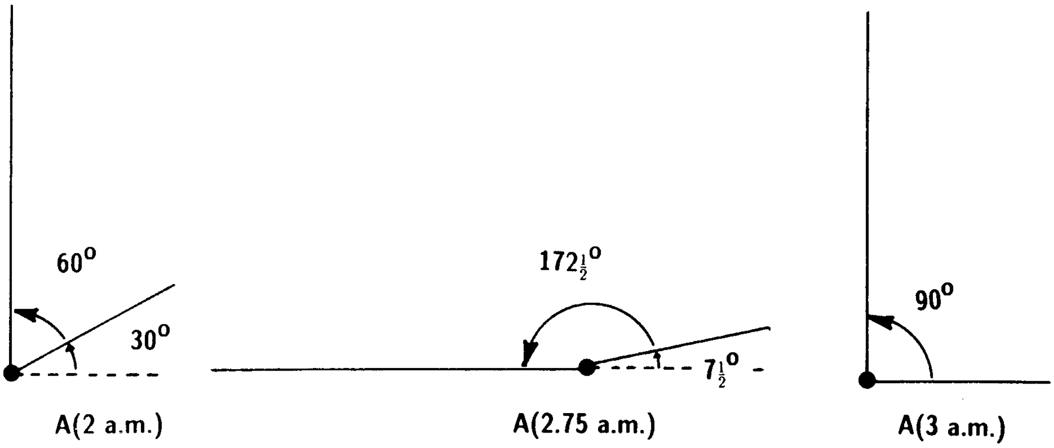


Figure III.88. A one-parameter family of IFS that tells the time!

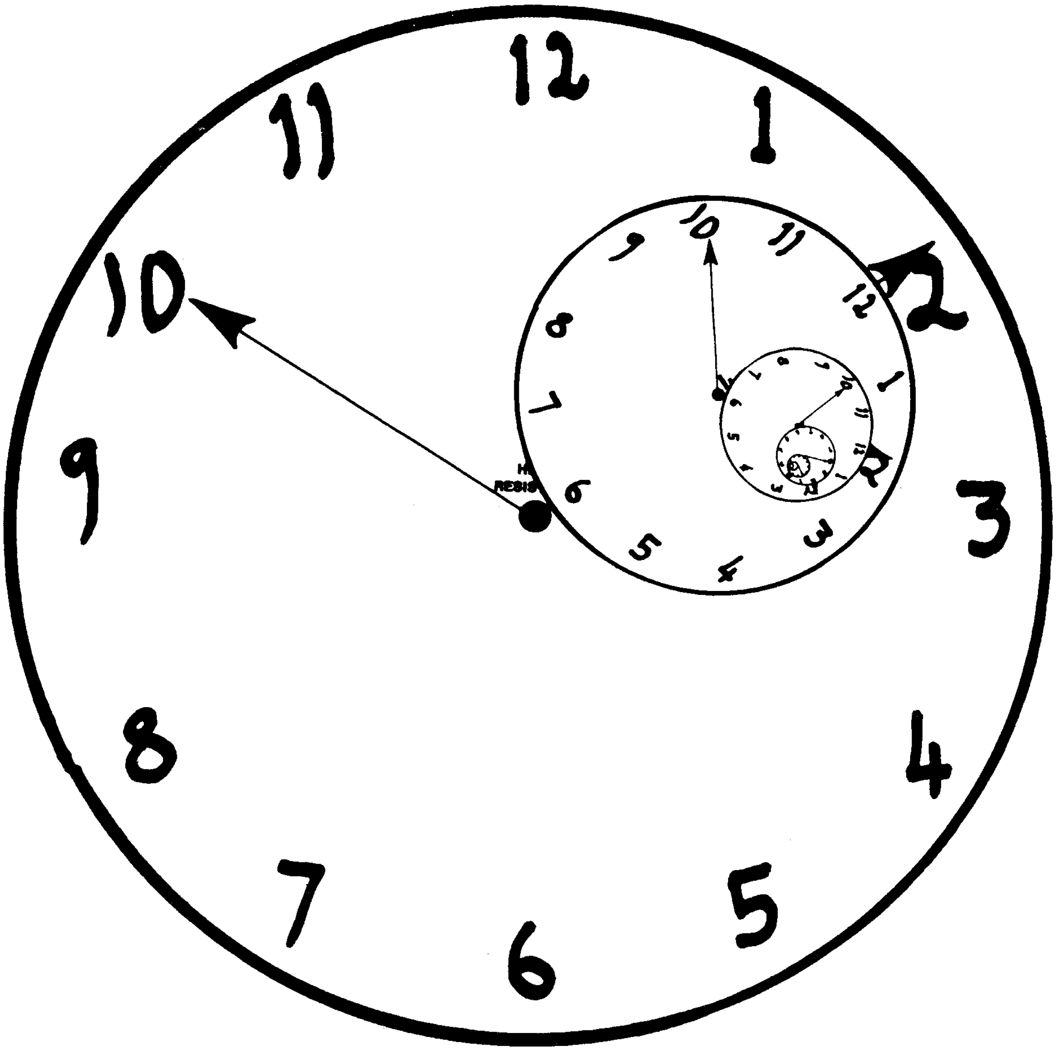


Figure III.89. This fractal clockface depends continuously on time in the Hausdorff metric.

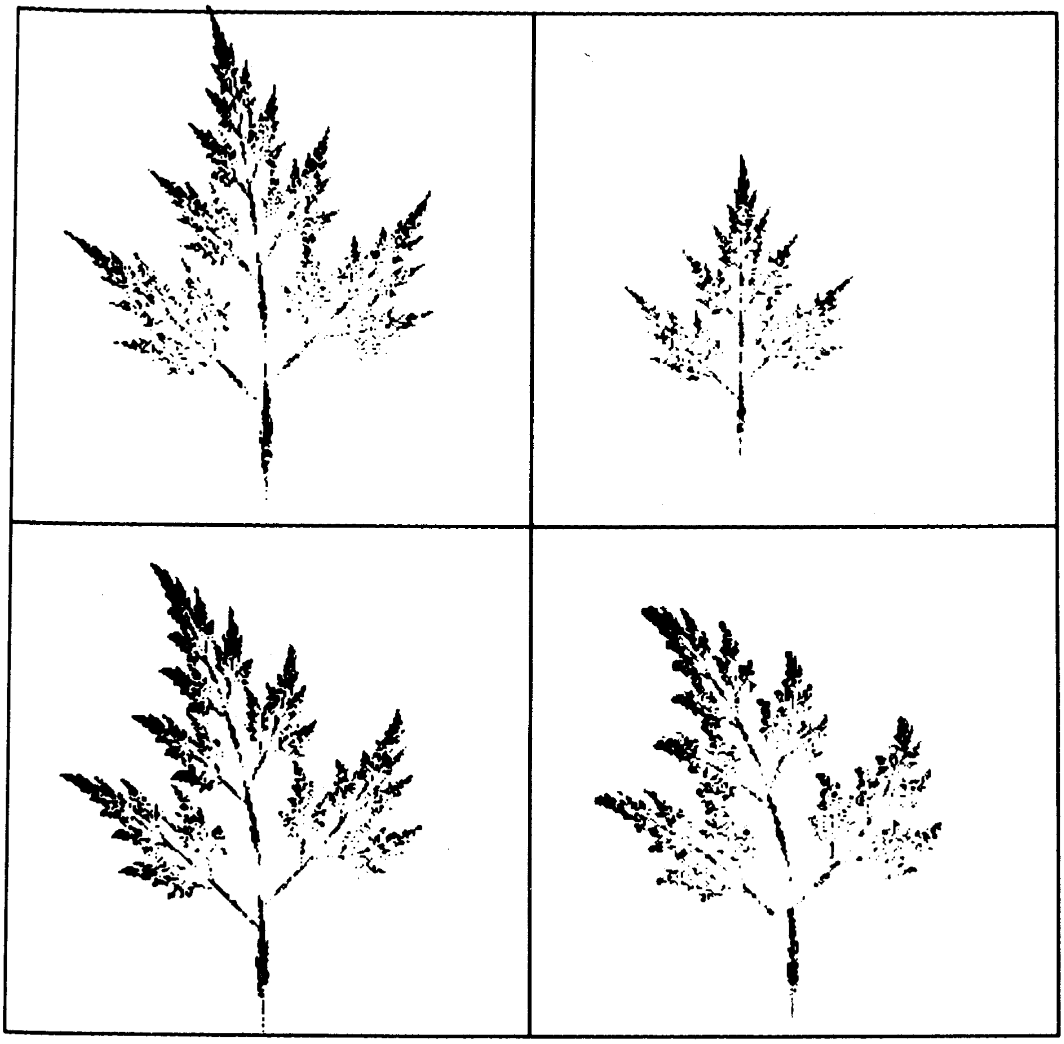


Figure III.90. Blowing in the wind. Find a one-parameter family of IFS whose attractors include the trees shown here. The Random Iteration Algorithm was used to compute these images.

(a) Find a one-parameter family collage of affine transformations.
(b) Find a "just-touching" collage of affine transformations.
(c) Find a collage using similitudes only. What is the smallest number of affine transformations in $\mathbb{R}^2$, such that the boundary is the attractor?
(d) Find a one-parameter family collage of affine transformations.
(e) Find a "just-touching" collage, using similitudes only, parameterized by the real number p.
(f) Find a collage for circles and disks.

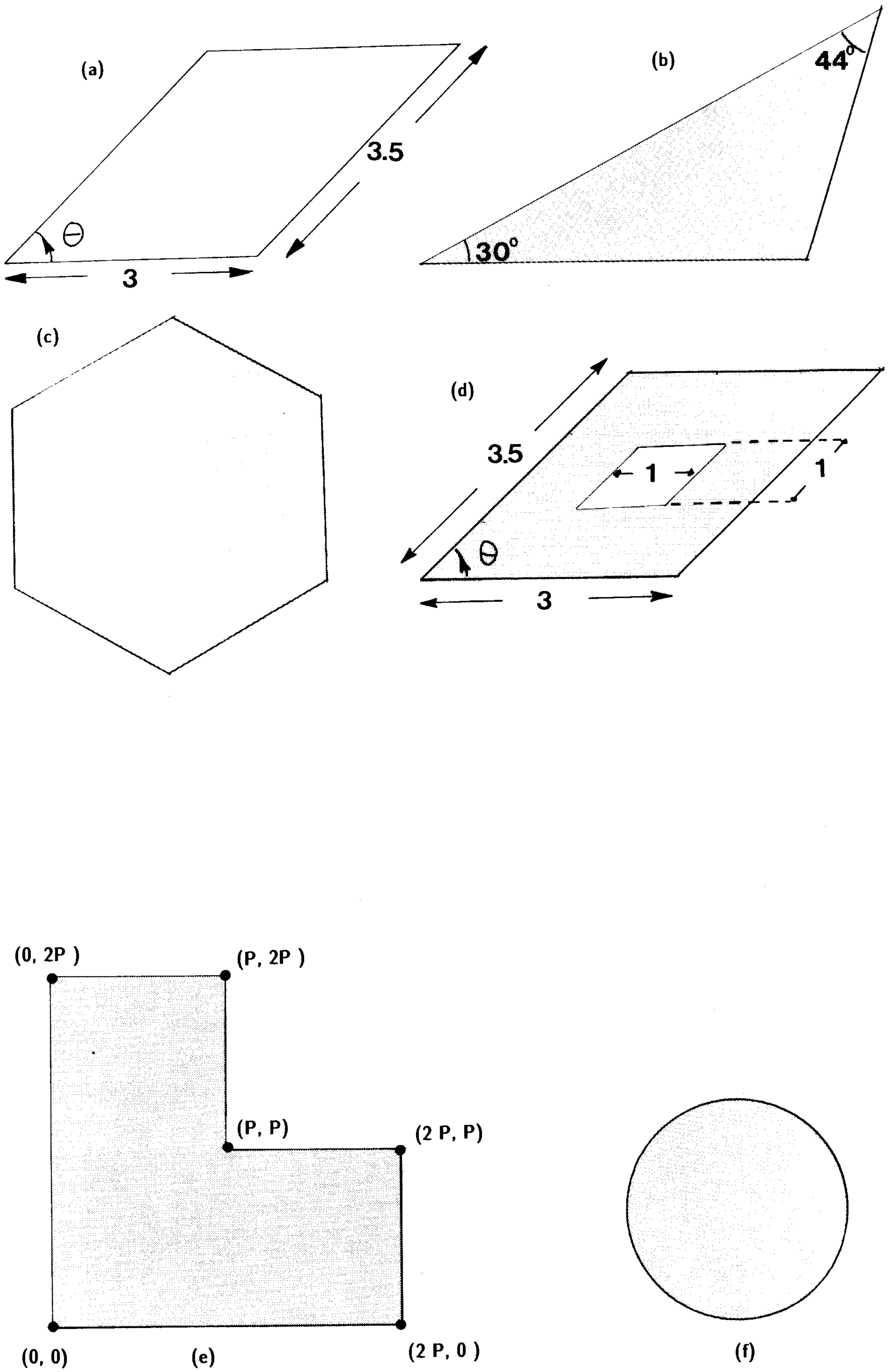


Figure III.91. Classical collages. Can you find an IFS corresponding to each of these classical geometrical objects?