Chapter 9 Iterated function systems—self-similar and self-affine sets

9.1 Iterated function systems

Many fractals are made up of parts that are, in some way, similar to the whole. For example, the middle third Cantor set is the union of two similar copies of itself, and the von Koch curve is made up of four similar copies. These self-similarities are not only properties of the fractals: they may actually be used to define them. Iterated function systems do this in a unified way and, moreover, often lead to a simple way of finding dimensions.

Let D be a closed subset of $\mathbb{R}^n$, often $D = \mathbb{R}^n$. A mapping $S : D \to D$ is called a *contraction* on D if there is a number c with $0 < c < 1$ such that $|S(x) - S(y)| \leqslant c|x - y|$ for all $x, y \in D$. Clearly any contraction is continuous. If equality holds, i.e. if $|S(x) - S(y)| = c|x - y|$, then S transforms sets into geometrically similar sets, and we call S a *contracting similarity*.

A finite family of contractions $\{S_1, S_2, \dots, S_m\}$, with $m \geqslant 2$, is called an *iterated function system* or IFS. We call a non-empty compact subset F of D an *attractor* (or *invariant set*) for the IFS if

$$F = \bigcup_{i=1}^{m} S_i(F).$$

The fundamental property of an iterated function system is that it determines a unique attractor, which is usually a fractal. For a simple example, take F to be the middle third Cantor set. Let $S_1, S_2 : \mathbb{R} \to \mathbb{R}$ be given by

$$S_1(x) = \tfrac{1}{3}x; \quad S_2(x) = \tfrac{1}{3}x + \tfrac{2}{3}.$$

Then $S_1(F)$ and $S_2(F)$ are just the left and right 'halves' of F, so that $F = S_1(F) \cup S_2(F)$; thus F is an attractor of the IFS consisting of the contractions $\{S_1, S_2\}$, the two mappings, which represent the basic self-similarities of the Cantor set.

Fractal Geometry: Mathematical Foundations and Application. Second Edition Kenneth Falconer
 ISBNs: 0-470-84861-8 (HB); 0-470-84862-6 (PB)

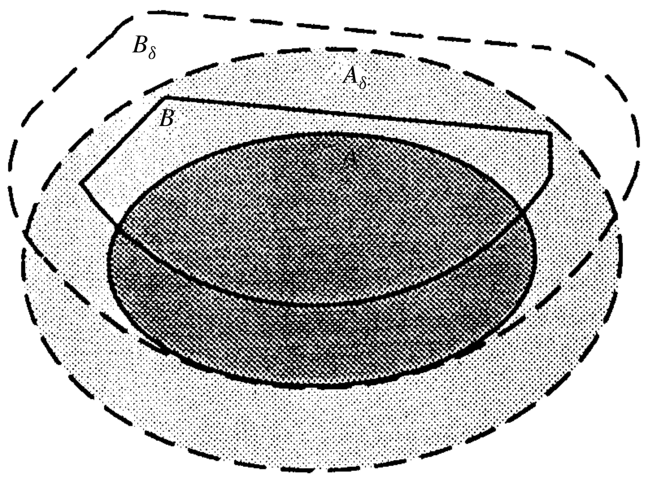


Figure 9.1 The Hausdorff distance between the sets A and B is the least $\delta > 0$ such that the δ-neighbourhood A_δ of A contains B and the δ-neighbourhood B_δ of B contains A

We shall prove the fundamental property that an IFS has a unique (non-empty compact, i.e. closed and bounded) attractor. This means, for example, that the middle third Cantor set is completely specified as the attractor of the mappings $\{S_1, S_2\}$ given above.

To this end, we define a metric or distance d between subsets of D. Let $\mathcal{S}$ denote the class of all non-empty compact subsets of D. Recall that the *δ-neighbourhood* of a set A is the set of points within distance δ of A, i.e. $A_\delta = \{x \in D : |x - a| \leqslant \delta \text{ for some } a \in A\}$. We make $\mathcal{S}$ into a metric space by defining the distance between two sets A and B to be the least δ such that the δ-neighbourhood of A contains B and vice versa:

$$d(A, B) = \inf\{\delta : A \subset B_\delta \text{ and } B \subset A_\delta\}$$

(see figure 9.1). A simple check shows that d is a metric or distance function, that is, satisfies the three requirements (i) $d(A, B) \geqslant 0$ with equality if and only if $A = B$, (ii) $d(A, B) = d(B, A)$, (iii) $d(A, B) \leqslant d(A, C) + d(C, B)$ for all $A, B, C \in \mathcal{S}$. The metric d is known as the *Hausdorff metric* on $\mathcal{S}$. In particular, if $d(A, B)$ is small, then A and B are close to each other as sets.

We give two proofs of the fundamental result on IFSs. The first depends on Banach's contraction mapping theorem, and the second is direct and elementary.

Theorem 9.1

Consider the iterated function system given by the contractions $\{S_1, \ldots, S_m\}$ on $D \subset \mathbb{R}^n$, so that

$$|S_i(x) - S_i(y)| \leqslant c_i|x - y| \quad (x, y) \in D \tag{9.1}$$

with $c_i < 1$ for each i. Then there is a unique attractor F, i.e. a non-empty compact set such that

$$F = \bigcup_{i=1}^{m} S_i(F). \tag{9.2}$$

Moreover, if we define a transformation S on the class $\mathcal{S}$ of non-empty compact sets by

$$S(E) = \bigcup_{i=1}^{m} S_i(E) \tag{9.3}$$

for $E \in \mathcal{S}$, and write S^k for the kth iterate of S (so $S^0(E) = E$ and $S^k(E) = S(S^{k-1}(E))$ for $k \geqslant 1$), then

$$F = \bigcap_{k=0}^{\infty} S^k(E) \tag{9.4}$$

for every set $E \in \mathcal{S}$ such that $S_i(E) \subset E$ for all i.

First proof. Note that sets in $\mathcal{S}$ are transformed by S into other sets of $\mathcal{S}$. If $A, B \in \mathcal{S}$ then

$$d(S(A), S(B)) = d\left(\bigcup_{i=1}^{m} S_i(A), \bigcup_{i=1}^{m} S_i(B)\right) \leqslant \max_{1 \leqslant i \leqslant m} d(S_i(A), S_i(B))$$

using the definition of the metric d and noting that if the δ-neighbourhood $(S_i(A))_\delta$ contains $S_i(B)$ for all i then $(\bigcup_{i=1}^{m} S_i(A))_\delta$ contains $\bigcup_{i=1}^{m} S_i(B)$, and vice versa. By (9.1)

$$d(S(A), S(B)) \leqslant (\max_{1 \leqslant i \leqslant m} c_i) d(A, B). \tag{9.5}$$

It may be shown that d is a *complete* metric on $\mathcal{S}$, that is every Cauchy sequence of sets in $\mathcal{S}$ is convergent to a set in $\mathcal{S}$. Since $0 < \max_{1 \leqslant i \leqslant m} c_i < 1$, (9.5) states that S is a contraction on the complete metric space $(\mathcal{S}, d)$. By Banach's contraction mapping theorem, S has a unique fixed point, that is there is a unique set $F \in \mathcal{S}$ such that $S(F) = F$, which is (9.2), and moreover $S^k(E) \to F$ as $k \to \infty$. In particular, if $S_i(E) \subset E$ for all i then $S(E) \subset E$, so that $S^k(E)$ is a decreasing sequence of non-empty compact sets containing F with intersection $\bigcap_{k=0}^{\infty} S^k(E)$ which must equal F. $\square$

Second proof. Let E be any set in $\mathcal{S}$ such that $S_i(E) \subset E$ for all i; for example $E = D \cap B(0, r)$ will do provided r is sufficiently large. Then $S^k(E) \subset S^{k-1}(E)$, so that $S^k(E)$ is a decreasing sequence of non-empty compact sets, which necessarily have non-empty compact intersection $F = \bigcap_{k=1}^{\infty} S^k(E)$. Since $S^k(E)$ is a decreasing sequence of sets, it follows that $S(F) = F$, so F satisfies (9.2) and is an attractor of the IFS.

To see that the attractor is unique, we derive (9.5) exactly as in the first proof. Suppose A and B are both attractors, so that $S(A) = A$ and $S(B) = B$. Since $0 < \max_{1 \leqslant i \leqslant m} c_i < 1$ it follows from (9.5) that $d(A, B) = 0$, implying $A = B$. $\square$

There are two main problems that arise in connection with iterated function systems. The first problem is to represent or 'code' a given set as the attractor of some IFS, and the second is to 'decode' the IFS by displaying its attractor. In both cases, we may wish to go on to analyse the structure and dimensions of the attractor, and the IFS can be a great aid in doing this.

Finding an IFS that has a given F as its unique attractor can often be done by inspection, at least if F is self-similar or self-affine. For example, the Cantor dust (figure 0.4) is easily seen to be the attractor of the four similarities which give the basic self-similarities of the set:

$$S_1(x, y) = (\tfrac{1}{4}x, \tfrac{1}{4}y + \tfrac{1}{2}), \quad S_2(x, y) = (\tfrac{1}{4}x + \tfrac{1}{4}, \tfrac{1}{4}y),$$
$$S_3(x, y) = (\tfrac{1}{4}x + \tfrac{1}{2}, \tfrac{1}{4}y + \tfrac{3}{4}), \quad S_4(x, y) = (\tfrac{1}{4}x + \tfrac{3}{4}, \tfrac{1}{4}y + \tfrac{1}{4}).$$

In general it may not be possible to find an IFS with a given set as attractor, but we can normally find one with an attractor that is a close approximation to the required set. This question of representing general objects by IFSs is considered in Section 9.5.

The transformation S introduced in Theorem 9.1 is the key to computing the attractor of an IFS; indeed (9.4) already provides a method for doing so. In fact, the sequence of iterates $S^k(E)$ converges to the attractor F for any initial set E in $\mathcal{S}$, in the sense that $d(S^k(E), F) \to 0$. This follows since (9.5) implies that $d(S(E), F) = d(S(E), S(F)) \leqslant cd(E, F)$, so that $d(S^k(E), F) \leqslant c^k d(E, F)$, where $c = \max_{1\leqslant i\leqslant m} c_i < 1$. Thus the $S^k(E)$ provide increasingly good approximations to F. If F is a fractal, these approximations are sometimes called *pre-fractals* for F.

For each k

$$S^k(E) = \bigcup_{\mathcal{I}_k} S_{i_1}\circ\cdots\circ S_{i_k}(E) = \bigcup_{\mathcal{I}_k} S_{i_1}(S_{i_2}(\cdots(S_{i_k}(E))\cdots)) \tag{9.6}$$

where the union is over the set $\mathcal{I}_k$ of all k-term sequences $(i_1, \dots, i_k)$ with $1 \leqslant i_j \leqslant m$; see figure 9.2. (Recall that $S_{i_1}\circ\cdots\circ S_{i_k}$ denotes the composition of mappings, so that $(S_{i_1}\circ\cdots\circ S_{i_k})(x) = S_{i_1}(S_{i_2}(\cdots(S_{i_k}(x))\cdots))$.) If $S_i(E)$ is contained in E for each i and x is a point of F, it follows from (9.4) that there is a (not necessarily unique) sequence $(i_1, i_2, \dots)$ such that $x \in S_{i_1}\circ\cdots\circ S_{i_k}(E)$ for all k. This sequence provides a natural coding for x, with

$$x = x_{i_1,i_2,\dots} = \bigcap_{k=1}^{\infty} S_{i_1}\circ\cdots\circ S_{i_k}(E), \tag{9.7}$$

so that $F = \bigcup\{x_{i_1,i_2,\dots}\}$.

This expression for $x_{i_1,i_2,\dots}$ is independent of E provided that $S_i(E)$ is contained in E for all i.

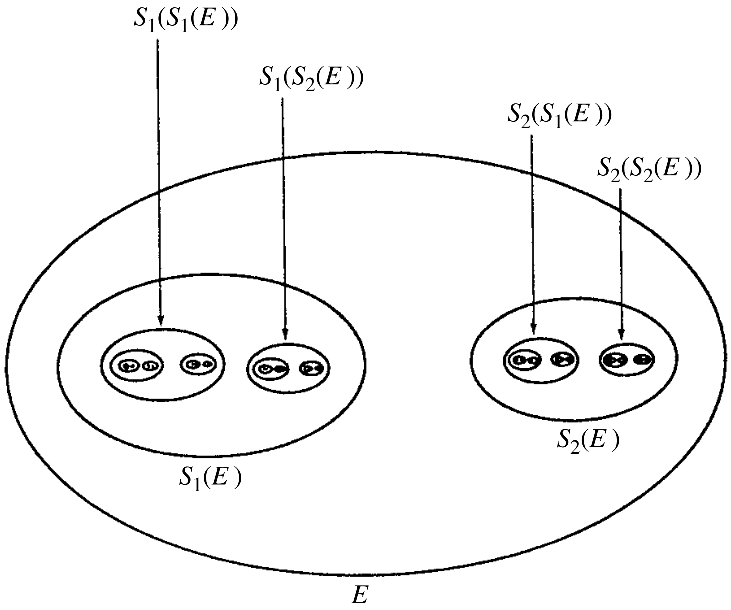


Figure 9.2 Construction of the attractor F for contractions S_1 and S_2 which map the large ellipse E onto the ellipses $S_1(E)$ and $S_2(E)$. The sets $S^k(E) = \bigcup_{i_i=1,2} S_{i_1} \circ \cdots \circ S_{i_k}(E)$ give increasingly good approximations to F

Notice that if the union in (9.2) is disjoint then F must be totally disconnected (provided the S_i are injections), since if $x_{i_1,i_2,\ldots} \neq x_{i'_1,i'_2,\ldots}$ we may find k such that $(i_1, \ldots, i_k) \neq (i'_1, \ldots, i'_k)$ so that the disjoint closed sets $S_{i_1} \circ \cdots \circ S_{i_k}(F)$ and $S_{i'_1} \circ \cdots \circ S_{i'_k}(F)$ disconnect the two points.

Again this may be illustrated by $S_1(x) = \frac{1}{3}x$, $S_2(x) = \frac{1}{3}x + \frac{2}{3}$ and F the Cantor set. If $E = [0, 1]$ then $S^k(E) = E_k$, the set of 2^k basic intervals of length 3^{-k} obtained at the kth stage of the usual Cantor set construction; see figure 0.1. Moreover, $x_{i_1,i_2,\ldots}$ is the point of the Cantor set with base-3 expansion $0 \cdot a_1 a_2 \ldots$, where $a_k = 0$ if $i_k = 1$ and $a_k = 2$ if $i_k = 2$. The pre-fractals $S^k(E)$ provide the usual construction of many fractals for a suitably chosen initial set E; the $S_{i_1} \circ \cdots \circ S_{i_k}(E)$ are called the *level-k sets* of the construction.

This theory provides us with two methods for computer drawing of IFS attractors in the plane, as indicated in figure 9.3. For the first method, take any initial set E (such as a square) and draw the kth approximation $S^k(E)$ to F given by (9.6) for a suitable value of k. The set $S^k(E)$ is made up of m^k small sets—either these can be drawn in full, or a representative point of each can be plotted. If E can be chosen as a line segment in such a way that $S_1(E), \ldots, S_m(E)$ join up to form a polygonal curve with endpoints the same as those of E, then the sequence of polygonal curves $S^k(E)$ provides increasingly good approximations to the fractal curve F. Taking E as the initial interval in the von Koch curve construction is an example of this, with $S^k(E)$ just the kth step of the construction (E_k in figure 0.2). Careful recursive programming is helpful when using this method.

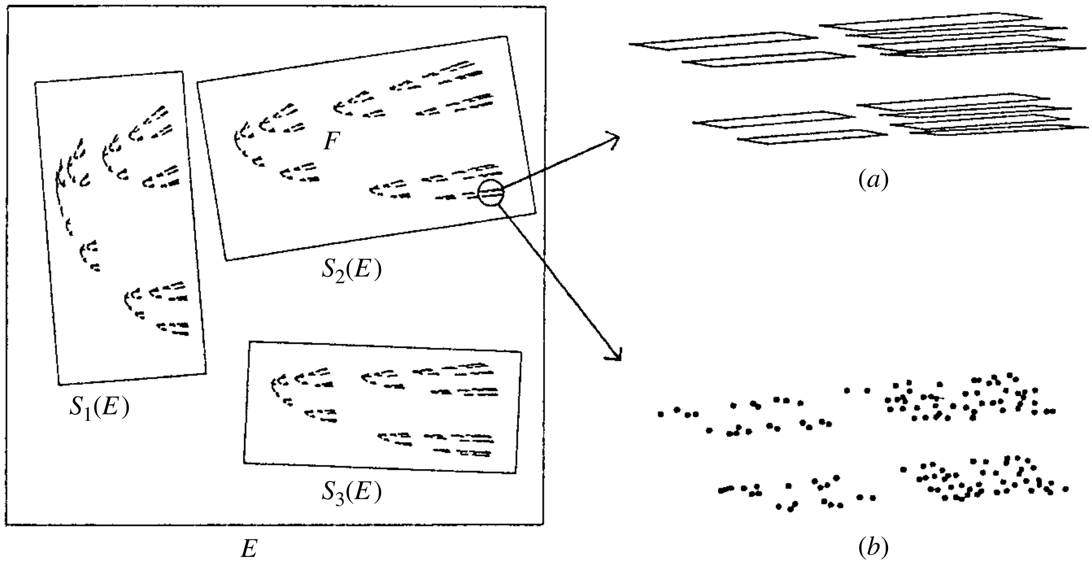


Figure 9.3 Two ways of computer drawing the attractor F of the IFS consisting of the three affine transformations S_1, S_2 and S_3 which map the square onto the rectangles. In method (a) the 3^k parallelograms $S_{i_1}(S_{i_2}(\cdots(S_{i_k}(E))\cdots))$ for $i_j = 1, 2, 3$ are drawn ($k = 6$ here). In method (b) the sequence of points x_k is plotted by choosing S_{i_k} at random from S_1, S_2 and S_3 for successive k and letting $x_k = S_{i_k}(x_{k-1})$

For the second method, take x_0 as any initial point, select a contraction S_{i_1} from $S_1, \ldots, S_m$ at random, and let $x_1 = S_{i_1}(x_0)$. Continue in this way, choosing S_{i_k} from $S_1, \ldots, S_m$ at random (with equal probability, say) and letting $x_k = S_{i_k}(x_{k-1})$ for $k = 1, 2, \ldots$. For large enough k, the points x_k will be indistinguishably close to F, with x_k close to $S_{i_k}\circ\cdots\circ S_{i_1}(F)$, so the sequence $\{x_k\}$ will appear randomly distributed across F. A plot of the sequence $\{x_k\}$ from, say, the hundredth term onwards may give a good impression of F. (It is a consequence of ergodic theory that, with probability 1, this sequence of points will fill F, in a manner that approximates a certain measure on F.)

9.2 Dimensions of self-similar sets

One of the advantages of using an iterated function system is that the dimension of the attractor is often relatively easy to calculate or estimate in terms of the defining contractions. In this section we discuss the case where $S_1, \ldots, S_m : \mathbb{R}^n \to \mathbb{R}^n$ are *similarities*, i.e. with

$$|S_i(x) - S_i(y)| = c_i|x - y| \quad (x, y \in \mathbb{R}^n) \tag{9.8}$$

where $0 < c_i < 1$ (c_i is called the *ratio* of S_i). Thus each S_i transforms subsets of $\mathbb{R}^n$ into geometrically similar sets. The attractor of such a collection of similarities is called a (*strictly*) *self-similar set*, being a union of a number of smaller similar copies of itself. Standard examples include the middle third Cantor set, the Sierpiński triangle and the von Koch curve, see figures 0.1–0.5. We show that, under certain conditions, a self-similar set F has Hausdorff and box dimensions equal to the value of s satisfying

$$\sum_{i=1}^{m} c_i^s = 1 \tag{9.9}$$

and further that F has positive and finite $\mathcal{H}^s$-measure. A calculation similar to the 'heuristic calculation' of Example 2.7 indicates that the value given by (9.9) is at least plausible. If $F = \bigcup_{i=1}^{m} S_i(F)$ with the union 'nearly disjoint', we have that

$$\mathcal{H}^s(F) = \sum_{i=1}^{m} \mathcal{H}^s(S_i(F)) = \sum_{i=1}^{m} c_i^s \mathcal{H}^s(F) \tag{9.10}$$

using (9.8) and Scaling property 2.1. On the assumption that $0 < \mathcal{H}^s(F) < \infty$ at the 'jump' value $s = \dim_{\mathrm{H}} F$, we get that s satisfies (9.9).

For this argument to give the right answer, we require a condition that ensures that the components $S_i(F)$ of F do no overlap 'too much'. We say that the S_i satisfy the *open set condition* if there exists a non-empty bounded open set V such that

$$V \supset \bigcup_{i=1}^{m} S_i(V) \tag{9.11}$$

with the union disjoint. (In the middle third Cantor set example, the open set condition holds for S_1 and S_2 with V as the open interval (0, 1).) We show that, provided that the similarities S_i satisfy the open set condition, the Hausdorff dimension of the attractor is given by (9.9).

We require the following geometrical result.

Lemma 9.2

Let $\{V_i\}$ be a collection of disjoint open subsets of $\mathbb{R}^n$ such that each V_i contains a ball of radius $a_1 r$ and is contained in a ball of radius $a_2 r$. Then any ball B of radius r intersects at most $(1 + 2a_2)^n a_1^{-n}$ of the closures $\overline{V}_i$.

Proof. If $\overline{V}_i$ meets B, then $\overline{V}_i$ is contained in the ball concentric with B of radius $(1 + 2a_2)r$. Suppose that q of the sets $\overline{V}_i$ intersect B. Then, summing the volumes of the corresponding interior balls of radii $a_1 r$, it follows that $q(a_1 r)^n \leqslant (1 + 2a_2)^n r^n$, giving the stated bound for q. □

The derivation of the lower bound in the following theorem is a little awkward. The reader may find it helpful to follow through the proof with the middle third Cantor set in mind, or by referring to the 'general example' of figure 9.2. Alternatively, the proof of Proposition 9.7 covers the case when the sets $S_1(F), \dots, S_m(F)$ are disjoint, and is rather simpler.

Theorem 9.3

Suppose that the open set condition (9.11) *holds for the similarities S_i on $\mathbb{R}^n$ with ratios $0 < c_i < 1$ for $1 \leqslant i \leqslant m$. If F is the attractor of the IFS $\{S_1, \dots, S_m\}$, that is*

$$F = \bigcup_{i=1}^{m} S_i(F), \tag{9.12}$$

then $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F = s$, *where s is given by*

$$\sum_{i=1}^{m} c_i^s = 1. \tag{9.13}$$

Moreover, for this value of s, $0 < \mathcal{H}^s(F) < \infty$.

Proof. Let s satisfy (9.13). Let $\mathcal{I}_k$ denote the set of all k-term sequences $(i_1, \dots, i_k)$ with $1 \leqslant i_j \leqslant m$. For any set A and $(i_1, \dots, i_k) \in \mathcal{I}_k$ we write $A_{i_1,\dots,i_k} = S_{i_1} \circ \dots \circ S_{i_k}(A)$. It follows, by using (9.12) repeatedly, that

$$F = \bigcup_{\mathcal{I}_k} F_{i_1,\dots,i_k}.$$

We check that these covers of F provide a suitable upper estimate for the Hausdorff measure. Since the mapping $S_{i_1} \circ \dots \circ S_{i_k}$ is a similarity of ratio $c_{i_1} \cdots c_{i_k}$, then

$$\sum_{\mathcal{I}_k} |F_{i_1,\dots,i_k}|^s = \sum_{\mathcal{I}_k} (c_{i_1} \cdots c_{i_k})^s |F|^s = \left(\sum_{i_1} c_{i_1}^s\right) \cdots \left(\sum_{i_k} c_{i_k}^s\right) |F|^s = |F|^s \tag{9.14}$$

by (9.13). For any $\delta > 0$, we may choose k such that $|F_{i_1,\dots,i_k}| \leqslant (\max_i c_i)^k |F| \leqslant \delta$, so $\mathcal{H}^s_\delta(F) \leqslant |F|^s$ and hence $\mathcal{H}^s(F) \leqslant |F|^s$.

The lower bound is more awkward. Let $\mathcal{I}$ be the set of all infinite sequences $\mathcal{I} = \{(i_1, i_2, \dots) : 1 \leqslant i_j \leqslant m\}$, and let $I_{i_1,\dots,i_k} = \{(i_1, \dots, i_k, q_{k+1}, \dots) : 1 \leqslant q_j \leqslant m\}$ be the 'cylinder' consisting of those sequences in $\mathcal{I}$ with initial terms $(i_1, \dots, i_k)$. We may put a mass distribution μ on $\mathcal{I}$ such that $\mu(I_{i_1,\dots,i_k}) = (c_{i_1} \cdots c_{i_k})^s$. Since $(c_{i_1} \cdots c_{i_k})^s = \sum_{i=1}^m (c_{i_1} \cdots c_{i_k} c_i)^s$, i.e. $\mu(I_{i_1,\dots,i_k}) = \sum_{i=1}^m \mu(I_{i_1,\dots,i_k,i})$, it follows that μ is indeed a mass distribution on subsets of $\mathcal{I}$ with $\mu(\mathcal{I}) = 1$. We may transfer μ to a mass distribution $\tilde{\mu}$ on F in a natural way by defining $\tilde{\mu}(A) = \mu\{(i_1, i_2, \dots) : x_{i_1,i_2,\dots} \in A\}$ for subsets A of F. (Recall that $x_{i_1,i_2,\dots} =$

$\bigcap_{k=1}^{\infty} F_{i_1,\ldots,i_k}$.) Thus the $\tilde{\mu}$-mass of a set is the μ-mass of the corresponding sequences. It is easily checked that $\tilde{\mu}(F) = 1$.

We show that $\tilde{\mu}$ satisfies the conditions of the Mass distribution principle 4.2. Let V be the open set of (9.11). Since $\overline{V} \supset S(\overline{V}) = \bigcup_{i=1}^{m} S_i(\overline{V})$, the decreasing sequence of iterates $S^k(\overline{V})$ converges to F; see (9.4). In particular $\overline{V} \supset F$ and $\overline{V}_{i_1,\ldots,i_k} \supset F_{i_1,\ldots,i_k}$ for each finite sequence $(i_1, \ldots, i_k)$. Let B be any ball of radius $r < 1$. We estimate $\tilde{\mu}(B)$ by considering the sets $V_{i_1,\ldots,i_k}$ with diameters comparable with that of B and with closures intersecting $F \cap B$.

We curtail each infinite sequence $(i_1, i_2, \ldots) \in \mathcal{I}$ after the first term i_k for which

$$\left(\min_{1 \leqslant i \leqslant m} c_i\right) r \leqslant c_{i_1} c_{i_2} \cdots c_{i_k} \leqslant r \tag{9.15}$$

and let $\mathcal{Q}$ denote the finite set of all (finite) sequences obtained in this way. Then for every infinite sequence $(i_1, i_2, \ldots) \in \mathcal{I}$ there is exactly one value of k with $(i_1, \ldots, i_k) \in \mathcal{Q}$. Since $V_1, \ldots, V_m$ are disjoint, so are $V_{i_1,\ldots,i_k,1}, \ldots, V_{i_1,\ldots,i_k,m}$ for each $(i_1, \ldots, i_k)$. Using this in a nested way, it follows that the collection of open sets $\{V_{i_1,\ldots,i_k} : (i_1, \ldots, i_k) \in \mathcal{Q}\}$ is disjoint. Similarly $F \subset \bigcup_{\mathcal{Q}} F_{i_1,\ldots,i_k} \subset \bigcup_{\mathcal{Q}} \overline{V}_{i_1,\ldots,i_k}$.

We choose a_1 and a_2 so that V contains a ball of radius a_1 and is contained in a ball of radius a_2. Then, for all $(i_1, \ldots, i_k) \in \mathcal{Q}$, the set $V_{i_1,\ldots,i_k}$ contains a ball of radius $c_{i_1} \cdots c_{i_k} a_1$ and therefore one of radius $(\min_i c_i) a_1 r$, and is contained in a ball of radius $c_{i_1} \cdots c_{i_k} a_2$ and hence in a ball of radius $a_2 r$. Let $\mathcal{Q}_1$ denote those sequences $(i_1, \ldots, i_k)$ in $\mathcal{Q}$ such that B intersects $\overline{V}_{i_1,\ldots,i_k}$. By Lemma 9.2 there are at most $q = (1 + 2a_2)^n a_1^{-n} (\min_i c_i)^{-n}$ sequences in $\mathcal{Q}_1$. Then

$$\begin{aligned}\tilde{\mu}(B) = \tilde{\mu}(F \cap B) &= \mu\{(i_1, i_2, \ldots) : x_{i_1,i_2,\ldots} \in F \cap B\}\\ &\leqslant \mu\left\{\bigcup_{\mathcal{Q}_1} I_{i_1,\ldots,i_k}\right\}\end{aligned}$$

since, if $x_{i_1,i_2,\ldots} \in F \cap B \subset \bigcup_{\mathcal{Q}_1} \overline{V}_{i_1,\ldots,i_k}$, then there is an integer k such that $(i_1, \ldots, i_k) \in \mathcal{Q}_1$. Thus

$$\begin{aligned}\tilde{\mu}(B) &\leqslant \sum_{\mathcal{Q}_1} \mu(I_{i_1,\ldots,i_k})\\ &= \sum_{\mathcal{Q}_1} (c_{i_1} \cdots c_{i_k})^s \leqslant \sum_{\mathcal{Q}_1} r^s \leqslant r^s q\end{aligned}$$

using (9.15). Since any set U is contained in a ball of radius $|U|$, we have $\tilde{\mu}(U) \leqslant |U|^s q$, so the Mass distribution principle 4.2 gives $\mathcal{H}^s(F) \geqslant q^{-1} > 0$, and $\dim_{\mathrm{H}} F = s$.

If $\mathcal{Q}$ is any set of finite sequences such that for every $(i_1, i_2, \ldots) \in \mathcal{I}$ there is exactly one integer k with $(i_1, \ldots, i_k) \in \mathcal{Q}$, it follows inductively from (9.13) that $\sum_{\mathcal{Q}} (c_{i_1} c_{i_2} \cdots c_{i_k})^s = 1$. Thus, if $\mathcal{Q}$ is chosen as in (9.15), $\mathcal{Q}$ contains at most $(\min_i c_i)^{-s} r^{-s}$ sequences. For each sequence $(i_1, \ldots, i_k) \in \mathcal{Q}$ we have $|\overline{V}_{i_1,\ldots,i_k}| = c_{i_1} \cdots c_{i_k} |\overline{V}| \leqslant r|\overline{V}|$, so F may be covered by $(\min_i c_i)^{-s} r^{-s}$ sets of diameter $r|\overline{V}|$ for each $r < 1$. It follows from Equivalent definition 3.1(iv) that $\overline{\dim}_{\mathrm{B}} F \leqslant s$; noting that $s = \dim_{\mathrm{H}} F \leqslant \underline{\dim}_{\mathrm{B}} F \leqslant \overline{\dim}_{\mathrm{B}} F \leqslant s$, using (3.17), completes the proof. $\square$

If the open set condition is not assumed in Theorem 9.3, it may be shown that we still have $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F$ though this value may be less than s.

Theorem 9.3 enables us to find the dimension of many self-similar fractals.

Example 9.4. Sierpiński triangle

The Sierpiński triangle or gasket F is constructed from an equilateral triangle by repeatedly removing inverted equilateral triangles; see figure 0.3. *Then* $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F = \log 3/\log 2$.

Calculation. The set F is the attractor of the three obvious similarities of ratios $\frac{1}{2}$ which map the triangle E_0 onto the triangles of E_1. The open set condition holds, taking V as the interior of E_0. Thus, by Theorem 9.3, $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F = \log 3/\log 2$, which is the solution of $3(\frac{1}{2})^s = \sum_1^3 (\frac{1}{2})^s = 1$. $\square$

The next example involves similarity transformations of more than one ratio.

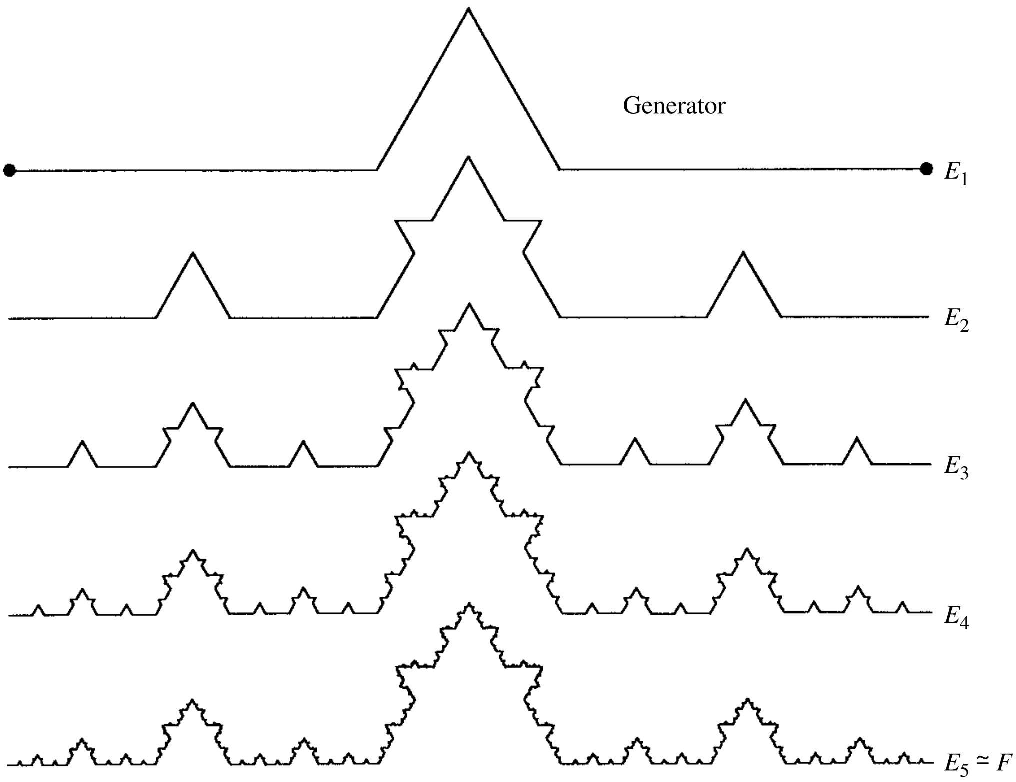


Figure 9.4 Construction of a modified von Koch curve—see Example 9.5

Example 9.5. Modified von Koch curve

Fix $0 < a \leqslant \frac{1}{3}$ *and construct a curve F by repeatedly replacing the middle proportion a of each interval by the other two sides of an equilateral triangle; see figure* 9.4. *Then* $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F$ *is the solution of* $2a^s + 2(\frac{1}{2}(1-a))^s = 1$.

Calculation. The curve F is the attractor of the similarities that map the unit interval onto each of the four intervals in E_1. The open set condition holds,

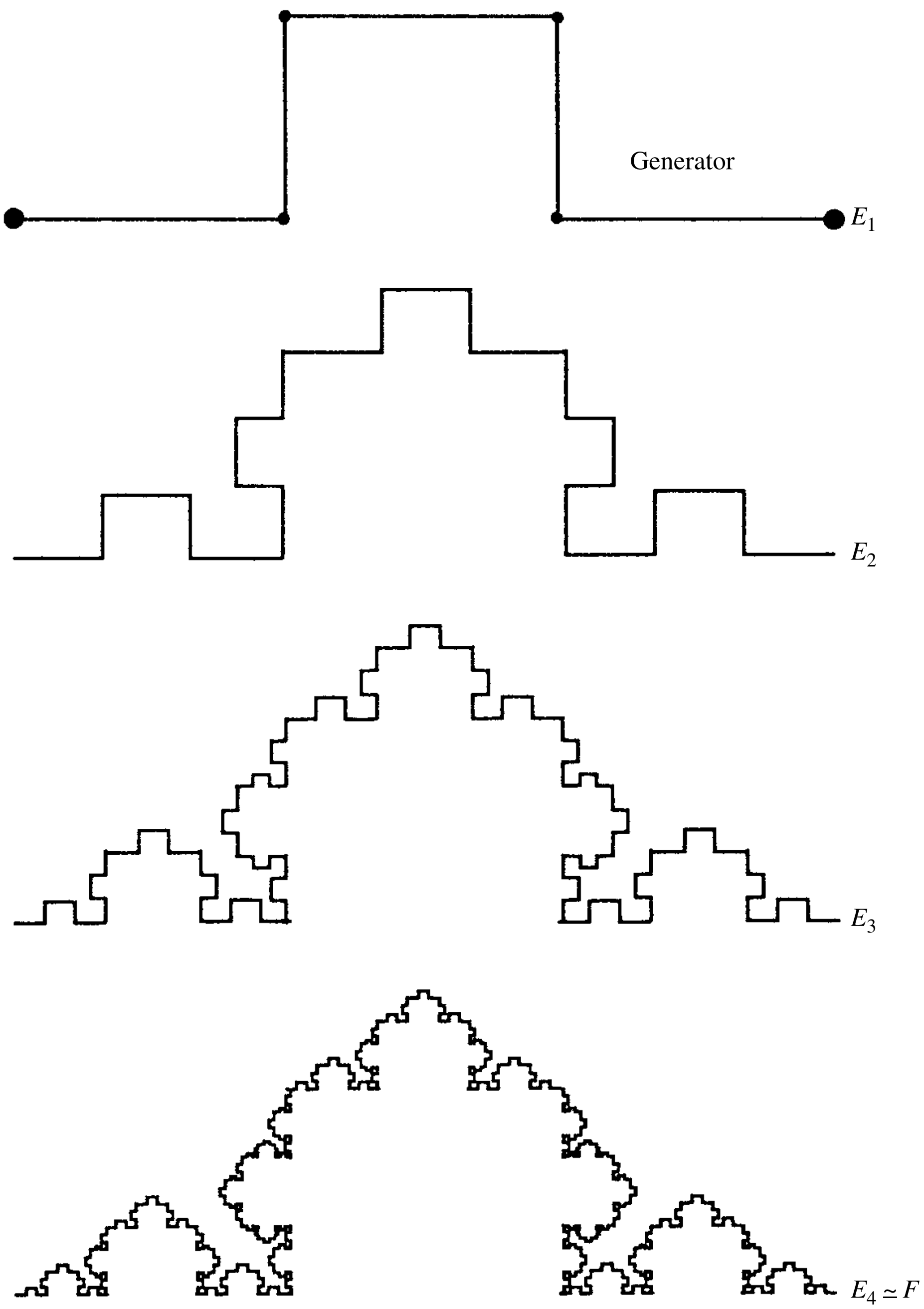


Figure 9.5 Stages in the construction of a fractal curve from a generator. The lengths of the segments in the generator are $\frac{1}{3}, \frac{1}{4}, \frac{1}{3}, \frac{1}{4}, \frac{1}{3}$, and the Hausdorff and box dimensions of F are given by $3(\frac{1}{3})^s + 2(\frac{1}{4})^s = 1$ or $s = 1.34\ldots$

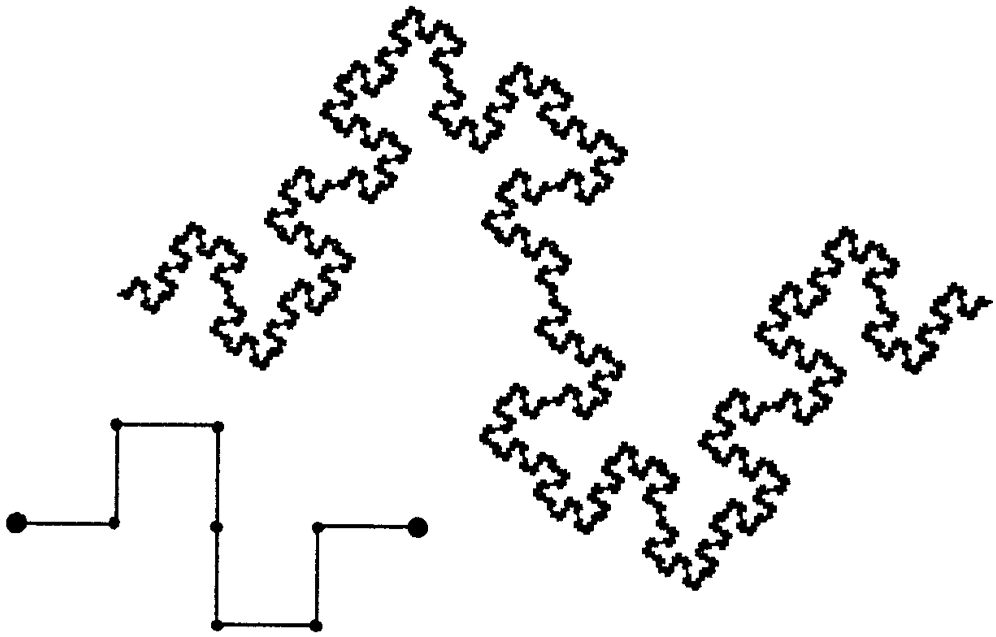


Figure 9.6 A fractal curve and its generator. The Hausdorff and box dimensions of the curve are equal to $\log 8/\log 4 = 1\frac{1}{2}$

taking V as the interior of the isosceles triangle of base length 1 and height $\frac{1}{2}a\sqrt{3}$, so Theorem 9.3 gives the dimension stated. □

There is a convenient method of specifying certain self-similar sets diagrammatically, in particular self-similar curves such as Example 9.5. A *generator*

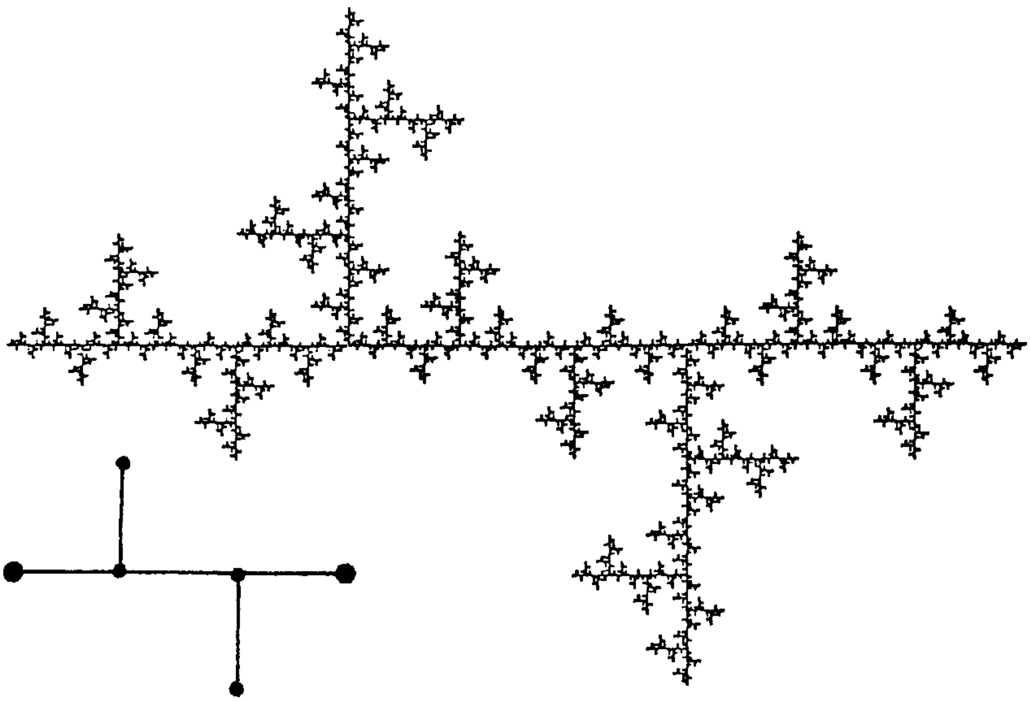


Figure 9.7 A tree-like fractal and its generator. The Hausdorff and box dimensions are equal to $\log 5/\log 3 = 1.465\ldots$

consists of a number of straight line segments and two points specially identified. We associate with each line segment the similarity that maps the two special points onto the endpoints of the segment. A sequence of sets approximating to the self-similar attractor may be built up by iterating the process of replacing each line segment by a similar copy of the generator; see figures 9.5–9.7 for some examples. Note that the similarities are defined by the generator only to within reflection and 180° rotation but the orientation may be specified by displaying the first step of the construction.

9.3 Some variations

The calculations underlying Theorem 9.3 may be adapted to estimate the dimension of the attractor F of an IFS consisting of contractions that are not similarities.

Proposition 9.6

Let F be the attractor of an IFS consisting of contractions $\{S_1, \dots, S_m\}$ on a closed subset D of $\mathbb{R}^n$ such that

$$|S_i(x) - S_i(y)| \leqslant c_i|x - y| \qquad (x, y \in D)$$

with $0 < c_i < 1$ for each i. Then $\dim_{\mathrm{H}} F \leqslant s$ and $\overline{\dim}_{\mathrm{B}} F \leqslant s$, where $\sum_{i=1}^m c_i^s = 1$.

Proof. These estimates are essentially those of the first and last paragraphs of the proof of Theorem 9.3, noting that we have the inequality $|A_{i_1,\dots,i_k}| \leqslant c_{i_1} \cdots c_{i_k}|A|$ for each set A, rather than equality. □

We next obtain a lower bound for dimension in the case where the components $S_i(F)$ of F are disjoint. Note that this will certainly be the case if there is *some* non-empty compact set E with $S_i(E) \subset E$ for all i and with the $S_i(E)$ disjoint.

Proposition 9.7

Consider the IFS consisting of contractions $\{S_1, \dots, S_m\}$ on a closed subset D of $\mathbb{R}^n$ such that

$$b_i|x - y| \leqslant |S_i(x) - S_i(y)| \qquad (x, y \in D) \tag{9.16}$$

with $0 < b_i < 1$ for each i. Assume that the (non-empty compact) attractor F satisfies

$$F = \bigcup_{i=1}^m S_i(F), \tag{9.17}$$

with this union disjoint. Then F is totally disconnected and $\dim_{\mathrm{H}} F \geqslant s$ *where*

$$\sum_{i=1}^{m} b_i^s = 1. \tag{9.18}$$

Proof. Let $d > 0$ be the minimum distance between any pair of the disjoint compact sets $S_1(F), \dots, S_m(F)$, i.e. $d = \min_{i\neq j} \inf\{|x - y| : x \in S_i(F), y \in S_j(F)\}$. Let $F_{i_1,\dots,i_k} = S_{i_1}\circ\cdots\circ S_{i_k}(F)$ and define μ by $\mu(F_{i_i,\dots i_k}) = (b_{i_1}\cdots b_{i_k})^s$. Since

$$\begin{aligned}\sum_{i=1}^{m} \mu(F_{i_1,\dots,i_k,i}) &= \sum_{i=1}^{m} (b_{i_1}\cdots b_{i_k} b_i)^s \\ &= (b_{i_1}\cdots b_{i_k})^s = \mu(F_{i_1,\dots,i_k}) \\ &= \mu\left(\bigcup_{i=1}^{k} F_{i_1,\dots,i_k,i}\right)\end{aligned}$$

it follows that μ defines a mass distribution on F with $\mu(F) = 1$.

If $x \in F$, there is a unique infinite sequence $i_1, i_2, \dots$ such that $x \in F_{i_1,\dots,i_k}$ for each k. For $0 < r < d$ let k be the least integer such that

$$b_{i_1}\cdots b_{i_k} d \leqslant r < b_{i_1}\cdots b_{i_{k-1}} d.$$

If $i'_1, \dots, i'_k$ is distinct from $i_1, \dots, i_k$, the sets $F_{i_1,\dots,i_k}$ and $F_{i'_1,\dots,i'_k}$ are disjoint and separated by a gap of at least $b_{i_1}\cdots b_{i_{k-1}} d > r$. (To see this, note that if j is the least integer such that $i_j \neq i'_j$ then $F_{i_j,\dots,i_k} \subset F_{i_j}$ and $F_{i'_j,\dots,i'_k} \subset F_{i'_j}$ are separated by d, so $F_{i_1,\dots,i_k}$ and $F_{i'_1,\dots,i'_k}$ are separated by at least $b_{i_1}\cdots b_{i_{j-1}} d$.) It follows that $F \cap B(x, r) \subset F_{i_1,\dots,i_k}$ so

$$\mu(F \cap B(x, r)) \leqslant \mu(F_{i_1,\dots,i_k}) = (b_{i_1} \dots b_{i_k})^s \leqslant d^{-s} r^s.$$

If U intersects F, then $U \subset B(x, r)$ for some $x \in F$ with $r = |U|$. Thus $\mu(U) \leqslant d^{-s}|U|^s$, so by the Mass distribution principle 4.2 $\mathcal{H}^s(F) > 0$ and $\dim_{\mathrm{H}} F \geqslant s$.

The separation indicated above implies that F is totally disconnected. □

Example 9.8. 'Non-linear' Cantor set

Let $D = \left[\frac{1}{2}(1 + \sqrt{3}), (1 + \sqrt{3})\right]$ *and let* $S_1, S_2 : D \to D$ *be given by* $S_1(x) = 1 + 1/x$, $S_2(x) = 2 + 1/x$. *Then* $0.44 < \dim_{\mathrm{H}} F \leqslant \underline{\dim}_{\mathrm{B}} F \leqslant \overline{\dim}_{\mathrm{B}} F < 0.66$ *where F is the attractor of* $\{S_1, S_2\}$. (*This example arises in connection with number theory*; *see Section* 10.2.)

Calculation. We note that $S_1(D) = \left[\frac{1}{2}(1 + \sqrt{3}), \sqrt{3}\right]$ and $S_2(D) = \left[\frac{1}{2}(3 + \sqrt{3}), 1 + \sqrt{3}\right]$ so we can use Propositions 9.6 and 9.7 to estimate $\dim_{\mathrm{H}} F$. By the mean value theorem (see Section 1.2) if $x, y \in D$ are distinct points then $(S_i(x) - S_i(y))/(x - y) = S_i'(z_i)$ for some $z_i \in D$. Thus for $i = 1, 2$,

$$\inf_{x \in D} |S_i'(x)| \leqslant \frac{|S_i(x) - S_i(y)|}{|x - y|} \leqslant \sup_{x \in D} |S_i'(x)|.$$

Since $S_1'(x) = S_2'(x) = -1/x^2$ it follows that

$$\tfrac{1}{2}(2 - \sqrt{3}) = (1 + \sqrt{3})^{-2} \leqslant \frac{|S_i(x) - S_i(y)|}{|x - y|} \leqslant (\tfrac{1}{2}(1 + \sqrt{3}))^{-2} = 2(2 - \sqrt{3})$$

for both $i = 1$ and $i = 2$. According to Propositions 9.6 and 9.7 lower and upper bounds for the dimensions are given by the solutions of $2(\frac{1}{2}(2 - \sqrt{3}))^s = 1$ and $2(2(2 - \sqrt{3}))^s = 1$ which are $s = \log 2/\log(2(2 + \sqrt{3})) = 0.34$ and $\log 2/\log(\frac{1}{2}(2 + \sqrt{3})) = 1.11$ respectively.

For a subset of the real line, an upper bound greater than 1 is not of much interest. One way of getting better estimates is to note that F is also the attractor of the *four* mappings on $[0, 1]$

$$S_i \circ S_j = i + 1/(j + 1/x) = i + x/(jx + 1) \qquad (i, j = 1, 2).$$

By calculating derivatives and using the mean-value theorem as before, we get that

$$(S_i \circ S_j)'(x) = (jx + 1)^{-2}$$

so

$$(j(1 + \sqrt{3}) + 1)^{-2}|x - y| \leqslant |S_i \circ S_j(x) - S_i \circ S_j(y)| \leqslant (\tfrac{1}{2}j(1 + \sqrt{3}) + 1)^{-2}|x - y|.$$

Lower and upper bounds for the dimensions are now given by the solutions of $2(2 + \sqrt{3})^{-2s} + 2(3 + 2\sqrt{3})^{-2s} = 1$ and $2(\frac{1}{2}(3 + \sqrt{3}))^{-2s} + 2(2 + \sqrt{3})^{-2s} = 1$, giving $0.44 < \dim_{\mathrm{H}} F < 0.66$, a considerable improvement on the previous estimates. In fact $\dim_{\mathrm{H}} F = 0.531$, a value that may be obtained by looking at yet higher-order iterates of the S_i. $\square$

*[The rest of this subsection may be omitted.]

The technique used in Example 9.8 to improve the dimension estimates is often useful for attractors of transformations that are not strict similarities. If F is the attractor for the IFS $\{S_1, \ldots, S_m\}$ on D then F is also the attractor for the IFS consisting of the m^k transformations $\{S_{i_1} \circ \cdots \circ S_{i_k}\}$ for each k. If the S_i are, say, twice differentiable on an open set containing F, it may be shown that when k is large, the contractions $S_{i_1} \circ \cdots \circ S_{i_k}$ are in a sense, close to similarities on D. In particular, for transformations on a subset D of $\mathbb{R}$, if $b = \inf_{x \in D} |(S_{i_1} \circ \cdots \circ S_{i_k})'(x)|$ and $c = \sup_{x \in D} |(S_{i_1} \circ \cdots \circ S_{i_k})'(x)|$, then

$$b|x - y| \leqslant |S_{i_1} \circ \cdots \circ S_{i_k}(x) - S_{i_1} \circ \cdots \circ S_{i_k}(y)| \leqslant c|x - y| \quad (x, y \in D)$$

If k is large then b/c will be close to 1, and applying Propositions 9.6 and 9.7 to the m^k transformations $S_{i_1}\circ\cdots\circ S_{i_k}$ gives good upper and lower estimates for the dimensions of F.

We may take this further. If the S_i are twice differentiable on a subset D of $\mathbb{R}$,

$$\frac{|S_{i_1}\circ\cdots\circ S_{i_k}(x) - S_{i_1}\circ\cdots\circ S_{i_k}(y)|}{|x-y|} \sim |(S_{i_1}\circ\cdots\circ S_{i_k})'(w)|$$

for large k, where x, y and w are any points of D. The composition of mappings $S_{i_1}\circ\cdots\circ S_{i_k}$ is close to a similarity on D, so by comparison with Theorem 9.3 we would expect the dimension of the attractor F to be close to the value of s for which

$$\sum_{\mathcal{I}_k} |(S_{i_1}\circ\cdots\circ S_{i_k})'(w)|^s = 1 \tag{9.19}$$

where the sum is over the set $\mathcal{I}_k$ of all k-term sequences. This expectation motivates the following theorem.

Theorem 9.9

Let $V \subset \mathbb{R}$ be an open interval. Let $S_1, \ldots, S_m$ be contractions on $\overline{V}$ that are twice differentiable on V with $a \leqslant |S_i'(w)| \leqslant c$ for all i and $w \in V$, where $0 < a \leqslant c < 1$ are constants. Suppose that the S_i satisfy the open set condition (9.11) *with open set V. Then the limit*

$$\lim_{k\to\infty}\left[\sum_{\mathcal{I}_k} |(S_{i_1}\circ\cdots\circ S_{i_k})'(w)|^s\right]^{1/k} = \varphi(s) \tag{9.20}$$

exists for each $s > 0$, is independent of $w \in V$, and is decreasing in s. If F is the attractor of $\{S_1, \ldots, S_m\}$ then $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F$ is the solution of $\varphi(s) = 1$, and F is an s-set, i.e. $0 < \mathcal{H}^s(F) < \infty$ *for this value of s.*

Note on Proof. The main difficulty is to show that the limit (9.20) exists—this depends on the differentiability condition on the S_i. Given this, the argument outlined above may be used to show that the value of s satisfying (9.19) is a good approximation to the dimension when k is large; letting $k \to \infty$ then gives the result.

Similar ideas, but involving the rate of convergence to the limit in (9.20), are needed to show that $0 < \mathcal{H}^s(F) < \infty$. □

There are higher-dimensional analogues of Theorem 9.9. Suppose that the contractions $S_1, \ldots, S_m$ on a domain D in the complex plane are complex analytic mappings. Then the S_i are conformal, or in other words are locally like similarity transformations, contracting at the same rate in every direction. We have

$$S_i(z) = S_i(z_0) + S_i'(z_0)(z - z_0) + \text{terms in } (z-z_0)^2 \text{ and higher powers}$$

so that if $z - z_0$ is small

$$S_i(z) \simeq S_i(z_0) + S_i'(z_0)(z - z_0) \tag{9.21}$$

where $S_i'(z_0)$ is a complex number with $|S_i'(z_0)| < 1$. But the right-hand side of (9.21) is just a similarity expressed in complex notation. In this setting, Theorem 9.9 holds, by the same sort of argument as in the 1-dimensional case.

Results such as these are part of the 'thermodynamic formalism', a body of theory that leads to dimension formulae for many attractors.

9.4 Self-affine sets

Self-affine sets form an important class of sets, which include self-similar sets as a particular case. An *affine transformation* $S : \mathbb{R}^n \to \mathbb{R}^n$ is a transformation of the form

$$S(x) = T(x) + b$$

where T is a linear transformation on $\mathbb{R}^n$ (representable by an $n \times n$ matrix) and b is a vector in $\mathbb{R}^n$. Thus an affine transformation S is a combination of a translation, rotation, dilation and, perhaps, a reflection. In particular, S maps spheres to ellipsoids, squares to parallelograms, etc. Unlike similarities, affine transformations contract with differing ratios in different directions.

If an IFS consists of affine contractions $\{S_1, \dots, S_m\}$ on $\mathbb{R}^n$, the attractor F guaranteed by Theorem 9.1 is termed a *self-affine set*. An example is given in figure 9.8: S_1, S_2 and S_3 are defined as the transformations that map the square E onto the three rectangles in the obvious way. (In the figure the attractor F is represented as the aggregate of $S_{i_1} \circ \cdots \circ S_{i_k}(E)$ over all sequences $(i_1, \dots, i_k)$ with $i_j = 1, 2, 3$ for suitably large k. Clearly F is made up of the three affine copies of itself: $S_1(F)$, $S_2(F)$ and $S_3(F)$.)

It is natural to look for a formula for the dimension of self-affine sets that generalizes formula (9.13) for self-similar sets. We would hope that the dimension depends on the affine transformations in a reasonably simple way, easily expressible in terms of the matrices and vectors that represent the affine transformation. Unfortunately, the situation is much more complicated than this—the following example shows that if the affine transformations are varied in a continuous way, the dimension of the self-affine set need not change continuously.

Example 9.10

Let S_1, S_2 be the affine contractions on $\mathbb{R}^2$ that map the unit square onto the rectangles R_1 and R_2 of sides $\frac{1}{2}$ and ε where $0 < \varepsilon < \frac{1}{2}$, as in figure 9.9. The rectangle R_1 abuts the y-axis, but the end of R_2 is distance $0 \leqslant \lambda \leqslant \frac{1}{2}$ from the y-axis. If F is the attractor of $\{S_1, S_2\}$, we have $\dim_H F \geqslant 1$ when $\lambda > 0$, but $\dim_H F = \log 2 / -\log \varepsilon < 1$ when $\lambda = 0$.

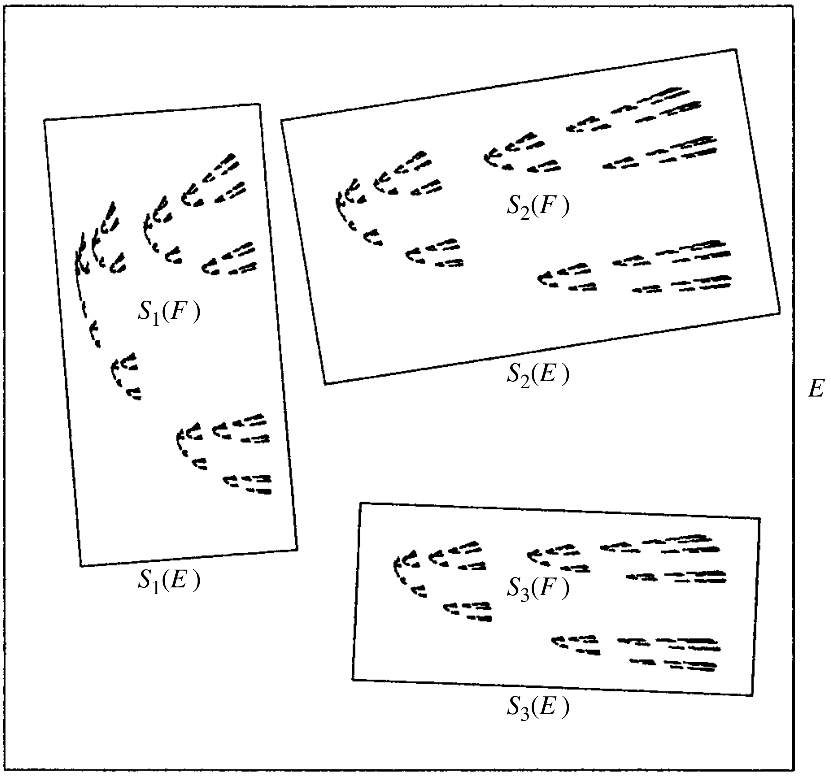


Figure 9.8 A self-affine set which is the attractor of the affine transformations that map the square E onto the rectangles shown

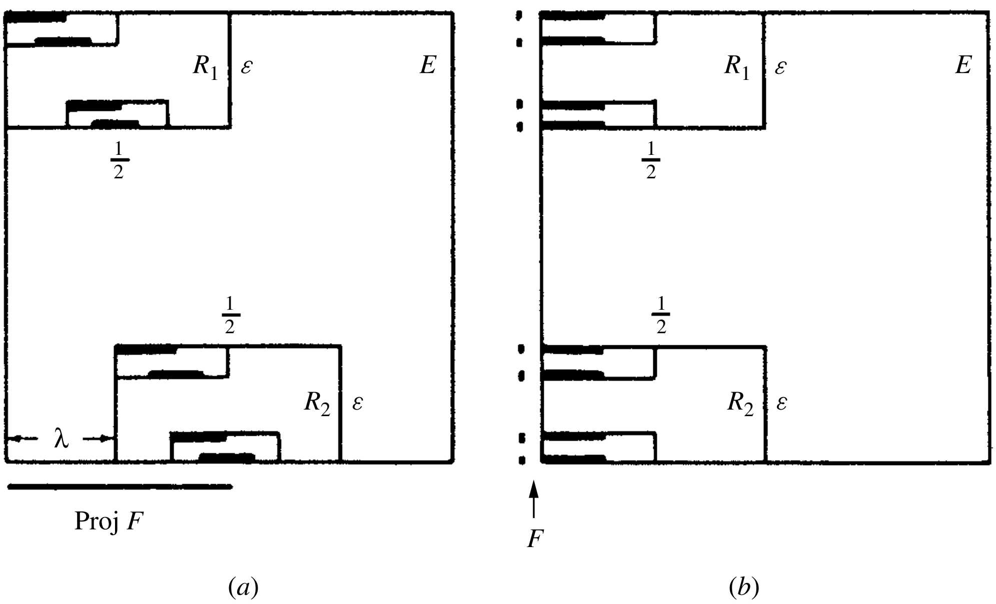


Figure 9.9 Discontinuity of the dimension of self-affine sets. The affine mappings S_1 and S_2 map the unit square E onto R_1 and R_2. In (a) $\lambda > 0$ and $\dim_H F \geqslant \dim_H \text{proj}\, F = 1$, but in (b) $\lambda = 0$, and $\dim_H F = \log 2 / -\log \varepsilon < 1$

Calculation. Suppose $\lambda > 0$ (figure 9.9(a)). Then the kth stage of the construction $E_k = \bigcup S_{i_1} \circ \cdots \circ S_{i_k}(E)$ consists of 2^k rectangles of sides 2^{-k} and ε^k with the projection of E_k onto the x-axis, $\text{proj}\, E_k$, containing the interval $[0, 2\lambda]$. Since $F = \bigcap_{i=1}^{\infty} E_k$ it follows that $\text{proj}\, F$ contains the interval $[0, 2\lambda]$. (Another way of seeing this is by noting that $\text{proj}\, F$ is the attractor of $\tilde{S}_1, \tilde{S}_2 : \mathbb{R} \to \mathbb{R}$ given by $\tilde{S}_1(x) = \frac{1}{2}x$, $\tilde{S}_2(x) = \frac{1}{2}x + \lambda$, which has as attractor the interval $[0, 2\lambda]$.) Thus $\dim_H F \geqslant \dim_H \text{proj}\, F = \dim_H[0, 2\lambda] = 1$.

If $\lambda = 0$, the situation changes (figure 9.9(b)). E_k consists of 2^k rectangles of sides 2^{-k} and ε^k which all have their left-hand ends abutting the y-axis, with E_k contained in the strip $\{(x, y) : 0 \leqslant x \leqslant 2^{-k}\}$. Letting $k \to \infty$ we see that F is a uniform Cantor set contained in the y-axis, which may be obtained by repeatedly removing a proportion $1 - 2\varepsilon$ from the centre of each interval. Thus $\dim_H F = \log 2 / -\log \varepsilon < 1$ (see Example 4.5). □

With such discontinuous behaviour, which becomes even worse for more involved sets of affine transformations, it is likely to be difficult to obtain a general expression for the dimension of self-affine sets. However, one situation which has been completely analysed is the self-affine set obtained by the following recursive construction; a specific case is illustrated in figures 9.10 and 9.11.

Example 9.11

Let the unit square E_0 be divided into a $p \times q$ array of rectangles of sides $1/p$ and $1/q$ where p and q are positive integers with $p < q$. Select a subcollection of these rectangles to form E_1, and let N_j denote the number of rectangles selected from the jth column for $1 \leqslant j \leqslant p$; see figure 9.10. Iterate this construction in the usual way, with each rectangle replaced by an affine copy of E_1, and let F be the limiting set obtained. Then

$$\dim_H F = \log\left(\sum_{j=1}^{p} N_j^{\log p/\log q}\right)\frac{1}{\log p}$$

and

$$\dim_B F = \frac{\log p_1}{\log p} + \log\left(\frac{1}{p_1}\sum_{j=1}^{p} N_j\right)\frac{1}{\log q}$$

where p_1 is the number of columns containing at least one rectangle of E_1.

Calculation. Omitted. □

Notice in this example that the dimension depends not only on the number of rectangles selected at each stage, but also on their relative positions. Moreover $\dim_H F$ and $\dim_B F$ are not, in general, equal.

*[The rest of this subsection may be omitted.]

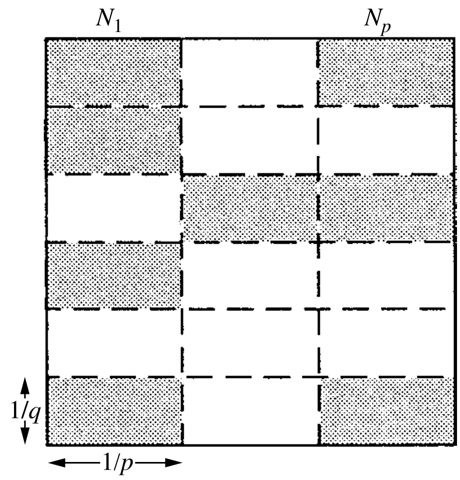


Figure 9.10 Data for the self-affine set of Example 9.11. The affine transformations map the square onto selected $1/p \times 1/q$ rectangles from the $p \times q$ array

The above example is rather specific in that the affine transformations are all translates of each other. Obtaining a dimension formula for general self-affine sets is an intractable problem. We briefly outline an approach which leads to an expression for the dimension of the attractor of the affine contractions $S_i(x) = T_i(x) + b_i (1 \leqslant i \leqslant m)$ for *almost all* sequences of vectors $b_1, \dots, b_m$.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear mapping that is contracting and non-singular. The *singular values* $1 > \alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n > 0$ of T may be defined in two ways: they are the lengths of the principal semi-axes of the ellipsoid $T(B)$ where

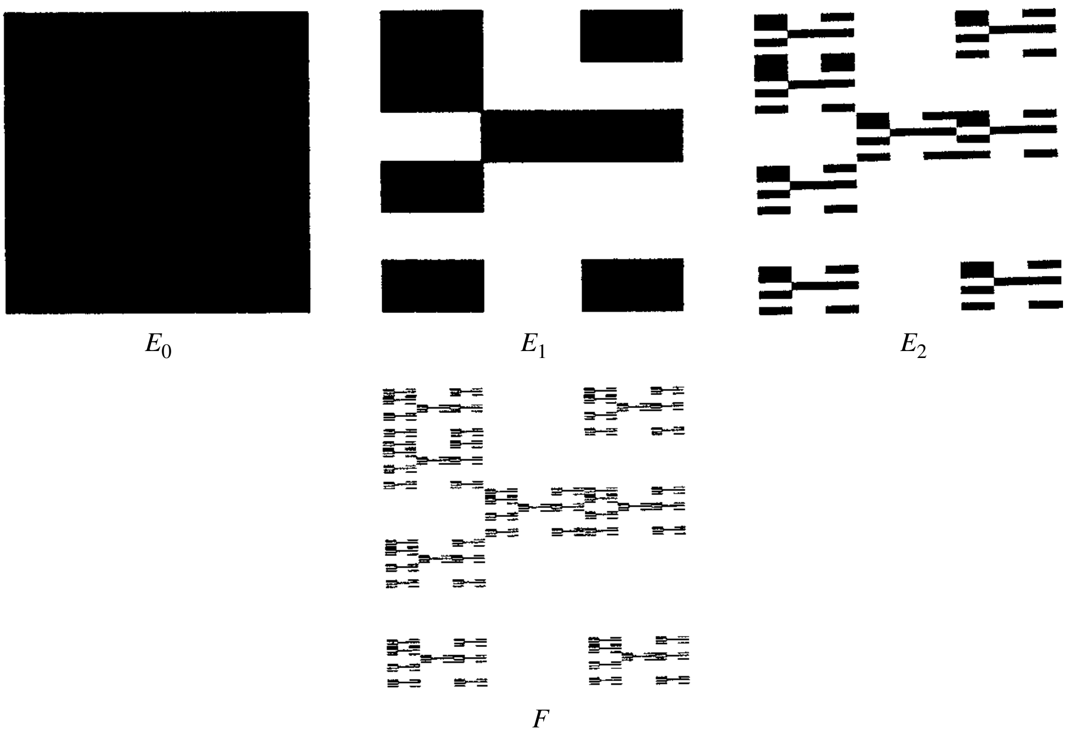


Figure 9.11 Construction of a self-affine set of the type considered in Example 9.11. Such sets may have different Hausdorff and box dimensions

B is the unit ball in $\mathbb{R}^n$, and they are the positive square roots of the eigenvalues of T^*T, where T^* is the adjoint of T. Thus the singular values reflect the contractive effect of T in different directions. For $0 \leqslant s \leqslant n$ we define the *singular value function*

$$\varphi^s(T) = \alpha_1\alpha_2 \cdots \alpha_{r-1}\alpha_r^{s-r+1} \tag{9.22}$$

where r is the integer for which $r - 1 < s \leqslant r$. Then $\varphi^s(T)$ is continuous and strictly decreasing in s. Moreover, for fixed s, φ^s may be shown to be submultiplicative, i.e.

$$\varphi^s(TU) \leqslant \varphi^s(T)\varphi^s(U)$$

for any linear mappings T and U. We introduce the kth level sums $\Sigma_k^s \equiv \sum_{\mathcal{I}_k} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_k})$ where $\mathcal{I}_k$ denotes the set of all k-term sequences $(i_1, \ldots, i_k)$ with $1 \leqslant i_j \leqslant m$. For fixed s

$$\begin{aligned}
\Sigma_{k+q}^s &= \sum_{\mathcal{I}_{k+q}} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_{k+q}}) \\
&\leqslant \sum_{\mathcal{I}_{k+q}} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_k})\varphi^s(T_{i_{k+1}} \circ \cdots \circ T_{i_{k+q}}) \\
&= \left(\sum_{\mathcal{I}_k} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_k}) \right) \left(\sum_{\mathcal{I}_q} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_q}) \right) = \Sigma_k^s \Sigma_q^s
\end{aligned}$$

i.e. the sequence Σ_k^s is submultiplicative in k. By a standard property of submultiplicative sequences, $(\Sigma_k^s)^{1/k}$ converges to a number Σ_∞^s as $k \to \infty$. Since φ^s is decreasing in s, so is Σ_∞^s. Provided that $\Sigma_\infty^n \leqslant 1$, there is a unique s, which we denote by $d(T_1, \ldots, T_m)$, such that $1 = \Sigma_\infty^s = \lim_{k\to\infty} \left(\sum_{\mathcal{I}_k} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_k})\right)^{1/k}$. Equivalently

$$d(T_1, \ldots, T_m) = \inf\left\{ s : \sum_{k=1}^{\infty} \sum_{\mathcal{I}_k} \varphi^s(T_{i_1} \circ \cdots \circ T_{i_k}) < \infty \right\}. \tag{9.23}$$

Theorem 9.12

Let $T_1, \ldots, T_m$ be linear contractions and let $y_1, \ldots, y_m \in \mathbb{R}^n$ be vectors. If F is the self-affine set satisfying

$$F = \bigcup_{i=1}^{m} (T_i(F) + y_i)$$

then $\dim_{\mathrm{H}} F = \dim_{\mathrm{B}} F \leqslant d(T_1, \ldots, T_m)$. *If* $|T_i(x) - T_i(y)| \leqslant c|x - y|$ *for all* i where $0 < c < \frac{1}{2}$, *then equality holds for almost all* $(y_1, \ldots, y_m) \in \mathbb{R}^{nm}$ *in the sense of nm-dimensional Lebesgue measure.*

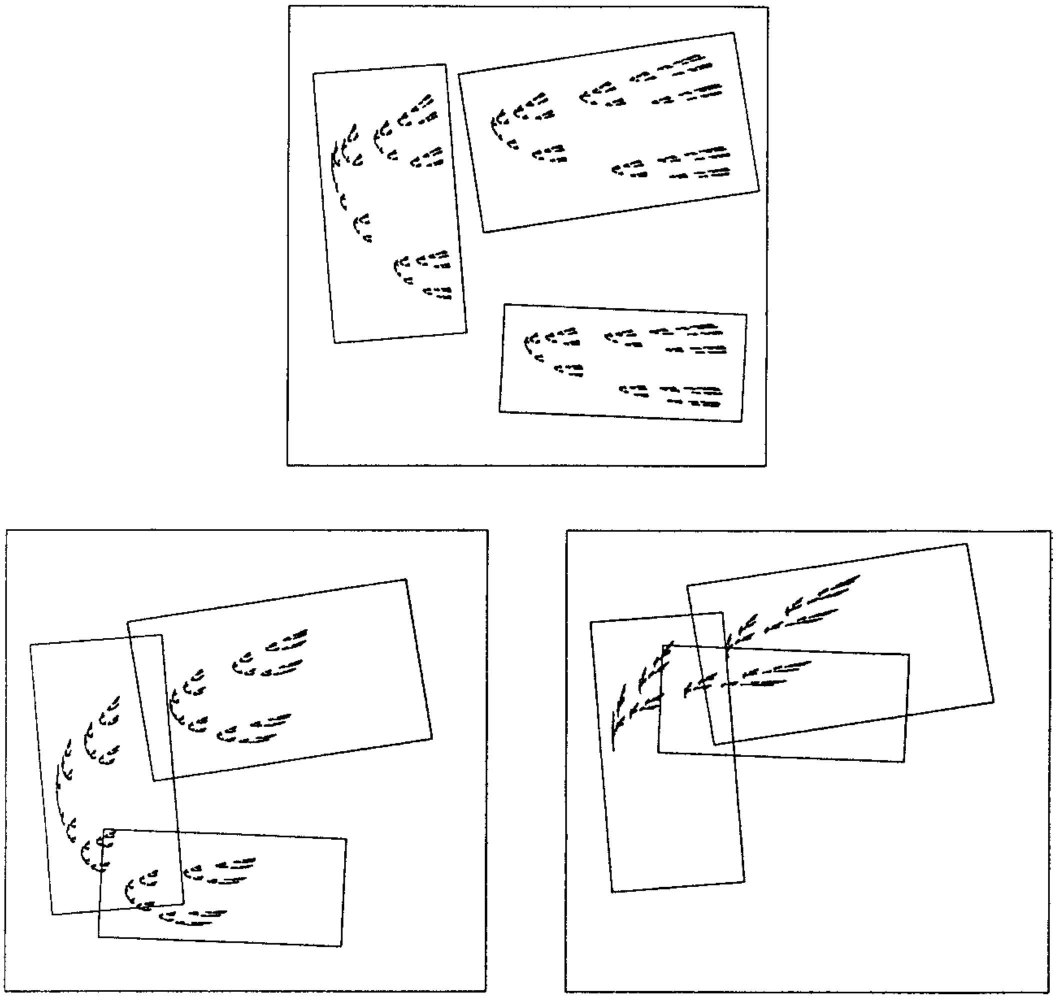


Figure 9.12 Each of the fractals depicted above is the attractor of the set of transformations that map the square onto the three rectangles. The affine transformations for each fractal differ only by translations, so by Theorem 9.12 the three fractals all have the same dimension (unless we have been very unlucky in our positioning!). A computation gives this common value of Hausdorff and box dimension as about 1.42

Partial proof. We show that $\dim_{\mathrm{H}} F \leqslant d(T_1, \ldots, T_m)$ for all $y_1, \ldots, y_m$. Write S_i for the contracting affine transformation $S_i(x) = T_i(x) + y_i$. Let B be a large ball so that $S_i(B) \subset B$ for all i. Given $\delta > 0$ we may choose k large enough to get $|S_{i_1} \circ \cdots \circ S_{i_k}(B)| < \delta$ for every k-term sequence $(i_1, \ldots, i_k) \in \mathcal{I}_k$. By (9.6) $F \subset \bigcup_{\mathcal{I}_k} S_{i_1} \circ \cdots \circ S_{i_k}(B)$. But $S_{i_1} \circ \cdots \circ S_{i_k}(B)$ is a translate of the ellipsoid $T_{i_1} \circ \cdots \circ T_{i_k}(B)$ which has principal axes of lengths $\alpha_1|B|, \ldots, \alpha_n|B|$, where $\alpha_1, \ldots, \alpha_n$ are the singular values of $T_{i_1} \circ \cdots \circ T_{i_k}$. Thus $S_{i_1} \circ \cdots \circ S_{i_k}(B)$ is contained in a rectangular parallelepiped P of side lengths $\alpha_1|B|, \ldots, \alpha_n|B|$. If $0 \leqslant s \leqslant n$ and r is the least integer greater than or equal to s, we may divide P into at most

$$\left(\frac{2\alpha_1}{\alpha_r}\right)\left(\frac{2\alpha_2}{\alpha_r}\right) \cdots \left(\frac{2\alpha_{r-1}}{\alpha_r}\right) \leqslant 2^n \alpha_1 \cdots \alpha_{r-1} \alpha_r^{1-r}$$

cubes of side $\alpha_r|B| < \delta$. Hence $S_{i_1}\circ\cdots\circ S_{i_k}(B)$ may be covered by a collection of cubes U_i with $|U_i| < \delta\sqrt{n}$ such that

$$\sum_i |U_i|^s \leqslant 2^n\alpha_1\cdots\alpha_{r-1}\alpha_r^{1-r}\alpha_r^s|B|^s$$
$$\leqslant 2^n|B|^s\varphi^s(T_{i_1}\circ\cdots\circ T_{i_k}).$$

Taking such a cover of $S_{i_1}\circ\cdots\circ S_{i_k}(B)$ for each $(i_1,\ldots,i_k)\in\mathcal{I}_k$ it follows that

$$\mathcal{H}^s_{\delta\sqrt{n}}(F) \leqslant 2^n|B|^s\sum_{\mathcal{I}_k}\varphi^s(T_{i_1}\circ\cdots\circ T_{i_k}).$$

But $k\to\infty$ as $\delta\to 0$, so by (9.23), $\mathcal{H}^s(F)=0$ if $s > d(T_1,\ldots,T_m)$. Thus $\dim_{\mathrm{H}} F \leqslant d(T_1,\ldots,T_m)$.

The lower estimate for $\dim_{\mathrm{H}} F$ may be obtained using the potential theoretic techniques of Section 4.3. We omit the somewhat involved details. $\square$

One consequence of this theorem is that, unless we have been unfortunate enough to hit on an exceptional set of parameters, the fractals in figure 9.12 all have the same dimension, estimated at about 1.42.

9.5 Applications to encoding images

In this chapter, we have seen how a small number of contractions can determine objects of a highly intricate fractal structure. This has applications to data compression—if a complicated picture can be encoded by a small amount of information, then the picture can be transmitted or stored very efficiently.

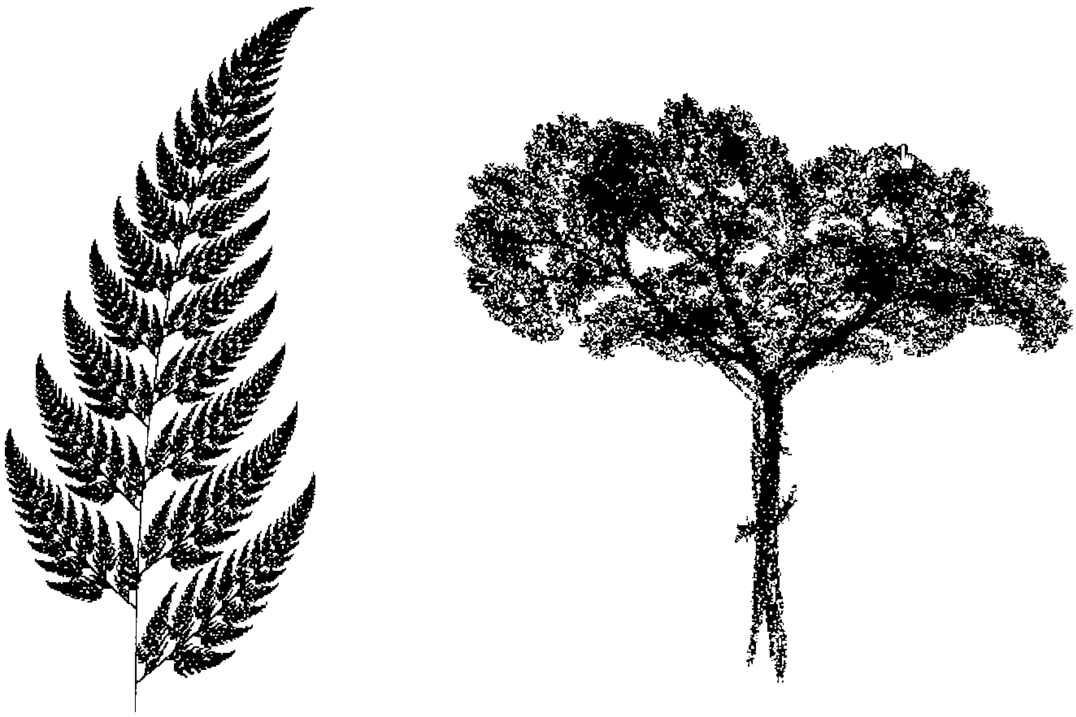
It is desirable to know which objects can be represented as, or approximated by, attractors of an iterated function system, and also how to find contractions that lead to a good representation of a given object. Clearly, the possibilities using, say, three or four transformations are limited by the small number of parameters at our disposal. Such sets are also likely to have a highly repetitive structure.

However, a little experimentation drawing self-affine sets on a computer (see end of Section 9.1) can produce surprisingly good pictures of naturally occurring objects such as ferns, grasses, trees, clouds, etc. The fern and tree in figure 9.13 are the attractors of just four and six affine transformations, respectively. Self-similarity and self-affinity are indeed present in nature.

The following theorem, sometimes known as the collage theorem, gives an idea of how good an approximation a set is to the attractor of an IFS.

Theorem 9.13

Let $\{S_1,\ldots,S_m\}$ be an IFS and suppose that $|S_i(x)-S_i(y)|\leqslant c|x-y|$ for all $x,y\in\mathbb{R}^n$ and all i, where $c<1$. Let $E\subset\mathbb{R}^n$ be any non-empty compact



(*a*) (*b*)

Figure 9.13 The fern (*a*) and tree (*b*) are the attractors of just four and six affine transformations, respectively

set. Then

$$d(E, F) \leqslant d\left(E, \bigcup_{i=1}^{m} S_i(E)\right) \frac{1}{(1-c)} \tag{9.24}$$

where F is the attractor for the IFS, and d is the Hausdorff metric.

Proof. Using the triangle inequality for the Hausdorff metric followed by the definition (9.2) of the attractor

$$\begin{aligned} d(E, F) &\leqslant d\left(E, \bigcup_{i=1}^{m} S_i(E)\right) + d\left(\bigcup_{i=1}^{m} S_i(E), F\right) \\ &= d\left(E, \bigcup_{i=1}^{m} S_i(E)\right) + d\left(\bigcup_{i=1}^{m} S_i(E), \bigcup_{i=1}^{m} S_i(F)\right) \\ &\leqslant d\left(E, \bigcup_{i=1}^{m} S_i(E)\right) + c\,d(E, F) \end{aligned}$$

by (9.5), as required. □

A consequence of Theorem 9.13 is that any compact subset of $\mathbb{R}^n$ can be approximated arbitrarily closely by a self-similar set.

Corollary 9.14

Let E be a non-empty compact subset of $\mathbb{R}^n$. Given $\delta > 0$ there exist contracting similarities $S_1, \dots, S_m$ with attractor F satisfying $d(E, F) < \delta$.

Proof. Let $B_1, \dots, B_m$ be a collection of balls that cover E and which have centres in E and radii at most $\frac{1}{4}\delta$. Then $E \subset \bigcup_{i=1}^m B_i \subset E_{\delta/4}$, where $E_{\delta/4}$ is the $\frac{1}{4}\delta$-neighbourhood of E. For each i, let S_i be any contracting similarity of ratio less than $\frac{1}{2}$ that maps E into B_i. Then $S_i(E) \subset B_i \subset (S_i(E))_{\delta/2}$, so $(\bigcup_{i=1}^m S_i(E)) \subset E_{\delta/4}$ and $E \subset \bigcup_{i=1}^m (S_i(E))_{\delta/2}$. By definition of the Hausdorff metric, $d(E, \bigcup_{i=1}^m S_i(E)) \leqslant \frac{1}{2}\delta$. It follows from (9.24) that $d(E, F) < \delta$ where F is the attractor. $\square$

The approximation by the IFS attractor given by the above proof is rather coarse—it is likely to yield a very large number of contractions that take little account of the fine structure of E. A rather more subtle approach is required to obtain convincing images with a small number of transformations. One method which often gives good results is to draw a rough outline of the object and then cover it, as closely as possible, by a number of smaller similar (or affine) copies. The similarities (or affinities) thus determined may be used to compute an attractor which may be compared with the object being modelled. Theorem 9.13 guarantees that the attractor will be a good approximation if the union of the smaller copies is close to the object. A trial and error process allows modification and improvements to the picture.

More complex objects may be built up by superposition of the invariant sets of several different sets of transformations.

Ideally, it would be desirable to have a ‘camera’ which could be pointed at an object to produce a ‘photograph’ consisting of a specified number of affine transformations whose attractor is a good approximation to the object. Obviously, the technical problems involved are considerable. One approach is to scan the object to estimate various geometric parameters, and use these to impose restrictions on the transformations.

For example, for a ‘natural fractal’ such as a fern, we might estimate the dimension by a box-counting method. The assumption that the similarities or affinities sought must provide an attractor of this dimension gives, at least theoretical, restrictions on the possible set of contractions, using results such as Theorem 9.3 or 9.12. In practice, however, such information is rather hard to utilize, and we certainly need many further parameters for it to be of much use.

Very often, attractors in the plane that provide good pictures of physical objects will have positive area, so will not be fractals in the usual sense. Nevertheless, such sets may well be bounded by fractal curves, a feature that adds realism to pictures of natural objects. However, fractal properties of boundaries of invariant sets seem hard to analyse.

These ideas may be extended to provide shaded or coloured images, by assigning a probability p_i to each of the contractions S_i, where $0 < p_i < 1$ and $\sum_{i=1}^m p_i = 1$. Without going into details, these data define a mass distribution

μ on the attractor F such that $\mu(A) = \sum_{i=1}^{m} p_i \mu(S_i^{-1}(A))$, and the set may be shaded, or even coloured, according to the local density of μ.

This leads to the following modification of the second method of drawing attractors mentioned at the end of Section 9.1. Let x_0 be any initial point. We choose S_{j_1} from $S_1, \cdots, S_m$ at random in such a way that the probability of choosing S_i is p_i, and let $x_1 = S_{j_1}(x_0)$. We continue in this way, so that $x_k = S_{j_k}(x_{k-1})$ where S_{j_k} equals S_i with probability p_i. Plotting the sequence $\{x_k\}$ (after omitting the first 100 terms, say) gives a rendering of the attractor F, but in such a way that a proportion $p_{i_1} \cdots p_{i_q}$ of the points tends to lie in the part $S_{i_1} \circ \cdots \circ S_{i_q}(F)$ for each $i_1, \ldots, i_q$. This variable point density provides a direct shading of F. Alternatively, the colour of a point of F can be determined by some rule, which depends on the number of $\{x_k\}$ falling close to each point. The computer artist may care to experiment with the endless possibilities that this method provides—certainly, some very impressive colour pictures have been produced using relatively few transformations.

It is perhaps appropriate to end this section with some of the 'pros and cons' of representing images using iterated function systems. By utilizing the self-similarity and repetition in nature, and, indeed, in man-made objects, the method often enables scenes to be described by a small number (perhaps fewer than 100) of contractions and probabilities in an effective manner. This represents an enormous compression of information compared, for example, with that required to detail the colour in each square of a fine mesh. The corresponding disadvantage is that there is a high correlation between different parts of the picture—the method is excellent for giving an overall picture of a tree, but is no use if the exact arrangement of the leaves on different branches is important. Given a set of affine contractions, reproduction of the image is computationally straightforward, is well-suited to parallel computation, and is stable—small changes in the contractions lead to small changes in the attractor. The contractions define the image at arbitrarily small scales, and it is easy to produce a close-up of a small region. At present, the main disadvantage of the method is the difficulty of obtaining a set of contractions to represent a given object or picture.

9.6 Notes and references

The first systematic account of what are now known as iterated function systems is that of Hutchinson (1981), though similar ideas were around earlier. The derivation of the formula for the dimension of self-similar sets was essentially given by Moran (1946). Computer pictures of self-similar sets and attractors of other IFSs are widespread, the works of Mandelbrot (1982), Dekking (1982), Peitgen, Jürgens and Saupe (1992) and Barnsley (1993) contain many interesting and beautiful examples.

For details of the thermodynamic formalism and material relating to Theorem 9.9, see Ruelle (1983), Bedford (1991), Beck and Schlögl (1993), Falconer (1997) and Pesin (1997).

A discussion of self-affine sets is given by Mandelbrot (1986) and a survey on their dimension properties by Peres and Solomyak (2000). Full details of Example 9.11 are given by McMullen (1984) and of Theorem 9.12 by Falconer (1988) and Solomyak (1998).

These ideas have been extended in many directions: for example, for IFSs with infinitely many transformations see Mauldin and Urbański (1996, 1999), and for graph directed constructions see Mauldin and Williams (1988).

Applications to image compression and encoding are described by Barnsley (1993), Barnsley and Hurd (1993) and Fisher (1995).

Exercises

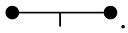
9.1 Verify that the Hausdorff metric satisfies the conditions for a metric.

9.2 Find a pair of similarity transformations on $\mathbb{R}$ for which the interval $[0, 1]$ is the attractor. Now find infinitely many such pairs of transformations.

9.3 Find sets of (i) four and (ii) three similarity transformations on $\mathbb{R}$ for which the middle third Cantor set is the attractor. Check that (9.13) has solution $\log 2/\log 3$ in each case.

9.4 Write down (using matrix notation) the four basic similarity transformations that define the von Koch curve (figure 0.2). Find an open set for which the open set condition holds and deduce from Theorem 9.3 that the von Koch curve does indeed have box and Hausdorff dimension of $\log 4/\log 3$.

9.5 Find an IFS for the set depicted in figure 0.5 and deduce that it has Hausdorff and box dimensions given by $4(\frac{1}{4})^s + (\frac{1}{2})^s = 1$.

9.6 Sketch the first few steps in the construction of a self-similar set with generator . What are the Hausdorff and box dimensions of this fractal? (The stem of the T is one quarter of the total length of the top.)

9.7 Let F be the set obtained by a Cantor-type construction in which each interval is replaced by two intervals, one of a quarter of the length at the left-hand end, and one of half the length at the right-hand end. Thus E_0 is the interval $[0, 1]$, E_1 consists of the intervals $[0, \frac{1}{4}]$ and $[\frac{1}{2}, 1]$, etc. Find an IFS with attractor F, and thus find the Hausdorff and box dimensions of F.

9.8 Describe the attractors of the following IFSs on $\mathbb{R}$.

(i) $S_1(x) = \frac{1}{4}x, \qquad S_2(x) = \frac{1}{4}x + \frac{3}{4};$
(ii) $S_1(x) = \frac{1}{2}x, \qquad S_2(x) = \frac{1}{2}x + \frac{1}{2};$
(iii) $S_1(x) = \frac{2}{3}x, \qquad S_2(x) = \frac{2}{3}x + \frac{1}{3}.$

9.9 Divide the unit square E_0 into p^2 squares of side $1/p$ in the obvious way and choose some m of these squares to form E_1. Let $S_i\,(1 \leqslant i \leqslant m)$ be similarity transformations that map E_0 onto each of these squares. Show that the attractor F of the IFS so defined has $\dim_H F = \dim_B F = \log m/\log p$.

9.10 Let $S_1, S_2 : [0, 1] \to [0, 1]$ be given by $S_1(x) = x/(2 + x)$ and $S_2(x) = 2/(2 + x)$. Show that the attractor F of this IFS satisfies $0.52 < \dim_H F < 0.81$.

9.11 Show that any self-similar set F satisfying the conditions of Theorem 9.3 has $c_1 \leqslant \underline{D}(F, x) \leqslant \overline{D}(F, x) \leqslant c_2$ for all $x \in F$, where c_1 and c_2 are positive constants. (See equations (5.2) and (5.3) for the definition of the densities.)

9.12 Let $S_1, \ldots, S_m$ be bi-Lipschitz contractions on a subset D of $\mathbb{R}^n$ and let F be the attractor satisfying (9.2). Show that, if V is any open set intersecting F, then F and $F \cap V$ have equal Hausdorff, equal upper box and equal lower box dimensions. Deduce from Corollary 3.9 that $\dim_{\mathrm{P}} F = \overline{\dim}_{\mathrm{B}} F$.

9.13 Verify the Hausdorff dimension formula in Example 9.11 in the cases (a) where $N_j = N$ for $1 \leqslant j \leqslant p$ and (b) where $N_j = N$ or 0 for $1 \leqslant j \leqslant p$, where N is an integer with $1 < N < q$. (Hint: see Example 7.13.)

9.14 Find the Hausdorff and box dimensions of the set in figure 9.11.

9.15 Write a computer program to draw self-similar sets in the plane, given a generator of the set.

9.16 Write a computer program to draw the attractor of a given collection of contractions of a plane region (see the end of Section 9.1). Examine the attractors of similarities, affinities and try some non-linear transformations. If you are feeling really enterprising, you might write a program to estimate the dimension of these sets using a box-counting method.