1 Introduction

We discuss a mathematical framework for deterministic and random (or non-deterministic) fractals. The approach is via *scaling laws* and *scaling operators*; the latter also being known as *iterated function systems*. This gives a very attractive theory with many applications, particularly to computer graphics and to image and data compression. Various applications of fractals are discussed elsewhere in this book.

The essential ideas will first be presented by means of a number of standard examples. Readers who wish to obtain a relatively informal overview of the material can accordingly restrict themselves to Sections 2 and 3, and to the less formal parts of the later Sections. Sections 4, 5 and 6 develop much of the mathematics behind these ideas. I have tried to keep the mathematics self-contained, and in particular have attempted to motivate and develop from first principles the relevant notions of metric spaces, measure theory, and probability theory. The later sections may perhaps serve as a brief introduction to some aspects of these subjects.

It is perhaps worth mentioning here one point that sometimes causes confusion. A "mathematical" fractal in a certain precise sense looks the same at all scales; i.e. when examined under a microscope *at no matter what the magnification* it will appear similar to the original object. On the other hand a "physical" fractal will display this "self-similarity" for only a range of magnifications or scales. The mathematical object will of course only be an accurate model within this particular range.

Examples of non-integer dimensional sets with scaling properties have long been known to mathematicians. It was Mandelbrot who introduced the term *fractal* and who in a series of papers and books (see [M] and the references there) developed the connections between these ideas and a range of phenomena in the physical, biological and social sciences.

In [H] we showed that to each Scaling Operator there corresponds a unique fractal set (or measure) in a natural manner. Approximations via the scaling operator, fractal coding, and dimension and density properties of fractals were also developed. The terminology "Iterated Function System" was introduced later by Barnsley and Demko [BD]. In earlier papers, Moran [Mo] proved the dimension results for (fractal) sets satisfying an "Open Set Condition" and Williams [W] developed properties of iterates of contraction maps. Applications to computer graphics were considered in Diaconis and Shahshahani [DS][1], in [BD] and in Barnsley, Ervin, Hardin and Lancaster [BEHL]. The Markov Process approach to generating fractals has been developed in [DS], [BD], and elsewhere.

The results for random fractal sets are due to Falconer [F2], Graf [G],

[1]See also: Esoteric Math Has Practical Result, *Science*, 1984 (**225**), 494–495.

and Mauldin and Williams [MW], and for random fractal measures are due to [MW] and Arbeiter [A]. A simple approach is given in Hutchinson and Rüschendorf [HR]. Zähle [Z] and Patzschke and Zähle [PZ] developed an "axiomatic" approach to random fractals and also established connections with the ideas developed by Falconer, Graf, Mauldin & Williams and Arbeiter.

General references at an introductory level are Barnsley [B], Falconer [F1], Peitgen, Jürgens and Saupe[PJS], and the book Fractal Geometry and Analysis [FG] (in particular, the article by Vrscray).

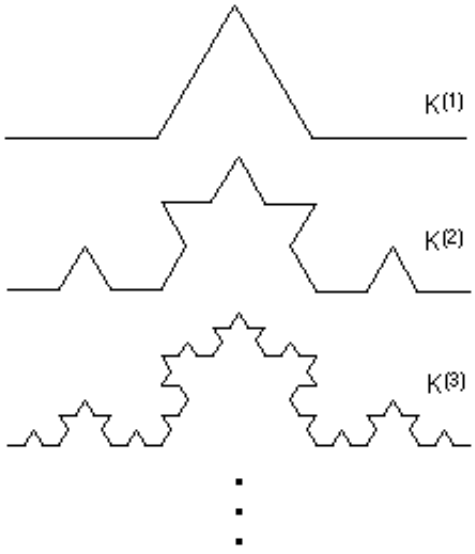
Acknowledgment I would particularly like to thank Jacki Wicks for many valuable discussions we had during the writing of her undergraduate honours thesis [Wi], and Andy Wood for various helpful suggestions in the section on probability.

2 Some important examples

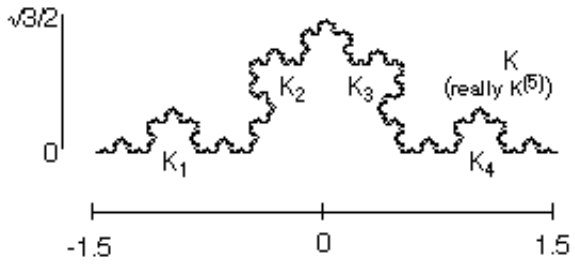
We begin with an informal discussion of some of the main ideas. Our intention is to develop the reader's intuition. The relevant notions will be defined later in a more precise manner.

2.1 The Koch curve

A *fractal set* is a set K in $\mathbb{R}^n$[2] with certain *scaling properties*. The *Koch curve* is an example. The following diagram shows certain sets in a sequence $K^{(1)}, K^{(2)}, \ldots, K^{(j)}, \ldots$ which approximates the Koch curve K. The set K is the limit of this sequence, but in practice we can only draw an approximation.



[2] $\mathbb{R}^n$ is n-dimensional Euclidean space. The important cases are the line $\mathbb{R} = \mathbb{R}^1$, the plane $\mathbb{R}^2$, and three-dimensional space $\mathbb{R}^3$.



From the diagram,

$$K = K_1 \cup K_2 \cup K_3 \cup K_4. \tag{1}$$

The important point here is that each K_i is congruent to a scaled version of K. It is also clear from the diagram that the scaling factor is $1/3$. More precisely, and significantly from our point of view, there are maps

$$S_i : \mathbb{R}^2 \to \mathbb{R}^2 \quad i = 1, \ldots, 4,$$

where each S_i is a composition of a translation, rotation, and scaling (with scaling factor $1/3$), such that (1) can be written

$$K = S_1(K) \cup S_2(K) \cup S_3(K) \cup S_4(K). \tag{2}$$

If we look again at the Koch curve, we see that we can also write it as the union of *two* sets (the left and right hand sides), each of which is obtained from K by means of a translation, rotation, *reflection* and scaling (in this case by $1/\sqrt{3}$, as is easy to calculate). Denoting the corresponding maps by T_i we have

$$K = T_1(K) \cup T_2(K). \tag{3}$$

2.2 A fractal measure on the line and the Koch measure

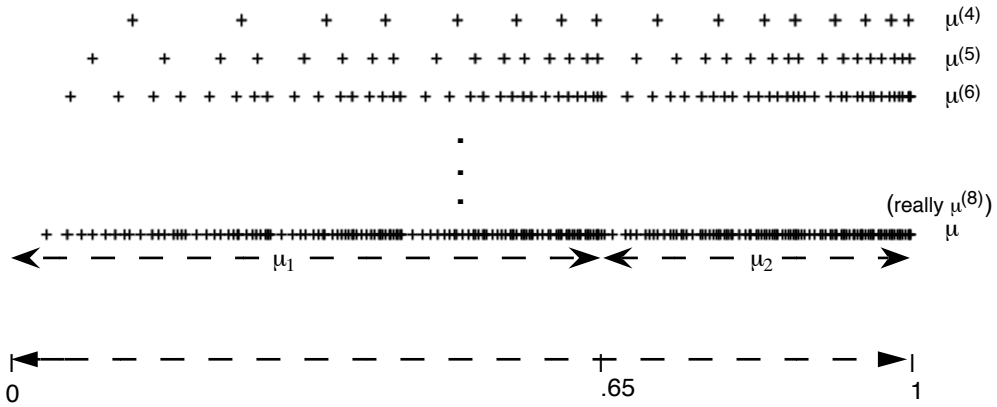
You should think of a *measure* μ in $\mathbb{R}^n$ as a "mass-distribution" and the measure $\mu(E)$ of a subset E of $\mathbb{R}^n$ as the total mass in that set. Simple examples are a unit mass concentrated at a point in $\mathbb{R}^n$ or a unit mass uniformly distributed along a curve in $\mathbb{R}^n$ having finite length. The total mass (i.e. the measure of $\mathbb{R}^n$) is called the *mass* of the measure.

A *fractal measure* μ is a measure in $\mathbb{R}^n$ with certain *scaling properties*. In the following diagram we have sketched various members in a sequence of approximations $\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(j)}, \ldots$ to a fractal measure μ on the interval $I = [0, 1] \subset \mathbb{R}$ (as noted in the previous footnote, $\mathbb{R}$ is the real line). Each cross represents a point mass of a certain magnitude. The number of point masses describing $\mu^{(j)}$ is denoted by $N^{(j)}$ and each such point mass is given

equal mass (or magnitude) $1/N^{(j)}$. If E is a set then the measure $\mu^{(j)}(E)$ of E is defined by

$$\begin{aligned}\mu^{(j)}(E) &= \text{total mass of all } j\text{th-level point masses in } E \\ &= (\text{total number of } j\text{th-level point masses in } E) \,/\, N^{(j)}\end{aligned}$$

(notice that $\mu^{(j)}$ has unit mass for every j). If j is large, then $\mu^{(j)}(E)$ is a very good approximation to $\mu(E)$ (at least for any "reasonable" set E and in particular for any interval).



In this example we see from the diagram that

$$\mu = \mu_1 + \mu_2, \tag{4}$$

where μ_1 is "supported" (c.f. Section 5.1) on the interval $[0, .65]$ and μ_2 is "supported" on $[.65, 1]$. The important point is that μ_1 can be obtained from μ by first "rescaling" by the factor .65 and then "reweighting" by the factor .5. A similar remark applies to μ_2 except that the rescaling factor is .35 and the reweighting factor is again .5. More precisely, and again significantly from our point of view, there are linear maps

$$S_i : \mathbb{R} \to \mathbb{R} \quad i = 1, 2 \ ,$$

where each S_i is a composition of a translation and scaling (with scaling factors .65 and .35), and weighting factors

$$\rho_1, \ \rho_2$$

(here both equal to .5) such that (4) can be written

$$\mu = \rho_1 S_1(\mu) + \rho_2 S_2(\mu) \ ^3 \tag{5}$$

In the present example the *support*[4] of the fractal measure is the entire interval $[0, 1]$. In many cases the support will be a more interesting fractal

[3]Think of $S_1(\mu)$ as the measure (mass-distribution) obtained by "pushing" μ forward with the map S_1. Similarly for $S_2(\mu)$. The measures $S_1(\mu)$ and $S_2(\mu)$ both have total mass one, as does μ, and so if we want equality to hold in (5) it is necessary to reweight by factors ρ_1 and ρ_2 whose sum is one.

[4]See Section (5.1) for the definition.

set. A fractal measure carries more information than its supporting set. *We can often usefully think of a fractal measure as a fractal set together with a grey scale, or weighting, at each point of the set.*

Another fractal measure, the *Koch measure*, can be constructed in a manner similar to the construction of the Koch curve. Thus (see the diagrams for the Koch curve in Section 2.1) we take $\mu^{(1)}$ to be a measure uniformly distributed along $K^{(1)}$ and with total mass one, so that the mass of each of the four line segments is 1/4. Similarly $\mu^{(2)}$ is uniformly distributed along $K^{(2)}$ again with total mass one, so that the mass of each of the 16 line segments is now 1/16. Likewise for $\mu^{(j)}$ with $j > 2$. Then μ is the limit, in a natural sense that can be made precise, of the sequence $\mu^{(1)}, \mu^{(2)}, \ldots, \mu^{(j)}, \ldots$. In this case

$$\mu = \rho_1 S_1(\mu) + \rho_2 S_2(\mu) + \rho_3 S_3(\mu) + \rho_4 S_4(\mu) \tag{6}$$

where each S_i is a composition of a translation, rotation and scaling (with scaling factor 1/3) and the scaling factors ρ_i all equal 1/4.

A fractal set or measure is often simply called a *fractal*. We also sometimes instead use the terminology *deterministic fractal* (*set, measure*) to distinguish the present notions from the *random* (or *statistical* or *non-deterministic*) versions which follow.

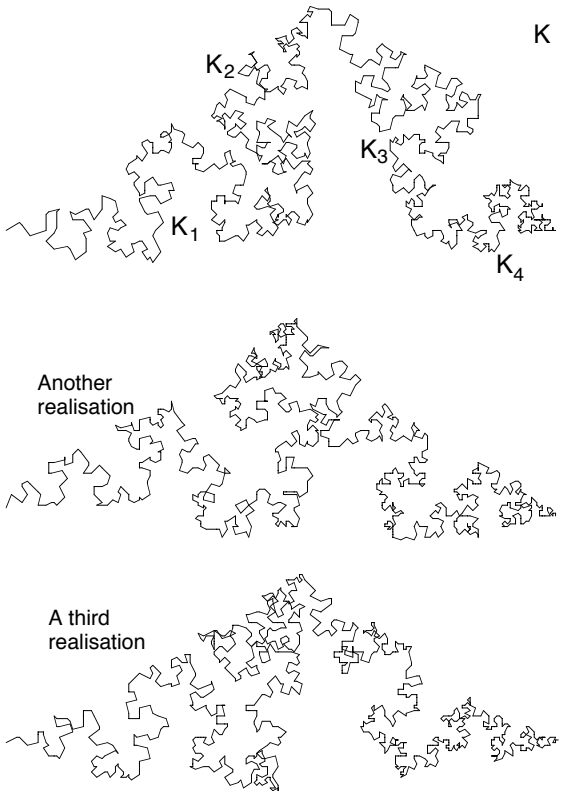
2.3 The random Koch curve

A *random fractal set* can be thought of as a set generated according to some probability distribution $\mathcal{K}$ on sets, analogous to the manner in which a random number is generated by a probability distribution on numbers. The probability distribution $\mathcal{K}$ will be required to have a certain type of *scaling property*. When we sketch a random fractal set we are in fact sketching a particular realisation of $\mathcal{K}$.

What is most important is the probability distribution $\mathcal{K}$ itself, rather than any particular realisation. In fact, it is more precise to call $\mathcal{K}$ the random fractal set, and to distinguish this from particular *realisations* of $\mathcal{K}$.[5]

In the following diagrams we give three realisations of the same random fractal, the *random Koch curve*.

[5]It is perhaps worth remarking that a particular realisation of a random number gives little, or no, idea of the underlying probability distribution. In contrast, because of scaling properties, even a single realisation of a random fractal gives a lot of information about the underlying distribution.



For each realisation K of $\mathcal{K}$ we have

$$K = K_1 \cup K_2 \cup K_3 \cup K_4,$$

where each K_i looks, "statistically" or "on the average", like a re-scaled version of K.

More precisely, and analogously to the deterministic case, we write

$$K \simeq S_1(K^{(1)}) \cup S_2(K^{(2)}) \cup S_3(K^{(3)}) \cup S_4(K^{(4)}). \tag{7}$$

Here $K^{(1)}, \ldots, K^{(4)}$ are chosen independently and at random via $\mathcal{K}$[6] and the single 4-tuple of maps (S_1, S_2, S_3, S_4) is chosen independently of the $K^{(i)}$ and at random via some probability distribution $\mathcal{S}$. Then the previous equation says that the probability distribution $\mathcal{K}$ on compact sets K is the same as the probability distribution on compact sets given by the right side of (7). In other words, (7) indicates equality in the sense of probability distributions.

Other physical examples are given by *Brownian motion*, the irregular oscillatory movement of microscopic particles in a limpid fluid.

2.4 The random Koch measure

A *random fractal measure* is a probability distribution $\mathcal{M}$ on measures which has a certain type of *scaling property*. An example can be constructed parallel

[6]Do not confuse $K^{(1)}, \ldots, K^{(4)}$ with the $K_1, \ldots, K_4$ in the diagram!

to the construction of the random Koch curve. Imagine a measure of unit mass distributed "uniformly" along the various realisations of the Koch curve in the previous diagram. This will give three realisations μ of the same *random Koch measure* $\mathcal{M}$.

The relevant property is

$$\mu \simeq \rho_1 S_1(\mu^{(1)}) + \rho_2 S_2(\mu^{(2)}) + \rho_3 S_3(\mu^{(3)}) + \rho_4 S_4(\mu^{(4)}) \tag{8}$$

Here $\mu^{(1)}, \ldots, \mu^{(4)}$ are chosen independently and at random via $\mathcal{M}$ and the single 8-tuple of maps and weights $\Big((S_1, \ldots, S_4), (\rho_1, \ldots, \rho_4)\Big)$ is chosen independently of the $\mu^{(i)}$ via some probability distribution $\mathcal{S}$. The previous equation indicates equality in the sense of probability distributions.

3 Discussion of main properties

We discuss some of the main properties of fractals as they apply to the examples from the previous section.

3.1 Scaling laws and coding

From equation (3) we see that K is the union of two sets, each of which is a scaled version of K itself. We can interpret (3) as saying that K satisfies the *scaling law* (determined by) $\mathbf{T} = (T_1, T_2)$. Similarly, (2) can be interpreted as saying that K satisfies the scaling law $\mathbf{S}$ where $\mathbf{S} = (S_1, S_2, S_3, S_4)$.

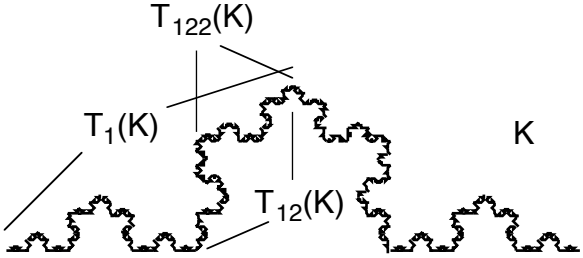
If we replace each occurrence of K on the right side of (3) by $T_1(K) \cup T_2(K)$ we obtain

$$\begin{aligned} K &= T_1\Big(T_1(K) \cup T_2(K)\Big) \cup T_2\Big(T_1(K) \cup T_2(K)\Big) \\ &= T_{11}(K) \cup T_{12}(K) \cup T_{21}(K) \cup T_{22}(K), \end{aligned} \tag{9}$$

where $T_{ij} = T_i \circ T_j$. Then (9) is in fact just the decomposition (2), even in the same order. Repeating the construction, we obtain

$$K = \bigcup_{1 \le \sigma_1, \sigma_2, \sigma_3 \le 2} T_{\sigma_1 \sigma_2 \sigma_3}(K),$$

where $T_{\sigma_1 \sigma_2 \sigma_3} = T_{\sigma_1} \circ T_{\sigma_2} \circ T_{\sigma_3}$. Thus K is the union of 8 sets, each of which is a rescaled version of K, the rescaling factor being $(1/\sqrt{3})^3$.



The previous construction can be iterated k times, writing K as the union of 2^k subsets each of which is a rescaled version of K with rescaling factor $(1/\sqrt{3})^k$:

$$K = \bigcup_{1\le\sigma_1,\ldots,\sigma_k\le 2} T_{\sigma_1\ldots\sigma_k}(K), \tag{10}$$

where $T_{\sigma_1\ldots\sigma_k} = T_{\sigma_1}\circ\cdots\circ T_{\sigma_k}$. For each infinite sequence $\sigma = \sigma_1, \sigma_2, \ldots, \sigma_k, \ldots$ the corresponding sequence

$$K \supseteq T_{\sigma_1}(K) \supseteq T_{\sigma_1\sigma_2}(K) \supseteq \cdots \supseteq T_{\sigma_1\ldots\sigma_k}(K) \supseteq \cdots \tag{11}$$

is a decreasing sequence of sets whose intersection is a single point. This point is assigned the *code* or *address* $\sigma = \sigma_1\sigma_2\ldots\sigma_k\ldots$. Every point in K has such a representation.

Similar remarks apply to fractal measures. Thus equation (5) can be interpreted as saying that μ satisfies the *scaling law* $\mathbf{S}$ where $\mathbf{S} = \Big((S_1, S_2), (\rho_1, \rho_2)\Big)$.

We can iterate (5) and obtain

$$\begin{aligned}\mu &= \rho_1 S_1\Big(\rho_1 S_1(\mu) + \rho_2 S_2(\mu)\Big) + \rho_2 S_2\Big(\rho_1 S_1(\mu) + \rho_2 S_2(\mu)\Big)\\ &= {\rho_1}^2 S_{11}(\mu) + \rho_1\rho_2 S_{12}(\mu) + \rho_2\rho_1 S_{21}(\mu) + {\rho_2}^2 S_{22}(\mu).\end{aligned}$$

Repeating this construction, we obtain

$$\begin{aligned}\mu &= \sum_{1\le\sigma_1,\sigma_2,\sigma_3\le 2} \rho_{\sigma_1\sigma_2\sigma_3} S_{\sigma_1\sigma_2\sigma_3}(\mu)\\ &= \sum_{1\le\sigma_1,\ldots,\sigma_k\le 2} \rho_{\sigma_1\ldots\sigma_k} S_{\sigma_1\ldots\sigma_k}(\mu),\end{aligned}$$

where $\rho_{\sigma_1\ldots\sigma_k} = \rho_{\sigma_1}\cdot\ldots\cdot\rho_{\sigma_k}$ and $S_{\sigma_1\ldots\sigma_k} = S_{\sigma_1}\circ\cdots\circ S_{\sigma_k}$.

Analogous ideas apply in the random case. Thus in Sections 2.3 and 2.4 the probability distributions $\mathcal{S}$ on the (S_1, S_2, S_3, S_4) and $\Big((S_1, \ldots, S_4), (\rho_1, \ldots, \rho_4)\Big)$ respectively, can be considered as *scaling laws* satisfied by the random Koch curve and the random Koch measure respectively.

3.2 Existence and uniqueness of fractals

A basic fact is that the scaling law satisfied by the Koch curve in fact characterises the Koch curve. This is a consequence of a general result which has had interesting applications and which we now discuss.

We first need a little notation. A map $S : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a *contraction map* if there exists some r satisfying $0 \le r < 1$ such that

$$|S(x_1) - S(x_2)| \le r|x_1 - x_2| \quad \text{for all } x_1, x_2 \in \mathbb{R}^n. \tag{12}$$

The number r is called a *contraction ratio* or *Lipschitz constant* for S.[7] The maps S_i and T_i in Section 2 are all contraction maps of a particularly simple kind in that *equality* holds in (12) with the appropriate choice of r.

If $\mathbf{S} = (S_1, \ldots, S_N)$ (where $N \geq 2$) is an N-tuple of contraction maps as above and K is a compact subset (c.f. Section 4.2) of $\mathbb{R}^n$, we say that K *satisfies the scaling law* $\mathbf{S}$ *if*

$$K = S_1(K) \cup \ldots \cup S_N(K). \tag{13}$$

Then we have the rather surprising result (Theorem 4.2) that *to each scaling law there exists exactly one compact set* (*fractal*) *satisfying that law.* Note however that different scaling laws may give the same fractal. For example, both $\mathbf{S} = (S_1, S_2, S_3, S_4)$ and $\mathbf{T} = (T_1, T_2)$ give the Koch curve; see Section 2.1.

Analogous results apply to scaling laws for fractal measures, and to scaling laws for random fractal sets and random fractal measures, c.f. Theorems 5.2, 6.4 and 6.8.

3.3 Approximating fractals

3.3.1 Deterministic approximations

Corresponding to the scaling law $\mathbf{S} = (S_1, \ldots, S_N)$ there is a *scaling operator*, also denoted by $\mathbf{S}$, such that for any compact set A the compact set $\mathbf{S}(A)$ is defined by

$$\mathbf{S}(A) = S_1(A) \cup \ldots \cup S_N(A).$$

Thus the fact (13) that K satisfies the scaling law $\mathbf{S}$ can be written

$$K = \mathbf{S}(K).$$

Beginning with any compact set A, sometimes called a *seed*, the scaling operator $\mathbf{S}$ can be iterated to obtain a sequence of sets

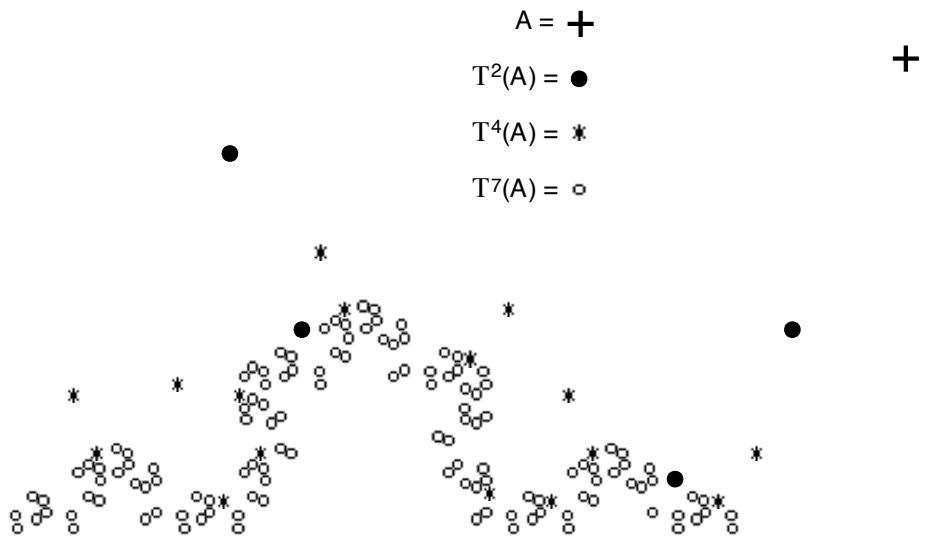
$$\mathbf{S}^1(A) = \mathbf{S}(A),\ \mathbf{S}^2(A) = \mathbf{S}(\mathbf{S}(A)),\ \mathbf{S}^3(A) = \mathbf{S}(\mathbf{S}(\mathbf{S}(A))), \ldots. \tag{14}$$

Then (Theorem 4.2) *the sequence* $\mathbf{S}^1(A)$, $\mathbf{S}^2(A), \ldots, \mathbf{S}^k(A), \ldots$ *converges to the unique compact set corresponding to the scaling law* $\mathbf{S}$. This is a constructive procedure that allows one to construct the fractal corresponding to the given scaling law to within any prescribed degree of accuracy.

In the diagram in Section 2.1 the sets $K^{(k)}$ equal $\mathbf{S}^k(A)$, where A is a horizontal line segment.

[7] Thus the distance between the *images* of any two points in space is less than the distance between the two points by a factor $r < 1$, where r is independent of the two chosen points.

In the following diagram another sequence of approximations is shown to the Koch curve. In this case we use the scaling operator $\mathbf{T} = (T_1, T_2)$ as in (3). Moreover, A contains a single point x and so $\mathbf{T}^k(A)$ consists of all points of the form $T_{\sigma_1 \ldots \sigma_k}(x)$.



Analogous results apply to fractal measures. In this case a *scaling law* is a $2N$-tuple

$$\mathbf{S} = \Big((S_1, \ldots, S_N), (\rho_1, \ldots, \rho_N)\Big)$$

where $N \geq 2$ is an integer, $S_1, \ldots, S_N : \mathbb{R}^n \to \mathbb{R}^n$ are contraction maps, and $\rho_1, \ldots, \rho_N$ are positive real numbers such that $\rho_1 + \cdots + \rho_N = 1$. Then we say μ satisfies the scaling law $\mathbf{S}$ if

$$\mu = \rho_1 S_1(\mu) + \cdots + \rho_N S_N(\mu). \tag{15}$$

The result (Theorem 5.2) is that *there is exactly one unit mass fractal which satisfies the scaling law* $\mathbf{S}$.

There is a corresponding *scaling operator* also denoted by $\mathbf{S}$ and defined by

$$\mathbf{S}(\nu) = \rho_1 S_1(\nu) + \cdots + \rho_N S_N(\nu),$$

where ν is any compactly supported unit mass measure in $\mathbb{R}^n$. Thus the fact (15) that μ satisfies the scaling law $\mathbf{S}$ can be conveniently written

$$\mu = \mathbf{S}(\mu).$$

Beginning from a (finite mass) measure ν we can iterate the scaling operator $\mathbf{S}$ to obtain a sequence of measures

$$\mathbf{S}^1(\nu) = \mathbf{S}(\nu),\ \mathbf{S}^2(\nu) = \mathbf{S}(\mathbf{S}(\nu)),\ \mathbf{S}^3(\nu) = \mathbf{S}(\mathbf{S}(\mathbf{S}(\nu))), \ldots. \tag{16}$$

Then (Theorem 5.2 again) *for any unit mass measure* ν *the sequence* $\mathbf{S}^1(\nu)$, $\mathbf{S}^2(\nu), \ldots, \mathbf{S}^k(\nu), \ldots$ *converges to the unique unit mass measure corresponding to the scaling law* $\mathbf{S}$.

Once again, analogous results to the preceding apply to the random case.

3.3.2 Random approximations

In (6) we saw that the Koch curve satisfies

$$\mu = \frac{1}{4}S_1(\mu) + \frac{1}{4}S_2(\mu) + \frac{1}{4}S_3(\mu) + \frac{1}{4}S_4(\mu),$$

where the S_i are certain maps $S_i : \mathbb{R}^2 \to \mathbb{R}^2$.

This leads to the following procedure, a *random* method to construct the *deterministic* Koch curve and measure.

(0) Begin with an arbitrary point (*seed*) $x_0 \in \mathbb{R}^2$.

(1) Choose S_{σ_1} from $(S_1, \ldots, S_4)$, where the probability of choice of any particular S_i is $1/4$. Let $x_1 = S_{\sigma_1}(x_0)$.

(2) Choose S_{σ_2} from $(S_1, \ldots, S_4)$ independently of S_{σ_1}, where again the probability of choosing any particular S_i is $1/4$. Let $x_2 = S_{\sigma_2}(x_1)$.

...

(k) Choose S_{σ_k} similarly and independently of the $S_{\sigma_1}, \ldots, S_{\sigma_{k-1}}$. Let $x_k = S_{\sigma_k}(x_{k-1})$.

...

In this way an *orbit* of points

$$x_0,\ x_1 = S_{\sigma_1}(x_0),\ x_2 = S_{\sigma_2}(x_1), \ldots,\ x_k = S_{\sigma_k}(x_{k-1}), \ldots$$

is constructed. Then with probability one (c.f. Section 6.1) the orbit will come arbitrarily close to every point in the Koch curve K. Moreover, see [E], if μ is the Koch measure then the measure $\mu(B)$ of any subset $B \subseteq \mathbb{R}^n$ is given by

$$\mu(B) = \lim_{k\to\infty} \frac{\mathcal{N}(k, B)}{k+1},$$ [8]

where

$$\mathcal{N}(k, B) = \text{number of points in } \{x_0, x_1, x_2, \ldots, x_k\} \cap B.$$

Thus the *relative visitation frequency* of the finite orbit $x_0, x_1, x_2, \ldots, x_k$ in the set B is a good approximation to $\mu(B)$ for k large.

In computer simulations using either the algorithm here or from the previous section, as successive points are plotted on the screen, increasingly accurate approximations to the Koch curve will be constructed. The density of points will indicate the distribution of mass according to the Koch measure. The Koch measure as we have defined it is in a certain sense distributed

[8] Technical condition: one needs to assume that $\mu(\partial B) = 0$, where ∂B is the boundary of B in the usual topological sense.

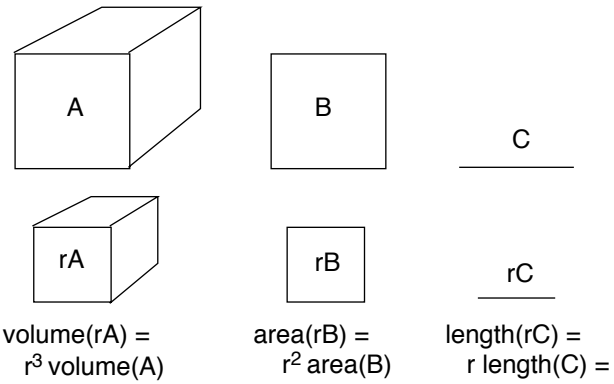
uniformly along the Koch curve, but in other cases such as the fractal measure on the line in Section 2.2, the mass is not uniformly distributed. The density of points constructed via appropriate modifications of either algorithm will closely approximate the mass distribution of the corresponding fractal measure.

3.4 Dimension of fractals

If S satisfies (12) with "$\leq$" there replaced by "$=$", we say S is a *similitude* with scaling factor r. There is a standard classical notion of "D-dimensional" volume, $D\text{-vol}\,(E)$, for certain sets $E \subset \mathbb{R}^n$. For $D = 1, 2, 3$ this gives the length, area, and usual volume, respectively. A standard property is

$$D\text{-vol}\,(S(E)) = r^D D\text{-vol}\,(E)$$

if S is a similitude with scaling factor r.



Now suppose

$$E = S_1(E) \cup \ldots \cup S_N(E)$$

where $S_1, \ldots, S_N$ are similitudes with scaling factors $r_1, \ldots, r_N$, and assume the $S_i(E)$ intersect each other on sets of lower dimension than D.[9] Then

$$D\text{-vol}\,(E) = {r_1}^D\, D\text{-vol}\,(E) + \cdots + {r_N}^D D\text{-vol}\,(E).$$

Thus we obtain the relation

$${r_1}^D + \cdots + {r_N}^D = 1, \tag{17}$$

assuming $0 < D\text{-vol}\,(E) < \infty$.

Motivated by this, from (2) we expect that the "dimension" D of the Koch curve K should satisfy

$$4\left(\frac{1}{3}\right)^D = 1. \tag{18}$$

[9]More precisely, the D-dimensional volume of the intersections is zero. In particular the intersections may be empty.

This gives $D = \log 4/\log 3 = 1.2615\ldots$ for the dimension of the Koch curve. If we use (3) we obtain

$$2\left(\frac{1}{\sqrt{3}}\right)^D = 1,$$

which leads to the same value for D.

There is a notion of dimension, the *Hausdorff dimension*, which assigns to any subset of $\mathbb{R}^n$ a real number in the range $[0, n]$. Then (Theorem 4.4) *if the scaling property* **S** *satisfies a certain "Open Set Condition", the corresponding fractal K will have Hausdorff dimension D given by* (17)[10]. In particular, the Hausdorff dimension of the Koch curve is indeed $\log 4/\log 3$. Analogous results hold for random fractal sets (Section 6).

Note that the dimension of the Koch curve is greater than one, and in particular the Koch curve will have "infinite length" in the sense that its one dimensional measure is infinite (c.f. Section 4.6). Any realisation (more precisely, with probability one as discussed in Section 6) of the random Koch curve will also have dimension greater than one.

A coastline can often be modelled by a kind of random fractal, c.f. [M]. The mathematical model will have dimension greater than one and hence infinite length, so in this sense we say that the coastline itself has dimension greater than one and has infinite length.

4 Fractal sets

In this and subsequent sections we develop the important ideas involved in the proofs of the main properties of fractals as discussed in Section 3. The mathematical level is accordingly a little higher, but we have attempted to keep the details to a minimum and hope that the ideas will be clear even to those to whom much of the background is new.

4.1 Contraction Maps

Euclidean n-space is denoted by $\mathbb{R}^n$. We have already introduced the notion of a *contraction map* on $\mathbb{R}^n$ and the corresponding *contraction ratio*. Recall also that S is a *similitude* with scaling factor r if S satisfies (12) with "$\leq$" there replaced by "$=$".

[10]Note that if we define $f(x) = {r_1}^x + \cdots + {r_N}^x$ then $f(0) = N$, $f(x) \to 0$ as $x \to \infty$, and $f(x)$ is strictly decreasing for $x \in \mathbb{R}$. It follows that there is indeed a unique value of D for which (17) is satisfied.

4.2 Metric spaces

A *metric space* (X, d) is a set X together with a notion of *distance* $d(a, b)$ between any two members $a, b \in X$ which satisfies the following properties:

1. $d(a, b) \geq 0$ and $d(a, b) = 0$ precisely when $a = b$,
2. $d(a, b) \leq d(a, c) + d(c, b)$,
3. $d(a, b) = d(b, a)$.

One also calls d a *metric*.

A simple example of a metric space is where $X = \mathbb{R}^n$ (e.g. $\mathbb{R}^2$) and $d(a, b)$ is the usual distance between the two points a and b. Another example is where $X = D$ is the *unit disc* consisting of points in the plane whose distance from the origin is *strictly* less than one, and $d(a, b)$ is again the usual distance between the points a and b (see the next diagram). We will soon meet some other very important, but more complicated examples.

A metric space (X, d) is *complete* if whenever a sequence $x_1, x_2, \ldots, x_k, \ldots$ of members of X has the property that

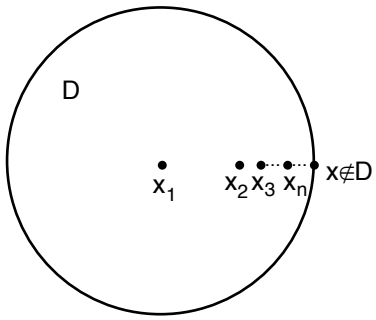
$$d(x_j, x_k) \to 0 \quad \text{as} \quad j, k \to \infty$$

then $d(x_k, x) \to 0$ for some $x \in X$. In this case one says x_k *converges to* x and writes $x_k \to x$ as $k \to \infty$.

The metric space $(\mathbb{R}^n, d)$ is complete but the metric space (D, d) discussed above is not complete. In the following diagram we show the sequence of points $x_k = (1 - 1/k, 0)$. We see that

$$d(x_j, x_k) = \left| \frac{1}{j} - \frac{1}{k} \right| \to 0 \quad \text{as} \quad j, k \to \infty.$$

But the limit $x = (1, 0)$ of this sequence is not in D.



As in (12), a map $F : X \to X$ is said to be a *contraction map* if there exists some $r \in (0, 1)$ such that

$$d(F(x_1), F(x_2)) \leq r d(x_1, x_2) \quad \text{for all } x_1, x_2 \in X.$$

The *Contraction Mapping Principle* asserts that: *if (X, d) is a complete metric space and $F : X \to X$ is a contraction map, then there exists exactly one $a \in X$, called the fixed point of F, such that $F(a) = a$. Moreover, for any $x \in X$, the sequence*

$$x,\ F^1(x) = F(x),\ F^2(x) = F(F(x)), F^3(x) = F(F(F(x))), \ldots$$

converges to a. Also, if $d(x, F(x)) \leq \varepsilon$ and r is a contraction ratio for F, then $d(x, a) \leq \varepsilon/(1 - r)$.

The point to the last claim is that if x is near $F(x)$ then in fact x is near the fixed point a. The proof of this claim is easy, since each application of F decreases the distance between points by the factor r. Thus

$$\begin{aligned} d(x, F^k(x)) &\leq d(x, F(x)) + d(F(x), F^2(x)) + \cdots + d(F^{k-1}(x), F^k(x)) \\ &\leq d(x, F(x))\,(1 + r + \cdots + r^{k-1}) \\ &\leq \varepsilon\,\frac{1 - r^k}{1 - r}. \end{aligned}$$

Now let $k \to \infty$ to see that $d(x, a) \leq \varepsilon/(1 - r)$.

4.3 The metric space $(\mathcal{C}, d_{\mathcal{H}})$ of compact subsets of $\mathbb{R}^n$

A set $A \subset \mathbb{R}^n$ is *closed* if whenever $x_1, x_2, \ldots, x_k, \ldots$ is a sequence of points from A and $x_k \to x$ for some $x \in \mathbb{R}^n$, then $x \in A$ (in other words, A contains all its limit points). A set $A \subset \mathbb{R}^n$ is *bounded* if the distance between any two points in A is less than some fixed finite number b (where b depends of course on A, but not on the particular points in A).

A *compact* subset of $\mathbb{R}^n$ is a closed and bounded subset. The set of all such compact subsets is denoted by $\mathcal{C}$. There is an important notion of distance between two members of $\mathcal{C}$, called the *Hausdorff distance* or *Hausdorff metric* and denoted by $d_{\mathcal{H}}$.

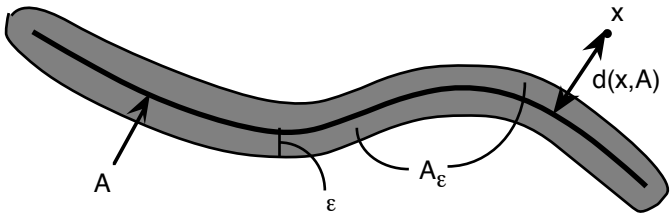
Before introducing the Hausdorff distance we need some preliminary definitions.

1. If $x \in \mathbb{R}^n$ and $A \in \mathcal{C}$ we define the distance $d(x, A)$ between the *point* x and the *set* A by

   $$d(x, A) = \min\{\, d(x, a) : a \in A \,\},$$ [11]

   where $d(x, a)$ is the usual distance between points.

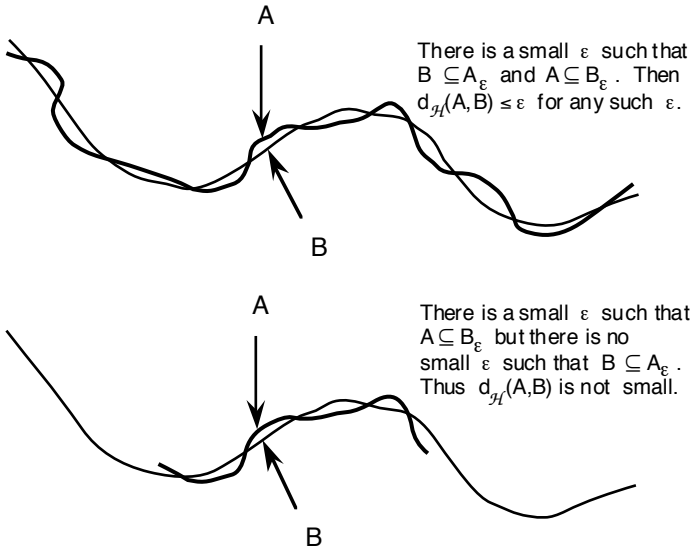
[11] Technical aside: one usually writes "inf" for infimum, instead of "min" for minimum, but this is equivalent in the present setting.



2. For a set $A \in \mathcal{C}$ and a real number $\varepsilon \geq 0$ define the *ε-enlargement* by

$$A_\varepsilon = \{\, x : d(x, A) \leq \varepsilon \,\}.$$

Thus $A = A_0$, and $A_{\varepsilon_1} \subseteq A_{\varepsilon_2}$ if $\varepsilon_1 \leq \varepsilon_2$.



3. For two sets $A, B \in \mathcal{C}$ define the *Hausdorff distance* between A and B by

$$d_{\mathcal{H}}(A, B) = \min\{\, \varepsilon : A \subseteq B_\varepsilon \text{ and } B \subseteq A_\varepsilon \,\}.^{12}$$

Then one can show that $(\mathcal{C}, d_{\mathcal{H}})$ *is a complete metric space.*

Two important properties of the Hausdorff metric are

$$d_{\mathcal{H}}\left(\bigcup_{i=1}^{N} A_i, \bigcup_{i=1}^{N} B_i\right) \leq \max_{1 \leq i \leq N} d_{\mathcal{H}}(A_i, B_i); \tag{19}$$

$$d_{\mathcal{H}}(F(A), F(B)) \leq r d_{\mathcal{H}}(A, B), \tag{20}$$

where in the second inequality $F : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction map with contraction ratio r.[13]

Finally, if $A \in \mathcal{C}$ then the *diameter* of A is defined by

$$\operatorname{diam}(A) = \max_{x,y \in A} d(x, y).$$

[12] As in the previous footnote, one could equivalently write "inf" instead of "min".

[13] One does not require $r \leq 1$ for this result.

4.4 Existence, uniqueness and approximation

Definition 4.1 A *scaling law for sets* is an N-tuple $\mathbf{S}$ of contraction maps $(S_1, \ldots, S_N)$ defined on $\mathbb{R}^n$. The corresponding *scaling operator* $\mathbf{S}$ is defined on compact sets by $\mathbf{S}(A) = \bigcup_{i=1}^{N} S_i(A)$. We say K *satisfies the scaling law* $\mathbf{S}$ *if* $K = \mathbf{S}(K)$.

We can now prove the *Existence and Uniqueness Result* for fractal *sets* discussed previously, see [H] 3.2(1).

Theorem 4.2 *There is a unique compact set K satisfying a given scaling law $\mathbf{S}$. If A is any compact subset of $\mathbb{R}^n$ then $\mathbf{S}^k(A)$[14] converges to K in the Hausdorff metric as $k \to \infty$.*

PROOF: Let $\mathbf{S} = (S_1, \ldots, S_N)$ and let r be the maximum contraction ratio of the S_i's.

For any $A_1, A_2 \in \mathcal{C}$,

$$\begin{aligned} d_{\mathcal{H}}(\mathbf{S}(A_1), \mathbf{S}(A_2)) &= d_{\mathcal{H}}\left(\bigcup_{i=1}^{N} S_i(A_1), \bigcup_{i=1}^{N} S_i(A_2)\right) \\ &\leq \max_{1 \leq i \leq N} d_{\mathcal{H}}(S_i(A_1), S_i(A_2)) \\ &\leq r d_{\mathcal{H}}(A_1, A_2), \end{aligned}$$

by (19) and (20). It follows that the scaling operator $\mathbf{S}: \mathcal{C} \to \mathcal{C}$ is a contraction map.[15]

The conclusion of the theorem now follows from the completeness of $(\mathcal{C}, d_{\mathcal{H}})$ and the Contraction Mapping Principle. ∎

Remark: The set K satisfying the scaling law $\mathcal{S}$ is usually called a *fractal (set)*.

A simple corollary of the Theorem is that if some set A satisfies the inequality $d_{\mathcal{H}}(A, \mathbf{S}(A)) \leq \varepsilon$ then $d_{\mathcal{H}}(A, K) \leq \varepsilon/(1-r)$, where K is the compact set satisfying the scaling law $\mathbf{S}$. This follows from the last part of the Contraction Mapping Principle in Section 4.2. Barnsley calls this the *Collage Theorem* since it implies that if A can be covered by scaled copies of itself to within some small Hausdorff distance ε then there is a corresponding "nearby" fractal K whose distance from A is at most $\varepsilon/(1-r)$.

The previous theorem also justifies the deterministic method for approximating fractals discussed in Section 3.3.1 in the context of the Koch curve.

[14] See (14) for notation.

[15] Do not confuse this with the fact that the S_i are contraction maps on $\mathbb{R}^n$.

4.5 Code space

Let K be the fractal set invariant under the scaling law $\mathbf{S} = (S_1, \ldots, S_N)$. Write

$$S_{\sigma_1 \ldots \sigma_k} = S_{\sigma_1} \circ \cdots \circ S_{\sigma_k}. \tag{21}$$

Then for any k, exactly as in (10) and (11), we can decompose K into "smaller and smaller" pieces:

$$K = \bigcup_{1 \le \sigma_1, \ldots, \sigma_k \le N} S_{\sigma_1 \ldots \sigma_k}(K), \tag{22}$$

$$K \supseteq S_{\sigma_1}(K) \supseteq S_{\sigma_1 \sigma_2}(K) \supseteq \cdots \supseteq S_{\sigma_1 \ldots \sigma_k}(K) \supseteq \cdots. \tag{23}$$

If r is the maximum contraction ratio of $S_1, \ldots, S_N$, then it is easy to see that

$$\operatorname{diam} S_{\sigma_1 \ldots \sigma_k}(K) \le r^k \operatorname{diam}(K).$$

Hence $\operatorname{diam} S_{\sigma_1 \ldots \sigma_k}(K) \to 0$ as $k \to \infty$. Thus there is a unique point which belongs to every member of the sequence (23). We give this point the *code* or *address*

$$\sigma = \sigma_1 \sigma_2 \ldots \sigma_k \ldots$$

and denote the point by

$$\Pi(\sigma). \tag{24}$$

Every point in K has a code, but the code is not necessarily unique. For example, in the diagram in Section 3.1 in which the Koch curve is generated by two similitudes T_1 and T_2, the top-most point has the two addresses $1222\ldots$ and $2111\ldots$. On the other hand, if the generating maps $S_1, \ldots, S_N$ satisfy $S_i(K) \cap S_j(K) = \emptyset$ for $i \ne j$, then the N^k sets in (22) are also mutually disjoint and it follows that every point in K does have a unique address.

The set C_N of all codes σ with each $\sigma_k \in \{1, \ldots, N\}$ is called *code space*. It is a metric space with metric

$$\delta(\sigma, \tau) = \sum_{k=1}^{\infty} \frac{|\sigma_k - \tau_k|}{N^k}.$$

Then one has ([H], Theorem 3.1):

Theorem 4.3 *For any $x \in \mathbb{R}^n$*

$$S_{\sigma_1 \ldots \sigma_k}(x) \to \Pi(\sigma) \in K \text{ as } k \to \infty.$$

Moreover, the map $\Pi : C_N \to K$ is a continuous map onto K.

PROOF: To prove the first result first note that if r is the maximum contraction ratio of the $S_1, \ldots, S_N$ then

$$d(S_{\sigma_1 \ldots \sigma_k}(x), S_{\sigma_1 \ldots \sigma_k}(K)) \le r^k d(x, K) \to 0 \text{ as } k \to \infty.$$

Since

$$\Pi(\sigma) \in S_{\sigma_1 \dots \sigma_k}(K)$$

and

$$\operatorname{diam} S_{\sigma_1 \dots \sigma_k}(K) \to 0 \text{ as } k \to \infty$$

the first claim follows.

For the second claim suppose $\sigma, \tau \in C_N$, that $\sigma_i = \tau_i$ for $i = 1, \dots, k$ and $\sigma_{k+1} \neq \tau_{k+1}$. Then

$$\delta(\sigma, \tau) \geq 1/N^k. \tag{25}$$

On the other hand

$$\Pi(\sigma), \Pi(\tau) \in S_{\sigma_1 \dots \sigma_k}(K) = S_{\tau_1 \dots \tau_k}(K),$$

and so

$$d(\Pi(\sigma), \Pi(\tau)) \leq \operatorname{diam} S_{\sigma_1 \dots \sigma_k}(K) \leq r^k \operatorname{diam}(K). \tag{26}$$

From (25), $\delta(\sigma, \tau)$ is small implies k is large, and from (26) this implies $d(\Pi(\sigma), \Pi(\tau))$ is small. Thus Π is continuous. ∎

One can also give a direct proof of Theorem 4.2 by arguing as in the previous proof, see [H] Theorem 3.1.

4.6 Dimension

4.6.1 Hausdorff dimension

Assume $E \subseteq \mathbb{R}^n$ and $d \geq 0$. If $\delta > 0$ we define the *d-dimensional δ-approximating measure* by

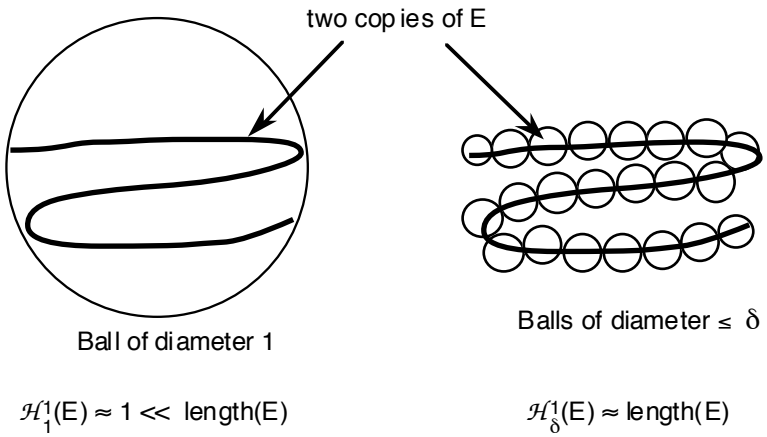
$$\mathcal{H}^d_\delta(E) = \inf \sum_{i \geq 1} (\operatorname{diam} B_i)^d, \tag{27}$$

where the infimum is taken over all (finite or infinite) sequences of balls[16] $(B_i)_{i \geq 1}$ such that (i) $\operatorname{diam} B_i \leq \delta$ and (ii) $E \subseteq \bigcup_{i \geq 1} B_i$.

If $\delta_1 \leq \delta_2$ then $\mathcal{H}^d_{\delta_1}(E) \geq \mathcal{H}^d_{\delta_2}(E)$ since there are *fewer* allowable families of balls for δ_1 than for δ_2. Thus $\mathcal{H}^d_\delta(E)$ is increasing as $\delta \to 0$. We define the *d-dimensional Hausdorff measure of E* by

$$\mathcal{H}^d(E) = \lim_{\delta \to 0} \mathcal{H}^d_\delta(E). \tag{28}$$

[16]Technical aside: One usually allows arbitrary subsets of $\mathbb{R}^n$. This will make no difference to the later definition of the Hausdorff dimension $\overline{d}$ but may change the value of $\mathcal{H}^{\overline{d}}(E)$.



Clearly,

$$0 \leq \mathcal{H}^d(E) \leq +\infty.$$

It is not too hard to show that there exists a *unique* $\overline{d} \in [0, n]$ such that

$$\mathcal{H}^d(E) = \begin{cases} \infty & \text{if } d < \overline{d} \\ 0 & \text{if } d > \overline{d}. \end{cases} \tag{29}$$

This unique $\overline{d}$ is called the *Hausdorff dimension* of E.

It is possible to construct examples of sets E where $\mathcal{H}^{\overline{d}}(E) = a$ for any given a satisfying $0 \leq a \leq \infty$. But if $\mathcal{H}^d(E)$ is *finite and non-zero* for some d then $\overline{d} = d$. If $\mathcal{H}^d(E) = 0$ then $\overline{d} \leq d$, if $\mathcal{H}^d(E) = \infty$ then $\overline{d} \geq d$.

In case E is a smooth curve in $\mathbb{R}^2$ or $\mathbb{R}^3$ then $\overline{d} = 1$ and $\mathcal{H}^1(E)$ is the length of E. In case E is a smooth surface in $\mathbb{R}^3$ then $\overline{d} = 2$ and $\mathcal{H}^2(E)$ is the area of E. For the fractal sets we are considering, $\overline{d}$ will not usually be an integer.

4.6.2 Scaling dimension and the Open Set Condition

Assume $\mathbf{S} = (S_1, \ldots, S_N)$ where the S_i are *similitudes* with scaling factors $r_1, \ldots, r_N$ respectively. Then the *scaling dimension* corresponding to $\mathbf{S}$ is defined to be the unique value of D such that

$$r_1^D + \cdots + r_N^D = 1, \tag{30}$$

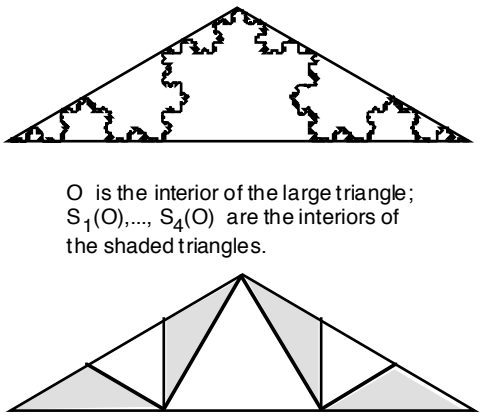
c.f. Section 3.4 and the final footnote there. Under certain conditions (Theorem 4.4) this scaling dimension will equal the Hausdorff dimension of the associated fractal set.

For this reason, we say that $\mathbf{S}$ satisfies an *Open Set Condition* if there exists a non-empty open set $O \subseteq \mathbb{R}^n$ such that

1. $\bigcup_{i=1}^N S_i(O) \subseteq O$,
2. $S_i(O)$ and $S_j(O)$ are disjoint if $i \neq j$.

The scaling operator $\mathbf{S} = (S_1, \ldots, S_4)$ for the Koch curve K satisfies the Open Set Condition, where O is the interior of the large triangle in the following diagram. Notice, on the other hand, that the sets $S_1(K), \ldots, S_4(K)$ are not disjoint.

The following result showing the equality of scaling dimension and Hausdorff dimension if the Open Set Condition is satisfied, is due to [Mo].



Theorem 4.4 *If* $\mathbf{S}$ *consists of similitudes and satisfies the Open Set Condition then the scaling dimension* D *of* $\mathbf{S}$ *equals the Hausdorff dimension of the associated fractal set* K*. Moreover,* $\mathcal{H}^D(K)$ *is finite and non-zero.*

PROOF: We will prove the easier result that the Hausdorff dimension is less than or equal to the scaling dimension D. This does *not* require the Open Set Condition, and so holds much more generally.

Using (30) first note that

$$\begin{aligned} 1 &= r_1^D + \cdots + r_N^D \\ &= r_1^D \left(r_1^D + \cdots + r_N^D\right) + \cdots + r_N^D \left(r_1^D + \cdots + r_N^D\right) \\ &= \sum_{i,j=1}^{N} r_i^D r_j^D. \end{aligned}$$

Similarly,

$$\sum_{i_1,\ldots,i_k=1}^{N} r_{i_1}^D \cdot \ldots \cdot r_{i_k}^D = 1 \tag{31}$$

for any k.

Now choose a ball B, with diameter b (say), such that $K \subset B$. From (22),

$$\begin{aligned} K &= \bigcup_{1\le\sigma_1,\ldots,\sigma_k\le N} S_{\sigma_1\ldots\sigma_k}(K) \\ &\subseteq \bigcup_{1\le\sigma_1,\ldots,\sigma_k\le N} S_{\sigma_1\ldots\sigma_k}(B). \end{aligned} \tag{32}$$

Notice that $S_{\sigma_1 \ldots \sigma_k}(B)$ is a ball of diameter $r_1 \cdot \ldots \cdot r_k \cdot b$.

Take any $\delta > 0$. By choosing k sufficiently large we may assume $r_1 \cdot \ldots \cdot r_k \cdot b \leq \delta$ (if $r = \max\{r_1, \ldots, r_n\}$ is the maximum contraction ratio just choose k so $r^k b \leq \delta$). It follows from (32) and (31) that K is covered by a finite number of balls of diameter $\leq \delta$ such that

$$\begin{aligned} \sum (\text{diameter of the balls})^D &= \sum_{i_1,\ldots,i_k=1}^{N} (r_{i_1} \cdot \ldots \cdot r_{i_k} \cdot b)^D \\ &= b^D \sum_{i_1,\ldots,i_k=1}^{N} r_{i_1}^D \cdot \ldots \cdot r_{i_k}^D \\ &= b^D. \end{aligned}$$

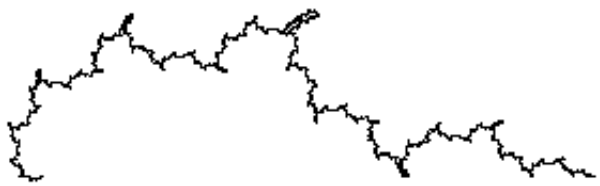
Hence $\mathcal{H}^D_\delta(K) \leq b^D$ from (27), for all $\delta > 0$. But then $\mathcal{H}^D(K) \leq b^D$ from (28). It follows from the remarks following (29) that if $\overline{d}$ is the Hausdorff dimension of K then $\overline{d} \leq D$.

The proof that $\overline{d} \geq D$ (and hence that $\overline{d} = D$) is more involved. The idea is to use the Open Set Condition to cover K by small balls that do not overlap "too much". We refer the reader to [Mo], or to [H] Theorem 5.3 where a result involving densities is also proved. The fact that $\mathcal{H}^D(K)$ is finite and non-zero is also a consequence of the argument. ∎

4.7 Parameter space

An *affine* map $S : \mathbb{R}^n \to \mathbb{R}^n$ is a composition of a linear map and a translation. Such an affine map is completely described by means of an $n \times n$ matrix and an n vector, and hence by a finite number of parameters. Thus a scaling operator $\mathbf{S} = (S_1, \ldots, S_N)$ of affine maps can be thought of as a point in some finite dimensional *parameter space* $\mathbb{R}^m$, c.f. [H] §5.5 and [O]. Not all points in $\mathbb{R}^m$ will actually correspond to a family of *contraction* maps and so parameter space will in fact correspond to a certain (open) subset of $\mathbb{R}^m$. It is often natural to work with certain restricted families of scaling operators, in which case the dimension m may be considerably reduced.

It follows from Theorem 4.2 and (3) that the Koch curve is uniquely characterised by the scaling law $\mathbf{T} = \{T_1, T_2\}$, which in turn corresponds to a point in $\mathbb{R}^m$ where $m = (2 \times 2 + 2) + (2 \times 2 + 2) = 12$. If T_1 is replaced by the same affine map *except* that no reflection is performed, and if T_2 is left unchanged, then the following *Dragon* fractal is obtained.

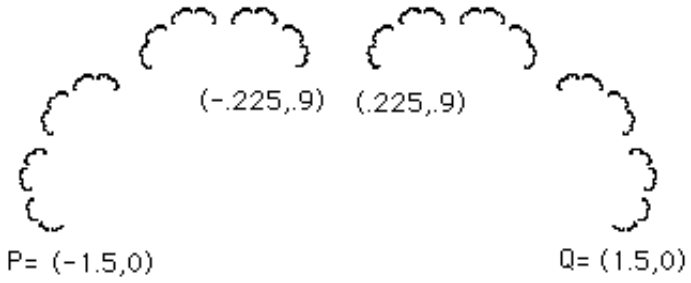


The same calculation as in (18) shows that the scaling dimension of the scaling law which gives the Dragon is $\log 4/\log 3$. Presumably the Hausdorff dimension of the Dragon is also $\log 4/\log 3$, but the Open Set Condition does not apply here.

If T_1 and T_2 are as for the Koch curve, except that no reflection is performed in either case, then the following *Brain* fractal is obtained.

The scaling dimension is again $\log 4/\log 3$, and the Hausdorff dimension is presumably the same.

If T_1 and T_2 are similar to the previous example except that now T_1 fixes $P = (-1.5, 0)$ and maps $Q = (1.5, 0)$ to $(-.225, .9)$ instead of to $(0, \sqrt{3}/2)$, and T_1 fixes Q and maps P to $(.225, .9)$ instead of to $(0, \sqrt{3}/2)$, then the following *Clouds* fractal is obtained. We can only show an approximation, and in fact the actual fractal in this case will in a certain precise sense be "totally disconnected".



The contraction ratios are here each $r = .5202\ldots$ and so the dimension of the Clouds is $\log 2/\log(1/r) = 1.0606\ldots$.

5 Fractal measures

5.1 Some notions from measure theory

A *finite measure*[17] ν on $\mathbb{R}^n$ assigns to every set $E \subseteq \mathbb{R}^n$ a positive real number such that (i) $\nu(\emptyset) = 0$ where $\emptyset$ is the empty set, (ii) if $E_1 \subseteq E_2$

[17] What we here call a finite measure is often called a *finite Radon outer measure*.

then $\nu(E_1) \leq \nu(E_2)$, and (iii) if $E \subseteq \bigcup_{i=1}^{\infty} E_i$ then $\nu(E) \leq \sum_{i=1}^{\infty} \nu(E_i)$. We also require that there be a class of subsets of $\mathbb{R}^n$, called the class of (ν-)measurable subsets, which includes the open and closed sets, is closed under complements and under finite and countably infinite intersections and unions, and satisfies $\nu(E) = \sum_{i=1}^{\infty} \nu(E_i)$ whenever $E = \bigcup_{i=1}^{\infty} E_i$ and the sets E_i are mutually disjoint and measurable. The *mass* of ν is $\nu(\mathbb{R}^n)$. If the mass of ν is one, we say that ν is a *unit mass* measure. The *integral* $\int f\, d\nu$ of various functions $f : \mathbb{R}^n \to \mathbb{R}$ is defined in such a way that, roughly speaking, "$\int f\, d\nu$ is a weighted sum of the values of f, where the weighting is done according to ν".

The *support* $\mathrm{spt}(\nu)$ of ν is the intersection of all closed sets whose complement has ν measure zero; and so $\mathrm{spt}(\nu)$ is the *smallest* closed set whose complement has ν measure zero. If $S : \mathbb{R}^n \to \mathbb{R}^n$ then the finite measure $S(\nu)$ is defined by $(S(\nu))(E) = \nu(S^{-1}(E))$; think of $S(\nu)$ as the measure ν "pushed forward" by the function S. If ρ is a positive number then the finite measure $\rho\nu$ is defined by $(\rho\nu)(E) = \rho \cdot \nu(E)$; think of the measure $\rho\nu$ as the measure ν reweighted by the factor ρ. Moreover, we have the important properties

$$\int f\, dS(\nu) = \int f \circ S\, d\nu \tag{33}$$

where $f \circ S$ is the usual composition of functions, and

$$\int f\, d(\rho\nu) = \rho \int f\, d\nu \tag{34}$$

where ρ is any positive number.

5.2 The metric space $(\mathcal{U}, d_U)$ of measures on $\mathbb{R}^n$ with unit mass and compact support

The set of all *unit mass measures*[18] *on* $\mathbb{R}^n$ *with compact support* is denoted by $\mathcal{U}$.[19]

For our purposes there is a useful notion of distance between two such measures, called the *Monge-Kantorovitch distance*, which is defined by

$$\begin{aligned} d_{\mathcal{MK}}(\mu,\nu) \;=\; \sup\Big\{ \Big| \int f\, d\mu - \int f\, d\nu \Big| \;\Big|\; & f : \mathbb{R}^n \to \mathbb{R} \text{ and} \\ & |f(x_1) - f(x_2)| \leq |x_1 - x_2| \text{ for all } x_1, x_2 \in \mathbb{R}^n \Big\} \end{aligned} \tag{35}$$

As a simple example, suppose P and Q are two point in $\mathbb{R}^2$. Let δ_P and δ_Q be unit measures concentrated at P and Q respectively—thus for $E \subseteq \mathbb{R}^n$,

[18] A unit mass measure μ is itself a probability distribution in a natural way, but we do not here think of μ in this way.

[19] If μ satisfies a scaling law as described in the next section, so does any multiple of μ. Thus there is no essential loss of generality in restricting to *unit mass* measures

$\delta_P(E) = 1$ if $P \in E$ and $\delta_P(E) = 0$ if $P \notin E$, and similarly for δ_Q and Q. Then with any f as in the definition of $d_{\mathcal{MK}}$, we see

$$\left| \int f \, d\delta_P - \int f \, d\delta_Q \right| = |f(P) - f(Q)| \leq |P - Q|.$$

Thus $d_{\mathcal{MK}}(\delta_P, \delta_Q) \leq |P - Q|$. On the other hand, it is easy to find a (in fact linear) function f as in the definition of $d_{\mathcal{MK}}$ such that $|\int f \, d\delta_P - \int f \, d\delta_Q| = |P - Q|$. It follows that $d_{\mathcal{MK}}(\delta_P, \delta_Q) = |P - Q|$, the usual distance between the *points* P and Q.

If X is a subset of $\mathbb{R}^n$, the set of measures in $\mathcal{U}$ whose support is a subset of X is denoted by $\mathcal{U}_X$. Although $(\mathcal{U}, d_{\mathcal{MK}})$ is not a complete metric space, it can be shown that *if $X \subset \mathbb{R}^n$ is compact then $(\mathcal{U}_X, d_{\mathcal{MK}})$ is a complete metric space.*

5.3 Existence, uniqueness and approximation

Definition 5.1 A $2N$-tuple $\mathbf{S} = \Big((S_1, \ldots, S_N), (\rho_1, \ldots, \rho_N)\Big)$ of contraction maps $S_i : \mathbb{R}^n \to \mathbb{R}^n$ and positive numbers ρ_i such that $\rho_1 + \cdots + \rho_N = 1$ is called a *scaling law for measures*. The corresponding *scaling operator* $\mathbf{S}$ is defined on measures in $\mathcal{U}$ by $\mathbf{S}(\nu) = \sum_{i=1}^N \rho_i S_i(\nu)$. We say ν *satisfies the scaling law* $\mathbf{S}$ if $\mathbf{S}(\nu) = \nu$.

We now prove the *Existence and Uniqueness Result* from [H] Theorem 4.4(1) for fractal *measures*.

Theorem 5.2 *There is a unique measure $\mu \in \mathcal{U}$ satisfying a given scaling law $\mathbf{S}$. If ν is any measure in $\mathcal{U}$ then $\mathbf{S}^k(\nu)$[20] converges to μ in the Monge-Kantorovitch metric as $k \to \infty$. If K is the support of μ, then K is the unique compact set satisfying the corresponding scaling law $(S_1, \ldots, S_N)$ for sets.*

PROOF: We outline the main points.

Because the S_i are all contraction maps, it is possible to find $R_0 > 0$ such that if B_R is the (compact) set consisting of all points whose distance from the origin is $\leq R$, then $\mathbf{S}(B_R)(= \bigcup_{i=1}^N S_i(B_R)\,) \subset B_R$ for any $R \geq R_0$.

For any such B_R, $\mathbf{S}$ is a contraction map on the complete metric space $(\mathcal{U}_{B_R}, d_{\mathcal{MK}})$. The main point in showing this fact is that for $\nu_1, \nu_2 \in \mathcal{U}_{B_R}$, if f is as in (35) and r is the maximum contraction ratio of the S_i's, then

$$\begin{aligned} &\left| \int f \, d\Big(\mathbf{S}(\nu_1)\Big) - \int f \, d\Big(\mathbf{S}(\nu_2)\Big) \right| \\ &= \left| \sum_{i=1}^N \left[\int f \, d(\rho_i S_i(\nu_1)) - \int f \, d(\rho_i S_i(\nu_2)) \right] \right| \end{aligned}$$

[20] See (16) for notation.

$$
\begin{aligned}
&\leq \sum_{i=1}^{N} \left| \int f \, d(\rho_i S_i(\nu_1)) - \int f \, d(\rho_i S_i(\nu_2)) \right| \\
&\leq \sum_{i=1}^{N} \left| \rho_i \int (f \circ S_i) \, d\nu_1 - \rho_i \int (f \circ S_i) \, d\nu_2 \right| \\
&= \sum_{i=1}^{N} \left| \rho_i r \int r^{-1}(f \circ S_i) \, d\nu_1 - \rho_i r \int r^{-1}(f \circ S_i) \, d\nu_2 \right| \\
&\leq \sum_{i=1}^{N} \rho_i r d_{\mathcal{MK}}(\nu_1, \nu_2) \\
&\leq r d_{\mathcal{MK}}(\nu_1, \nu_2).
\end{aligned}
$$

For the second inequality we have used (33) and (34). For the penultimate inequality we have used the fact that since f is as in (35) then

$$
\left| r^{-1}(f \circ S_i)(x_1) - r^{-1}(f \circ S_i)(x_2) \right| \leq |x_1 - x_2|,
$$

as can be readily checked, and hence (35) can now be applied with f there replaced by $r^{-1}(f \circ S_i)$.

It follows from the Contraction Mapping Principle that $\mathbf{S}$ has a unique fixed point $\mu \in \mathcal{U}_{B_R}$, and if $\nu \in \mathcal{U}_{B_R}$ then $\mathbf{S}^k(\nu)$ converges to μ in the Hausdorff metric as $k \to \infty$. The first two claims in the Theorem follow since we can choose *any* $R \geq R_0$.

The final claim follows from elementary properties of the support of a measure. Namely,

$$
\begin{aligned}
\operatorname{spt} \mu &= \operatorname{spt} \mathbf{S}(\mu) = \operatorname{spt} \sum_{i=1}^{N} \rho_i S_i(\mu) \\
&= \bigcup_{i=1}^{N} \operatorname{spt} S_i(\mu) = \bigcup_{i=1}^{N} S_i(\operatorname{spt} \mu).
\end{aligned}
$$

Thus the compact set $\operatorname{spt} \mu$ satisfies the scaling law $(S_1, \ldots, S_N)$ and so is the unique such compact set by Theorem 4.2. ∎

Remark: The measure μ satisfying the scaling law $\mathcal{S}$ is usually called a *fractal measure*.

The method from Section 3.3.2 generalises to any fractal measure, see [E]. At the kth stage in the iteration, the map S_{σ_k} is chosen from $(S_1, \ldots, S_N)$, where the probability of choice of any particular S_i is p_i.

6 Random fractals

6.1 Probability theory

We begin with some *Examples*:

1. If a fair dice is thrown, the set of possible outcomes is $X = \{1, 2, \ldots, 6\}$. The associated probability distribution or measure P assigns to each of these numbers the equal measure (or probability) 1/6. If $E = \{1\}$ then $P(E) = 1/6$, if $E = \{1, 5\}$ then $P(E) = 1/3$, and of course $P(X) = 1$.

2. Consider an experiment in which a dart is thrown at a dartboard, all points are "equally likely" to be hit, and the dart never misses the board. If we take the dart board to be the disc of radius R given by

$$B_R = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} \leq R\},$$

then the probability measure P is a unit mass uniformly distributed over B_R. Thus we write

$$P = \tfrac{1}{\pi R^2}\, dx\, dy|_{B_R} = \tfrac{1}{\pi R^2}\mathcal{L}^2|_{B_R},$$

by which we mean the usual Lebesgue (i.e. "unit density") measure on $\mathbb{R}^2$ restricted to the disc B_R, and reweighted so the total mass (measure) for B_R is one. For any event E (which we identify with a subset of B_R) the probability of E (i.e. the probability that the dart lands in the set E) is given by

$$P(E) = \int_E \tfrac{1}{\pi R^2}\, dx\, dy = \tfrac{1}{\pi R^2}\, \mathcal{L}^2(E).$$

By $\int_E$ we mean integration over the set E.

3. Suppose a number x is selected according to the normal distribution with mean a and variance σ^2. Then the frequency function is

$$f(x) = \tfrac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-a)^2}{2\sigma^2}\right),$$

and the corresponding probability measure is denoted by

$$P = f(x)\, dx.$$

The probability of an event E, i.e. the probability that a number selected by this normal distribution lies in the set $E \subseteq \mathbb{R}$, is

$$P(E) = \int_E f(x)\, dx.$$

Motivated in part by these examples, one defines a *probability distribution* or *probability measure* P to be a unit mass measure on the set X of possible outcomes of an experiment. The definition of a finite measure is as in the first sentence of Section 5.1 with $\mathbb{R}^n$ there replaced by X, but we also now require that $P(X) = 1$ (so $0 \leq P(E) \leq 1$ if $E \subseteq X$).

If E is an *event*, i.e. E is a set of possible outcomes of the experiment and so $E \subseteq X$, then $P(E)$ is interpreted as the *probability of E occurring*. We might also think of $P(E)$ as the "measure" of E via P.

One point that sometimes causes confusion is the following. To say that an event E *occurs with probability zero* does not mean that it cannot occur. For example, in (b) above the mathematical probability of hitting a predesignated point on the board with an idealised dart whose tip is also a (mathematically ideal) point is zero. One sometimes says that E occurs *almost never*. Similarly, if an event *occurs with probability one* we say it occurs *almost surely*. For example, the (idealised) dart almost surely will *not* hit the origin.

In the previous three examples we have the situation where a *number*, or a point in $\mathbb{R}^2$, is "selected" via some probability distribution (or measure) P. The set X of outcomes was identified with a subset of $\mathbb{R}$ or of $\mathbb{R}^2$.

We, however, will be interested in conceptually more difficult cases where the set X of outcomes is either the set $\mathcal{C}$ of compact subsets of $\mathbb{R}^n$ (as in the case of random fractal sets), or the set $\mathcal{U}$ of compactly supported unit mass measures on $\mathbb{R}^n$ (as in the case of random fractal measures), or certain sets of N-tuples $(S_1, \ldots, S_N)$ of contraction maps S_i on $\mathbb{R}^n$, or certain sets of $2N$-tuples $\big((S_1, \ldots, S_N), (\rho_1, \ldots, \rho_N)\big)$ where the S_i are again contraction maps and the "weights" ρ_i are positive numbers.

6.2 Random fractal sets

We first make precise the ideas from Section 2.3.

Definition 6.1 A *random set* $\mathcal{E}$ is a set generated according to a a probability distribution on the set $\mathcal{C}$ of compact subsets of $\mathbb{R}^n$.

More precisely, a random set $\mathcal{E}$ is the actual probability distribution itself, and one distinguishes this from particular "realisations" of $\mathcal{E}$.

Thus when we think of a random set we usually consider, somewhat imprecisely, various "typical" realisations (i.e. outcomes) selected according to the underlying probability distribution $\mathcal{E}$, as in the diagrams in Section 2.3.

Definition 6.2 A *scaling law* for random sets is a probability distribution $\mathcal{S}$ on the set of N-tuples $(S_1, \ldots, S_N)$ of contraction maps $S_i : \mathbb{R}^n \to \mathbb{R}^n$, for some fixed $N \geq 2$. Moreover, we assume that for some fixed compact set $B \subset \mathbb{R}^n$, $S_i(B) \subseteq B$ for $1 \leq i \leq N$ whenever $(S_1, \ldots, S_N)$ is selected via $\mathcal{S}$.[21]

Example In the case of the Koch curve as described in (3), recall that $T_1(1.5, 0) = T_2(-1.5, 0) = (0, \sqrt{3}/2)$.

[21] Note that this last condition implies that if $(S_1, \ldots, S_N)$ is selected via $\mathcal{S}$ then the fixed point of each S_i is contained in B (to be precise, with probability one).

Suppose now that instead of the point $(0, \sqrt{3}/2)$ we select a point $(0, a)$ according to the uniform probability distribution on the vertical line segment joining the points $(0, \sqrt{3}/2 - .15)$ and $(0, \sqrt{3}/2 + .15)$, say. This induces in a natural way a probability distribution $\mathcal{T}$ on pairs of maps (T_1, T_2).[22]

Definition 6.3 Suppose compact sets $E^{(1)}, \ldots, E^{(N)}$ are each selected independently according to the probability distribution (i.e. random set) $\mathcal{E}$, and a single N-tuple $\mathbf{S} = (S_1, \ldots, S_N)$ of contraction maps is selected independently of the $E^{(i)}$ and according to the scaling law $\mathcal{S}$ (remember that $\mathcal{S}$ is a probability distribution)[23]. Consider the compact set

$$\bigcup_{i=1}^{N} S_i(E^{(i)}). \tag{36}$$

The method of selection of the $E^{(i)}$ and of $\mathbf{S} = (S_1, \ldots, S_N)$ induces in this way a natural probability distribution on compact sets. This probability distribution on compact sets is denoted by

$$\mathcal{S}(\mathcal{E}).$$

Since $\mathcal{S}$ when applied to a random set $\mathcal{E}$ yields another random set $\mathcal{S}(\mathcal{E})$, we also say $\mathcal{S}$ is a *scaling operator*.

We say the random set $\mathcal{K}$ *satisfies the scaling law* $\mathcal{S}$ iff

$$\mathcal{K} = \mathcal{S}(\mathcal{K}).$$

The following result is due independently to [F2], [G] and Mauldin & Williams (1986).

Theorem 6.4 *For each scaling law $\mathcal{S}$ on random sets, there exists a unique random set $\mathcal{K}$ which satisfies $\mathcal{S}$. Moreover, if $\mathcal{E}$ is any random set then $\mathcal{S}^k\mathcal{E}$ converges*[24] *to $\mathcal{K}$ as $k \to \infty$.*

PROOF: (*Ideas and remarks only*) Completely disregarding significant technical difficulties, we make the following comments.

The method in [F2] is to define a certain metric on the set of random sets and then to show that $\mathcal{S}$ is a contraction map with respect to this metric. This is analogous to the proof of Theorem 4.2.

The method in [G] and [MW] is to suitably randomise the construction in Section 4.5 and in the proof of Theorem 4.3. Informally the idea is to build a (*random*) *construction tree* as follows:

[22] More precisely, after selecting $(0, a)$ according to the prescribed uniform distribution, T_1 is obtained by reflecting in the x-axis, followed by a contraction and rotation about $(-1.5, 0)$ so that $(1.5, 0)$ is mapped to $(0, a)$. One similarly obtains T_2, but in this case the contraction and rotation are about the point $(1.5, 0)$, and $(-1.5, 0)$ is mapped to $(0, a)$.

[23] Note that the S_i are not selected "individually". The entire N-tuple is chosen according to $\mathcal{S}$.

[24] That is, the sequence $\mathcal{E}$, $\mathcal{S}(\mathcal{E})$, $\mathcal{S}^2(\mathcal{E}) = \mathcal{S}(\mathcal{S}(\mathcal{E}))$, $\mathcal{S}^3(\mathcal{E}) = \mathcal{S}(\mathcal{S}^2(\mathcal{E}))$, ... converges in the technical sense of weak convergence of probability measures.

(0) choose an N-tuple (of contraction maps) $\mathbf{S} = (S_1, \ldots, S_N)$ according to $\mathcal{S}$;

(1) for each $1 \leq i_1 \leq N$ choose N-tuples (of contraction maps) $\mathbf{S}^{i_1} = (S_1^{i_1}, \ldots, S_N^{i_1})$ according to $\mathcal{S}$, independently of one another and of the $\mathbf{S}$ chosen in (0);

(2) for $1 \leq i_1, i_2 \leq N$ choose N-tuples (of contraction maps) $\mathbf{S}^{i_1 i_2} = (S_1^{i_1 i_2}, \ldots, S_N^{i_1 i_2})$ according to $\mathcal{S}$, independently of one another and of the previously chosen N-tuples;

(3) for $1 \leq i_1, i_2, i_3 \leq N$ choose N-tuples $\mathbf{S}^{i_1 i_2 i_3} = (S_1^{i_1 i_2 i_3}, \ldots, S_N^{i_1 i_2 i_3})$ according to $\mathcal{S}$, independently of one another and of the previously chosen N-tuples;

...

Define

$$S_{\sigma_1 \ldots \sigma_k} = S_{\sigma_1} \circ S_{\sigma_2}^{\sigma_1} \circ S_{\sigma_3}^{\sigma_1 \sigma_2} \circ \cdots \circ S_{\sigma_k}^{\sigma_1 \ldots \sigma_{k-1}}. \tag{37}$$

The reasons for this definition are as follows:

1. First note the correspondence with (21), with the difference that here there is no longer a single *fixed* N-tuple $\mathbf{S} = (S_1, \ldots, S_N)$. Analogously to (24), there is associated to each code $\sigma = \sigma_1 \sigma_2 \ldots \sigma_k \ldots$ a point

$$\Pi(\sigma) = \lim_{k \to \infty} S_{\sigma_1 \ldots \sigma_k}(x),$$

which is independent of x.[25]

2. The set $\Pi(C_N)$[26] of all points $\Pi(\sigma)$ will form a compact set analogous to those in each of the three diagrams in Section 2.3[27]. For a fixed $x \in \mathbb{R}^n$ and a fixed large k, the set of points of the form $S_{\sigma_1 \ldots \sigma_k}(x)$ will be a good approximation to $\Pi(C_N)$, analogous to the diagram in Section 3.3.1 for the deterministic case (see also the sentence preceding that diagram).

Since Π depends on the choices made in (0), (1), (2), ..., we actually have a *random* set $\Pi(C_N)$, or more precisely a probability distribution on the set of compact subsets of $\mathbb{R}^n$. This random set is, moreover, invariant

[25] The independence follows as in the deterministic case if we assume that for some *fixed* $r < 1$ the contraction maps in any selected N-tuple have contraction ratio less than or equal to r; c.f. (12). But more general conditions are also possible.

[26] Recall from Section 4.5 that C_N is code space consisting of all codes $\sigma = \sigma_1 \sigma_2 \ldots \sigma_k \ldots$ where each $\sigma_i \in \{1, \ldots, N\}$

[27] More precisely, the diagrams in Section 2.3 are three different realisations of $\mathcal{T}^9\mathcal{E}$, where $\mathcal{E}$ is (with probability one) the line segment joining the endpoints $P = (-1.5, 0)$ and $Q = (1.5, 0)$, and $\mathcal{T}$ is a probability distribution on pairs (T_1, T_2) similar to the one in the Example near the beginning of this section.

under the scaling law $\mathcal{S}$. The point is that if $(S_1^*, \ldots, S_N^*)$ is chosen via $\mathcal{S}$ and independently of all the choices made in (0), (1), (2), ..., then $\Pi(C_N)$ and $\bigcup_{i=1}^N S_i^*(\Pi(C_N))$ will have the same probability distribution. ∎

6.3 Random fractal measures

We next make precise the ideas from Section 2.4.

Definition 6.5 A *random measure* $\mathcal{N}$ is a measure generated according to a a probability distribution on the set $\mathcal{U}$ of compactly supported unit mass measures in $\mathbb{R}^n$.

More precisely, a 'random measure $\mathcal{E}$ is the actual probability distribution itself, and one distinguishes this from particular "realisations" of $\mathcal{E}$.

We think of a random measure as a measure "selected according to" the distribution $\mathcal{N}$.

Definition 6.6 A *scaling law* for random measures is a probability distribution $\mathcal{S}$ on the set of $2N$-tuples $\Big((S_1, \ldots, S_N), (\rho_1, \ldots, \rho_N)\Big)$ of contraction maps $S_i : \mathbb{R}^n \to \mathbb{R}^n$, and of positive numbers $\rho_1, \ldots, \rho_N$ satisfying $\rho_1 + \cdots + \rho_N = 1$, for some fixed $N \geq 2$.[28] Moreover, we assume that for some fixed compact set B, $S_i(B) \subseteq B$ for $1 \leq i \leq N$ whenever $(S_1, \ldots, S_N)$ is selected via $\mathcal{S}$.

Example Consider the Example of the random Koch curve in the previous section. If the same probability distribution $\mathcal{T}$ on pairs of contraction maps (T_1, T_2) is again used, and if the weights (ρ_1, ρ_2) are each always taken to be $1/2$, then this induces a probability distribution $\mathcal{T}^*$ on 4-tuples $\Big((T_1, T_2), (\rho_1, \rho_2)\Big)$ which satisfies the previous definition with $N = 2$.

Definition 6.7 Suppose measures $\nu^{(1)}, \ldots, \nu^{(N)}$ are each selected independently according to the probability distribution (i.e. random measure) $\mathcal{N}$, and a single $2N$-tuple $\mathbf{S} = \Big((S_1, \ldots, S_N), (\rho_1, \ldots, \rho_N)\Big)$ is selected independently of the $\nu^{(i)}$ and according to the scaling law $\mathcal{S}$. Consider the measure

$$\sum_{i=1}^{N} \rho_i S_i(\nu^{(i)}). \tag{38}$$

The method of selection of the $\nu^{(i)}$ and of $\mathbf{S}$ induces in a natural way a probability distribution on measures (selecting in particular, measures of the form (38) with probability one). This probability distribution (i.e. random measure) is denoted by

$$\mathcal{S}(\mathcal{N}).$$

[28] Weaker conditions are sometimes used, in particular that the *expected value* (or *average value*) of $\rho_1 + \cdots + \rho_N$ equals 1.

Since $\mathcal{S}$ when applied to a random measure $\mathcal{N}$ yields another random measure $\mathcal{S}(\mathcal{N})$, we also say $\mathcal{S}$ is a *scaling operator*.

We say the random measure $\mathcal{M}$ *satisfies the scaling law* $\mathcal{S}$ iff

$$\mathcal{M} = \mathcal{S}(\mathcal{M}).$$

The following result is due to [MW] and [A]. See also [Z] and [PZ] for related ideas and [Wi] for connections between these approaches.

Theorem 6.8 *For each scaling law $\mathcal{S}$ on random measures, there exists a unique random measure $\mathcal{M}$ which satisfies $\mathcal{S}$. Moreover, if $\mathcal{N}$ is any random measure then $\mathcal{S}^k\mathcal{N}$ converges*[29] *to $\mathcal{M}$ as $k \to \infty$.*

PROOF: (*Some ideas only*) Given a scaling operator $\mathcal{S}$ for random measures, one can proceed parallel to the selection of the $S_j^{i_1 \ldots i_p}$ in the proof of Theorem 6.4 and simultaneously obtain weights $\rho_j^{i_1 \ldots i_p}$. This leads to a natural (random) measure on code space C_N.

Namely, for each k and for each $\sigma_1, \ldots, \sigma_k$, the set of all codes of the form $\sigma_1 \ldots \sigma_k \ldots$ is given the measure

$$\rho_{\sigma_1 \ldots \sigma_k} = \rho_{\sigma_1} \cdot \rho_{\sigma_2}^{\sigma_1} \cdot \rho_{\sigma_3}^{\sigma_1 \sigma_2} \cdot \ldots \cdot \rho_{\sigma_k}^{\sigma_1 \ldots \sigma_{k-1}}.$$

One can check that for fixed k this defines the measure of sets in a partition of C_N and that the total measure of these sets is one. As k increases this partition becomes "finer". One can show that there is a unique corresponding measure on C_N which will be denoted by Ψ[30].

If Π is the map constructed in the proof of Theorem 6.4 then $\Pi(\Psi)$ is a measure on $\mathbb{R}^n$. Since Π and Ψ depend on the choices in (0), (1), (2), ..., the measure $\Pi(\Psi)$ is in fact a *random* measure on $\mathbb{R}^n$. Moreover, it follows from the construction process that $\Pi(\Psi)$ satisfies the scaling law $\mathcal{S}$. ∎

Remark: In [HR] we give a simple proof of generalisations of the previous theorem using the Contraction Mapping Principle.

Final Remark: Conditions under which the dimension of random fractals can be computed have been extensively investigated. See [A], [F2], Graf (1987), [GMW], [MW] and Patzschke & Zähle (1990).

References

[A] ARBEITER, M. A. 1991 Random recursive constructions of self-similar fractal measures. The noncompact case. *Prob. Th. Rel. Fields* **88**, 497–520.

[29] In the technical sense of "weak convergence of probability measures".

[30] If we assume that the expected value of $\rho_1 + \cdots + \rho_N$ is one, rather than $\rho_1 + \cdots + \rho_N = 1$ with probability one, then the definition of the measure is technically more complicated.

[B] Barnsley, M. F. 1988 Fractals Everywhere. Academic Press.

[BD] Barnsley, M. F. & Demko, S. 1985 Iterated function systems and the global construction of fractals. *Proc. Roy. Soc. London* **A399**, 243–275.

[BEHL] Barnsley, M.F., Ervin, V., Hardin, D., & Lancaster, J. 1986 Solution of an inverse problem for fractals and other sets. *Proc. Nat. Acad. Sci.* **83**, 1975–1977.

[DS] Diaconis, P & Shahshahani, M 1984 Products of Random Matrices and Computer Image Generation. Preprint, Dept. of Stats. Stanford.

[E] Elton, J. 1987 An ergodic theorem for iterated maps. *Ergod. Th. Dynam. Syst.* **7**, 481–488.

[F1] Falconer, K. J. 1985 The Geometry of Fractal Sets. Cambridge University Press.

[F2] Falconer, K. J. 1986 Random fractals. *Math. Proc. Camb. Philos. Soc.* **100**, 559–582.

[F3] Falconer, K. J. 1990 Fractal Geometry : mathematical foundation and applications. John Wiley.

[FG] Bélair, J. & Dubuc, S. (eds) 1991 Fractal Geometry and Analysis. Proceedings of the NATO Advanced Study Institute, Montréal, Canada, 1989. Kluwer.

[G] Graf, S. 1987 Statistically self-similar fractals. *Prob. Th. Rel. Fields* **74**, 357–392.

[GMW] Graf, S., Mauldin, R. D. & Williams, S. C. 1988 The exact Hausdorff dimension in random recursive constructions. *Mem. Amer. Math. Soc.* **71**, no. 381.

[H] Hutchinson, J. E. 1981 Fractals and self similarity. *Indiana. Univ. Math. J.* **30**, 713–747.

[HR] Hutchinson, J. E. & Rüschendorf, L. A note on random fractals. In preparation.

[M] Mandelbrot, B. B. 1982 The Fractal Geometry of Nature. W.H. Freeman.

[MW] Mauldin, R. D. & Williams, S. C. 1986 Random recursive constructions; asymptotic geometric and topological properties. *Trans. Amer. Math. Soc.* **295**, 325–346.

[Mo] Moran, P. A. P. 1946 Additive functions of intervals and Hausdorff measure. *Proc. Camb. Phil. Soc.* **42**, 15–23.

[O] Oppenheimer, P. 1980 Constructing an atlas of self similar sets, Undergraduate thesis, Princeton.

[PZ] Patzschke, N. & Zähle, U. 1990 Self-similar random measures. IV — The recursive constuction model of Falconer, Graf, and Mauldin and Williams. *Math. Nachr.* **149**, 285–302.

[PJS] Peitgen, H-O., Jürgens, H. & Saupe, D. 1991 Fractals for the Classroom. Springer-Verlag.

[Wi] Wicks, J. 1993 Deterministic and random self-similar fractals and fractal measures, Undergraduate honours thesis, Aust. Nat. Univ.

[W] Williams, R. F. 1971 Compositions of contractions. *Bol. Soc. Brasil Mat.* **2**, 55–59.

[Z] Zähle, U. 1988 Self-similar random measures. I — Notion, carrying Hausdorff dimension, and hyperbolic distribution. *Prob. Th. Rel. Fields* **80**, 79–100.