Fractals: A Mathematical Framework

John E. Hutchinson
Department of Mathematics
School of Mathematical Sciences
Australian National University
(e-mail: John.Hutchinson@anu.edu.au)

Abstract.
We survey some of the mathematical aspects of deterministic and non-deterministic (random) fractals that have been useful in applications. Sets and measures (or grey-scales) are included. Some new results on random fractals are presented. Directions that may be worth further exploration in image compression are marked with a *.

Some of the underlying mathematics is explained in more detail, but still at an informal level, in [8]. Other general references at an elementary level are [4] and [12].

Examples of sets with scaling properties and whose dimension is not an integer have been known to mathematicians for a long time. But it was Mandelbrot who introduced the term *fractal* and who developed the connections between these ideas and a range of phenomena in the physical, biological and engineering sciences (see [10]).

One point that sometimes causes confusion is the following. A "mathematical" fractal in a certain precise sense looks the same at all scales; 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 only display this "self-similarity" for a range of magnifications or scales. The mathematical object will be an accurate model only within this particular range.

1. Fractal Sets

To fix our ideas, let us begin with the following *Brain* fractal, Figure 1, which we call B.

The set B is *self-similar* in that it is the union of two sets B_1 and B_2, the left and right brains, each of which is a scaled version of B (actually, if we draw a vertical line through the centre of B, a little of B_1 is to the right of this line, and a little of B_2 is to the left of this line). Thus we can write

$$B_1 = S_1(B), \quad B_2 = S_2(B),$$

and

$$B = S_1(B) \cup S_2(B). \tag{1}$$

The scaling maps S_1 and S_2 can be thought of as maps which apply to the entire plane, not just to the set B. Explicitly, S_1 consists of an anticlockwise rotation through 30° (i.e. $\pi/6$ radians) about the point $(-1, 0)$, followed by a contraction centred at $(-1, 0)$ with contraction ratio $1/\sqrt{3}$ (as a little calculation shows). Similarly, S_2 consists of a clockwise rotation through 30° about the point $(1, 0)$, followed by a contraction

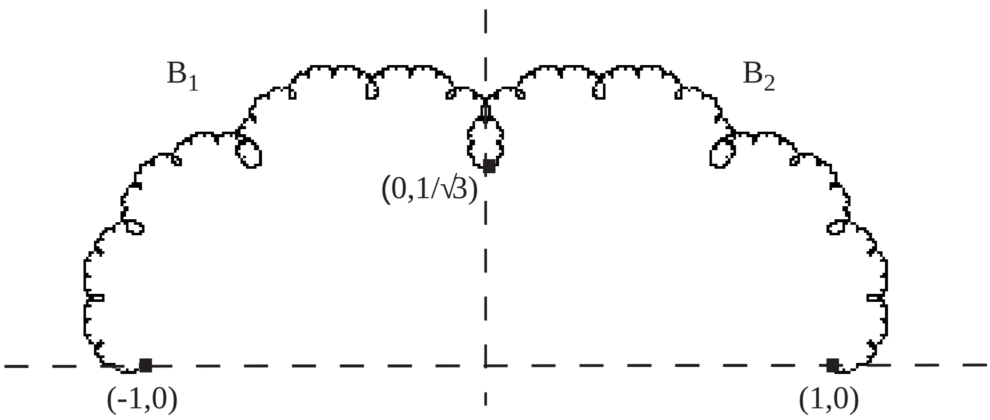

Figure 1: The Fractal Brain B

centred at $(1,0)$ with contraction ratio $1/\sqrt{3}$. We see after a simple calculation that

$$S_1(x,y) = \left(\frac{x}{2} - \frac{y}{2\sqrt{3}} - \frac{1}{2},\ \frac{x}{2\sqrt{3}} + \frac{y}{2} + \frac{1}{2\sqrt{3}}\right), \tag{2}$$

$$S_2(x,y) = \left(\frac{x}{2} + \frac{y}{2\sqrt{3}} + \frac{1}{2},\ -\frac{x}{2\sqrt{3}} + \frac{y}{2} + \frac{1}{2\sqrt{3}}\right). \tag{3}$$

We can regard $\mathbf{S} = (S_1, S_2)$ as a *Scaling Law* and we interpret (1) as stating that *the Brain B satisfies the scaling law* $\mathbf{S}$.

Notation The following notation will be important.

1. The plane is denoted by $I\!R^2$ and space is denoted by $I\!R^3$; thus the numbers 2 and 3 refer to dimension.
2. By a *dilation map* is meant a map $S\colon I\!R^2 \to I\!R^2$ (or $S\colon I\!R^3 \to I\!R^3$) such that

$$|S(a) - S(b)| \leq r|a-b| \tag{4}$$

   for all $a, b \in I\!R^2$ (or $I\!R^3$) and some fixed number r. Thus the dilation map S increases distance between points by a factor of at most r, where r is called the *dilation ratio* of S. If $r < 1$ we say S is a *contraction map*, since in this case distances are actually decreased by at least the factor r. Thus dilation maps are more general than contraction maps.
3. A simple but important example of a dilation map is given by

$$S(x) = a_0 + rO(x - a_0),$$

   where a_0 is a given point, O is a rotation (perhaps followed by a reflection) and r is a positive real number. Notice that $S(a_0) = a_0$, so that a_0 is *fixed* by the map S. We see that

$$S(a) - S(b) = rO(a-b),$$

using elementary properties of rotations. Thus

$$|S(a) - S(b)| = r|a - b|,$$

since rotations leave distances unchanged. In particular, S is a (rather special) dilation map with dilation ratio r. If $r < 1$ then S is a contraction map. Maps such as in this example where in fact one has "=" rather than "≤" in (4) are called *similitudes*.

We can now give some important definitions, which can be best understood by referring to the examples which follow.

Definition 1

1. A *Scaling Operator* $\mathbf{S}$ is a set $(S_1, \ldots, S_N)$ of contraction maps operating on either the plane $I\!R^2$ or space $I\!R^3$.

2. The image of a set E under $\mathbf{S}$ is given by

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

   We can iterate this construction to form an infinite sequence of sets

$$E,\ \mathbf{S}(E),\ \mathbf{S}^2(E) = \mathbf{S}(\mathbf{S}(E)),\ \mathbf{S}^3(E) = \mathbf{S}(\mathbf{S}^2(E)), \ldots.$$

3. A set K is *invariant under the Scaling Operator* $\mathbf{S}$, or *satisfies the scaling law* $\mathbf{S}$, if

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

We remark that what we call here a Scaling Operator, is often called an *Iterated Function System*, c.f. [4].

Examples

1. The "Brain" satisfies the Scaling Law $\mathbf{S} = (S_1, S_2)$ where S_1, S_2 are as in (2) and (3).

2. The well known *Koch Curve* can be described by another Scaling Law $\mathbf{S} = (S_1, S_2)$. In Figure 2 we show various iterations of this $\mathbf{S}$ applied to a set E consisting of a single point marked by a + . Notice how the sets $\mathbf{S}^k(E)$ converge to the Koch curve as k increases.

More Notation Informally, a set is *closed* if it contains all its "limit points". Thus the set of points on the straight line joining two points P and Q and which lie strictly between P and Q is not closed; but it becomes closed if we also include P and Q in the set. A set is *bounded* if it has finite diameter; thus the set consisting of the entire plane is *not* bounded.

In this Section it is convenient to consider the following type of sets.

Definition 2 A set is *compact* if it is both closed and bounded.[1]

[1]In order to avoid any possible confusion, we note that in more abstract settings one needs to give a different definition of *compact*.

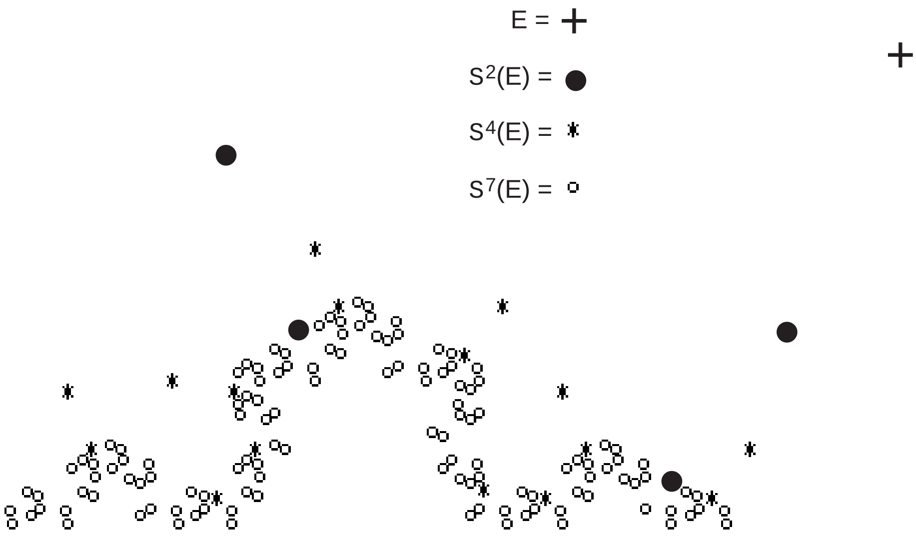

Figure 2: Approximating the Koch Curve

The "Brain" is compact, being both closed and bounded!

The following result from [7], which characterises certain types of "fractal sets", has turned out to be quite useful in image compression work.

Theorem 1

1. *To each scaling law* $\mathbf{S} = (S_1, \ldots, S_N)$ *there is one, and only one, compact set* K *which satisfies* $\mathbf{S}$. *The set* K *is called the* **fractal set** *corresponding to* $\mathbf{S}$.

2. *If* E *is any compact set, then the sequence of (compact) sets*

$$E,\ \mathbf{S}(E),\ \mathbf{S}^2(E),\ \mathbf{S}^3(E), \ldots.$$

   converges to K *(in the "Hausdorff metric" sense).*

3. *If the* S_i *are similitudes, and the so-called* Open Set Condition *holds, then the dimension* D *of* K *can be computed from*

$$r_1^D + \cdots + r_N^D = 1,$$

   where the r_i *are the dilation ratios of the* S_i.

Technical Remark The first two parts of the theorem can be proved either by explicit construction using an N-branching tree to code up the members of the resulting fractal, or by using the Contraction Mapping Principle for the Hausdorff metric on compact sets.

Examples

1. Thus the Brain and the Koch Curve are the *unique* compact sets invariant under the respective Scaling Operators.

2. The convergence result of the above theorem is indicated in Figure 2.

3. The Open Set Condition corresponds to a sort of "minimal overlap condition" of the parts of the associated fractal. This holds for the Koch curve but not for the Brain. In each case the contraction ratios are $1/\sqrt{3}$, and in the case of the Koch curve this gives
$$(1/\sqrt{3})^D + (1/\sqrt{3})^D = 1,$$
where D is the dimension. Taking logs of both sides gives
$$D = \frac{\log 4}{\log 3}.$$
It seems likely that in the case of the Brain, the dimension has the same value.

From the point of view of image compression the significance of Scaling Operators is that the information describing a complicated fractal set K can be "coded up", at least approximately, by a set of maps which can often be described by a finite set of parameters. An important problem is the *Inverse Problem* of finding a Scaling Operator which will, to some prescribed degree of accuracy, approximate a given image, or part thereof. A simple consequence of the Contraction Mapping proof of Theorem 1 is that if a compact set C can be approximated by scaled images of itself, that is if

$$\mathbf{S}(C) \approx C,$$

then also
$$K \approx C,$$
where K is the fractal generated by $\mathbf{S}$ and "$\approx$" means "approximately equals". The degree of approximation can be made precise and is usually measured in the Hausdorff metric. Barnsley calls this the *Collage Theorem*, c.f. [3], [4] and [8].

2. Fractal Measures

For purposes of image compression, and for a natural mathematical treatment, it is convenient to work with fractal measures rather than fractal sets. Whereas a set can be considered as either black ("in the set") or white ("not in the set"), a measure corresponds to an image with a grey scale. A set is a particular type of measure, where there are only two scales of grey, rather than a continuum of scales.

Another way to think of a measure is as a distribution of matter. Thus one can imagine matter "uniformly" distributed along the *Brain* or the *Koch Curve*. More generally, one can imagine a highly non-uniform distribution of matter along the same underlying sets. The density of matter at a point corresponds to the scale of grey at the point. The *total* amount of matter is called the *mass* of the measure.

Corresponding to each measure there is an underlying set called the *support* of the measure. The support can be thought of (loosely speaking) as the grey part of the plane (or space) given by the measure, i.e. the set of points with at least some colouring. The support of the measures noted in the previous paragraph are just the *Brain* and the *Koch Curve* respectively.

It is convenient to restrict considerations to the following type of measures.

Definition 3 A *standard measure* is a measure for which

1. the mass is one, and

2. somehat imprecisely, the amount of mass outside large balls of radius R approaches zero faster than $1/R$ as R approaches infinity.[2]

The requirement that the total mass be one is a normalisation requirement and amounts to a choice of units for mass. The second requirement is, in particular, satisfied if the support of the measure is bounded.

We have seen that a fractal set is self-similar in a manner which is described by a set of contraction maps. A fractal measure is *self-similar* in a manner which is described by a set of *dilation* maps $(S_1, \ldots, S_N)$ *together with* a set of *weights* $(w_1, \ldots, w_N)$ satisfying certain *extra* conditions. The following make this precise.

Definition 4

1. A *Scaling Operator* $\mathbf{S}$ for measures is a set $(S_1, \ldots, S_N, w_1, \ldots, w_N)$ of dilation maps S_i with dilation ratios r_i, and weights w_i, such that

$$0 < w_1, \ldots, w_N < 0, \quad w_1 + \cdots + w_N = 1, \tag{5}$$

and

$$w_1 r_1 + \cdots + w_N r_N < 1. \tag{6}$$

2. The image of a measure ν under $\mathbf{S}$ is given by

$$\mathbf{S}(\nu) = w_1 S_1(\nu) + \ldots + w_N S_N(\nu).^3$$

We can iterate this construction to form an infinite sequence of measures

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

3. A measure μ is *invariant under the Scaling Operator* $\mathbf{S}$, or *satisfies the scaling law* $\mathbf{S}$, if

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

Analogous to the case of fractals sets, one has the following result [7], which can perhaps be best understood by considering the examples which follow.

Theorem 2

1. To each scaling operator $\mathbf{S}$ *for measures there is one, and only one, standard measure* μ *which satisfies* $\mathbf{S}$*. The measure* μ *is called the* **fractal measure** *corresponding to* $\mathbf{S}$.

2. If ν *is any standard measure, then the sequence of (standard) measures*

$$\nu,\ \mathbf{S}(\nu),\ \mathbf{S}^2(\nu),\ \mathbf{S}^3(\nu), \ldots.$$

converges to μ *(in the sense of the "Monge-Kantorovitch metric").*

[2] More precisely, the measure μ satisfies $\int |x|\, d\mu(x) < \infty$.

[3] By $S_1(\nu)$ is simply meant the image of the measure ν under the map S_1. Similarly for $S_2(\nu), \ldots, S_N(\nu)$. Each of these is a measure of mass one. We then "reweight" these measures by factors $w_1, \ldots, w_N$, so that the total combined mass is $w_1 + \cdots + w_N$, which by (5) is just 1.

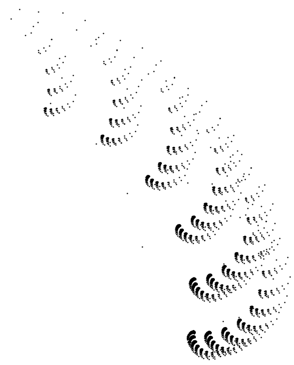

Figure 3: An Unbounded Fractal

3. If K is the support of the fractal measure μ, then K satisfies the scaling law $(S_1, \ldots, S_N)$

Examples

1. In Figure 2 let ν be a unit mass measure concentrated on the point $+$. Let (S_1, S_2) be unchanged, and choose $w_1 = w_2 = 1/2$. Since $r_1 = r_2 = 1/\sqrt{3}$ it is easy to see that (5) and (6) are true. Then $\mathbf{S}^2(\nu)$ is a unit mass measure with 1/4 unit mass concentrated on each of the four points $\bullet$. Similar remarks apply to $\mathbf{S}^4(\nu)$ and $\mathbf{S}^7(\nu)$. The sequence
$$\nu,\ \mathbf{S}(\nu),\ \mathbf{S}^2(\nu),\ \mathbf{S}^3(\nu), \ldots, \tag{7}$$
converges to a unit mass distributed "uniformly" along the Koch curve.

2. If in the previous example, the weights w_1 and w_2 are taken to be unequal, then $\mathbf{S}^2(\nu)$ is still a unit mass measure, but the masses of each $\bullet$ will no longer be equal, and will in fact be ${w_1}^2$, $w_1 w_2$, $w_2 w_1$ (which is of course the same as $w_1 w_2$) and ${w_2}^2$. Similar remarks apply to other members of the sequence (7).

3. If the S_i are all contraction maps, so each $r_i < 1$, then condition (6) follows from condition (5). But in fact it is only necessary that *one* of the S_i be a contraction map, in the sense that it is then always possible to choose weights w_i such that (5) and (6) are true.

   For example in Figure 3 we show a fractal measure generated by two dilation maps (S_1, S_2) where one of the dilation ratios is a little larger than one and the two weights satisfy $w_1 = w_2 = 1/2$. The mass distribution is indicated by the density of the points. What is shown is (as always) only an approximation and the actual fractal is unbounded, although most of the mass does lie in the region shown.

* **Remarks** One of the problems with working with sets and the Hausdorff metric is the so-called problem of "outliers". Two sets may be close, or even coincide, except for a few points which are far apart. We think of such sets as being close, yet as measured

by the Hausdorff metric they are *far apart*. However, if we consider the sets as measures in an appropriate sense, then they will only differ on a small amount of mass, and they will indeed be close in the Monge-Kantorovitch sense. Thus using measures and the Monge-Kantorovitch metric is often more natural, even when we are not interested in grey scales.

A related point is as follows. In Figure 3, the support of the fractal measure is an *unbounded* but closed set satisfying the scaling law (S_1, S_2). By working with measures, we can thus analyse *unbounded* fractal sets in a natural setting.

There are also generalisations of the Monge-Kantorovitch metric that may prove useful in applications, see [13] and [9].

There are at least two other methods of obtaining fractal like (deterministic) objects along the lines of Theorems 1 and 2, and which may have further applications. One can work with parametrised curves or surfaces in the uniform, and other, metrics, c.f. [7], §3.5. One can also use the notion of "flat distance" between curves or surfaces as in [7], §6.3.

3. Random Fractals

Many objects occurring in nature can more realistically be described using the notion of a random, or non-deterministic, fractal. Such a fractal is a set or measure generated or selected according to a probability distribution. The *self-similarity*, strictly speaking, applies to the underlying probability distribution, rather than to a particular realisation of the random fractal.

Consider a Brownian particle, a microscopic particle in a fluid. The particle is buffeted by the molecules with which it comes in contact and it traces out an irregular path, called a *Brownian path*. The paths traced out between times $t = 0$ and $t = 1$, and between times $t = 1$ and $t = 2$ are, after rescaling, "statistically" similar to the path traced out between times $t = 0$ and $t = 2$. In a sense which can be made precise, a Brownian path is a realisaton of a random fractal set.

In Figure 4 we show three different realisations of a kind of *random Koch curve*. For each realisation K we have

$$K = K_1 \cup K_2,$$

where K_1 and K_2 are the "left" and "right" halves of K. The sets K_1 and K_2 look "statistically" like rescaled versions of K. It is important to realise that, as in Figure 1, the diagram is only an approximation. More precisely, each small straight line segment should be replaced by a (different) scaled realisation of the random Koch curve, and this process should be repeated at smaller and smaller scales.

It again turns out to be more natural to work with (standard) random measures, rather than random sets. If, for example, we imagine a unit mass distributed uniformly along each of the three realisations shown, then we have three different realisations of a certain underlying probability distribution on measures. More generally, it is possible to have the masses themselves distributed in a random (but self-similar) manner along the shown curves.

We make this precise as follows. We consider random (standard) measures ν which

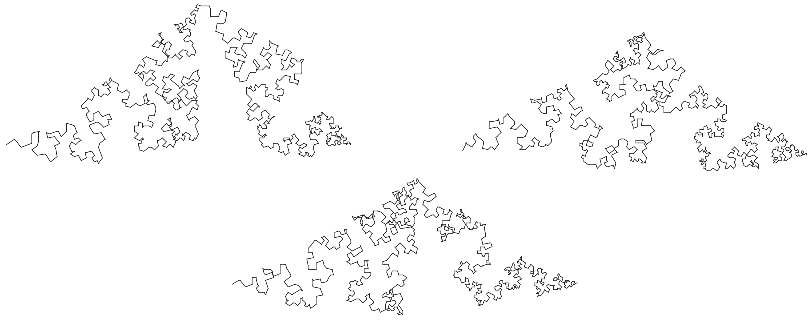

Figure 4: Realisations of a Random Koch Curve

are generated by a probability distribution $\mathcal{N}$.

Definition 5

1. A *Scaling Operator* $\mathcal{S}$ for random measures is a probability distribution $\mathcal{S}$ on sets $\mathbf{S} = (S_1, \ldots, S_N, w_1, \ldots, w_N)$ of dilation maps S_i with dilation ratios r_i, and weights w_i, such that

$$\begin{aligned}&0 < w_1, \ldots, w_N < 0, \quad w_1 + \cdots + w_N = 1\\ &\mathbb{E}(w_1 r_1 + \cdots + w_N r_N) < 1\\ &\mathbb{E}\left(w_1 |S_1(0)| + \cdots + w_N |S_N(0)|\right) < \infty.\end{aligned}$$

   By $\mathbb{E}(w_1 r_1 + \cdots + w_N r_N)$ is meant the *expected* or *average* value of $w_1 r_1 + \cdots + w_N r_N$ selected according to $\mathcal{S}$. Similarly for $\mathbb{E}\left(w_1 |S_1(0)| + \cdots + w_N |S_N(0)|\right)$.

2. The image under $\mathcal{S}$ of the probability distribution $\mathcal{N}$ is denoted by

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

   and is another probability distribution which generates random measures as follows:

   (a) first select $\nu^{(1)}, \ldots, \nu^{(N)}$ independently and at random via the distribution $\mathcal{N}$,

   (b) then select $\mathbf{S} = (S_1, \ldots, S_N, w_1, \ldots, w_N)$ independently of the $\nu^{(1)}, \ldots, \nu^{(N)}$ via the probability distribution $\mathcal{S}$,

   (c) construct the measure

$$w_1 S_1(\nu^{(1)}) + \cdots + w_N S_N(\nu^{(N)}). \tag{8}$$

   We say that the measure in (8) is constructed via the probability distribution $\mathcal{S}(\mathcal{N})$. With a somewhat sloppy notation, we sometimes refer to a random measure generated by $\mathcal{S}(\mathcal{N})$ as a random measure of the form

$$\mathbf{S}(\nu).$$

3. We can iterate this construction to form an infinite sequence of probability distributions

$$\mathcal{N},\ \mathcal{S}(\mathcal{N}),\ \mathcal{S}^2(\mathcal{N}) = \mathcal{S}(\mathcal{S}(\mathcal{N})),\ \mathcal{S}^3(\mathcal{N}) = \mathcal{S}(\mathcal{S}^2(\mathcal{N})), \ldots .$$

4. A probability distribution $\mathcal{M}$ is *invariant under the Scaling Operator* $\mathcal{S}$, or *satisfies the scaling law* $\mathcal{S}$, if

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

While all this may seem rather abstract, it gives in many instances a more natural model of physical reality than the notion of a deterministic fractal.

Once again, one has the following analogue of Theorems 1 and 2.

Theorem 3

1. *To each scaling operator $\mathcal{S}$ for random (standard) measures there is one, and only one, probability distribution $\mathcal{M}$ which satisfies $\mathcal{S}$. A realisation of $\mathcal{M}$ is called a* **random fractal measure** *corresponding to $\mathcal{S}$.*

2. *If $\mathcal{N}$ is any probability distribution on standard measures then the sequence*

$$\mathcal{N},\ \mathcal{S}(\mathcal{N}),\ \mathcal{S}^2(\mathcal{N}),\ \mathcal{S}^3(\mathcal{N}), \ldots$$

converges to $\mathcal{M}$ (in the sense of the appropriate "Monge-Kantorovitch" metric).

3. *The supporting sets* (**random fractal sets**) *of the random measures have a probability distribution which is self-similar in a natural sense.*

4. *If the S_i are all similitudes, then under certain conditions each supporting set has dimension D with probability one, where D is determined by the relation*

$$\mathbb{E}({r_1}^D + \cdots + {r_N}^D) = 1.[4]$$

The above theorem, under the assumption that all the S_i are contraction maps, was proved in [1], [5], [6] and [11], under the extra condition that the S_i be all contraction maps, but with the weaker assumption that $\mathbb{E}(w_1 + \cdots + w_N) = 1$ rather than $w_1 + \cdots + w_N = 1$. In [9] we give a very simple proof of the first three parts using the Contraction Mapping Principle.

* **Remarks** Figure 4 took 1.2 megabytes to store as a postscript file. But the probability distribution $\mathcal{S}$ required to generate the appropriate similitudes takes only a few lines of code.

Consider now, the *Inverse Problem* of finding $\mathcal{S}$, given one or more realisations of the corresponding random fractal set. For a probability distribution on random numbers (say), it is in fact impossible to gain much information about the underlying distribution from one, or a small number, of realisations of the distribution. But in the case of a random self-similar set, the situation is much better. By taking small parts of a single realisation and rescaling, we can obtain a large number of realisations. From these, it is at least in principle possible to obtain a good approximation to the underlying distribution of the generating maps $\mathbf{S} = (S_1, \ldots, S_N)$.

---

[4]Recall that the r_i are the dilation ratios of the S_i

Once one knows the underlying probability distribution, it is possible to generate realisations of the self-similar process. If one wants to store a good approximation to a *particular* realisation, it would be necessary to condition the process by various constraints.

It would not be simple to implement such a procedure (to solve the Inverse Problem) in an efficient manner, but it may be well worth exploring the possibility.

References

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

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

[3] 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**, pp. 1975–1977.

[4] Barnsley, M. F. (1988), *Fractals Everywhere*, Academic Press.

[5] Falconer, K. J. (1986), "Random fractals," *Math. Proc. Camb. Philos. Soc.*, **100**, pp 559–582.

[6] Graf, S. (1987) "Statistically self-similar fractals," *Prob. Th. Rel. Fields*, **74**, pp 357–392.

[7] Hutchinson, J. E. (1981), "Fractals and self similarity," *Indiana. Univ. Math. J.*, **30**, pp 713–747.

[8] Hutchinson, J. E., "Deterministic and Random Fractals," to appear in *Complex Systems*, eds T. Bossomaier and D. Green, Cambridge Univ. Press.

[9] Hutchinson J. E. & Rüschendorf L., "A Note on Random Fractals," preprint, Aust. Nat. Univ.

[10] Mandelbrot B. B. (1982), *The Fractal Geometry of Nature*, W.H. Freeman.

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

[12] Peitgen H-O., Jürgens H. & Saupe D (1991), *Fractals for the Classroom*, Springer-Verlag.

[13] Rachev S. T. & Rüschendorf L. (1994), "Probability Metrics and Recursive Algorithms," preprint.