The Interaction of Theory and Procedure in Fractal Geometry

Stephen D. Casey
Department of Mathematics and Statistics
The American University
4400 Massachusetts Ave., N.W.
Washington, D.C. 20016–8050
e-mail: scasey@american.edu

Abstract

This article discusses the interplay in fractal geometry occurring between computer programs for developing (approximations of) fractal sets and the underlying dimension theory. The computer is ideally suited to implement the recursive algorithms needed to create these sets, thus giving us a laboratory for studying fractals and their corresponding dimensions. Moreover, this interaction between theory and procedure goes both ways. Dimension theory can be used to classify and understand fractal sets. This allows us, given a fixed generating pattern, to describe the resultant images produced by various programs. We will also tie these two perspectives in with the history of the subject. Three examples of fractal sets developed around the turn of the century are introduced and studied from both classical and modern viewpoints. Then, definitions and sample calculations of fractal and Hausdorff–Besicovitch dimension are given. Finally, dimension theory is used to classify images.

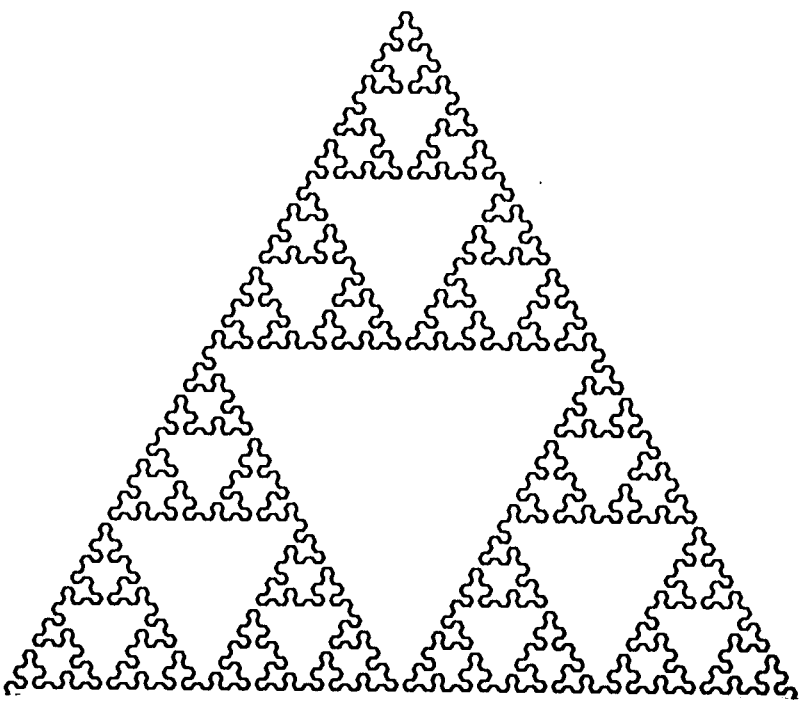


Figure 1 – A Curve Approximating the Sierpinski Gasket

1 Introduction

Fractal geometry lies at an intersection of many areas of science – mathematics, physics, computer science, electrical and mechanical engineering, economics, etc. The theory has evolved into an extension (but by no means a replacement) of our current models of the world, which are based on Euclidean geometry and smooth generalizations of that geometry, e.g., differentiable manifolds. It includes highly detailed objects produced by repeated application of maps, while Euclidean geometry includes objects made by the conglomeration of Euclidean shapes – circles, lines, polygons, etc. – each of which can be generated by simple maps. Both visual inspection and statistical analysis[1] lead us to believe that much of the natural world (e.g., clouds, mountains, coastlines, trees (see Feder [10] and Mandelbrot [15])) and many man–made objects (e.g., spatial distributions of communication networks, temporal flow of messages in these networks, economic distribution of resources (see Anthony and Casey[2] and Mandelbrot [15])) can be modeled by this theory.

This article discusses a very interesting interplay in the theory of fractal geometry, occurring between computer programs for developing (approximations of) fractal sets and the underlying dimension theory. The computer is ideally suited to implement the recursive algorithms needed to create these sets, thus giving us a laboratory for studying fractals and their corresponding dimensions. Moreover, this interaction between theory and procedure goes both ways. Dimension theory can be used to classify and understand fractal sets. This allows us, given a fixed generating pattern, to describe the resultant images produced by various programs.

The structure of the article reflects this interconnection. We first discuss an efficient algorithm for creating fractals. Classical examples of fractals are described in the next section, giving us a convenient framework under which to discuss dimension theory in section 3. Section 4 contains the heart of the paper. We first use fractal theory to partially classify fractals. Then we show how the theory of self–similar fractals can be used to predict and classify the output of programs used to generate self–similar sets.

1.1 The *Fracmkr* Algorithm

Figures for the article were generated by an algorithm (*Fracmkr*) developed by the author and Nicholas Reingold (see Casey and Reingold [6]). This is an efficient algorithm which produces (approximations of) self–similar fractal sets. These sets are produced by a repeated scaling, translation, reflection, and/or rotation of a fixed pattern or set of patterns. The resultant sets strictly preserve fractal scaling. The algorithm complements Mandelbrot's *The Fractal Geometry of Nature* [15], in that it can be developed into a program which reproduces the self–similar fractals in Mandelbrot (approximately 45% of the figures in the book). The most recent addition to the algorithm is a routine which computes the dimension of the fractals from the pattern generator. Theoretical justification for this new routine is given in section 4.

Fracmkr is a "pattern rewriting system" in which a given geometric pattern is drawn repeatedly after suitable scaling and placement. The system uses two patterns – an initial configuration, or base, and a generating pattern, or seed. Generating schemes are patterns with some built–in information on orientation and connection in later levels of iteration. This is exactly how one would begin to study self–similar patterns with a paper and pencil. The system is a cousin of the string rewriting or L–systems introduced by A. Lindenmayer and developed by P. Prusinkiewicz (see Peitgen and Saupe [16]), which are themselves cousins of LOGO, the product of the LOGO Group of the MIT Artificial Intelligence Laboratory. (See *Turtle Geometry* by H. Abelson and A. diSessa [1].) However, it differs from these systems in some important ways. The generating schemes for patterns are themselves patterns. In L–systems and in LOGO, a pattern is generated by "teaching the turtle" where to move based on some fundamental set of instructions. In our system, if you can draw seed and base patterns on a piece of paper, identifying vertices and orientation schemes, then you can produce iterates of the seed on the base in the computer. Also, there is the matter of efficiency and simplicity. In our routine, the only overhead between levels of recursion is a single boolean variable. In L–systems, the higher the level of iteration, the longer the string produced.

[1] Robust methods based on Pareto–Levy distributions can be used to measure data that exhibits a fractal structure (see Casey [5]).

The engine of the system is a recursive routine. This works with the information in the seed. It is passed the endpoints $(x_1, y_1), (x_2, y_2)$ of a line segment and the number of iterations to perform. It then "draws" a scaled, rotated, and translated copy of each segment of the seed in such a way that this copy starts at the beginning of the given line segment and ends at the end of the given line segment. However, when it "draws" a segment it really just calls itself with the number of iterations reduced by one (unless the number of iterations is 1, in which case it really does draw the segment). This involves almost no bookkeeping – only a single boolean variable for keeping track of reflections is needed between levels of recursion.

The calculations used to handle the rotation, scaling, and translation are done in two parts. The calculation of the general directions and scaling of the seed (as determined by the base) are done globally. The directions and scaling factors are passed via global variables. Within the pattern generator, we calculate the coefficients of a rotation matrix (times a contraction or expansion matrix), which determines the coordinates of the next endpoint to aim for. The matrix is used to rotate (and scale) the current segments being drawn.

The pattern generator also takes arguments used to handle orientation and connection features. This information is built into the seed, and is used by the generator as it draws. This gives the flexibility to produce objects with many components, e.g. Cantor sets, and orientations, e.g., plates 31 and 146 from [15]. See Casey and Reingold [6] for a more complete description of the procedure and a listing of the code.

1.2 Some Definitions

We now must establish some terminology. Let $\mathbf{R}^n$ denote n–dimensional Euclidean space with norm $\|\cdot\|$.

We say a mapping $f : \mathbf{R}^n \to \mathbf{R}^m$ is a *contraction* if there exists an α, $0 \leq \alpha < 1$, such that

$$\|f(x) - f(y)\| \leq \alpha \|x - y\|$$

for all x, y in the domain of f. We say that a set S is *invariant* under a collection of contraction mappings, say $\{m_1, m_2, \ldots, m_n\}$, if

$$S = \bigcup_{i=1}^{n} m_i S .$$

Using these last two definitions, we define a *self–similar* set as a set which is invariant under a collection of contraction maps such that $m_i S \cap m_j S$ is at most a finite number of points when $i \neq j$ (see Hutchinson[2] [14]).

2 Fractal Geometry: Three Classical Examples

The theory of fractal geometry goes back at least to K. Weierstrass, who constructed a fractal set in his development of a continuous but nowhere differentiable curve[3] in 1872 (see Edgar [7]). The first fractal sets were developed as examples and counterexamples necessary in the natural development of mathematics, e.g., a totally disconnected perfect set, a curve of infinite length without tangent contained in a finite region, or a space–filling curve (as developed by G. Peano in 1890). From our modern perspective, we see that they all exhibit both the scale invariance and intricate detail which group them together as fractals. It is important to note the years during which these mathematical constructs appeared. Although we may think of fractals as modern objects generated by computers, the mathematical foundations of fractal geometry are over one hundred years old. Each of the examples below are important in the development of the theory in analysis, topology, and geometry (see Falconer [8]–[9], Federer [11], Hocking and Young [12], and Huriewicz and Wallman [13]). All are currently being used in the mathematical modeling of natural and man–made phenomena (see Anthony and Casey [2], Feder [10], Mandelbrot [15], and West [18]).

The three sets described below are all self–similar. We include the collection of contraction mappings associated with each set in the description, and remark that the fractal sets may be constructed directly from the mappings by using M. Barnsley's iterated function systems (see Barnsley [3]). We also note that the dimensions of these sets can be computed directly from these maps (see section 4).

[2] More sophisticated overlap conditions are possible (see [14]). This will be sufficient for our purposes.

[3] A curve with similar properties was developed by B. Bolzano in 1830. He did not publish his result.

2.1 The Cantor Set

G. Cantor (1883)

The ancient Greeks felt that in order to be indefinitely subdivisible a body had to be a continuum. In 1883, Cantor produced a set showing that this was not the case. This set, the Middle Thirds Set C, is constructed as follows. Given $[0,1]$, remove the open middle third segment $(\frac{1}{3},\frac{2}{3})$. Then, remove the open middle thirds of the remaining two segments, i.e., remove $(\frac{1}{9},\frac{2}{9})$ from $[0,\frac{1}{3}]$ and $(\frac{7}{9},\frac{8}{9})$ from $[\frac{2}{3},1]$. Repeat. At the n^{th} stage of this process, there will be 2^n segments, each of length $(1/3)^n$. Continue ad infinitum.

The total length of the removed set is $(1/3)\sum_{i=0}^{\infty}(2/3)^n$, which equals one. However, the remaining set C is a perfect, and thus an uncountable set which is totally disconnected[4] (see Hocking and Young [12]).

The Cantor set can be generated by the maps

$$m_1(x) = \frac{1}{3}x \ , \ m_2(x) = \frac{1}{3}x + \frac{2}{3} \ .$$

2.2 Snowflake Curves

H. von Koch (1904)

Koch's snowflake curve K is an example of a simple closed curve of infinite length fitting inside a bounded region. The limit curve also has a tangent nowhere.

The Koch construction is to replace the segment $[0,1]$ with four segments of length $1/3$ with endpoints

$$(0,0), (\frac{1}{3},0), (\frac{1}{2},\frac{\sqrt{3}}{6}), (\frac{2}{3},0), (1,0).$$

Then, each of the four segments is replaced by the seed pattern, and so on. At the n^{th} stage the curve consists of 4^n segments of length $(1/3)^n$. As n goes to infinity, the total length also goes to infinity.

Let $\vec{x} = [x\ y]^T$ be a vector in the Cartesian plane $\mathbf{R}^2$. Let $\theta \in [0, 2\pi)$ be an angle, and let

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

be the rotation matrix for angle θ. Then, the contraction mappings under which K is invariant are

$$m_1(\vec{x}) = \tfrac{1}{3}\vec{x} \ , \qquad m_2(\vec{x}) = \tfrac{1}{3}R_{(\pi/3)}\vec{x} + [0\ \tfrac{1}{3}]^T \ ,$$
$$m_3(\vec{x}) = \tfrac{1}{3}R_{(-\pi/3)}\vec{x} + [\tfrac{1}{2}\ \tfrac{\sqrt{3}}{6}]^T \ , \quad m_4(\vec{x}) = \tfrac{1}{3}\vec{x} + [\tfrac{2}{3}\ 0]^T \ .$$

2.3 Sierpinski Gasket

W. Sierpinski (1915)

Starting with a filled-in equilateral triangle, perform a Cantor-like removal process by removing the interior of the middle equilateral triangle whose vertices are the midpoints of the three edges. Repeat this process on the three remaining filled-in triangles, and then the nine remaining ones, and so on. All of the points in the remaining set are vertices of some removed triangle, and thus are branch, or ramification, points.

This construction is due to Sierpinski. The set S_g is invariant under

$$m_1(\vec{x}) = \frac{1}{2}\vec{x} \ , \ m_2(\vec{x}) = \frac{1}{2}\vec{x} + [\frac{1}{2}\ 0]^T \ , \ m_3(\vec{x}) = \frac{1}{2}\vec{x} + [\frac{1}{4}\ \frac{\sqrt{3}}{4}]^T \ .$$

These three examples provide a good framework in which to introduce the definition of a fractal, given in the next section. In order to do this, we need to discuss dimension theory. Understanding even the definitions in this theory requires some technical background, which can be found in Hocking and Young [12].

[4] Infinitely many other Cantor constructions are possible, e.g., removing the second and four sections of an interval divided into equal fifths.

3 Scaling, Fractals, and Dimension Theory

We can "define" a *fractal* as a set which :

(i.) *has detailed fine structure.* Fractals have the property that all magnifications of some or all of the set reveals intricate detail. This is unlike the differentiable curves and surfaces we encounter in calculus. Repeated enlargements of every differentiable curve or surface reveal a line or a plane, respectively.

(ii.) *has scale invariance.* All magnifications of the fractal reveal a set which is exactly the same (self-similarity) or statistically or asymptotically the same (statistical or quasi- self-similarity). This is related to the first property. However, there are self-similar sets with no detailed fine structure, e.g. a line.

(iii.) *is produced recursively.* We can generate these sets by repeated application of some collection of maps. These maps provide a simple encoding of these complicated sets.

(iv.) *can not be described easily in terms of Euclidean geometry.* Although the sets can be constructed from the objects studied under classical geometry, the structure of the limit set is an uncountable collection of points in a complicated arrangement.

(v.) *you know is a fractal when you see it.*

The mathematical definition is more complicated and very technical.

Definition 3.1 *A* fractal *is a set whose Hausdorff-Besicovitch dimension strictly exceeds its topological dimension* (B. Mandelbrot (1975) [15], pg. 15).

In order to understand the definition of a fractal, we first need to calculate the *dimension* of a given set. Dimension is the primary tool used in classifying fractal sets. The topological and Hausdorff-Besicovitch dimensions (denoted D_T and D_{HB}, respectively) are very technical. Topological dimension can be thought of as an integer n which describes which Euclidean n-space $\mathbf{R}^n$ that a given set resembles locally. This dimension is invariant under homeomorphisms[5], and therefore is an accurate determination of topological information about the set. It does not, however, measure detailed geometry and scaling structure. For example, every Cantor set, any countable set, and a set consisting of a single point all have topological dimension 0. All non-intersecting curves have $D_T = 1$. Proof of these facts are difficult, and involve machinery from algebraic topology. Topological dimension is discussed in greater detail in Casey and Reingold [6], Falconer [8]-[9], and Huriewicz and Wallman [13].

Fortunately, there is a more accessible and simplified version of dimension — the *fractal dimension* D_{KM}. It was introduced by L. Pontrjagin (1932), and studied by A. Kolmogorov (1950's), L. Richardson (1960's), and Mandelbrot (1970's), among others (see [15] for more information). The fractal dimension of a set is a key ingredient in identifying a fractal set. Mandelbrot [15] assigns a D_{KM} value with most of the images in his book. For self-similar and nearly self-similar sets with an intricate geometry, fractal dimension can not only measure the structure of a set, but can also be determined experimentally (see Barnsley [3], Casey [5], and Mandelbrot [15]). Assuming this definition, we define a *fractal* as a set whose fractal dimension strictly exceeds its topological dimension. In most cases, the fractal dimension of a fractal is a non-integer number. We shall see that the definition of a fractal given above reflects properties (i.), (ii.), and (iv.) in the list above. A theorem of M. Barnsley and S. Demko gives that all fractals can be approximated arbitrarily closely by repeated applications of a set of maps (see Barnsley [3], chapter 3).

Six examples will be discussed. The first two are the unit interval $I = [0,1]$ and the unit square $I^2 = [0,1] \times [0,1]$. The next three were developed in the previous section - C, K, and S_g. The final example is the set $H = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\} \cup \{0\}$.

[5] A homeomorphism is a one-to-one onto continuous mapping with a continuous inverse.

Consider any closed and bounded set in n–dimensional Euclidean space, $\mathbf{R}^n$. Let r be any positive number, and let $\mathcal{N}(r)$ be the *minimal* number of closed line segments, balls, or spheres of radius r needed to cover the set. Then we want to calculate a number D such that as $r \to 0$,

$$\mathcal{N}(r) \cdot r^D \sim 1 \quad \text{(Scaling Relationship)}.$$

Consider first a single point. For any $r > 0$, we need only one line segment to cover. Thus, $\mathcal{N}(r) \cdot r^D = 1 \cdot r^D = 1$ for $D_{KM} = 0$.

Next consider $[0,1]$. Let $\lceil x \rceil$ be the smallest integer greater than or equal to any real number x. Given any $r > 0$, we need $\lceil 1/r \rceil$ line segments of length r to cover $[0,1]$. Thus $\mathcal{N}(r) \cdot r^D = \lceil 1/r \rceil \cdot r^D \sim 1$, for $D_{KM} = 1$.

The set $[0,1] \times [0,1]$ can be covered by $\lceil 1/r \rceil^2$ balls of radius r, so in this case D_{KM} is 2.

In these first three examples, D_{KM} agrees with the topological dimension. The next example breaks from this notion.

Consider the Cantor Middle Thirds set constructed in 2.1. At the n^{th} stage of its construction, we need 2^n line segments of radius $(1/3)^n$ to cover. As this is true for r restricted to $\{(1/3)^n\}$ and for all n, $n = 1, 2, \ldots$,

$$\mathcal{N}(r) \cdot r^D = 2^n (1/3)^{nD}.$$

Therefore,

$$\begin{array}{rrcl} & \mathcal{N}(r) \cdot r^D & = & 1 \\ \Leftrightarrow & \log(2^n (1/3)^{nD}) & = & \log 1 \\ \Leftrightarrow & n(\log 2 - D \log 3) & = & 0. \end{array}$$

(where log denotes the natural logarithm). Thus, for r restricted to $\{(1/3)^n\}$, $n = 1, 2, \ldots$, $D_{KM}(C) = \log 2 / \log 3$.

These examples motivate the following:

Definition 3.2 The Kolmogorov–Mandelbrot, or fractal, dimension $D_{KM}(X)$ *of a set* X *is:*

$$D_{KM}(X) = \lim_{r \to 0} \frac{\log(\mathcal{N}(r))}{\log(1/r)}.$$

The number D_{KM} is the value necessary to preserve the scaling relationship $\mathcal{N}(r) \cdot r^D \sim 1$ as $r \to 0$. Calculation of D_{KM} is simplified by the fact that the continuous variable r may be replaced by the discrete variable $r_n = \rho^n$, $0 < \rho < 1$, $n = 1, 2, \ldots$. That is, if

$$\Delta = \lim_{n \to \infty} \frac{\log(\mathcal{N}(r_n))}{\log(1/r_n)},$$

then $\Delta = D_{KM}$ (see Barnsley [3], pg. 176). This allows us to derive the fractal dimension of a self–similar fractal set directly from its generating seed. Let's look at the dimension of some of the sets from the previous section.

First we see that for the Cantor Middle thirds set, $D_{KM}(C) = \log 2 / \log 3$. Next, consider the Koch curve described in section 2.2. At the n^{th} stage it consists of 4^n segments of length $(1/3)^n$. Therefore, for $r = r_n = (1/3)^n$,

$$\begin{array}{rrcl} & \mathcal{N}(r) \cdot r^D & = & 1 \\ \Leftrightarrow & \log(4^n (1/3)^{nD}) & = & \log 1 \\ \Leftrightarrow & n(\log 4 - D \log 3) & = & 0. \end{array}$$

Thus,

$$D_{KM}(K) = \frac{\log 4}{\log 3}.$$

At the n^{th} stage of construction, the Sierpinski gasket consists of 3^n equilateral triangles with sides of length $(1/2)^n$. Therefore, calculating as above, we see that

$$D_{KM}(S_g) = \frac{\log 3}{\log 2}.$$

We also introduce the Hausdorff–Besicovitch dimension. This dimension is an essential tool of geometric measure theory. This measurement requires sophisticated mathematical tools. We give only the definition and the dimensions of the examples, referring the interested reader to Falconer [8]–[9], Federer [11], and Taylor [17] for details. These references also include detailed discussions of other related dimensions, including packing dimension.

Definition 3.3 The Hausdorff–Besicovitch Dimension $D_{HB}(X)$ *of a set* X*: Let* X *be a subset of* $\mathbf{R}^n$*, and consider the set of all countable covers of* X *by sets of diameter less than* $r > 0$*. Let* r_n *be the diameter of the* n^{th} *set in any given cover. Then, for* $\beta > 0$*, the* β^{th} Hausdorff–Besicovitch outer measure of X *is*

$$\mu^*_\beta(X) = \lim_{r\to 0}\left[\inf_{\{\text{covers}\}}\left[\sum_{n=1}^{\infty} r_n^\beta\right]\right].$$

The Hausdorff–Besicovitch dimension is then

$$D_{HB}(X) = \inf\{\beta : \mu^*_\beta(X) = 0\}.$$

(F. Hausdorff (1919), A. Besicovitch (1928) [8], pp. 7–10)

Equivalently, we could define

$$D_{HB}(X) = \sup\{\beta : \mu^*_\beta(X) = \infty\}.$$

Clearly, the difference between D_{HB} and D_{KM} is the choice of the covering sets. The flexibility of choosing arbitrary sets of diameter less than r, as opposed to line segments, disks, or balls of radius r, makes the HB dimension a much finer measurement, albeit more difficult to calculate. However, for self–similar sets, we have the following.

Theorem 3.1 (Hutchinson [14]) *If* X *is self–similar, then*

$$D_{KM}(X) = D_{HB}(X).$$

Therefore, we see that for the examples above, we have that $D_{HB}(I) = 1$, $D_{HB}(I^2) = 2$, $D_{HB}(C) = \frac{\log 2}{\log 3}$, $D_{HB}(K) = \frac{\log 4}{\log 3}$, and $D_{HB}(S_g) = \frac{\log 3}{\log 2}$. Also, as self–similar sets strictly preserve the scaling relationship given in the previous section, D_{HB} will reflect the scale invariant structure for these sets. It is also sufficiently robust to show this structure for statistically and quasi– self–similar sets.

In general,

$$D_{HB}(X) \le D_{KM}(X).$$

The set H is an example of a set for which $D_{HB}(X) < D_{KM}(X)$. In fact, $D_{HB}(H) = 0$, while $D_{KM}(H) = 1/2$ (see Falconer [9], pg. 45)[6]. Note that in this case, as H has only one "scaling point" (namely the point $\{0\}$), H is not a set which would be considered as a fractal. Thus, we see that the finer measurements used in calculating D_{HB} are needed to define a fractal.

4 Dimension and the Classification of Fractals

4.1 Classifying General Fractal Sets

We commented in the previous section that dimension is the primary tool for identifying fractals. In this section, we mention a few results in which dimension is used to classify fractals.

A set F with a countable number of elements has HB dimension 0. If F is uncountable and has HB dimension less than one, it has qualities in common with the Cantor set.

[6] For another example, let $T = \{1, \frac{1}{3}, \frac{1}{5}, \ldots\} \cup \{0\}$. Then, $D_{HB}(T) = 0$, while $D_{KM}(T) = 1/2$.

Theorem 4.1 ([9], pg. 30) *A set $F \subset \mathbf{R}^n$ with $D_{HB}(F) < 1$ is totally disconnected.*

Curves with dimension between 1 and 2 all share a common property. The first curve developed with this property was the Koch curve.

Theorem 4.2 ([9], pg. 80) *If C is a curve in $\mathbf{R}^n$ with $1 < D_{HB}(C) < 2$, then C contains a subcurve C_s such that at almost all points of C_s, no tangent exists.*

We will say that a curve is space–filling if it fills in any planar area.

Theorem 4.3 ([15], II, 7) *If C is a curve in $\mathbf{R}^n$ such that $D_{HB}(C) = 2$, then C is space–filling.*

Theorem 4.2 generalizes to higher dimensions. If we have a surface of topological dimension n, we say that it has a tangent at a point x if there exists an n–dimensional hyperplane approximating the surface at that point.

Theorem 4.4 ([11], section 3.3) *If S is an n–dimensional surface in $\mathbf{R}^m$, $n < m$, with $n < D_{HB}(S) < n+1$, then S contains a subsurface S_s such that at almost all points of S_s, no tangent exists.*

It becomes increasingly difficult to classify fractals as D_{HB} increases. Fractal sets can have small subsets exhibiting one type of behavior, other subsets exhibiting another, and so on. The dimension of the set is given by the component with the highest dimension. Thus, a curve could have $D_{HB} > 1$, and yet be differentiable along most of its length. The theorems in this section are just a few of the results in the area of geometric measure theory. Classification of fractals using geometric measure theory is by no means complete, nor is the study of fractal sets the only topic of research in this area. H. Federer's *Geometric Measure Theory* [11] is a fundamental work in the field. Also see Falconer [8]–[9] and Taylor [17].

4.2 Self–similar Fractals

We have seen that the rigorous study of fractal sets involves sophisticated mathematical tools. We have also seen that fractals can be created via recursion, and therefore computers are an ideal tool for studying these sets. The computers can be used as tools in proving theorems about these sets, generating examples and checking calculations, but cannot prove theorems. However, the theory can be used to predict the output of a computer program, as we will now demonstrate. Our discussion in this section is limited to self–similar fractals.

Recall that the definition of self–similar set given in the first section. The first theorem we mention is a theorem of J. Hutchinson given above. It tells us that for self–similar sets X,

$$D_{HB}(X) = D_{KM}(X)\,.$$

Thus for self–similar sets, the "proper" form of dimension can be calculated by the methods used to calculate D_{KM}. Furthermore, this number can be calculated directly from the maps used to define the set.

Theorem 4.5 (Hutchinson [14]) *Let X be a self–similar set defined by contraction mappings $\{m_i\}_{i=1}^n$. Let $\{\alpha_i\}_{i=1}^n$ be the set of contractions associated with these mappings. Then $D_{HB}(X) = D$, where D is the solution to the equation*

$$\sum_{i=1}^{n} \alpha_i^D = 1\,.$$

An immediate consequence of this theorem is the dimensions of our classical examples. We can see that $D_{HB}(C) = \frac{\log 2}{\log 3}$, $D_{HB}(K) = \frac{\log 4}{\log 3}$, and $D_{HB}(S_g) = \frac{\log 3}{\log 2}$ directly from the generating maps.

We can also use Hutchinson's theorems to calculate the HB dimension of a self–similar set X approximated by the pattern rewriting system directly from the generating seed. The set X we refer to is the theoretical limit set which is produced by iterating the pattern rewriting system infinitely often. Each seed

needs to be normalized by drawing it so that it begins at $(0,0)$ and ends at $(1,0)$. We also have to guarantee that the set we produce is self–similar. Our seed cannot have line segments in it of length greater than or equal to one. (If you gave the pattern rewriting system such a seed, it would produce a pattern – it just would not converge to a new self–similar pattern.[7]) We also cannot have the line segments overlap each other except at the endpoints. This can be determined by examining two generations of the pattern.

Theorem 4.6 *Let S be a normalized seed pattern consisting of n line segments which generates a self–similar set X. Let ℓ_i denote the length of the i^{th} line segment in S, $i = 1, \ldots, n$. Let D be the solution to the equation*

$$\sum_{i=1}^{n} \ell_i^D = 1 .$$

Then $D_{HB}(X) = D$.

The theorem follows directly from Hutchinson's second theorem. Given a normalized seed, we can construct the contraction mappings associated with the pattern that this seed generates by mapping the unit interval $[0,1]$ onto each line segment in the seed. Each map is the composition of a contraction, a rotation, and a translation. The contraction factor α_i for the i^{th} map equals the length ℓ_i of the i^{th} line segment.

If we have a normalized seed and we have that $\sum_{i=1}^{n} \ell_i^D = 1$ has solution $D > 2$, then X will be be space–filling. Moreover, iterates of the seed will overwrite each other. In fact, we can think of the number $\frac{D}{2}$ as the number of times that the iterates will overwrite each other. We discuss an example of this below. We note that the need for an overlap condition is clear from the theorem. If we allowed for overlapping segments in our defining seed, then we would be counting segments more than once in our formula. We also remark that in general we cannot solve the equation given in Theorem 4.6 in closed form unless all of the lengths of the line segments are equal. However, a numerical solution can be easily calculated, using, say, the Newton–Raphson algorithm.

The natural question to ask is what does this number get us. The answer is that it can tell us what sort of pattern will emerge after repeated iterations of a given seed. If $D_{HB} = 0$, we will see only isolated dots. For $0 < D_{HB} < 1$, we will approximate a type of Cantor set. If $D_{HB} = 1$, then we will produce a piecewise–linear curve. For $1 < D_{HB} < 2$, we will approximate a curve or a set of curves with the property that the curves have tangents almost nowhere, similar to the Koch arc. If $D_{HB} = 2$, iterates of the seed will, in the limit, be space–filling. Finally, if $D > 2$, iterates of the seed will quickly fill up regions, "overlapping" many points $\frac{D}{2}$ times.

We give the following examples. A "fat" Cantor set can be produced by a seed consisting of the line segments

$$[0, \frac{3-\alpha}{6}] , [\frac{3+\alpha}{6}, 1] , \; 0 < \alpha < 1 .$$

Here, we are removing middle interval of length $\frac{\alpha}{3}$. If X is the set that is produced by iteration of the seed, then

$$D_{HB}(X) = \frac{\log(2)}{\log(\frac{6}{3-\alpha})} .$$

Since $0 < D_{HB}(X) < 1$ for all α, $0 < \alpha < 1$, X is totally disconnected. If the base is $[0,1]$, the total mass of X, as measured by Lebesgue measure,[8] is $1 - \alpha$. Note that we can make α arbitrarily close to zero.

The next examples come from Mandelbrot [15]. The "monkey's tree," plates 31 and 146 from [15], is made up of six segments of length $\frac{1}{3}$ and five segments of length $\frac{\sqrt{3}}{9}$, with various orientations. Let M_1 be the set which results from iteration of the seed (see Figure 2). Then, $D_{HB}(M_1) = D$, where D is the solution of

$$6(\frac{1}{3})^D + 5(\frac{\sqrt{3}}{9})^D = 1 .$$

[7] There is, of course, the case where we have $[0,1]$ as a seed. Repeated iteration of the pattern produces nothing new.

[8] See Federer [11] for reference.

By noting that $(\frac{1}{\sqrt{3}})^2 = \frac{1}{3}$ and $(\frac{1}{\sqrt{3}})^3 = \frac{\sqrt{3}}{9}$, we see that $(\frac{1}{3})^{D/2} = x_0$, where x_0 is the positive solution of $5x^3 + 6x^2 - 1 = 0$. Solving, we get that

$$D = \frac{2\log(\frac{10}{\sqrt{21}-1})}{\log(3)} = 1.868726\ldots .$$

Thus, M_1 will be a curve which has a tangent nowhere, but is not space filling (in spite of the fact that iterates of the seed will "fill up" patches of the computer screen). The "snowflake sweep," plate 68 from [15], is made up of six segments of length $\frac{1}{3}$ and one segment of length $\frac{\sqrt{3}}{3}$, with various orientations. Let M_2 be the set which results from iteration of the seed (see Figure 3). Then, $D_{HB}(M_1) = 2$, because

$$6(\frac{1}{3})^2 + (\frac{\sqrt{3}}{3})^2 = 1 .$$

The curve M_2 fills the interior of the Koch snowflake.

Finally, we can produce curves with variable dimension using the two sides of a isosceles triangle. We let these be the line segments with endpoints $(0,0)$, $(\frac{1}{2},x)$, $(1,0)$, $(\frac{1}{2},x)$, respectively. With the orientation produced by connecting the line segments in the order they were written, and $x = \frac{1}{2}$, iterates of this produce Mandelbrot's "dragon sweep" (see plate 66 of [15]), which is a space–filling curve. For variable x, $0 \le x \le \frac{1}{2}$, the general seed pattern will produce self-similar curves. With x in this range, we have that

$$D_{HB}(X) = \frac{\log(2)}{\log(\frac{1}{\sqrt{x^2+\frac{1}{4}}})} .$$

If $\frac{1}{2} < x < \frac{\sqrt{3}}{2}$, then iterates of this pattern will produce an image by overwriting it. In fact, if ℓ is the length of each side of the triangle, and if D is the solution of $2\ell^D = 1$, then we can produce a set with $\frac{D}{2}$ overlap by setting

$$x = \sqrt{\frac{1}{2^{\frac{2}{D}}} - \frac{1}{4}}$$

for $2 < D < \infty$, which gives a range of $\frac{1}{2} < x < \frac{\sqrt{3}}{2}$. Note that we can make D arbitrarily large. For example, we can let $x = \sqrt{\frac{1}{\sqrt[50]{2}} - \frac{1}{4}}$, which then makes $\ell = \frac{1}{\sqrt[100]{2}}$ and $D = 100$. Iterates of the seed overwrite each other 50 times.

Acknowledgements

The author would like to thank Lawrence Crone and Richard Holzsager of American University and Nicholas Reingold of AT&T Bell Labs for mathematical and programming advice.

References

[1] H. Abelson and A. A. diSessa, *Turtle Geometry*, MIT Press, Cambridge, Mass. (1980).

[2] R. A. Anthony and S. D. Casey, "Protype fractal representation and analysis of C^3 systems (Pro–Fracs)", Institute of Defense Analysis Technical Report (1991).

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

[4] S. D. Casey, "Formulating fractals," *Computer Language* 4, 4, 28-40, cover (1987).

[5] S. D. Casey, "Analysis of fractal and Pareto–Levy sets: theory and application," *Proceedings of EFTF*, 205-211 (1990).

[6] S. D. Casey and N. F. Reingold, "Self–similar fractal sets: theory and procedure," *IEEE Computer Graphics and Applications* **14**, 3, 73-82 (1994).

[7] G. A. Edgar, ed., *Classics on Fractals*, Addison–Wesley, Reading, MA (1993).

[8] K. Falconer, *The Geometry of Fractal Sets*, Cambridge University Press, Cambridge (1985).

[9] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley and Sons, Chichester (1990).

[10] J. Feder, *Fractals*, Plenum Press, New York (1988).

[11] H. Federer, *Geometric Measure Theory*, Springer–Verlag, New York (1969).

[12] J. Hocking and G. Young, *Topology*, Dover Publications, New York (1988).

[13] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, Princeton, N.J. (1941).

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

[15] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York (1983).

[16] H.–O. Peitgen and D. Saupe, *The Science of Fractal Images*, Springer–Verlag, New York (1988).

[17] S. J. Taylor, "The measure theory of random fractals," *Math. Proc. Camb. Phil. Soc.* **100**, 383-406 (1986).

[18] B. West, *An Essay on the Importance of Being Non–Linear*, Springer–Verlag, Berlin (1985).

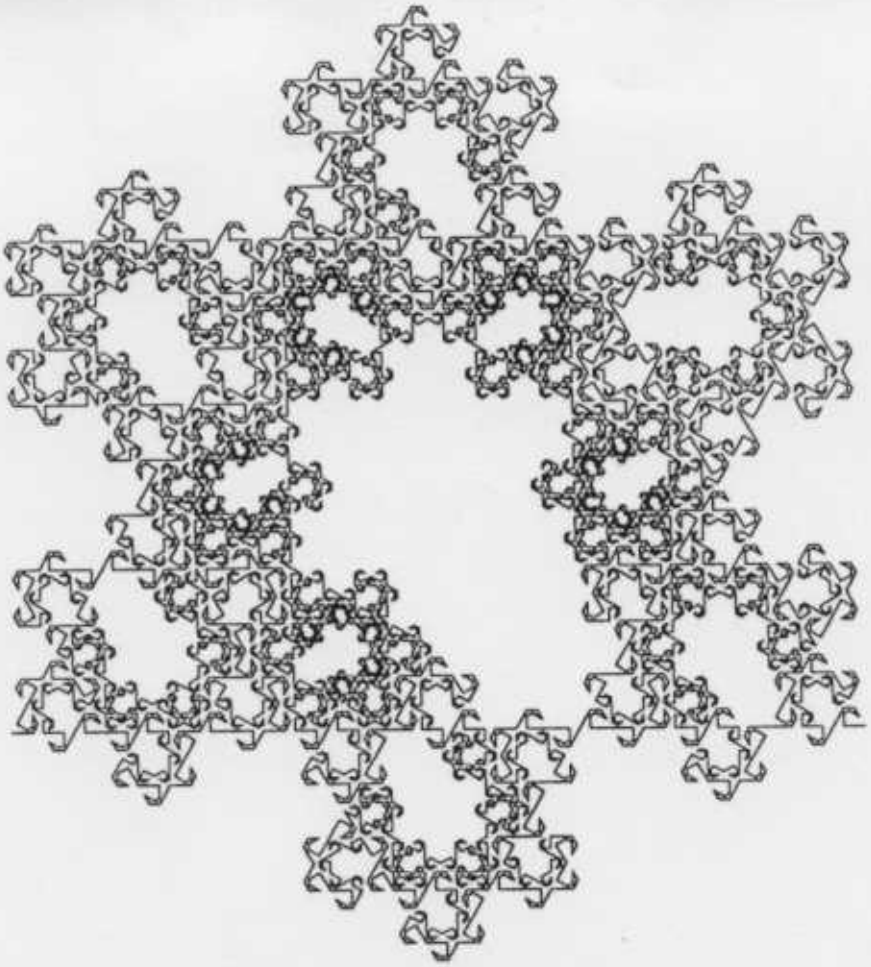


Figure 2 – Mandelbrot's "Monkey's Tree"

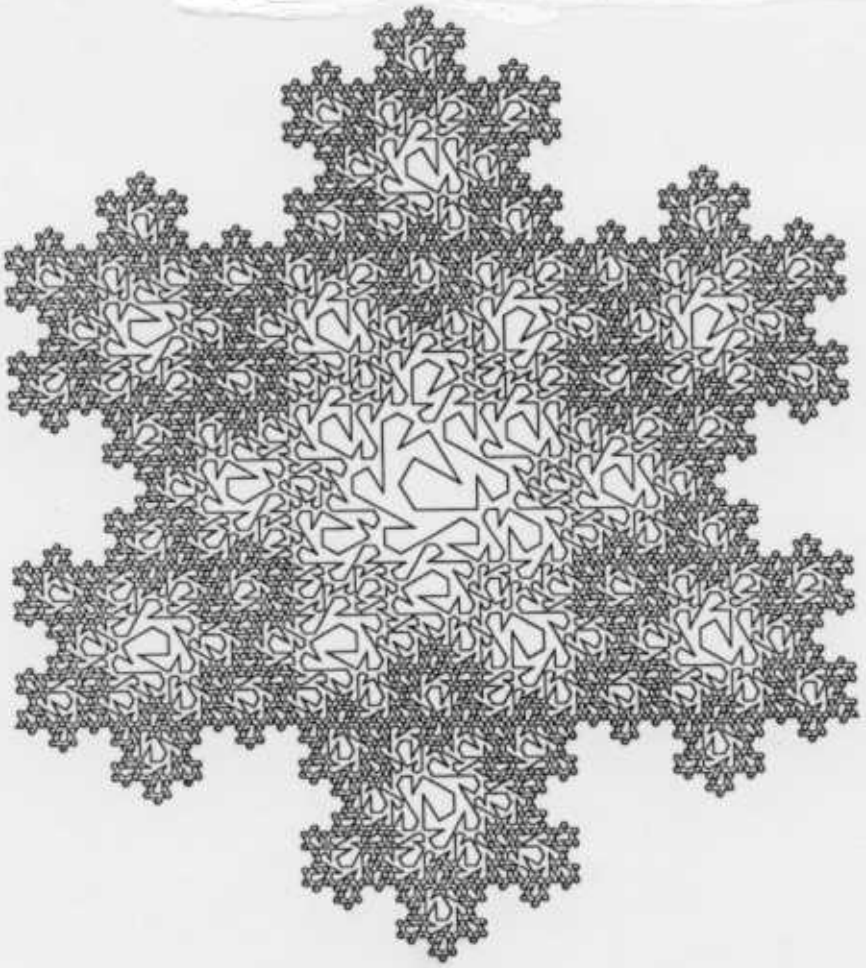
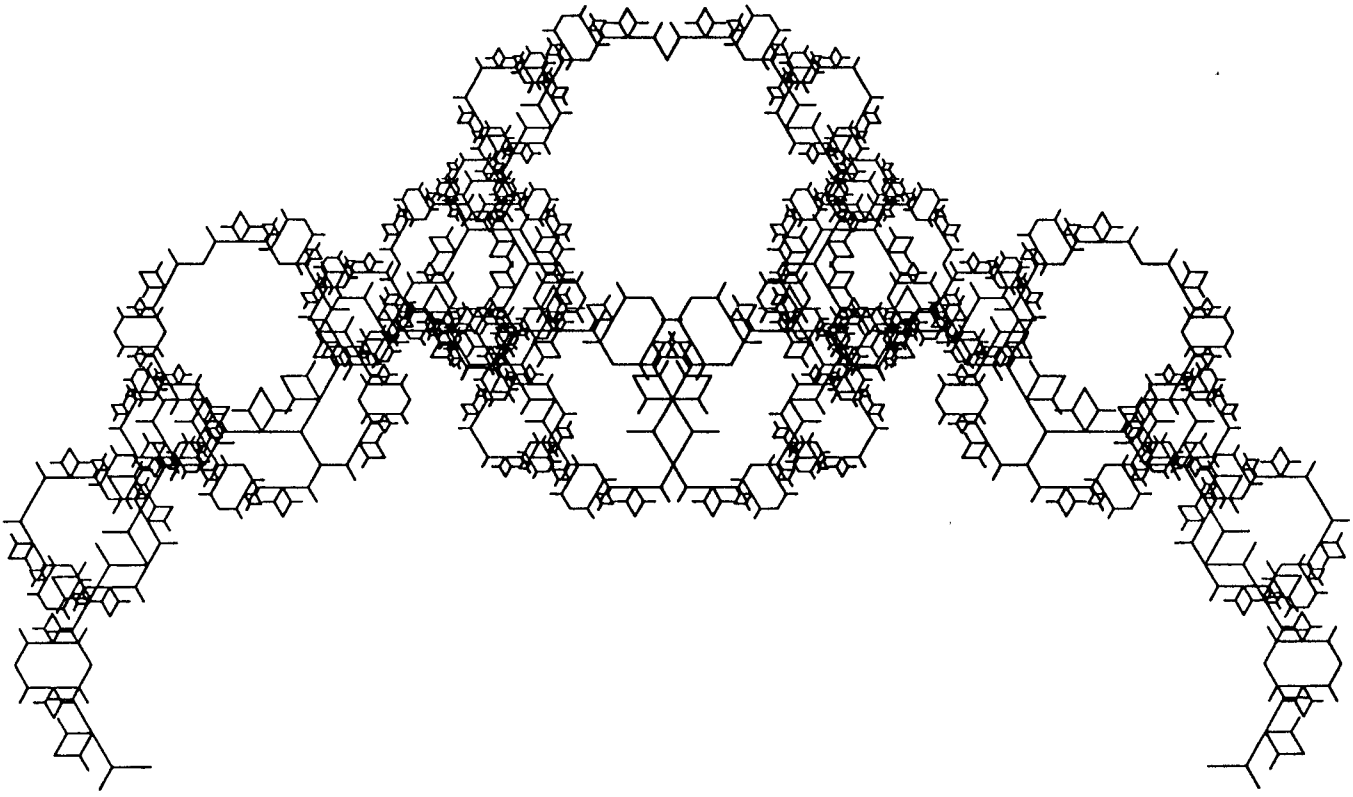


Figure 3 – Mandelbrot's Snowflake Sweep



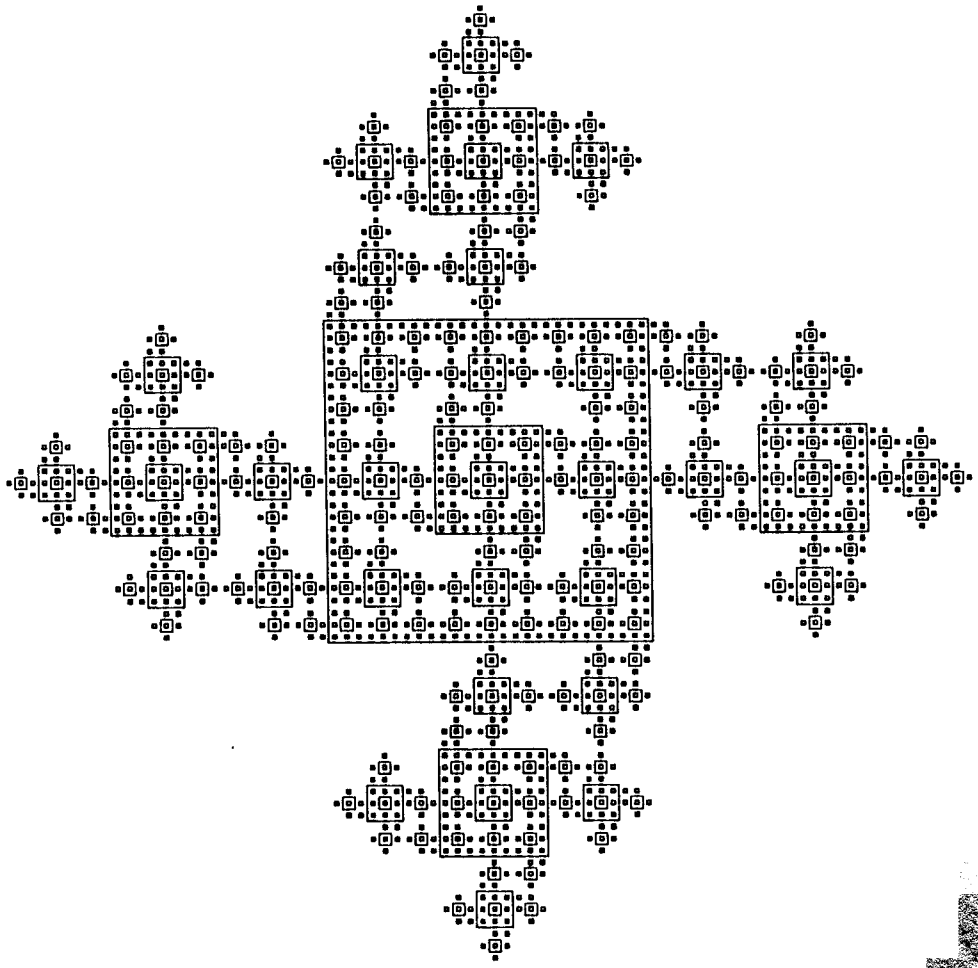


Figure 4 – Various Figures Generated by *Fracmkr*