A Formal Treatment of Deterministic Fractals

J.M. Curry

May 19, 2006

Abstract

We explore in depth the theory behind deterministic fractals by investigating transformations on metric spaces and the contraction mapping theorem. In doing so we introduce the notion of the Hausdorff distance metric and its connection to the space of fractals. In order to understand how deterministic fractals are generated, we develop the concept of an iterated function system (IFS) and what it means for these fractals to be an attractor of the IFS. Finally, we give creedance to our notion of fractals as objects having fractional dimension, by introducing a simplified version of the Hausdorff Dimension.

Introduction

Fractals as a mathematical object of study is in many respects, still in its infancy. Before the term *fractal* was ever coined, Karl Weierstrass began a formalized train of thought that would eventually give birth to the mathematical study of fractals. In 1872, Weierstrass sought a curve which was continuous everywhere and differentiable nowhere[1]. Although this famous counterexample may have served originally as a powerful result in mathematical analysis, what was really discovered was a class of *rough* objects that would be characterized as fractal.

Impelled by Weierstrass's example, Helge Von Koch developed one of history's first fractals by prescribing a simple geometric algorithm for manipulating a line[2].

Pursuing almost an entirely different line of thought, Georg Cantor provided a construction for a perfect set with uncountably many points, whose recursive algorithm, when given a real line could provide a self-similar fractal of the utmost simplicity[3].

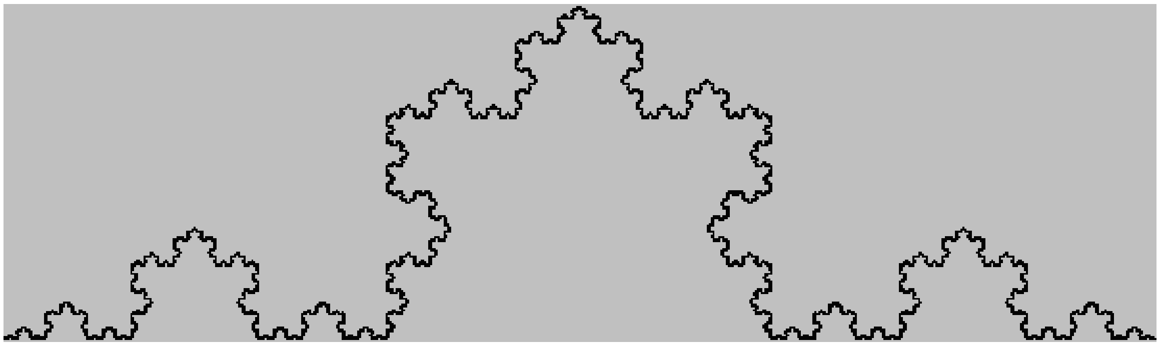

Figure 1: Koch's Curve – Continuous everywhere, differentiable nowhere. – From Wikipedia

We will discover that this self-similar nature of many fractals comes from the prescribed algorithm for construction of the fractal, and can actually be viewed as a transformation of some metric space. The Cantor Set, for example, can be viewed as the contraction and translation of an interval on the real line. This notion of transforming a metric space has proved extremely useful in multiple branches of mathematics and we will find it of particular use in the development of the space of fractals.

The Space of Fractals

Now that we understand some of the basic principles behind contraction mappings on general complete metric spaces, we may now introduce the metric space where fractals live, often called $(\mathcal{H}(X), h(d))$. Here $\mathcal{H}$ refers to the space of nonempty compact subsets of X, and $h(d)$ refers to the Hausdorff metric. In order to build up these two concepts, let us introduce the following definitions and theorems[4]:

> **Definition:** Let (X, d) be a complete metric space. Then $\mathcal{H}(X)$ denotes the space whose points are the compact subsets of X, other than the empty set.
>
> **Definition:** Let (X, d) be a complete metric space, $x \in X$ and $B \in \mathcal{H}(X)$. Let $d(x, B)$ be the distance from the point x to the set B, where
>
> $$d(x, B) = min\{d(x, y) : y \in B\}.$$
>
> **Definition:** Let (X, d) be a complete metric space, and $A, B \in \mathcal{H}(X)$.

Let $d(A, B)$ be the distance between the set A and the set B, where

$$d(A, B) = max\{d(x, B) : x \in A\}.$$

It is clear that the distance metric $d(x, B)$ is well-defined by the fact that the minimum can be viewed as the intersection of decreasing subsets, which must contain a single point since B is compact. The last definition requires careful consideration. Firstly, one must note that, as defined, the distance $d(A, B)$ does not constitute a metric. In particular one should observe

$$d(A, B) \neq d(B, A).$$

To provide an informal justification for this, let our two sets A and B be the United States of America (USA) and France.If we take Washington D.C. as our element x of the USA, then $d(x, France)$ is approximately the distance from Washington D.C. and the western most coastal city of France: Brest[5]. However, in evaluating $d(USA, France)$ then we want the distance between Brest, France and some city in the USA to be largest. We then see that

$$d(USA, France) \approx d(Seattle, Brest).$$

If instead we consider $d(France, USA)$, we would find that first we would pick the closest city in the USA (Danforth, Maine) to France, and then maximize the distance from that city, by picking the eastern most inland city of France (Strasbourg, France)[5]. We would then conclude that

$$d(France, USA) \approx d(Danforth, Strasbourg).$$

We now have our desired counterexample and can safely conclude that the distance function, as defined, *does not constitute a metric!* In particular,

$$d(USA, France) \neq d(France, USA).$$

Now that we have spent all this time developing these distance functions between sets, which sadly do not constitute a metric, we would like something which could operate as a metric between sets. Such a metric is termed the *Hausdorff distance* and is described below[4]:

Definition: Let (X, d) be a complete metric space. Then the *Hausdorff distance* between two points $A, B \in \mathcal{H}(X)$ is defined by

$$h(A, B) = max\{d(A, B), d(B, A)\}.$$

Theorem: The Hausdorff distance is a metric on $\mathcal{H}(X)$.

We will demonstrate that $h(d)$ suffices as a metric by verifying the three axioms needed for a distance function to qualify as a metric. Let $A, B, C \in \mathcal{H}(X)$. We then see that:

$h(A, A) = max\{d(A, A), d(A, A)\} = d(A, A) = max\{d(x, A) : x \in A\} = 0$. This proves that the distance between a set and itself is zero.

It is also clear that $h(A, B) = max\{d(A, B), d(B, A)\} = max\{d(B, A), d(A, B)\} = h(B, A)$. This gives us that distance between two points isn't dependent on the order in which we compare them.

To show that $h(A, B) \leq h(A, C) + h(C, B)$ we first show that $d(A, B) \leq d(A, C) + d(C, B)$. We see that for every $a \in A$:

$$\begin{aligned}
d(A, B) &= min\{d(a, b) : b \in B\} \\
&\leq min\{d(a, c) + d(c, b) : b \in B\} \forall c \in C \\
&= d(a, c) + min\{d(c, b) : b \in B\} \forall c \in C \\
d(a, b) &\leq min\{d(a, c) : c \in C\} + max\{min\{d(c, b) : b \in B\} : c \in C\} \\
&= d(a, C) + d(C, B) \\
d(A, B) &\leq d(A, C) + d(C, B)
\end{aligned}$$

We can similarly conclude that $d(B, A) \leq d(B, C) + d(C, A)$, and thus that:

$$\begin{aligned}
h(A, B) &= max\{d(A, B), d(B, A)\} \\
&\leq max\{d(A, C), d(C, A)\} + max\{d(B, C), d(C, B)\} \\
&= h(A, C) + h(C, B)
\end{aligned}$$

This shows that the triangle inequality holds, and now h satisfies the three axioms of a metric.

Transformations and Contraction Mappings

Spaces can be mapped to spaces, just as the function $f(x) = \frac{1}{2}x$ maps the unit interval to $[0, \frac{1}{2}]$. Before we develop fractals formally, we must first understand basic principles of how certain mappings work, and what a transformation is, and what it means for a mapping to have a fixed point.

We offer the following definitions[4]:

> **Definition:** Let (X, d) be a metric space. A *transformation* on X is a function $f : X \mapsto X$, which assigns exactly one point $f(x) \in X$ to each point $x \in X$.
>
> **Definition:** Let $f : X \mapsto X$ be a transformation on a metric space. The *forward iterates* of f are transformations $f^{\circ n} : X \mapsto X$ defined by $f^{\circ 0}(x) = x$, $f^{\circ 1}(x) = f(x)$,..., $f^{\circ(n+1)}(x) = f(f^{\circ n}(x))$ for $n = 0, 1, 2, ...$ If f is invertible, then the *backward iterates* of f are transformations defined analogously with $f^{\circ -m}(x) : X \mapsto X$.

Transformations, in general are an easy concept to understand, but their power should not be doubted. They can take on many forms, but we will be concerned with a specific type of mapping or transformation. We offer the following definition and theorem[6]:

> **Definition:** Let X be a metric space, with metric d. If φ maps X into X and if there is a number $0 < c < 1$ such that
>
> $$d(\varphi(x), \varphi(y)) \leq cd(x, y)$$
>
> for all $x, y \in X$, then φ is said to be a *contraction* of X into X.
>
> **Theorem:** If X is a complete metric space, and if φ is a contraction of X into X, then there exists one and only one $x \in X$ such that $\varphi(x) = x$.

We call x a *fixed point* if $\varphi(x) = x$. It is easy to see that this fixed point is unique. If there happened to be two fixed points $x, y \in X$ then

$$d(\varphi(x), \varphi(y)) \leq cd(x, y) \Rightarrow d(x, y) \leq cd(x, y)$$

is only true if

$$d(x, y) = 0 \Rightarrow x = y.$$

Although we have established that such a fixed point, if it exists, must be unique, the existence is still in question. We will show that such a fixed point can actually be constructed as follows:

Pick $x_0 \in X$ at random, and construct a sequence $\{x_n\}$ such that

$$x_{n+1} = \varphi(x_n) \qquad (n = 0, 1, 2, ...).$$

We know that since φ is a contraction, there exists a c such that for $n \geq 1$

$$d(x_{n+1}, x_n) = d(\varphi(x_n), \varphi(x_{n-1})) \leq cd(x_n, x_{n-1}).$$

By induction we observe that

$$d(x_{n+1}, x_n) \leq c^n d(x_1, x_0) \qquad (n = 0, 1, 2, ...).$$

If $n < m$, we have that

$$\begin{aligned} d(x_n, x_m) &\leq \textstyle\sum_{i=n+1}^{m} d(x_i, x_{i-1}) \\ &\leq (c^n + c^{n+1} + \cdots + c^{m-1}) d(x_1, x_0) \\ &\leq \tfrac{d(x_1, x_0)}{1-c} c^n. \end{aligned}$$

The last inequality satisfies the *Cauchy Criterion* and we call $\{x_n\}$ a Cauchy sequence in X. Since X is complete (by assumption) we know that $\lim_{n\to\infty} x_n = x$ for $x \in X$. Finally, since φ is a contraction, it is continuous and thus

$$\varphi(x) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

Contractions and Fractals

Now that we are intimately familiar with the space of fractals and contraction mappings, we are now in a position to define exactly what kind of set a fractal is. We shall call connect the idea of an *attractor* for a set of maps on a metric space with a *deterministic fractal*. Once we have proven some key results, we will be left with a picture of just how deterministic fractals are created, and what their connection to real world dynamics is. We offer the following set of lemmas, definitions and theorems[4]:

> **Lemma:** Let $w : X \mapsto X$ be a contraction mapping on the metric space (X, d). Then w is uniformly continuous.

Let $\epsilon > 0$ and $s > 0$ be a contractivity factor for w. Then there exists a $\delta = \frac{\epsilon}{s}$ such that when $d(x, y) < \delta$

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

Lemma: Let $w : X \mapsto X$ be a contraction mapping on the metric space (X, d). Then w maps $\mathcal{H}(X)$ into itself.

Let K be a nonempty compact subset of X. Since $\mathcal{H}(X)$ is the space whose points are all nonempty compact subsets of (X, d), the lemma follows immediately if we show that $w(K)$ is compact[6]. Let $\{V_\alpha\}$ be an open cover of $w(K)$. Since w is continuous we know that $w^{-1}(V_\alpha)$ is open. Since K is compact, we know that there exist finitely many indices $\alpha_1, ..., \alpha_n$ such that

$$K \subset w^{-1}(V_{\alpha_1}) \cup \cdots \cup w^{-1}(V_{\alpha_n}).$$

Since $w(w^{-1}(E)) \subset E$ for every $E \subset X$, we have then our result that every open cover of $w(S)$ can be covered by a finite subcover:

$$w(S) \subset V_{\alpha_1} \cup \cdots \cup V_{\alpha_n}.$$

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

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

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

From the previous two lemmas we know that w is continuous and it maps $\mathcal{H}(X)$ into itself. Now consider $B, C \in \mathcal{H}(X)$, then

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

We can similarly conclude that $d(w(C), w(B)) \leq sd(C, B)$, and thus that:

$$\begin{aligned} h(w(B), w(C)) &= max\{d(w(B), w(C)), d(w(C), w(B))\} \\ &\leq max\{sd(B, C), sd(C, B)\} \\ &= sh(B, C) \end{aligned}$$

Proposition: For all $B, C, D, E \in \mathcal{H}(X)$ and h the Hausdorff metric:

$$h(A \cup C, D \cup E) \leq max\{h(B, D), h(C, E)\}$$

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

$$\begin{aligned} W(B) &= w_1(B) \cup w_2(B) \cup \cdots \cup w_N(B) \\ &= \textstyle\bigcup_{n=1}^{N} w_n(B) \qquad \forall B \in \mathcal{H}(X) \end{aligned}$$

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

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

$$\begin{aligned} h(W(B), W(C)) &= h(w_1(B) \cup w_2(B), w_1(C) \cup w_2(C)) \\ &\leq max\{h(w_1(B), w_1(C)), h(w_2(B), w_2(C))\} \\ &\leq max\{s_1 h(B, C), s_2 h(B, C)\} \\ &\leq sh(B, C) \end{aligned}$$

Now that we understand some of the formal notions about contraction mappings on the space of fractals, we can develop what is meant by an *iterated function system*. The key notion in the next few definitions and theorems will be to realize that in order to create a fractal object, whether it be the Cantor set or the Sierpinski triangle, one merely starts with some nonempty compact set, twist, turn, and shift pieces in some prescribed order, then iterate that same process until one reaches a level of detail desired. To make this notion precise we offer the following definiton and theorems[4]:

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

Theorem: Let $\{X; w_n, n = 1, 2, ..., N\}$ be an IFS with contractivity factor s. Then the transformation $W : \mathcal{H}(X) \mapsto \mathcal{H}(X)$ defined by

$$W(B) = \bigcup_{n=1}^{N} w_n(B) \qquad \forall B \in \mathcal{H}(X)$$

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

$$h(W(B), W(C)) \leq sh(B, C) \qquad \forall B, C \in \mathcal{H}(X).$$

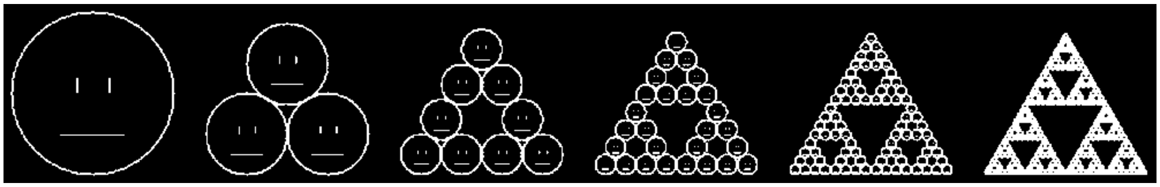

Figure 2: A smiley face under the influence of an IFS

Its unique fixed point, $A \in \mathcal{H}(X)$ satisfies the condition

$$A = W(A) = \bigcup_{n=1}^{N} w_n(A).$$

and is given by

$$\lim_{n\to\infty} W^{\circ n}(B) \qquad \forall B \in \mathcal{H}(X).$$

Definition: The fixed point A, described above, is called an *attractor* of the IFS.

The notion of an attractor, as presented, can be taken as a definition of determinsitic fractals. It is important to understand the power of this characterization. A deterministic fractal and be generated by *any* nonempty compact set, and then by iterating a finite set of contraction mappings, we can generate our fractals. The figure demonstrates how a smiley-face can be used as our initial set, and applying a prescribed sequence of contraction mappings, the Sierpinski triangle is created.

Fractional Dimensions

We have made considerable progress in the formal development of deterministic fractals, but the take-home message has not been delivered. It would be terrible if our best answer to the question "What is a fractal?" would go something like "Oh, it's the attractor of an iterated function system on some compact subset of a complete metric space." Rather, we would like to connect our current understanding of fractals with the notion of a *fractional number of dimensions.*

The intuitive notion of dimension, seems so deeply ingrained in us that defining it appears to be nonsense. This intuition is so deeply held that the concept of a non-integer dimension can be mind-boggling. However, if we give our concept of dimension a formal definition, we will see fractals nicely fall out.

In 1977, Benoit Mandlbrot introduced notion of sets with fractional number of dimensions and called them fractals. The definition provided is actually a simplification of a more formal concept of *Hausdorff Dimension* that Mandelbrot used to base his definition on. We will focus on this definition as it is more intuitive than the rigorous definition offered by Hausdorff[4].

> **Definition:** Let $A \in \mathcal{H}(X)$ where (X, d) is a metric space. For each $\epsilon > 0$ let $\mathcal{N}(A, \epsilon)$ denote the smallest number of closed balls of radius $\epsilon > 0$ needed to cover A. If
>
> $$D = \lim_{\epsilon \to 0} \frac{\log \mathcal{N}(A, \epsilon)}{\log \frac{1}{\epsilon}}$$
>
> exists, then D is called the *fractal dimension* of A.

In order to gain an intuitive understanding of how this definition could be possibly capture our understanding of dimension, consider the following series of observations:

If you have a line and you double it, you have a line twice as long, in which you have two copies of the original line. If you have a square and you double all of its sides, you have four times the area of the original square, and thus four copies of the original square. If you have a cube you can double each of its sides and then you have an object with eight times the volume of the original cube. It appears that the following pattern emerges:

$$2 = 2^1 \qquad 4 = 2^2 \qquad 8 = 2^3 \qquad N = 2^d$$

Where N refers to the number of copies of the original object, and d refers to the dimension of the object. If we then consider an such as the Sierpinski triangle, we see that if we double the length of each leg of the triangle, removing the middle portion leaves us with 3 copies of the original Sierpinski triangle. Our generalization leads us to believe that the dimension of the Sierpinski triangle must satisfy the following series of equations:

$$3 = 2^d \Rightarrow d = \frac{\log 3}{\log 2} = 1.58496...$$

One might object that the situation so far considered has only a superficial relationship with the definition provided. The definition speaks of balls and covering sets, where we have been speaking vaguely about *copies of the original*. The difference can be reconciled quite easily. We cannot speak of *doubling* a set and asking how many copies of the original set are in the old one, this would greatly sacrifice the kind of generality and accuracy that we require. Instead imagine, we keep our original line and instead we try to cover it by a ball (a circle in this case) of radius equal to have the length of the line. It clearly requires only one circle to do this. Now let us reduce our circle's radius by a half, then it takes two circles, by a fourth, four circles, and so on. For a square, reducing the radius of a circle originally covering the entire square by a half means we need four circles to cover the entire square, and so on. For the case of the cube we imagine a sphere (a literal ball for once!) covering the original sphere and having its radius shrunk progressively. This algorithm is extremely useful for calculating the fractal dimension of almost any object and when boxes serve the role of balls, we have the following theorem[4]:

> **The Box Counting Theorem:** Let $A \in \mathcal{H}(\mathbb{R}^m)$, where the Euclidean metric is used. Cover $\mathbb{R}^m$ by closed just-touching square boxes of side length $\frac{1}{2^n}$. Let $\mathcal{N}_n(A)$ denote the number of boxes of side length $\frac{1}{2^n}$ which intersect the attractor. If
>
> $$D = \lim_{n\to\infty} \frac{\log \mathcal{N}_n(A)}{\log 2^n}$$
>
> exists, then D is called the *fractal dimension* of A.

One can think of this algorithm as placing your object on a grid, counting the number of grid-boxes necessary to cover the entire object. We then increase the resolution on our grid by decreasing the spacing of the grid by a half, and then repeating this cycle while counting the number of boxes necessary to cover the object, during each iteration.The figure demonstrates a version of this procedure that doesn't use a grid as visual crutch, but rather just decreases the size of the boxes and then covers the object. In the figure involving Koch's curve, the geometry suggests that we decrease the size of our squares by a $\frac{1}{3}$. Thus for our first iteration, we require 3 squares, then 12, then 48, following the relationship that logarithm of our number of boxes, divided by their inverse size approaches $\frac{\log 4}{\log 3} = 1.26185...$ The calculation of a fractional number of dimensions is now no longer a theoretical possibility, but a practical algorithm, applied in numerous situations[8].

Conclusions and Applications

Fractals are beautiful mathematical creatures which have played an important role in analysis for over a century. However the mathematical principles behind deterministic

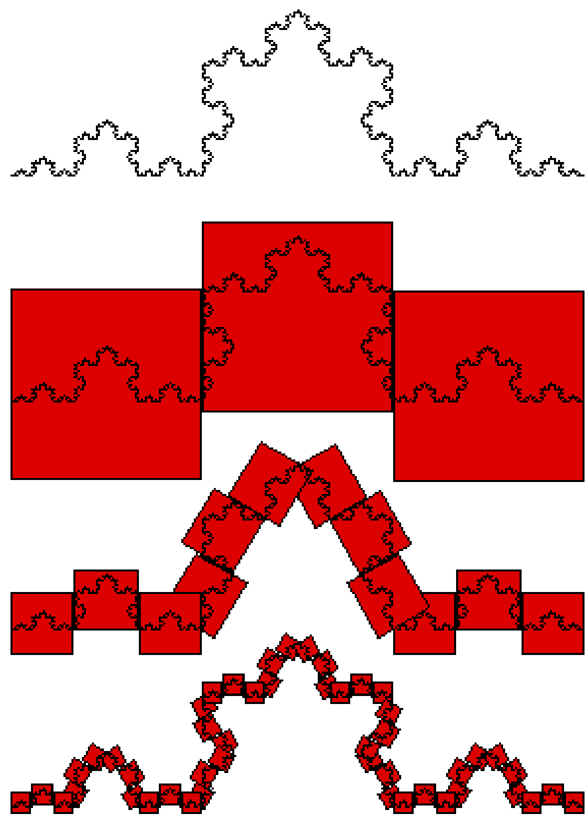

Figure 3: Box Counting Dimension of Koch's Curve

fractals have only recently been understood. In order to actually build up a fractal as simple as the Sierpinski triangle, we had to cover a great deal of mathematical territory. We first introduced the notion of a space of fractals, with a metric known as the Hausdorff distance. From there we considered general transformations on general metric spaces, focusing in particular on contractive mappings. All the ingredients were then in place to cook up the notion of a sequence of contractive mappings known as an iterated function system or IFS. Finally we were able to characterize deterministic fractals as the attractor of an IFS. This understanding of fractals, although technically correct, is intuitively unsatisfying, so a weaker notion of the Hausdorff dimension was introduced in order to understand fractals as objets with a non-integer number of dimensions. The notion of box-counting has provided a very practical method for calculating the dimension of a wide range of objects.

However these pursuits would only subsist as mathematical oddities if it weren't for a boom in research inspired by Benoit Mandelbrot. His understanding of the world told him that a fractal would mathematically model a cloud better than a circle or other smooth object every could. After developing many of the tools presented above, researchers were in a unique position to ask seemingly strange questions about the fractal nature of the world. The question of "How long is the Coast of Britain" may seem trivial, but when one considers the scale of our measurement the question is not so easy. From an airplane one could make a rough approximation, but if a beetle were to hung the shore, following, every nuance, every stream and river as it runs deep into the British Isle, the measured length tends toward infinity. Characterizing the coastline as having a fractal dimension of 1.2, proved to be a more accurate answer to the question. By the late 70s, it became clear that fractal geometry is a powerful tool for understanding the structure of everyday things

The geometry proved useful, but its relationship with dynamics proved even more profound. While modelling certain physical systems with differential equations, some researchers found that the phase space of *physical, real-world systems*, could only be characterized by "stange attractors" and the connection with fractals was made. The complexity revolution collided with full force into the fractal revolution, and people began to see fractals everywhere.

References

[1] Karl Weierstrass "On Continuous Functions of a Real Argument that do not have a Well-Defined Differential Quotient" in *Classics on Fractals* (Westview Press, 2004) pp.3-9.

[2] Helge Von Koch "On a Continuous Curve without Tangent Constructible from Elementary Geometry" in *Classics on Fractals* (Westview Press, 2004) pp.25-45.

[3] Georg Cantor "On the Power of Perfect Sets of Points" in *Classics on Fractals* (Westview Press, 2004) pp.11-23.

[4] Michael Barnsley, *Fractals Everywhere* (Academic Press, 1988)

[5] Map of France. *Wikipedia* (http://en.wikipedia.org/wiki/Image:Fr-map.png)

[6] Walter Rudin *Principles of Mathematical Analysis, 3rd Ed.* (McGraw-Hill, 1976) pp.220-221

[7] IFS on a Smiley Face (http://www.sissa.it/~jrlaguna/lamaga/schain.html)

[8] Box Dimension of Koch's Curve (http://classes.yale.edu/fractals/FracAndDim/BoxDim/BoxDim.html)