Exploration of the Fractal Dimension

André Khalil

A Thesis

in

The Department

of

Mathematics & Statistics

Presented in Partial Fulfilment of the Requirements

for the Degree of Master of Science at

Concordia University

Montréal, Québec, Canada

August 1998

0-612-39464-6

ABSTRACT

Exploration of the Fractal Dimension

André Khalil

This thesis studies the theoretical and experimental determination of the fractal dimension of different sets. It contains both pure and applied topics. After establishing the ground rules, we look at ways to calculate the fractal dimension of fractal interpolation functions and we end the study with an experiment consisting in the estimation of the fractal dimension of two graphs representing the activities of two companies at the Toronto Stock Exchange. In fact, we will see that there is a certain cohesion between the fractal dimension of a cloud of points and the absolute value of the slope of the regression line approximating the same cloud of points.

ACKNOWLEDGEMENTS

I wish to thank my family and my friends for their support throughout my studies. Also many thanks to my supervisor, Dr. Pawel Góra, for his constant support and encouragement and who always finds a way to make problems seem simpler.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figures

Tables

CHAPTER I: Fractal Dimension

After introducing our subject, we begin by elaborating on what a fractal is, where it lives (in what space). This is followed by the definition of the fractal dimension and its theoretical and experimental determination. Chapter one ends with an overview of the Hausdorff-Besicovitch fractal dimension.

I.1 What is a Fractal?

To many, geometry is the study of the objects and shapes that surround us in nature. It is important to study the circle, the cylinder and the sphere as they are figures that can look like a water puddle, a tree trunk, and the earth or the moon. It is always a good idea to have a model of the objects that we are studying.

But in nature, the vast majority of forms and shapes do not look at all like a cube, a cone, or an ellipse... Assuming that standard geometry studies objects of one, two, and three dimensions, *Fractal Geometry* can be seen as an extension of the standard geometry, where shapes and objects may have a non-integer dimension.

For example, we know that a straight line has dimension one, and that a plane has dimension two. If we consider the line, or should we say, the figure, corresponding to the boundary of a cloud, or the outline of a rock, the dimension of this figure will be between one and two. If we zoom in very closely on this figure, it will never be piecewise linear. So the dimension is more than one, and on the other hand, it is less than two as it does not fill the whole area of the plane.

A figure such as the one mentioned above could be called a *Fractal*. We will avoid attempting to give a formal definition of a fractal. If needed, we could define a *deterministic fractal.* The space where fractals are found, $\mathcal{H}(X)$, is the collection of the nonempty compact subsets of the complete metric space X. In order to define a deterministic fractal, we could say that it is the fixed-point of a contraction transformation on $\mathcal{H}(X)$ (with the Hausdorf metric). But by doing so, we would require that the underlying metric space (X, d) be 'geometrically simple'.

I assume the reader is familiar with basic analysis and topics such as compactness, completeness, balls, metrics, etc...

I.2 Fractal Dimension

We consider the complete metric space (X, d) (in most cases considered in the theory and for the experiment, X will denote the real plane $\Re^2$ and d will represent the Euclidean metric). Consider $A \in \mathcal{H}(X)$. A is a nonempty compact subset of X. Next, let $B(x, \epsilon)$ denote a closed ball of radius $\epsilon > 0$, centered at $x \in X$. Finally, let $\mathcal{N}(A, \epsilon)$ be the smallest integer M such that

$$A \subset \bigcup_{i=0}^{M} B(x_i, \epsilon)$$

(union of balls of radius ϵ). So $\mathcal{N}(A, \epsilon)$ is the least number of radius-ϵ balls needed to cover A. The existence of $\mathcal{N}(A, \epsilon)$ is a consequence of the compactness of A (as any cover of A by open sets will possess a finite subcover).

Definition I.2.1 Let (X, d) be a metric space. Let $A \in \mathcal{H}(X)$. $D = D(A)$ is called the *Fractal Dimension* of A, and

$$D = \lim_{\epsilon \to 0} \frac{\ln(\mathcal{N}(A, \epsilon))}{\ln(1/\epsilon)},$$

when the limit exists.

So in the real plane with the Eucledian metric, if we consider a single $a \in X$, so that $A = \{a\}$, then $D(A) = 0$. Also, if (X, d) is a metric space and we have $a, b, c \in X$ so that $A = \{a, b, c\}$, then $D(A) = 0$. The fractal dimension of a closed line segment is 1. After the following two theorems (I.2.1 and I.2.2), we will see other examples for A, and the tools given by the theorems will help us find $D(A)$.

In Def. I.2.1, we can replace ϵ, which can be seen as a continuous variable, by a discrete variable $\epsilon_n = Cr^n$, where $C > 0$, and in the case of the following theorem, where $0 < r < 1$.

Theorem I.2.1 Let (X, d) be a metric space. Let $A \in \mathcal{H}(X)$. Let $\epsilon_n = Cr^n$, where $C > 0$, $n = 1, 2, 3, \ldots$. If

$$D = \lim_{n \to \infty} \frac{\ln(\mathcal{N}(A, \epsilon_n))}{\ln(1/\epsilon_n)},$$

then $D = D(A)$.

Proof ([B]): Let us define a function $f(\epsilon) = \max\{\epsilon_n : \epsilon_n \leq \epsilon\}$. So $f(\epsilon) = \epsilon_{n-1}$, for $\epsilon_n \leq \epsilon \leq \epsilon_{n-1}$. Assume that $\epsilon \leq r$. Then

$$f(\epsilon) \leq \epsilon \leq f(\epsilon)/r,$$

and

$$\mathcal{N}(A, f(\epsilon)) \geq \mathcal{N}(A, \epsilon) \geq \mathcal{N}(A, f(\epsilon)/r).$$

(Note: $\ln(x)$ is an increasing positive function of x for $x \geq 1$.) Then, we have

$$\frac{\ln(\mathcal{N}(A, f(\epsilon)/r))}{\ln(1/f(\epsilon))} \leq \frac{\ln(\mathcal{N}(A, \epsilon))}{\ln(1/\epsilon)} \leq \frac{\ln(\mathcal{N}(A, f(\epsilon)))}{\ln(r/f(\epsilon))}.$$

Let us assume that $\mathcal{N}(A, \epsilon) \to \infty$ as $\epsilon \to 0$; if not, then the theorem is true. (Setting up the Sandwich Theorem of Calculus). From the right-hand side inequality, we have;

$$\lim_{\epsilon \to 0} \frac{\ln(\mathcal{N}(A, f(\epsilon)))}{\ln(r/f(\epsilon))} = \lim_{n \to \infty} \frac{\ln(\mathcal{N}(A, \epsilon_n))}{\ln(r/\epsilon_n)} = \lim_{n \to \infty} \frac{\ln(\mathcal{N}(A, \epsilon_n))}{\ln(r) + \ln(1/\epsilon_n)} =$$

$$= \lim_{n\to\infty} \frac{\ln(\mathcal{N}(A,\epsilon_n))}{\ln(1/\epsilon_n)} = D.$$

And from the left-hand side inequality, we have;

$$\lim_{\epsilon\to 0} \frac{\ln(\mathcal{N}(A,f(\epsilon)/r))}{\ln(1/f(\epsilon))} = \lim_{n\to\infty} \frac{\ln(\mathcal{N}(A,\epsilon_{n-1}))}{\ln(1/\epsilon_n)} = \lim_{n\to\infty} \frac{\ln(\mathcal{N}(A,\epsilon_{n-1}))}{\ln(1/r)+\ln(1/\epsilon_{n-1})} =$$
$$= \lim_{n\to\infty} \frac{\ln(\mathcal{N}(A,\epsilon_n))}{\ln(1/\epsilon_n)} = D.$$

So as $\epsilon \to 0$, both the left-hand side and the right-hand side of the above inequalities approach the same value, claimed in the theorem. End of proof.

The next theorem (the box-counting theorem) gives us a very practical method to estimate the fractal dimension of a set. Instead of covering the set A with balls of radius ϵ (from Def. I.2.1), or radius Cr^n (from Thm I.2.1), we cover it with square boxes of side length $1/2^n$.

Theorem I.2.2 Let $A \in \mathcal{H}(\Re^m)$, (Euclidean Metric). Cover $\Re^m$ by the mesh of closed square boxes of side length $1/2^n$. Let $\mathcal{N}_n(A)$ denote the number of boxes which intersect the attractor A. If

$$D = \lim_{n\to\infty} \frac{\ln(\mathcal{N}_n(A))}{\ln(2^n)},$$

then $D = D(A)$.

Proof: First, let us consider the following statements:

1. A circle of radius $1/2^n$ intersects at most 4 squares of side length $1/2^{n-1}$.

2. (Generalization) A ball of radius $1/2^n$ intersects at most 2^m boxes of side length $1/2^{n-1}$.

Statement 2 tells us that $\#\text{balls} \geq \frac{\#\text{squares}}{2^m}$, which gives us;

$$\mathcal{N}(A, 1/2^n) \geq 2^{-m}\mathcal{N}_{n-1}(A).$$

Let us put a square in a circle (or a box in a ball). If we let x be the side length of the box and r, the radius of the ball, then we have

$$(x/2)^2 + (x/2)^2 = r^2 \qquad , m = 2,$$

$$(x/2)^2 + (x/2)^2 + (x/2)^2 = r^2 \qquad , m = 3,$$

$$(...)$$

$$(x/2)^2 + (x/2)^2 + ... + (x/2)^2 = q(x/2)^2 = r^2 \qquad , m = q,$$

So if the dimension is m, we have;

$$m(x/2)^2 = r^2, \tag{2.2.1}$$

and

$$x = \frac{2r}{\sqrt{m}}.$$

If $x \leq \frac{2r}{\sqrt{m}}$, the box of side length x will fit inside a ball of radius r. Thus we can say that $\mathcal{N}(A, 1/2^n) \leq \mathcal{N}_{k(n)}(A)$ for some $k(n)$. What is $k(n)$? We can use equation (2.2.1) in the following way: For a fixed n, what $k(n)$ will satisfy

$$m(\frac{1/2^{k(n)}}{2})^2 \leq (1/2^n)^2?$$

If we solve for $k(n)$, we get

$$k(n) \geq (1/2)\log_2 m + n - 1.$$

Then,

$$2^{-m}\mathcal{N}_{n-1}(A) \leq \mathcal{N}(A, 1/2^n) \leq \mathcal{N}_{k(n)}(A), \qquad m, n = 1, 2, 3, ...$$

and

$$\frac{2^{-m}\mathcal{N}_{n-1}(A)}{2^n} \leq \frac{\mathcal{N}(A, 1/2^n)}{2^n} \leq \frac{\mathcal{N}_{k(n)}(A)}{2^n}.$$

(Setting up the Sandwich Theorem). From the left hand side inequality, we have;

$$\lim_{n\to\infty} \frac{\ln(2^{-m}\mathcal{N}_{n-1}(A))}{\ln(2^n)} = \lim_{n\to\infty} \frac{\ln(2^{-m}) + \ln(\mathcal{N}_{n-1}(A))}{\ln(2) + \ln(2^{n-1})} = D.$$

And from the right-hand side inequality, we have;

$$\lim_{n\to\infty} \frac{\ln(\mathcal{N}_{k(n)}(A))}{\ln(2^n)} = \lim_{n\to\infty} \frac{\ln(\mathcal{N}_{k(n)}(A))}{\ln(2^{k(n)})} . \frac{\ln(2^{k(n)})}{\ln(2^n)} = \lim_{n\to\infty} \frac{\ln(\mathcal{N}_{k(n)}(A))}{\ln(2^{k(n)})} = D.$$

Note that as $n \to \infty$, we have $\frac{k(n)}{n} \to 1$, and therefore, $\frac{\ln(2^{k(n)})}{\ln(2^n)} \to 1$, justifying the last step. By the Sandwich Theorem we have the desired result. End of proof.

As an example, we can consider A = 'Unit Square' in the real plane. A is the square of side length 1 located at coordinates $\{(0,0),(0,1),(1,0),(1,1)\}$. As n increases, the side length of each box is getting smaller and smaller. In fact, each time n increases by one, all the needed boxes to cover A get divided into 4 boxes of half side length. So the number of boxes $\mathcal{N}_n(A)$ will always be 4^n. And $\frac{\ln(4^n)}{\ln(2^n)} = 2$ (for any n). Hence D(Unit Square) $= 2$.

Similarly, let us consider the Sierpinski triangle in the real plane. The Sierpinski triangle (sometimes called the Sierpinski gasket) is a triangle that we divide into 4 smaller triangles, and we cut out the middle one (Fig. I.1). In this case, each time n increases by one, only 3 out of a possible 4 covering square boxes are needed to cover A. So $\mathcal{N}_n(A) = 3^n$. And $\ln(3^n)/\ln(2^n) = \ln 3/\ln 2$ (for any n). Hence D(Sierpinski triangle) $= \ln 3/\ln 2 \approx 1.58$.

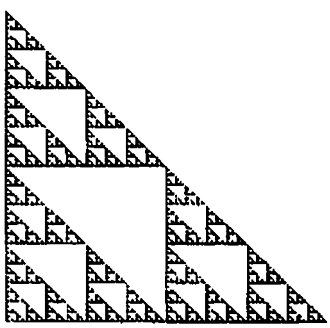

FIGURE I.1: Sierpinski's Gasket.

It is important to mention here that although the Box-Counting Theorem is set up for square boxes of side length $1/2^n$, using $c/2^n$, where c is a constant, will still work. In Chapter III $c = E - S$ ((E)nd - (S)tart) will measure an interval of time (in minutes).

Definition I.2.2 Two metrics d_1 and d_2 on a space X are *equivalent* if there exist constants $0 < c_1 < c_2 < \infty$ such that

$$c_1 d_1(x,y) \le d_2(x,y) \le c_2 d_2(x,y), \qquad \forall (x,y) \in X^2.$$

In the following theorem, we will see that metrically equivalent sets have the same fractal dimension.

Theorem I.2.3 Let (X_1, d_1) and (X_2, d_2) be metrically equivalent, and let $\theta : X_1 \to X_2$ where θ provides the equivalence of the spaces. Take $A_1 \in \mathcal{H}(X_1)$, and let $D(A_1) = D$. If $A_2 = \theta(A_1)$ then $D(A_2) = D$. That is, $D(A_1) = D(\theta(A_1))$.

Proof: ([B]) As (X_1, d_1) and (X_2, d_2) are equivalent under θ, there exist positive constants e_1 and e_2 such that

$$e_1 d_2(\theta(x), \theta(y)) < d_1(x,y) < e_2 d_2(\theta(x), \theta(y)) \qquad \forall x, y \in X_1.$$

We may assume that $e_1 < 1 < e_2$, giving;

$$d_2(\theta(x), \theta(y)) \le \frac{d_1(x,y)}{e_1} \qquad \forall x, y \in X_1.$$

This implies that

$$\theta(B(x,\epsilon)) \subset B(\theta(x), \epsilon/e_1) \qquad \forall x \in X_1. \tag{2.3.1}$$

Now, considering $\mathcal{N}(A_1, \epsilon)$, we can find $\{x_1, x_2, ..., x_N\} \subset X_1$, where $N = \mathcal{N}(A_1, \epsilon)$, such that the closed balls $\{B(x_n, \epsilon) : n = 1, 2, ..., N)\}$ cover A_1. It follows that $\{\theta(B(x_n, \epsilon)) : n =$

$1, 2, ..., N\}$ covers A_2. But equation (2.3.1) implies that $\{B(\theta(x_n), \epsilon/e_1) : n = 1, 2, ..., N\}$ also covers A_2. Hence;

$$\mathcal{N}(A_2, \epsilon/e_1) \leq \mathcal{N}(A_1, \epsilon),$$

so that when $\epsilon < 1$,

$$\frac{\ln(\mathcal{N}(A_2, \epsilon/e_1))}{\ln(1/\epsilon)} \leq \frac{\ln(\mathcal{N}(A_1, \epsilon))}{\ln(1/\epsilon)},$$

from which it follows that:

$$\overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_2, \epsilon))}{\ln(1/\epsilon)} = \overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_2, \epsilon/e_1))}{\ln(1/\epsilon)} \leq \lim_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_1, \epsilon))}{\ln(1/\epsilon)} = D(A_1).$$

(There is no 'sup' on the right-hand side limit as $D(A_1)$ was assumed to exist, and so $\overline{\lim}$ and $\underline{\lim}$ coincide.) Now let us find an inequality in the opposite direction. By altering the inequality given for the definition of equivalent spaces, we can say;

$$d_1(\theta^{-1}(x), \theta^{-1}(y)) \leq e_2 d_2(x, y) \qquad \forall x, y \in X_2.$$

This tells us that

$$\theta^{-1}(B(x, \epsilon)) \subset B(\theta^{-1}(x), e_2\epsilon)) \qquad \forall x \in X_2,$$

Hence;

$$\mathcal{N}(A_1, e_2\epsilon) \leq \mathcal{N}(A_2, \epsilon),$$

so that when $\epsilon < 1$,

$$\frac{\ln(\mathcal{N}(A_1, e_2\epsilon))}{\ln(1/\epsilon)} \leq \frac{\ln(\mathcal{N}(A_2, \epsilon))}{\ln(1/\epsilon)},$$

from which it follows that

$$\underline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_2, \epsilon))}{\ln(1/\epsilon)} \geq \lim_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_1, e_2\epsilon))}{\ln(1/\epsilon)} = \lim_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_1, \epsilon))}{\ln(1/\epsilon)} = D(A_1).$$

So we can combine our inequalities to obtain;

$$\underline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_2, \epsilon))}{\ln(1/\epsilon)} = D(A_1) = \overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A_2, \epsilon))}{\ln(1/\epsilon)}$$

And finally;

$$D(A_2) = \lim_{\epsilon \to 0} \frac{\ln(\mathcal{N}(A_2, \epsilon))}{\ln(1/\epsilon)} = D(A_1).$$

Thus we get $D(\theta(A_1)) = D(A_1)$. End of proof.

Let C denote the Cantor set in [0,1]. A Cantor set can be seen as a set consisting of repeatedly removed middle third line segments (Fig. I.2). Let $\hat{C}$ denote the Cantor set in [0,3]. Then C and $\hat{C}$ are metrically equivalent, and Theorem I.1.3 says that $D(C) = D(\hat{C})$.

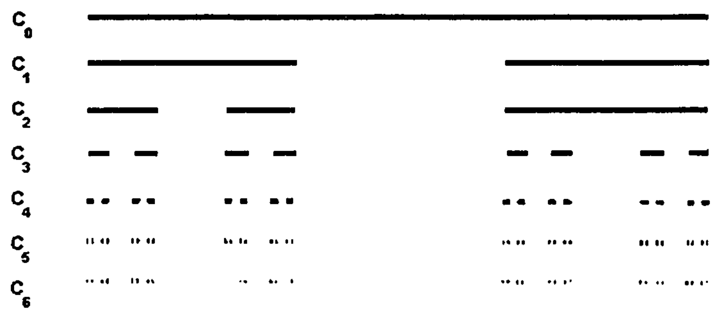

FIGURE I.2: The Cantor Set.

We define the Manhattan metric in the following way: For any points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in the space $X = \Re^2$, $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$. When using this metric, the path joining two points will consist only of horizontal and vertical lines, thus justifying the name 'Manhattan' for the comparison of the path with the streets of a city.

Theorem I.2.3 says that for a set $A \in \mathcal{H}(\Re^2)$, if we let $D(A) = D_1$ in the Manhattan metric, and $D(A) = D_2$ in the Euclidean metric, then as the two metrics are equivalent, $D_1 = D_2$. To show that these metrics are equivalent, we can consider d_1 to be the Manhattan metric and d_2 to be the Euclidean metric, and use, for instance,

$c_1 = 1/2$ and $c_2 = 2$ in definition I.2.2.

I.3 The Theoretical Determination of the Fractal Dimension

Let us start this section by giving a broader definition of $D(A)$, the fractal dimension of A. This new definition provides a fractal dimension in some cases where the previous definition makes no assertion.

Definition I.3.1 Let (X, d) be a complete metric space. Let $A \in \mathcal{H}(X)$. Let $\mathcal{N}(\epsilon)$ be the smallest number of balls of radius ϵ needed to cover A. If

$$D = \overline{\lim}_{\epsilon \to 0} \frac{\ln(\mathcal{N}(\hat{\epsilon}))}{\ln(1/\hat{\epsilon})},$$

where $\hat{\epsilon} \in (0, \epsilon)$, then $D = D(A)$ and D is called the *Fractal Dimension* of A. (Note that the limit has to exist for D to equal $D(A)$.)

All of our theorems apply with either definition I.2.1 or I.3.1.. Indeed, one can show that if a set has fractal dimension D according to definition I.2.1 then it has the same dimension according to definition I.3.1.. We use the broader definition I.3.1 in proving Theorem I.3.1, which states an intuitively logical result, namely that a set containing another set has higher fractal dimension than that of the contained set.

Theorem I.3.1 Consider $\Re^m$ with the Euclidean metric. Assume that for any $A \in \mathcal{H}(\Re^m)$, $D(A)$ exists. Let $B \in \mathcal{H}(\Re^m)$ be such that $A \subset B$. Then $D(A) \leq D(B)$. In particular, $0 \leq D(A) \leq m$.

Proof ([B]) (for $m = 2$): We are in $\Re^2$. We can assume that $A \subset$ (Unit Square.).

This implies that

$$\mathcal{N}(A,\epsilon) \leq \mathcal{N}(\text{Unit Square}, \epsilon) \qquad \forall \epsilon > 0.$$

We can also assume that $\epsilon < 1$. We get

$$0 \leq \frac{\ln(\mathcal{N}(A,\epsilon))}{\ln(1/\epsilon)} \leq \frac{\ln(\mathcal{N}(\text{Unit Square},\epsilon))}{\ln(1/\epsilon)}$$

and

$$\overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A,\epsilon))}{\ln(1/\epsilon)} \leq \overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(\text{Unit Square},\epsilon))}{\ln(1/\epsilon)} = 2.$$

So we have $0 \leq D(A) \leq 2$.

Now if $A, B \in \mathcal{H}(\Re^2)$ (where $A \subset B$), then $D(A)$ and $D(B)$ both exist. So replacing (Unit Square) by the set B will show that $D(A) \leq D(B)$. End of proof.

Note on the above proof (of Thm I.3.1): Going from $m = 2$ to $m = k$ can be done easily as the 2-dimensional unit square has the same characteristics as the k-dimensional cube.

Theorem I.3.2 Consider $\Re^m$ with the Euclidean metric. Let $A, B \in \mathcal{H}(\Re^m)$, and let $D(A)$ be defined as in Def. I.2.1. Suppose $D(B) < D(A)$. Then $D(A \cup B) = D(A)$.

Proof: ([B]) Let us consider the desired equality as two inequalities, and let us get rid of the first part by saying that $D(A \cup B) \geq D(A)$, which follows from Thm I.3.1.

Part 2: (Show that $D(A \cup B) \leq D(A)$). $\forall\ \epsilon > 0$, we have $\mathcal{N}(A \cup B, \epsilon) \leq \mathcal{N}(A,\epsilon) + \mathcal{N}(B,\epsilon)$. This implies

$$D(A \cup B) = \overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A\cup B,\epsilon))}{\ln(1/\epsilon)} \leq \overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A,\epsilon)+\mathcal{N}(B,\epsilon))}{\ln(1/\epsilon)} \leq$$

$$\leq \overline{\lim}_{\epsilon\to 0}\frac{\ln(\mathcal{N}(A,\epsilon))}{\ln(1/\epsilon)} + \overline{\lim}_{\epsilon\to 0}\frac{\ln(1+\frac{\mathcal{N}(B,\epsilon)}{\mathcal{N}(A,\epsilon)})}{\ln(1/\epsilon)}.$$

So,

$$D(A \cup B) \leq D(A) + \overline{\lim}_{\epsilon \to 0} \frac{\ln(1 + \frac{\mathcal{N}(B,\epsilon)}{\mathcal{N}(A,\epsilon)})}{\ln(1/\epsilon)}.$$

We want to have the above $\overline{\lim}$ equal to 0. To do it, we want to show that;

$$\lim_{\epsilon \to 0^+} \frac{\mathcal{N}(B,\epsilon)}{\mathcal{N}(A,\epsilon)} < 1$$

Notice that if we define the function g in the following way;

$$g(\epsilon) = \sup_{\hat{\epsilon} < \epsilon} \frac{\ln(\mathcal{N}(B,\hat{\epsilon}))}{\ln(1/\hat{\epsilon})},$$

we can see that g is a decreasing function of the positive variable ϵ. So for a sufficiently small $\epsilon > 0$,

$$\Rightarrow \frac{\ln(\mathcal{N}(B,\epsilon))}{\ln(1/\epsilon)} < D(A),$$

$$\Rightarrow \frac{\ln(\mathcal{N}(B,\epsilon))}{\ln(1/\epsilon)} < \frac{\ln(\mathcal{N}(A,\epsilon))}{\ln(1/\epsilon)},$$

$$\Rightarrow \frac{\ln(\mathcal{N}(B,\epsilon))}{\ln(\mathcal{N}(A,\epsilon))} < 1.$$

End of proof.

Consider a kiwi fruit. Try to picture the kiwi as two images; one is the fruit itself, but without its little hairs, and the other image is the hairs only. Let A represent the whole kiwi, let K_b denote the image representing the 'bald' kiwi, and let K_h denote the image representing the 'hairs' of the kiwi. Theorem I.3.2 says that since the contribution to $\mathcal{N}(A,\epsilon)$ of the hairs, K_h, becomes exponentially small compared to the contribution from the bald kiwi, K_b. The condition $D(K_h) < D(K_b)$ holds and $D(A) = D(K_b \cup K_h) = D(K_b) = D(\text{bald kiwi})$. Other examples: consider a peach and its tiny hairs, or a cactus and its needles.

We will soon see a very important and useful theorem (I.3.4) which provides the fractal dimension of an important class of iterated function systems (IFS). Before the theorem, we formally define what is an IFS, an attractor and the three classes of IFS's.

Definition I.3.2 A *(hyperbolic) iterated function system (IFS)* consists of a complete metric space (X, d) with a finite set of contraction mappings $w_n : X \to X$, with respective contraction factors s_n, for $n = 1, 2, ..., N$.

The word hyperbolic is put in parenteses as it is sometimes dropped in practice. The notation (from Barnsley) used to define an IFS is $\{X; w_n, n = 1, ..., N\}$, and $s = \max\{s_n : n = 1, ..., N\}$ is the contraction factor for the IFS. Next we only state a theorem from which we will define what is an attractor.

Theorem I.3.3 Let $\{X; w_n, n = 1, ..., N\}$ be a hyperbolic IFS with contractivity factor s. Then the transformation $W : \mathcal{H}(X) \to \mathcal{H}(X)$ defined by

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

for all $B \in \mathcal{H}(X)$, is a contraction mapping on the complete metric space $(\mathcal{H}(X), h(d))$ with contractivity factor s. Its unique fixed-point, $A \in \mathcal{H}(X)$, obeys

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

and is given by $A = \lim_{n\to\infty} W^{\circ n}(B)$ for any $B \in \mathcal{H}(X)$.

Definition I.3.3 The fixed-point $A \in \mathcal{H}(X)$ described in the above theorem is called the *attractor* of the IFS.

Here the word 'attractor' could be replaced by 'deterministic fractal'. Dynamical systems often possess attractors, and when these are interesting, they are called 'strange atractors'.

An IFS can be *totally disconnected* (e.g. the Cantor set), *just-touching* (e.g. the Sierpinski triangle) or *overlapping*. An example of an overlapping IFS in $\Re$ is

$$w_1(x) = x/3, \qquad w_2(x) = x/2 + 1/4.$$

Before seeing the formal definition, it is important to point out that the terminologies 'totally disconnected', 'just-touching' and 'overlapping' refer to the IFS itself, and not the attractor of the IFS. The reason for this is that an attractor may satisfy several different hyperbolic IFS's. For example, one could easily come up with two IFS's, one just-touching, the other overlapping, and both having attractor, say, $[0, 1]$.

Definition I.3.4 Let $\{X; w_1, ..., w_N\}$ be a hyperbolic IFS. The *code space associated with the IFS*, (Σ, d_C), is defined to be the code space on N symbols $\{1, 2, ..., N\}$, with the metric d_C given by

$$d_C(\omega, \sigma) = \sum_{n=1}^{\infty} \frac{|\omega_n - \sigma_n|}{N + 1)^n}, \qquad \forall \omega, \sigma \in \Sigma.$$

Definition I.3.5 Let $\{X; w_n, n = 1, 2, ..., N\}$ be a hyperbolic IFS with associated code space Σ. For each $\sigma \in \Sigma$, $n \in N$, and $x \in X$, consider the continuous and onto function $\phi : \Sigma \to A$ defined by

$$\phi(\sigma, n, x) = \omega_{\sigma_1} \circ \omega_{\sigma_2} \circ ... \omega_{\sigma_n}(x)$$

where

$$\phi(\sigma) = \lim_{n \to \infty} \phi(\sigma, n, x)$$

exists, belongs to A, and is independent of $x \in X$. An *address* of a point $a \in A$ is any member of the set

$$\phi^{-1}(A) = \{\omega \in \Sigma : \phi(\omega) = a\}.$$

The IFS is said to be *totally disconnected* if each point of its attractor possesses a unique address. The IFS is said to be *just-touching* if it is not totally disconected and yet, its attractor contains an open set O such that

(1) $w_i(\mathcal{O}) \cap w_j(\mathcal{O}) = \emptyset \qquad \forall i, j \in \{1, 2, ..., N\},\ i \neq j;$

(2) $\cup_{i=1}^{N} \lim w_i(\mathcal{O}) \subset \mathcal{O}.$

An IFS whose attractor obeys (1) and (2) is said to obey the *open set* condition. The IFS is said to be *overlapping* if it is neither just-touching nor disconnected.

Theorem I.3.4 Let $\{\Re^m; w_1, ..., w_N\}$ be a hyperbolic IFS, and let A denote its attractor. For $n = 1, .., N$, w_n is a similitude of scaling factor s_n. If the IFS is totally disconnected or just-touching, then the attractor has fractal dimension $D(A)$, which is given by the unique solution of

$$\sum_{n=1}^{N} |s_n|^{D(A)} = 1,$$

where $D(A) \in [0, m]$. If the IFS is overlapping then $D(A) \leq D'$, where D' is the solution of

$$\sum_{n=1}^{N} |s_n|^{D'} = 1,$$

and $D' \in [0, \infty)$.

Proof: There are two parts to this proof. The first part is concerned with an IFS that is just-touching or totally disconnected. We will omit the first part as it is well taken care of in the proof of theorem I.5.3. ([F]). So here, we will assume the first part and concern ourselves only with the second part, which is about the overlapping IFS.

Let D' satisfy the following equation;

$$\sum_{n=1}^{N} |s_n|^{D'}, \qquad D' \in [0, \infty). \tag{3.4.1}$$

For any set E, we write $E_{i_1,...,i_k} = w_{i_1} \circ ... \circ w_{i_k}(E)$.

Let J_k denote the set of all k-term sequences $(i_1, ..., i_k)$, with $1 \le i_j \le N$. By keeping in mind that we are dealing with an overlapping IFS, and by using $A \supset \bigcup_{n=1}^{N} w_n(A)$ repeatedly, it follows that;

$$A \supset \bigcup_{J_k} A_{i_1,...,i_k}.$$

Since the mapping $w_{i_1} \circ ... \circ w_{i_k}$ is a similitude of scaling factor $s_{i_1}...s_{i_k}$, then;

$$\sum_{J_k} |A_{i_1...i_k}|^{D'} \le \sum_{J_k} (s_{i_1}...s_{i_k})^{D'} |A|^{D'} = (\sum_{i_1} s_{i_1}^{D'})...(\sum_{i_k} s_{i_k}^{D'})|A|^{D'} = |A|^{D'}. \tag{3.4.2}$$

The above inequality is true by equation (3.4.1). This implies that there exists a $D(A) \le D'$ such that inequality (3.4.2) becomes an equality. And therefore,

$$\sum_{n=1}^{N} |s_n|^{D'} = 1, \qquad D(A) \le D'.$$

End of proof.

Consider Sierpinski's Triangle on the real plane. It is the attractor of a just-touching IFS of three similitudes; $w_1(x, y) = (x/2, y/2)$, $w_2(x, y) = (x/2 + 1/2, y/2)$, and $w_3(x, y) = (x/2 + 1/4, y/2 + 1/2)$. The three scaling factors are equal, and $s_1 = s_2 = s_3 = 1/2$. Using Thm I.3.4, we have $(1/2)^D + (1/2)^D + (1/2)^D = 1$, which means that $(1/2)^D = 1/3$, and thus, $D = \ln 3/\ln 2 \approx 1.58$.

Similarly, if we consider the Cantor set on [0,1], its two similitudes are $w_1(x) = x/3$, and $w_2(x) = x/3 + 1/3$. The two scaling factors are equal; $s_1 = s_2 = 1/3$. Again, using Thm I.3.4, we have $(1/3)^D + (1/3)^D = 1$, which means that $(1/3)^D = 1/2$, and thus, $D = \ln 2/\ln 3 \approx 0.63$.

If we consider the overlapping IFS given as an example above (p. 14), we see that the application of the Theorem I.3.4 is not so easy. Indeed, we would get $(1/3)^{D'} + (1/2)^{D'} = 1$, where D' is the upper bound for the fractal dimension, i.e. $D(A) \le D'$. Solving for D' here would be difficult.

I.4 The Experimental Determination of the Fractal Dimension

A topic that I have found interesting while doing some research, is that practically everyone agrees on the following fact: When associating a fractal dimension with a certain set of data, the method used to obtain D is not unique. And so far, no general conjecture has been set up for it. It is therefore customary to specify with the data, how its fractal dimension was calculated.

In our application of the box-counting theorem for the experiment of chapter III, we actually use rectangles. The data we use will fit perfectly in the rectangles, as it would in square boxes. More on that in section III.1.

The fractal dimension is calculated by associating it with the slope of the regression line that approximates the log-log plot of $(1/\epsilon_n)$ and $\mathcal{N}_n(A)$ ($\ln(1/\epsilon_n)$ on the x-axis and $\ln(\mathcal{N}_n(A))$ on the y-axis). In our case, we will see later that $\epsilon_n = (E-S)/2^n$, where S is the minute at which the experiment starts, and E is the minute at which the experiment ends. The slope is approximated by using the least-squares method.

As an example, let us consider the results of a fictional experiment (Table I.1). For simplicity, we will assume that $S = 0$ and $E = 1$. Then, $\epsilon_n = 1/2^n$.

n	eps_n	$N_n(A)$	$\ln(N_n(A))$	$\ln(1/eps_n)$
0	1	1	0	0
1	1/2	3	1.10	0.69
2	1/4	7	1.95	1.38
3	1/8	10	2.30	2.08
4	1/16	19	2.94	2.77
5	1/32	33	3.50	3.46
6	1/64	58	4.06	4.16

TABLE I.1: Sample results as an example.

As we can see in Figure I.3, the slope of the regression line that approximates the points given by the log-log plot is where we get our fractal dimension.

Now in the experiment of chapter III, in order to get a better regression line approximation, we actually throw out the first dot (the dot located at the origin of the graph in Figure I.3). What does this dot represent? It is associated with the first (and thus unique) covering box. It represents the existence of the set being covered. Indeed, if this dot were not at the origin, it would mean that the contribution to $N(A, \epsilon)$, for when ϵ is still its largest, is null. So if we have nothing to cover with our biggest box, why continue and try to cover A with smaller boxes.

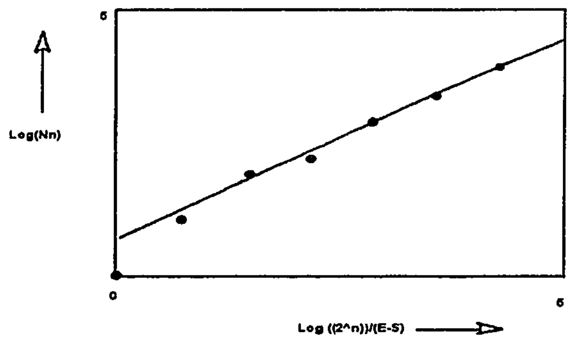

FIGURE I.3: Slope of Regression = Fractal Dimension.

Furthermore, this first dot would fit perfectly in a linear function whose y-intercept is 0, if only for all of the next smaller ϵ's, all the covering boxes would actually be needed to cover a piece of A. For our second ϵ, (i.e., when the original covering box is subdivided into 4 smaller boxes), it is still possible for A to be covered by the 4 boxes. However after the next subdivision, with 16 boxes, it is quite unlikely that all of the 16 boxes would be needed to cover A. So considering the first dot or not certainly has an effect on the slope of the regression line, and thus on the fractal dimension.

I.5 The Hausdorff-Besicovitch (H-B) Fractal Dimension

The H-B dimension is much more complex and subtle in its definition than that of the fractal dimension seen above. Its importance is revealed when comparing the sizes of sets that have the same fractal dimension.

We will need the following notations and specifications to understand the definitions and theorems in the present section.

$\rightarrow X = (\Re^m, d),\ m \in \mathcal{Z}^+,\ d$-Euclidean metric.

$\rightarrow A \subset \Re^m$ is bounded.

$\rightarrow \mathrm{diam}(A) = \sup_{x,y \in A} d(x,y)$. We will also denote the diameter of a set A by $|A|$.

For $0 < \epsilon < \infty$ and $0 \leq p < \infty$, $A_i \subset A$ and $A \subset \bigcup_{i=1}^{\infty} A_i$.

$\rightarrow M(A,p,\epsilon) = \inf_{i=1,2,\ldots}\{\sum_{i=1}^{\infty}(\mathrm{diam}(A_i))^p : A \subset \bigcup_{i=1}^{\infty} A_i, \mathrm{diam}(A_i) < \epsilon\}$

→Convention: $(\text{diam}(A_i))^0 = 0$ when $A_i = \emptyset$. $M(a, p, \epsilon) \in [0, \infty]$.

→ $M(a, p) = \sup_{\epsilon > 0} M(A, p, \epsilon)$ $\qquad$ $(\Rightarrow \forall p \in [0, \infty), M(a, p) \in [0, \infty])$

Definition I.5.1 Let m be a positive integer, and let A be defined as above. For each $p \in [0, \infty)$, $M(A, p)$ is called the *Hausdorff p-dimensional measure of A.*

As we will see in the proof of Thm I.5.1, $M(A, p, \epsilon)$ is a nonincreasing function of ϵ. In the same proof, we see that $M(A, p)$ is a nonincreasing function of $p \in [0, \infty]$.

Consider again the Cantor set C in [0,1] and the Sierpinski triangle $\triangle$. Then $M(C, 0) = \infty$, $M(C, 1) = 0$ and $M(\triangle, 1) = \infty$, $M(\triangle, 2) = 0$.

$M(A,p)$ is quite remarkable as its range can consist of only one, two or three values, i.e., 0, ∞, and a finite positive number. So for C and $\triangle$, the 'shift' from ∞ to 0 is made respectively between 0 and 1, and between 1 and 2 (Fig. I.4). And if p is an integer, $M(A, p)$ is simply the Lebesgue measure.

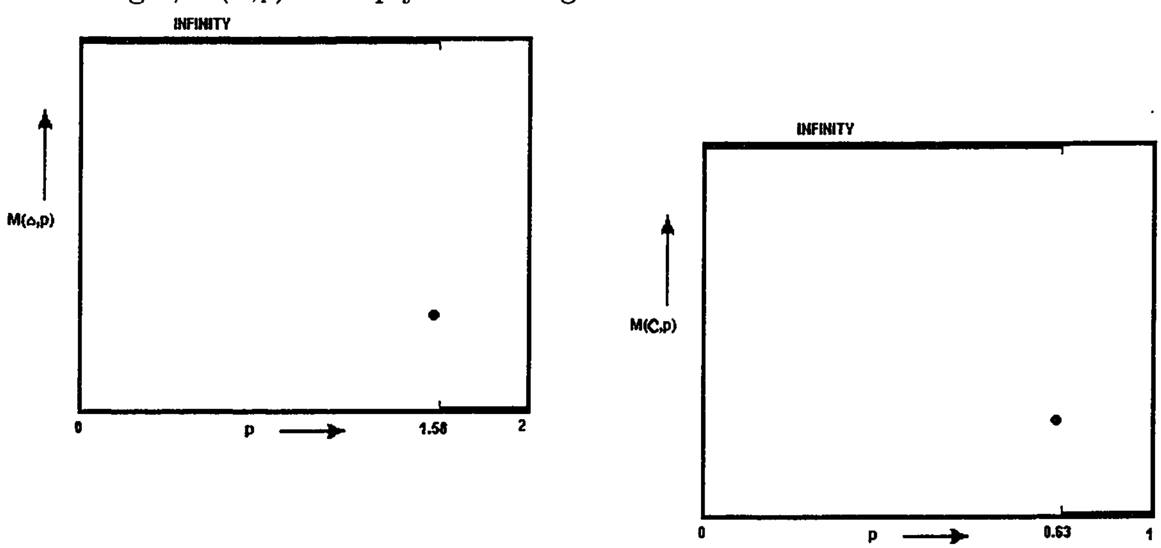

FIGURE I.4: Graphs of $M(\triangle, p)$ and $M(C, p)$.

Theorem I.5.1 Let m be a positive integer, and let A be a bounded set in $\Re^2$. For $p \in [0, \infty)$, there exists a unique real number $D_H \in [0, m]$ such that $M(A,p) = \infty$ if $p < D_H$ and $M(A,p) = 0$ if $p > D_H$.

Proof ([FA]): First, we will note that $M(A,p)$ is a nonincreasing function as p goes from 0 to ∞. From there, we will go through the steps that complete the proof.

Assuming ϵ to be less than 1, it follows that ϵ^p will decrease (go to 0) as p increases (goes to ∞). Let us consider the following important fact which helps us to obtain this above result;

$$\lim_{\epsilon \to 0^+} \lim_{p \to 0} \epsilon^p = 1,$$

and

$$\lim_{\epsilon \to 0^+} \lim_{p \to \infty} \epsilon^p = 0.$$

Now consider $p < p_o$. We claim that

$$M(A,p,\epsilon) \geq M(A,p_o,\epsilon)\epsilon^{p-p_o}.$$

Proof (of Claim): $\text{diam}(A_i) < \epsilon$.

$$\sum_{i=1}^{\infty}(\text{diam}(A_i))^{p_o} = \sum_{i=1}^{\infty}(\text{diam}(A_i))^{p_o-p+p} \leq \sum_{i=1}^{\infty}\epsilon^{p_o-p}(\text{diam}(a_i))^p = \epsilon^{p_o-p}\sum_{i=1}^{\infty}(\text{diam}(a_i))^p.$$

Take the infima over all possible coverings $\{A_i\}$ of both sides of the above inequality to obtain

$$M(A,p_o,\epsilon) \leq \epsilon^{p_o-p}M(A,p,\epsilon).$$

End of proof (of Claim).

Using this claim, we can say that if $M(A,p_o) > 0$, then $M(A,p) = \infty$. Thus we conclude: there exists a unique real number D_H such that;

$$M(A,p) = \infty \qquad ,0 \leq p < D_H,$$

and

$$M(A,p) = 0 \qquad , D_H < p < \infty.$$

End of proof.

Definition I.5.2 In the above theorem (Thm I.5.1), the real number D_H is called the *Hausdorff-Besicovitch (H-B) Dimension* of A. It is also denoted by $D_H(A)$.

This next theorem states that the H-B dimension of a set is not greater than the fractal dimension of the same set.

Theorem I.5.2 Let m be a positive integer, and let A be a set in $\Re^m$. $D(A)$ represents the fractal dimension of A and $D_H(A)$ represents the H-B dimension of A. Then $0 \leq D_H(A) \leq D(A) \leq m$.

Proof: From Def. I.2.1, $D(A) = \lim_{\epsilon \to 0} \frac{\ln(\mathcal{N}_\epsilon)}{\ln(1/\epsilon)}$. Fix $\epsilon > 0$.

$$D(A)\ln(1/\epsilon) = \ln(\mathcal{N}_\epsilon)$$

$$\Rightarrow \frac{1}{\epsilon^{D(A)}} = \mathcal{N}_\epsilon$$

$$\Rightarrow 1 = \mathcal{N}_\epsilon \epsilon^{D(A)} = \sum_{i=1}^{N} \epsilon^{D(A)}$$

$(N = \mathcal{N}_\epsilon)$. Let $\{B_i : i = 1, ..., N\}$ represent the covering boxes used in the box-counting algorithm (and keep in mind that $\text{diam}(B_i) \leq \epsilon$), and let $\{A_i : i = 1, ..., N\}$ represent those sets used as a covering for the Hausdorff measure. Choose A_i such that $A_i = B_i \cap A$ (A_i is the intersection of each box with the set A). We have $\text{diam}(A_i) \leq \text{diam}(B_i)$. It follows that with this ϵ, $M(A, D(A), \epsilon) < \infty$ (since ϵ must be less than or equal to the diameter of this choice of A_i). Next, as $D(A)$ exists, this must be true for every ϵ, and we get;

$$\lim_{\epsilon \to 0} \sum_{i=1}^{\infty} (\text{diam}(A_i))^p \leq 1,$$

for $p = D(A)$. So we have $D_H(A) = D(A)$, except when the above sum is 0, in which case $D_H(A) \leq D(A)$. End of proof.

Our next theorem (Thm I.5.3) is an extension of Thm I.3.4. It mentions that for a totally disconnected or just-touching IFS, the fractal and Hausdorff dimensions coincide.

Theorem I.5.3 Let m be a positive integer. Let $\{\Re^m; w_1, ..., w_N\}$ be a hyperbolic IFS, and let A denote its attractor. For $n = 1, ..., N$, w_n is a similitude of scaling factor s_n. If the IFS is totally disconnected or just-touching, then $D(A) = D_H(A) = D$, where D is given by the unique solution of

$$\sum_{n=1}^{N} |s_n|^D = 1,$$

where $D(A) \in [0, m]$. If $D > 0$, then the Hausdorff D-dimensional measure $M(A, D_H(A))$ is a positive real number.

Before we discuss the proof of this theorem (it is by far the most complex proof encountered in this paper), we state and prove a small lemma that will be used near the end of the proof of the theorem. Also in the proof of this theorem, we use the Mass Distribution Principle. It is stated and proved after the Lemma.

Lemma I.5.3 Let $\{V_i\}$ be a collection of disjoint open subsets of $\Re^m$ such that each V_i contains a ball of radius $a_1 r$ and is contained in a ball of radius $a_2 r$. Then any ball B of radius r intersects at most $(1+2a_2)^m a_1^{-m}$ of the closures $\hat{V}_i$.

Proof (of Lemma): ([F]) If $\hat{V}_i$ meets B, then $\hat{V}_1$ is contained in the ball concentric with B of radius $(1 + 2a_2)r$.

Suppose that q of the sets $\hat{V}_i$ intersect B. Then, summing the volumes of the corresponding interior balls of radii $a_1 r$, it follows that

$$q(a_1 r)^m \leq (1 + 2a_2)^m r^m,$$

giving;

$$q \leq (1 + 2a_2)^m r^m (a_1 r)^{-m} = (1 + 2a_2)^m r^{-m}.$$

End of proof (of Lemma).

Definition I.5.3 A measure μ on a bounded subset of $\Re^n$ for which $0 < \mu(\Re^n) < \infty$ will be called a *mass distribution.*

Mass Distribution Principle ([F]) Let μ be a mass distribution on A, and suppose that for some D, there are numbers $c > 0$, and $\epsilon > 0$, such that

$$\mu(U) \leq c|U|^D$$

for all sets U with $|U| \leq \epsilon$. Then $M(A, D) \geq \mu(A)/c$, and

$$D \leq D_H(A) \leq \underline{D(A)} \leq \overline{D(A)},$$

where $\underline{D(A)}$ and $\overline{D(A)}$ represent respectively the limit inf and the limit sup from the definition of the fractal dimension of A.

Proof (of Mass Dist. Princ.): If $\{U_i\}$ is any cover of A, then

$$0 < \mu(A) = \mu(\cup_i U_i) \leq \sum_i \mu(U_i) \leq c \sum_i |U_i|^s.$$

Taking infima, $M(A, D, \epsilon) \geq \mu(A)/c$ if ϵ is small enough, so $M(A, D) \geq \mu(A)/c$. End of proof (of Mass Dist. Princ.).

Proof (of Theorem I.5.3): ([F]) Let D satisfy the following equation;

$$\sum_{n=1}^{N} |s_n|^D = 1, \qquad D \in [0, m]. \tag{5.3.1}$$

For any set E, we write $E_{i_1,...,i_k} = w_{i_1} \circ ... \circ w_{i_k}(E)$.

Let J_k denote the set of all k-term sequences $(i_1, ..., i_k)$, with $1 \leq i_j \leq N$. It follows, by using $A = \bigcup_{n=1}^{N} w_n(A)$ repeatedly, that;

$$A = \bigcup_{J_k} A_{i_1,...,i_k}.$$

We check that these covers (above) of A provide a suitable **upper estimate** for the Hausdorff measure.

Since the mapping $w_{i_1} \circ ... \circ w_{i_k}$ is a similitude of scaling factor $s_{i_1}...s_{i_k}$, then;

$$\sum_{J_k} |A_{i_1...i_k}|^D = \sum_{J_k} (s_{i_1}...s_{i_k})^D |A|^D = (\sum_{i_1} s_{i_1}^D)...(\sum_{i_k} s_{i_k}^D)|A|^D = |A|^D.$$

The above is true by equation (5.3.1). For any $\epsilon > 0$, we may choose k such that $|A_{i_1...i_k}| \leq (\max_i s_i)^k \leq \epsilon$, so $M(A, D_H(A), \epsilon) \leq |A|^D$ and hence $M(A, D_H(A)) \leq |A|^D$. (We have the upper estimate.)

We now check for the **lower estimate**. This task is more awkward...

Let I be the set of all infinite sequences $I = \{(i_1, i_2, i_3, ...) : 1 \leq i_j \leq N\}$, and let $I_{i_1,...,i_k} = \{(i_1, ..., i_k, q_{k+1}, ...) : 1 \leq q_j \leq N\}$ be the *cylinder* consisting of these sequences in I with initial terms $(i_1, ..., i_k)$.

We may put a mass distribution μ on I such that

$$\mu(I_{i_1,...,i_k}) = (s_{i_1}...s_{i_k})^D.$$

Since $(s_{i_1}...s_{i_k})^D = \sum_{i=1}^{N}(s_{i_1}...s_{i_k}s_i)^D$, i.e., $\mu(I_{i_1,...,i_k}) = \sum_{i=1}^{N} \mu(I_{i_1,...,i_k,i})$, it follows that μ is indeed a mass distribution on subsets of I with $\mu(I) = 1$.

Now that we have a mass distribution μ on I, we may transfer it to a mass distribution $\hat{\mu}$ on A in a natural way by defining $\hat{\mu}(E) = \mu\{(i_1, i_2, i_3, ...) : x_{i_1,i_2,...} \in E\}$ for subsets E of A. (Note: $\mathrm{x}_{i_1,i_2,...} = \bigcap_{k=1}^{\infty} A_{i_1,...,i_k}$.) It can be checked that $\hat{\mu}(A) = 1$.

We show that $\hat{\mu}$ satisfies the conditions of the Mass Distribution Principle.

Let V be an open set satisfying: $V \supset \bigcup_{i=1}^{N} w_i(V)$ (a disjoint union), V is nonempty and bounded.

Let $\hat{V}$ denote the closure of the open set V. Since $\hat{V} \supset W(\hat{V}) = \bigcup_{i=1}^{N} w_i(\hat{V})$, the decreasing sequence of iterates $W^k(\hat{V})$ converges to A. ($\bigcap_{k=1}^{\infty} W^k(H) = A$ for any set $H \subseteq \Re^m$.)

In particular, $\hat{V} \supset A$, and $\hat{V}_{i_1,...,i_k} \supset A_{i_1,...,i_k}$ for each finite sequence $(i_1, ..., i_k)$.

Let B be any ball of radius $r < 1$. We estimate $\hat{\mu}(B)$ by considering the sets $V_{i_1,...,i_k}$ with diameters comparable with that of B and with closures intersecting $A \bigcap B$.

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

$$(\min_i s_i) r \leq s_{i_1} s_{i_2} ... s_{i_k} \leq r \tag{5.3.2}$$

and let Q denote the finite set of all (finite) sequences obtained in this way. Then, for every infinite sequence $(i_1, i_2, ...) \in I$, there is exactly one value of k with $(i_1, ..., i_k) \in Q$.

Since $V_1, ..., V_N$ are disjoint, so are $V_{i_1,...,i_k,1}, ..., V_{i_1,...,i_k,N}$ for each $(i_1, ..., i_k)$. Using this in a nested way, it follows that the collection of open sets $\{V_{i_1,...,i_k} : (i_1, ..., i_k) \in Q\}$ is disjoint.

Similarly,

$$A \subset \bigcup_Q A_{i_1,\ldots,i_k} \subset \bigcup_Q \hat{V}_{i_1,\ldots,i_k}.$$

We now choose a_1 and a_2 so that V contains a ball of radius a_1 and is contained in a ball of radius a_2. Then, for $(i_1,\ldots,i_k) \in Q$, the set $V_{i_1,\ldots,i_k}$ contains a ball of radius $s_{i_1}\ldots s_{i_k}a_1$, and therefore, it also contains one of radius $(\min_i s_i)a_1 r$ and is contained in a ball of radius $s_{i_1}\ldots s_{i_k}a_2$, and hence, in a ball of radius $a_2 r$.

Let Q_1 denote those sequences $(i_1,\ldots,i_k)$ in Q such that B intersects $\hat{V}_{i_1,\ldots,i_k}$.

By the above Lemma (I.5.3), there are at most $q = (1+2a_2)^m a_1^{-m}(\min_i s_i)^{-m}$ sequences in Q. Then

$$\hat{\mu}(B) = \hat{\mu}(A\bigcap B) \le \mu\{(i_1,i_2,\ldots) : x_{i_1,i_2,\ldots} \in A\bigcap B\} \le \mu\{\bigcup_{Q_1} I_{i_1,\ldots,i_k}\}.$$

So $\hat{\mu}(B) \le \mu\{\bigcup_{Q_1} I_{i_1,\ldots,i_k}\}$ since, if $x_{i_1,i_2,\ldots} \in A\bigcap B \subset \bigcup_{Q_1} \hat{V}_{i_1,\ldots,i_k}$, then there is an integer k such that $(i_1,\ldots,i_k) \in Q_1$. Thus,

$$\hat{\mu}(B) \le \sum_{Q_1} \mu(I_{i_1,\ldots,i_k}) = \sum_{Q_1}(s_{i_1}\ldots s_{i_k})^D \le \sum_{Q_1} r^D \le r^D q$$

(using equation (5.3.2)).

Since any set U is contained in a ball of radius $|U|$, we have $\hat{\mu}(U) \le |U|^D q$, so the Mass Distribution Principle gives;

$$M(A, D_H(A)) \ge q^{-1} > 0, \qquad \& \ D(A) = D.$$

If Q is any set of infinite sequences such that for every $(i_1,i_2,\ldots) \in I$, there is exactly one integer k with $(i_1,\ldots,i_k) \in Q$. It follows inductively that

$$\sum_Q (s_{i_1}\ldots s_{i_k})^D = 1.$$

Thus if Q is chosen as in equation (5.3.2), Q contains at most $(\min_i s_i)^{-D} r^{-D}$ sequences.

For each sequence $(i_1, ..., i_k)$ in Q, we have

$$|\hat{V}_{i_1,...,i_k}| = s_{i_1} ... s_{i_q} |\hat{V}| \leq r|\hat{V}|,$$

so A may be covered by $(\min_i s_i)^{-D} r^{-D}$ sets of diameter $r|\hat{V}|$ $(\forall r < 1)$.

Now $\overline{D(A)} = \overline{\lim}_{\epsilon \to 0} \frac{\ln N_\epsilon(A)}{\ln 1/\epsilon}$, where $N_\epsilon(A)$ is the smallest number of sets of diameter at most ϵ that cover A.

$\Rightarrow \overline{D(A)} \leq D$. Since the Hausdorff dimension is also D, this completes the proof. End of proof (of theorem I.5.3)

When considering two sets of different fractal dimension, the set with the higher fractal dimension is called the 'larger' one. For two fractals that have the same fractal dimension, we can still differentiate them with the aid of the Hausdorff measure. The larger $M(A, D_H(A))$, the larger the fractal.

CHAPTER II: Fractal Interpolation

In this chapter, we talk about a function that interpolates a certain set of data. This function is also the attractor of a specific hyperbolic IFS. A fractal interpolation function (F.I.F.) (defined below), has a graph that can be used to approximate non-regular shapes such as; top of clouds, river beds, horizons over forests or mountains, etc... The graph of the function (F.I.F.) can be made close to the data in the Hausdorff metric.

Theorem II.3.1 gives a surprisingly neat way to calculate the fractal dimension of a F.I.F.. We will see that this gets even better if the points over which the function is interpolating are equally spaced.

The chapter ends after introducing the concept of the hidden variable. It is a generalization of the scaling factors. As the scaling factors, hidden variables are used to adjust the graph of the F.I.F., but they offer more flexibility.

II.1 Applications for Fractal Functions.

Let us consider the line segment representing a river drawn on a map. A F.I.F. will interpolate coordinates at different points on the line. By looking at the map and at the graph of the F.I.F., we can play with the scaling factors (defined below) to adjust how the graph of the F.I.F. behaves between two interpolation points until it ressembles the river drawn on the map.

II.2 Fractal Interpolation Functions (FIF's)

We consider a collection of data of the form: $\{(x_i, F_i) \in \Re^2, i = 0, 1, ..., N\}$, where $x_0 < x_1 < ... < x_N$.

Definition II.2.1 An *interpolation function* corresponding to the above set of data is a continuous function $f : [x_0, x_N] \rightarrow \Re$ such that $f(x_i) = F_i$ for $i = 1, 2, ..., N$. The points (x_i, F_i) are called the *interpolation points*. So f interpolates the data, and the graph of f passes through the interpolation points.

As an example, consider the function $f(x) = 1 + x$. f is an interpolation function for the points $\{(0,1), (1,2)\}$. Consider w_1 and w_2;

$$w_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x/2 \\ (y+1)/2 \end{pmatrix}, \qquad w_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (x+1)/2 \\ (y+2)/2 \end{pmatrix}.$$

The graph of the attractor of the IFS corresponding to these two similitudes is exactly the straight line joining the points (0,1) to (1,2), i.e., the graph of $f(x)$ over the interval [0,1].

More generally, we can write (and this is the notation used in our next theorem (II.2.1)):

$$w_n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x_n - x_{n-1}}{x_N - x_0} & 0 \\ \frac{F_n - F_{n-1}}{x_N - x_0} - d_n \frac{(F_N - F_0)}{x_N - x_0} & d_n \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{x_N x_{n-1} - x_0 x_n}{x_N - x_0} \\ \frac{x_N F_{n-1} - x_0 F_n}{x_N - x_0} - d_n \frac{(x_N F_0 - x_0 F_N)}{x_N - x_0} \end{pmatrix}$$

or, to simplify;

$$w_n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_n & 0 \\ c_n & d_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \end{pmatrix},$$

where

$$a_n = \frac{x_n - x_{n-1}}{x_N - x_0}, \qquad e_n = \frac{x_N x_{n-1} - x_n x_0}{x_N - x_0},$$

$$c_n = \frac{F_n - F_{n-1} - d_n(F_N - F_0)}{x_N - x_0}, \qquad f_n = \frac{x_n F_{n-1} - x_0 F_n}{x_N - x_0} - d_n \frac{(x_N F_0 - x_0 F_N)}{x_N - x_0}.$$

Here, d_n is called the vertical scaling factor, where $0 < d_n < 1$ and $n = 1, 2, ..., N$. d_n can be seen as a 'distortion knob'. If we are trying to approximate say, a sample of the boundary of a lake, the graph corresponding to the w_n's will definitely go through the interpolation points, but it is the choice of the scaling factors that determines where, or rather how, the line representing the boundary of the lake behaves between two interpolation points. If $d_n = 0$, then the F.I.F. is simply a piecewise linear interpolation function.

Note that a similitude w_n (described above) does not necessarily have to be a contraction, even though $|d_n| < 1$.

These next two theorems establish the theoretical justifications for the above descriptions.

Theorem II.2.1 Let N be a positive integer (> 1). Let $\{\Re^2; w_n, n = 1, ..., N\}$ denote the IFS defined above, associated with the data set $\{(x_n, F_n) : n = 1, ..., N\}$. Let the vertical scaling factor d_n obey $0 < d_n < 1$, for $n = 1, ..., N$. Then there is a metric d on $\Re^2$, equivalent to the Euclidean metric, such that the IFS is hyperbolic with respect to d. In particular, there is a unique nonempty compact set $G \in \Re^2$ such that

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

Proof: ([B]) Let us define a metric d on $\Re^2$ by;

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + \theta|y_1 - y_2|, \qquad \theta \in \Re^+$$

(θ is defined below.) $(\Re^2, d)$ is equivalent to ($\Re^2$, Euclidean). For $n \in \{1, 2, ..., N\}$, let a_n, c_n, e_n and f_n be defined as above (bottom of p.30).

$$\Rightarrow d(w_n(x_1, y_1), w_n(x_2, y_2)) =$$

$$d((a_n x_1 + e_n, c_n x_1 + d_n y_1 + f_n), (a_n x_2 + e_n, c_n x_2 + d_n y_2 + f_n)) =$$

$$a_n|x_1 - x_2| + \theta|c_n(x_1 - x_2) + d_n(y_1 - y_2)| \leq$$

$$(|a_n| + \theta|c_n|)|x_1 - x_2| + \theta|d_n||y_1 - y_2|.$$

Note that $|a_n| = \frac{|x_n - x_{n-1}|}{|x_N - x_0|} < 1$ as $N \geq 2$. Choose

$$\theta = \frac{\min\{(2 - |a_n|) : n = 1, 2, ..., N\}}{\max\{2|c_n| : n = 1, 2, ..., N\}},$$

and if all of the c_n's are 0, then simply choose $\theta = 1$.

$$\Rightarrow d(w_n(x_1, y_1), w_n(x_2, y_2)) \leq (|a_n| + \theta|c_n|)|x_1 - x_2| + \theta|d_n||y_1 - y_2| \leq$$

$$a|x_1 - x_2| + \theta\delta|y_1 - y_2| \leq \max\{a, \delta\} d((x_1, y_1), (x_2, y_2)),$$

where $a = 1 + a_n - (1/2)\max\{|a_n|_n = 1, 2, ..., N\}$ $(a < 1)$,

and $\delta = \max\{|d_n| : n = 1, 2, ..., N\}$ $(\delta < 1)$. As both a and δ are < 1, the IFS is hyperbolic. End of proof.

Theorem II.2.2 Let N be a positive integer (> 1). Let $\{\Re^2; w_n, n = 1, ..., N\}$ denote the same IFS defined above, associated with the data set $\{(x_n, F_n) : n = 1, ..., N\}$. Let the vertical scaling factor d_n obey $0 < d_n < 1$, for $n = 1, ..., N$, so that the IFS is hyperbolic. Let G denote the attractor of the IFS. Then G is the graph of a continuous function $f : [x_0, x_N] \to \Re$, which interpolates the data $\{(x_i, F_i) : i = 1, ..., N\}$. That is,

$$G = \{(x, f(x)) : \ x \in [x_0, x_N]\},$$

where

$$f(x_i) = F_i, \ i = 0, 1, ..., N.$$

Proof: ([B]) Let F denote the set of continuous functions

$$f : [x_0, x_N] \to \Re, s.t. \ f(x_0) = F_0 \ \& \ f(x_N) = F_N.$$

Define a metric d on F by;

$$d(f,g) = \max\{|f(x) - g(x)| : x \in [x_0, x_N]\} \qquad \forall f,\ g \in F.$$

Claim #1: (F, d) is a complete metric space. (I think it's a good idea to prove this fact as it uses a good basis of analysis.)

Proof (of Claim #1): First we can prove that (F, d) is a metric space.

Axiom 1: $d(f,g) \geq 0$ as it is the max of a set of nonnegative numbers. Also, if $f(x) = g(x)\ \forall x \in [x_0, x_N]$,

$$\Rightarrow \max_{x \in [x_0, x_N]} |f(x) - g(x)| = 0, \qquad f, g \in F$$

Conversely, the above max is equal to 0 only if $|f(x) - g(x)| = 0\ \forall x \in [x_0, x_N]$.

$$\Rightarrow f(x) = g(x) \qquad \forall x \in [x_0, x_N].$$

Axiom 2 This part is obvious as $|f(x) - g(x)| = |g(x) - f(x)|\ \forall x \in [x_0, x_N]$.

Axiom 3 Let $f,\ g,\ h \in F$. For any $x_1 \in [x_0, x_N]$, we use the triangle inequality to say;

$$|f(x_1) - h(x_1)| \leq |f(x_1) - g(x_1)| + |g(x_1) - h(x_1)|$$

$$\leq \max_{x \in [x_0, x_N]} |f(x) - g(x)| + \max_{x \in [x_0, x_N]} |g(x) - h(x)|$$

$$= d(f,g) + d(g,h).$$

So, $d(f,g) + d(g,h)$ is an upper bound for the set

$$S = \{|f(x_1) - h(x_1)| : x_1 \in [x_0, x_N]\}$$

$$\Rightarrow d(f,g) + d(g,h) \geq \max_{x \in [x_0, x_N]} S = d(f,h)$$

and the 3rd axiom is proved.

Now that we know that (F, d) is a metric space, we can prove that it is a complete metric space:

Consider $\{f_n\}$, $(f_n \in F)$ a sequence of real-valued functions defined on $[x_0, x_N]$. Now suppose we only consider sequences $\{f_n\}$ which are uniformly Cauchy, i.e.,

$$\forall \epsilon > 0, \exists N s.t.\ |f_n(x) - f_m(x)| < \epsilon \qquad \forall m, n \geq N \quad \& \quad \forall x \in [x_0, x_N].$$

To show: $\{f_n\}$ converges uniformly.

Part A $\rightarrow$ Show that $\{f_n\}$ is bounded. Fix $\epsilon > 0$. $\exists N$ such that $m, n, \geq N$

$$\Rightarrow |f_n(x) - f_m(x)| < \epsilon \qquad (\forall x \in [x_0, x_N]).$$

So for any $n \geq N$, $\quad |f_n(x) - f_N(x)| < \epsilon,$

$$\Rightarrow |f_n(x)| < \epsilon + |f_N(x)| \qquad (\forall x \in [x_0, x_N])$$

and

$$|f_n(x)| \leq \max_n\{|f_1|, |f_2|, ..., |f_{N-1}|, \epsilon + |f_N|\} \qquad (\forall x \in [x_0, x_N]).$$

Thus, the sequence is bounded.

Part B $\rightarrow$ Construct a monotonic sequence out of $\{f_n\}$. For each m, let S_m denote the set of members of the sequence of functions $\{f_n\}$, from the m^{th} stage onwards;

$$S_m = \{f_n(x) : n \geq m\}$$

Let $g_m(x) = \sup_x\{f_n(x) : n \geq m\} = \sup_x S_m$.

$$S_{m+1} \subset S_m \Rightarrow \sup_x S_{m+1} \leq \sup_x S_m.$$

Thus, $\{g_m\}$ is monotonic and decreasing. Also, $g_m(x) \geq f_m(x)$, and S_1 is bounded below. Therefore, so is $\{g_m\}$. Say $\{g_m\} \rightarrow l$. (The sequence $\{g_m\}$ converges to l.)

Part C $\rightarrow$ Show that $\{f_n\}$ also converges to l. Given $\epsilon > 0$, $\exists\ N_1\ s.t.\ |f_m - f_n| < \epsilon$, $m, n \geq N_1$, and $\exists N_2$ s.t. $|l - g_m| < \epsilon$, $m \geq N_2$. Let $N = \max\{N_1, N_2\}$. $g_N - \epsilon$ is not an upper bound of $S_N = \{f_n(x) : n \geq N\}$

$$\Rightarrow \exists M \geq N\ s.t.\ f_M(x) > g_N(x) - \epsilon\ \&\ f_M(x) \leq g_N(x)$$

since g_N is an upper bound for S_N. Now $\forall n \geq N$,

$$|f_n - l| = |f_n - f_M + f_M - g_N + g_N - l| \leq |f_n - f_M| + |f_M - g_M| + |g_N - l| < 3\epsilon.$$

$\Rightarrow \{f_n\} \rightarrow l$. End of proof (of Claim #1).

Let a_n, c_n, e_n, f_n be defined as before. Define a mapping $T : F \rightarrow F$ by;

$$(Tf)(x) = c_n l_n^{-1}(x) + d_n f(l_n^{-1}(x)) + f_n \qquad x \in [x_{n-1}, x_n], n = 1, 2, ..., N.$$

$l_n : [x_0, x_N] \rightarrow [x_{n-1}, x_n]$ is the invertible transformation

$$l_n(x) = a_n x + e_n.$$

1. We verify that T does indeed take F into itself. Let $f \in F$. Then we see that

$$(Tf)(x_0) = c_1 l_1^{-1}(x_0) + d_1 f(l_1^{-1}(x_0)) + f_n = c_1 x_0 + d_1 f(x_0) + f_n = c_1 x_0 + d_1 F_0 + f_n = F_0$$

and,

$$(Tf)(x_N) = c_N l_N^{-1}(x_N) + d_N f(l_N^{-1}(x_N)) + f_N = c_N x_N + d_N f(x_N) + f_N = c_N x_N + d_N F_N.$$

$\Rightarrow$ The function $(Tf)(x)$ obeys the endpoint conditions.

Claim #2: $(Tf)(x)$ is continuous on $[x_{n-1}, x_n]$, for $n = 1, 2, ..., N$.

Proof (of Claim #2): ([B]) Let us fix an $\epsilon > 0$. We have to show that $\exists \delta > 0\ s.t.\ |x - y| < \delta \Rightarrow |(Tf)(x) - (Tf)(y)| < \epsilon,\ x, y, \in [x_{n-1}, x_n]$.

$$(Tf)(x) = c_n l_n^{-1}(x) + d_n f(l_n^{-1}(x)) + f_n$$

where $l_n^{-1}(x) = (x - e_n)/a_n$ as $l_n(x) = a_n x + e_n$. a_n, c_n, d_n, e_n, and f_n are all fixed constants. So $l_n(x)$ and $l_n^{-1}(x)$ are both continuous. And therefore, $c_n l_n^{-1}(x)$, f_n, and $d_n f(l_n^{-1}(x))$ are all continuous as f is continuous. Hence, $(Tf)(x)$ is continuous on $[x_{n-1}, x_n]$, for $n = 1, 2, ..., N$. End of proof (of Claim #2).

To continue in the proof of Theorem II.2.2, it remains to be demonstrated that $(Tf)(x)$ is continuous at each of the points $x_1, x_2, ..., x_{N-1}$. At each of these points the value of $(Tf)(x)$ is defined in two different ways:

For $n \in \{1, 2, ..., N-1\}$, we have

$$(Tf)(x_n) = c_{n+1} l_{n+1}^{-1}(x_n) + d_{n+1} f(l_{n+1}^{-1}(x_n)) + f_{n+1}$$

$$= c_{n+1} x_0 + d_{n+1} f(x_0) + f_{n+1} = F_n,$$

and

$$(Tf)(x_n) = c_n l_n^{-1}(x_n) + d_n f(l_n^{-1}(x_n)) + f_n$$

$$= c_n x_N + d_n f(x_N) + f_n = F_n.$$

So both methods of evaluation lead to the same result. We conclude that T does indeed take F into F.

2. We show that T is a contraction mapping on (F, d). Let $f, g \in F$. Let $n \in \{1, 2, ..., N\}$ and let $x \in [x_{n-1}, x_n]$. Then

$$|(Tf)(x) - (Tg)(x)| = |d_n||f(l_n^{-1}(x)) - g(l_n^{-1}(x))| \leq |d_n| d(f, g).$$

$$\Rightarrow d(Tf, Tg) \leq \delta d(f, g), \qquad \delta = \max\{|d_n| : n = 1, 2, ..., N\} \ (\delta < 1).$$

We conclude that $T : F \to F$ is a contraction mapping.

The Contraction Mapping Theorem implies that T possesses a unique fixed-point in F. That is, there exists a function $f \in F$ such that

$$(Tf)(x) = f(x) \qquad \forall x \in [x_0, x_N].$$

So f passes through the interpolation points!

Let $\hat{G}$ denote the graph of f. Notice that the equations that define T can be rewritten;

$$(Tf)(a_n x + e_n) = c_n x + d_n f(x) + f_n, \qquad x \in [x_0, x_N],\ n = 1, 2, ..., N.$$

$$\Rightarrow \hat{G} = \bigcup_{n=1}^{N} w_n(\hat{G}).$$

But $\hat{G}$ is a nonempty compact subset of $\Re^2$. By Thm II.2.1, there is only one nonempty compact set G, the attractor of the IFS, which obeys the latter equation. It follows that $G = \hat{G}$. End of proof.

Note that in definition II.2.1, we defined what is an interpolation function. We finish this section with this next definition, where we formally state what is a fractal interpolation function (F.I.F.).

Definition II.2.2 The function $f(x)$ (whose graph is the attractor of an IFS as described in Theorems II.2.1 and II.2.2 above), is called a *fractal interpolation function, (F.I.F)* corresponding to the data $\{(x_i, F_i) : i = 1, ..., N\}$.

II.3 The Fractal Dimension of Fractal Functions.

Before stating the first theorem of this section, let us recall the variables defined in the proof of Thm II.2.1:

$$a_n = \frac{x_n - x_{n-1}}{x_N - x_0}, \qquad e_n = \frac{x_N x_{n-1} - x_n x_0}{x_N - x_0},$$

$$c_n = \frac{F_n - F_{n-1} - d_n(F_N - F_0)}{x_N - x_0}, \qquad f_n = \frac{x_n F_{n-1} - x_0 F_n}{x_N - x_0} - d_n \frac{(x_N F_0 - x_0 F_N)}{x_N - x_0}$$

Theorem II.3.1 Let N be a positive integer, and consider the set of data $\{(x_n, F_n) : n = 1, ..., N\}$. Let $\{\Re^2; w_n, n = 1, ..., N\}$ be an IFS associated with the data, where

$$w_n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_n & 0 \\ c_n & d_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \end{pmatrix}$$

for $n = 1, 2, ..., N$. The vertical scaling factors d_n obey $0 < d_n < 1$; and the constants a_n, c_n, e_n, and f_n are given as above. Let G denote the attractor of the IFS, so that. G is the graph of a F.I.F. associated with the data. If

$$\sum_{n=1}^{N} |d_n| > 1, \tag{3.1.1}$$

and the interpolation points do not all lie on a single straight line, then the fractal dimension of G is the unique real solution D of

$$\sum_{n=1}^{N} |d_n| a_n^{D-1} = 1$$

Otherwise, the fractal dimension of G is 1.

Proof: ([B]) (This is an informal demonstration. The notation is the same)

Let $\epsilon > 0$. Consider G in the following way: We place a grid on the graph of G, and ϵ is the length of each side of the squares in the grid where G is superimposed.

Let $N(\epsilon)$ be the number of square boxes of side ϵ which intersect G. (These boxes are similar to the ones used in the box-counting theorem.)

Suppose G has fractal dimension D, where $N(\epsilon) \approx C\epsilon^{-D}$, as $\epsilon \to 0$. (C is a constant.)

To show: We want to estimate D from this assumption.

Let $n = 1, 2, ..., N$. Let $N_n(\epsilon)$ be the number of square boxes of side ϵ which intersect $w_n(G)$, for $n = 1, 2, ..., N$. Suppose $\epsilon \ll |x_N - x_0|$. ($\ll$ is 'much smaller than')

Then, as the IFS is just-touching, it is reasonable to say:

$$N(\epsilon) \approx N_1(\epsilon) + N_2(\epsilon) + N_3(\epsilon) + ... + N_N(\epsilon) \tag{3.1.2}$$

Can we find a relationship between $N(\epsilon)$ and $N_n(\epsilon)$? The answer is yes, and here is how:

Let $\{c_j(\epsilon)\}_{j=1}^{k(\epsilon)}$ be the set of columns of boxes of side ϵ which intersect G, where $k(\epsilon) =$ # of columns.

By equation (3.1.1), we can say that the minimum number of boxes in a column will go to infinity as ϵ goes to 0. Assume that $|d_n| > a_n$ for $n = 1, 2, ..., N$ (this is a stronger assumption than eq. (3.1.1)). Note that

$$\sum_{n=1}^{N} a_n = \sum_{n=1}^{N} \frac{(x_n - x_{n-1})}{x_N - x_0} = 1.$$

Then the column of boxes $c_j(\epsilon)$ becomes a column of parallelograms when w_n is applied to it. The width of the new column is $a_n\epsilon$ and the height of the new column is $|d_n|$\{height of old column\}.

Let $N(c_j(\epsilon))$ be the number of boxes in the column $c_j(\epsilon)$. Then the new column $w_n(c_j(\epsilon))$ can be thought of as being made up of square boxes of side length $a_n\epsilon$, each of which intersect $w_n(G)$. The new columns are made up of small parallelograms, but the number of square boxes of side length $a_n\epsilon$ which they contain is readily estimated.

The number of boxes of side $a_n\epsilon$ in $w_n(c_j(\epsilon))$ is approximately equal to $|d_n|N(c(\epsilon))/a_n$.

This implies;

$$N_n(a_n\epsilon) \approx \sum_{j=1}^{k(\epsilon)} |d_n| N(c_j(\epsilon))/a_n = \frac{|d_n|}{a_n} \sum_{j=1}^{k(\epsilon)} N(c_j(\epsilon)) = \frac{|d_n|}{a_n} N(\epsilon).$$

Note that $\epsilon \ll |x_N - x_0|$.

$$\Rightarrow N_n(\epsilon) \approx \frac{|d_n|}{a_n} N(\epsilon/a_n), \qquad n = 1, 2, ..., N. \tag{3.1.3}$$

Next, if we substitute equation (3.1.3) into equation (3.1.2), we obtain;

$$N(\epsilon) \approx \frac{d_1}{a_1} N(\epsilon/a_1) + \frac{d_2}{a_2} N(\epsilon/a_2) + ... + \frac{d_N}{a_N} N(\epsilon/a_N) \tag{3.1.4}$$

Now we subsitute $N(\epsilon) \approx C\epsilon^{-D}$ into equation (3.1.4);

$$\epsilon^{-D} \approx |d_1| a_1^{D-1} \epsilon^{-D} + |d_2| a_2^{D-1} \epsilon^{-D} + ... + |d_N| a_N^{D-1} \epsilon^{-D} \tag{3.1.5}$$

$$\Rightarrow \sum_{n=1}^{N} |d_n| a_n^{D-1} = 1.$$

(We still have to show that if $\sum_{n=1}^{N} |d_n| \leq 1$, then $D = 1$.)

If $\sum_{n=1}^{N} |d_n| \leq 1$, then $N(\epsilon) \approx C\epsilon^{-1}$, as we can say that $|d_n| \leq a_n$, and thus that $|d_n|/a_n \leq 1$. Substituting in the approximation (3.1.5) gives us the desired result.

$$\Rightarrow D = 1.$$

End of proof.

In this theorem, the equation $\sum_{n=1}^{N} |d_n| a_n^{D-1} = 1$ expresses D in a fairly simple manner (in view of all the variables involved). In particular, when we consider equally spaced interpolation points, the result is very beautiful. When considering equally spaced points, we can say that $x_n = x_0 + n\Delta(x)$, where $\Delta(x) = (x_N - x_0)/N$. Therefore,

$$a_n = \frac{x_n - x_{n-1}}{x_n - x_0} = \frac{\Delta(x)}{x_N - x_0} = \frac{1}{N}.$$

Then, if $\sum_{n=1}^{N} |d_n| > 1, D$ will obey;

$$\sum_{n=1}^{N} |d_n| a_n^{D-1} = \sum_{n=1}^{N} |d_n| (\frac{1}{N})^{D-1} = (\frac{1}{N})^{D-1} \sum_{n=1}^{N} |d_n| = 1.$$

(by Thm II.3.1). If we solve for D, we get;

$$D = 1 + \frac{\ln(\sum_{n=1}^{N} |d_n|)}{\ln(N)}.$$

We can see that if $\sum_{n=1}^{N} |d_n| < N$, then the fractal dimension of the F.I.F. will be < 2. Also, if $\sum_{n=1}^{N} |d_n| > 1$, then the fractal dimension of the F.I.F. will be > 1.

A remarkable fact is that D is totally independent of the values of F_i (for $i = 1, ..., N$). This is very useful, and namely, it is practical when building an algorithm that calculates D. All we need to calculate the fractal dimension D (when the interpolation points are equally spaced), is the number of interpolation points, N, and the scaling factors, d_n, for $n = 1, ..., N$.

We can therefore study a collection of F.I.F.'s that all have the same dimension D. To do so, we give a simple constraint on the d_n's:

$$\sum_{n=1}^{N} |d_n| = N^{D-1}.$$

II.4 Hidden Variable Fractal Interpolation

This section is a generalization of section II.2. Here, other than the scaling factors d_n, we will use new, *hidden variables.* The variables are in fact the 'left-over'variables resulting from a projection of $\Re^3$ into $\Re^2$.

We consider (Y, d_Y) to be a complete metric space.

Definition II.4.1 Let $I \subset \Re$. Let $f : I \rightarrow Y$ be a function. The *graph* of f is the set of points $G = \{(x, f(x)) \in \Re \times Y : x \in I\}$

We consider a set of generalized data of the form: $\{(x_i, F_i) \in \Re \times Y, i = 0, 1, ..., N\}$, where $x_0 < x_1 < ... < x_N$.

Definition II.4.2 (Compare with Def II.2.1) An interpolation function corresponding to the above set of data is a continuous function $f : [x_0, x_N] \to \Re \times Y$ such that $f(x_i) = F_i$ for $i = 1, 2, ..., N$. The points $(x_i, F_i) \in \Re \times Y$ are called the *interpolation points.* So f interpolates the data, and the graph of f passes through the interpolation points.

Now, we want to generalize the notions covered in Thm II.2.1. So the desired generalization is to go from ($\Re^2$, Euclidean metric) to (X, d =any metric). To set this up, we will consider some conditions that will be assumed to hold for the theorem.

Let X denote the cartesian product space $\Re \times Y$, and let θ denote a positive number. Define d on X by;

$$(Cond1) \quad d(X_1, X_2) = |x_1 - x_2| + \theta d_Y(y_1, y_2), \quad \forall X_1 = (x_1, y_1),\ X_2 = (x_2, y_2),\ X_1, X_2 \in X.$$

((X, d) is complete.) Let N be a positive integer. Let $\{(x_n, F_n) \in X : n = 0, 1, ..., N\}$. Define $L_n : \Re \to \Re$ by;

$$(Cond2) \quad L_n(x) = a_n x + e_n,$$

for $n \in \{1, 2, ..., N\}$ so that $L_n([x_0, x_N]) = [x_{n-1}, x_n]$, where a_n and e_n are the constants defined in section II.3. Let c, s be two real numbers such that $0 < s < 1$, and $c > 0$. Let $M_n : X \to Y$ be a function that obeys

$$(Cond3) \quad d(M_n(a, y), M_n(b, y)) < c|a - b| \quad \forall a, b \in \Re,$$

and

$$(Cond4) \quad d(M_n(x, a), M_n(x, b)) < s d_Y(a, b) \quad \forall a, b \in Y.$$

Next, we want a transformation that 'projects' $\Re^3$ to $\Re^2$. In order to obtain such a transformation, we will go through these next steps. More specifically, we will state and prove two theorems, and then give the definition of a *generalized* fractal interpolation function before obtaining the desired transformation.

Consider a transformation $w_n : X \to X$ defined by

$$w_n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} L_n(x) \\ M_n(x,y) \end{pmatrix} \ \forall (x,y) \in X, n = 1, ..., N.$$

As an extension of Thm II.2.1., we can state Thm II.4.1 as follows;

Theorem II.4.1 Let the IFS $\{X; w_n, n = 1, ..., N\}$ be defined as above. Assume that there are real constants c, s such that $0 < s < 1$, and $c > 0$, and that conditions 3 and 4 are satisfied. Let θ (from condition 1) be defined by

$$\theta = (1-a)/2c,$$

where $a = \max\{a_n : n = 1, ..., N\}$. Then the IFS is hyperbolic with respect to the metric d.

Proof: Define the metric d on $\Re$ x Y by;

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + \theta d_Y(y_1, y_2),$$

where $a = \max\{a_i : i = 1, ..., N\}$.

$$d_Y(M_n(a,y), M_n(b,y)) \leq c|a-b|$$

$$d_Y(M_n(a,y), M_n(b,y)) \leq c|a-b|$$

$$d_Y(M_n(x,a), M_n(b,y)) \leq s d_Y(a,b)$$

and $\theta = (1-a)c$, and s, $c > 0$, and under the usual assumptions about the ordering of x_i, $1 > a > 0$.

Under each map, $(x,y) \to (L_n(x), M_n(x,y))$

$$\Rightarrow d(w_n(x_1,y_1), w_n(x_2,y_2)) = d((L_n(x_1), M_n(x_1,y_1)), (L_n(x_2), M_n(x_2,y_2)))$$

$$= |L_n(x_1) - L_n(x_2)| + \theta d_Y(M_n(x_1,y_1), M_n(x_2,y_2))$$

$$= |a_n x_1 - a_n x_2| + \theta d_Y(M_n(x_1,y_1), M_n(x_2,y_2))$$

$$= |a_n||x_1 - x_2| + \theta d_Y(M_n(x_1,y_1), M_n(x_2,y_2))$$

$$\leq a|x_1 - x_2| + \theta d_Y(M_n(x_1,y_1), M_n(x_2,y_2)).$$

We can use the triangle inequality on the last term of the above expression;

$$d_Y(M_n(x_1,y_1), M_n(x_2,y_2)) \leq d_Y(M_n(x_1,y_1), M_n(x_2,y_1)) + d_Y(M_n(x_2,y_1), M_n(x_2,y_2))$$

$$\leq c|x_1 - x_2| + s d_Y(y_1,y_2).$$

$$\Rightarrow d(w_n(x_1,y_1), w_n(x_2,y_2)) \leq a|x_1 - x_2| + \theta c|x_1 - x_2| + s d_Y(y_1,y_2)$$

$$\leq \frac{2a+1-a}{2}|x_1 - x_2| + s d_Y(y_1,y_2)$$

$$\leq \frac{1+a}{2}|x_1 - x_2| + s\theta d_Y(y_1,y_2).$$

Since $0 < a < 1$, we have $1 + a < 2$, so if we let

$$s' = \max\{(1+a)/2, s) < 1,$$

then we get;

$$d(w_n(x_1,y_1), w_n(x_2,y_2)) \leq s'|x_1 - x_2| + s'\theta d_Y(y_1,y_2) = s' d((x_1,y_1),(x_2,y_2)).$$

Hence the IFS is hyperbolic. End of proof.

Similarly, there is also a generalization that can be made for Thm II.2.2.. For our extension, we will need a new condition that will constrain the IFS used, so that its attractor includes the set of generalized data. Assume that

$(Cond5)$ $\quad M_n(x_0, F_0) = F_{n-1}, \qquad M_n(x_N, F_N) = F_n, n = 1, ..., N.$

From this assumption, it follows that

$$w_n(x_0, F_0) = (L_n(x_0), M_n(x_0, y_0)) = (x_{n-1}, F_{n-1}),$$

and

$$w_n(x_N, F_N) = (L_n(x_N), M_n(x_N, y_N)) = (x_n, F_n), \;\; n = 1, ..., N.$$

Theorem II.4.2 Let N be a positive integer. Let $\{X; w_n, n = 1, ..., N\}$ denote the same IFS defined above, associated with the data set $\{(x_i, F_i) \in \Re \text{ x } Y : i = 1, ..., N\}$. Assume that there are real constants c, s such that $0 < s < 1$, and $c > 0$, and that conditions 3, 4 and 5 (above) are satisfied. Let $G \in H(X)$ denote the attractor of the IFS. Then G is the graph of a continuous function $f : [x_0, x_N] \to Y$, which interpolates the data $\{(x_i, F_i) : i = 1, ..., N\}$. That is,

$$G = \{(x, f(x)) : x \in [x_0, x_N]\},$$

where

$$f(x_i) = F_i, \;\; i = 0, 1, ..., N.$$

Proof: ([B]) First note that by Thm II.4.1, the attractor exists. As we did in Thm II.2.2, we devise the operator $T : F \to F$ to be;

$$(Tf)(x) = M_n(f(L_n^{-1}(x))), \qquad x \in [x_{n-1}, x_n].$$

(F is the family of continuous functions $f : [x_0, x_N] \to Y$.)

$\to L_N(x) = a_n x + e_n$ ('L_n'was 'l_n'in Thm II.2.2)

$\to L_n$ is continuous, and so is its inverse.

$\to f$ is assumed to be continuous.

$\to M_n$ is linear, so that Tf is continuous on each $[x_{n-1}, x_n]$.

1. We verify that T does indeed take F into itself. We have

$$L_1(x_0) = a_1 x_0 + e_1 = \frac{x_1 - x_0}{x_N - x_0} x_0 + \frac{x_n x_0 - x_0 x_1}{x_N - x_0} = x_0.$$

$$L_N(x_N) = a_N x_N + e_N = \frac{x_N - x_{N-1}}{x_N - x_0} x_N + \frac{x_N x_{N-1} - x_N x_0}{x_N - x_0} = x_N.$$

$$L_n(x_{n-1}) = a_n x_{n-1} + e_n = \frac{x_n - x_{n-1}}{x_N - x_0} x_{n-1} + \frac{x_N x_{n-1} - x_0 x_{n-1}}{x_N - x_0} = x_{n-1}.$$

$$L_n(x_n) = a_n x_n + e_n = \frac{x_n - x_{n-1}}{x_N - x_0} x_n + \frac{x_N x_n - x_0 x_n}{x_N - x_0} = x_n.$$

$\Rightarrow$ The end points of each $[x_{n-1}, x_n]$ get mapped to the next and last,

$\Rightarrow$ x_0 and x_N get mapped to themselves.

$\Rightarrow$ All these points are fixed under L_n, and thus under L_n^{-1}.

By definition,

$$(Tf)(x_0) = M_n(x_0, f_0) = (x_{n-1}, F_{n-1}),$$

$$(Tf)(x_N) = M_n(x_N, f_N) = (x_n, F_n).$$

So $T : F \to F$ as desired.

2. We show that T is a contraction mapping. Let $f, g \in F$. Then;

$$(Tf)(x) = M_n(L_n^{-1}(x), f(L_n^{-1}(x)))$$

$$(Tg)(x) = M_n(L_n^{-1}(x), g(L_n^{-1}(x)))$$

Next, from the definition of a metric, and the restrictions on M_n, we have;

$$D(Tf, Tg) = \max_{x \in [x_{n-1}, x_n]} d_Y(M_n(L_n^{-1}(x), f(L_n^{-1}(x))), M_n(L_n^{-1}(x), g(l_n^{-1}(x))))$$

$$\leq \max_{x\in[x_{n-1},x_n]} s d_Y(f(L_n^{-1}(x)), g(L_n^{-1}(x))) \leq d(f,g).$$

Hence, T is a contraction. End of proof.

Definition II.4.3 The function f defined in Theorems II.4.1 & II.4.2, where its graph is the attractor of an IFS as described is called a *generalized fractal interpolation function (G.F.I.F.)*, corresponding to the generalized data $\{(x_i, F_i) : i = 1, ..., N\}$.

As one might expect, G.F.I.F.'s produce more flexible interpolation functions. We use affine transformations acting on $\Re^3$, and we project the graph of the G.F.I.F. into $\Re^2$. The extra degrees of freedom provided by working in $\Re^3$ give us hidden variables. As with the scaling factors, these can be used to adjust the shape and fractal dimension of the interpolation function.

If the use of scaling factors can be compared to a shareware version of a software, then the set of hidden variables could be compared to the full version of the software... In the following paragraph, note where the d_n is located in the matrix.

We introduce a set of real parameters $\{H_i : i = 0, 1, ..., N\}$ (they are assumed to be fixed for now). Next, define a generalized set of data to be $\{(x_i, F_i, H_i) \in \Re$ x $\Re^2 : i = 0, 1, ..., N\}$. So we are considering the case $(Y, d_Y) = (\Re^2$,Euc.). Now consider an IFS $\{\Re^3; w_n, n = 1, ..., N\}$, where $w_n : \Re^3 \to \Re^3$ is affine of the special structure:

$$w_n \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_n & 0 & 0 \\ c_n & d_n & h_n \\ k_n & l_n & m_n \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \\ g_n \end{pmatrix}$$

Here, $a_n, c_n, d_n, e_n, f_n, g_n, h_n, k_n, l_n, m_n$ are all real numbers. Assume they obey

$$w_n \begin{pmatrix} x_o \\ F_0 \\ H_0 \end{pmatrix} = w_n \begin{pmatrix} x_{n-1} \\ F_{n-1} \\ H_{n-1} \end{pmatrix}$$

and

$$w_n \begin{pmatrix} x_N \\ F_N \\ H_N \end{pmatrix} = w_n \begin{pmatrix} x_n \\ F_n \\ H_n \end{pmatrix}, \qquad n = 1, 2, ..., N.$$

So we now have the desired transformation that projects $\Re^3$ to $\Re^2$ as we can say

$$w_n \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} L_n(x) \\ M_n(x,y,z) \end{pmatrix} \qquad (x,y,z) \in \Re^3, \ n = 1,2,...,N,$$

where $L_n(x) = a_n x + e_n$ (as before) and

$$M_n \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A_n \begin{pmatrix} y \\ z \end{pmatrix} + \begin{pmatrix} f_n + c_n x \\ g_n + k_n x \end{pmatrix} \qquad (A_n = \begin{pmatrix} d_n & h_n \\ l_n & m_n \end{pmatrix}).$$

In the condition #5, replace F_n by (F_n, H_n). (This implies that M_n obeys cond. #5.) Define

$$c = \max\{\max\{c_i, k_i\} : i = 1,...,N\}.$$

(This implies that cond. #3 is true). Now assume the linear A_n is contractive, with contraction factor s, with $0 < s < 1$. (This implies that cond. #4 is true).

In conclusion, under the conditions seen here, the IFS satisfies the assumptions of Thm II.4.2. Therefore, the attractor of the IFS is the graph of a continuous function $f : [x_0, x_N] \to \Re^2$ such that

$$f(x_i) = (F_i, H_i), \qquad i = 1,...,N.$$

Now if we write $f(x) = (f_1(x), f_2(x))$, then $f_1 : [x_0, x_N] \to \Re$ is continuous, and

$$f_1(x_i) = F_i, \qquad i = 1,...,N.$$

Definition II.4.4 The function f_1 constructed above is called a *hidden variable fractal interpolation function* associated with the set of data $\{(x_i, F_i) \in \Re^2 : i = 1,...,N\}$.

Let A denote the attractor of the IFS we are working with. Remark that although $A = \cup_{n=1}^{N} w_n(A)$, it is not the case if 'attractor'is replaced by 'projection of the attractor'. The graph of f_1 is not self-similar.

CHAPTER III: The Experiment

The two sections of this chapter establish how the graphical analysis of two stocks is done. As mentioned in Chapter I, the method used to calculate the fractal dimension has to be optimal for the data considered. So the more we know about how the data is made available, better are our chances of using the right tool. The first section explains how the data was received and processed, and in the second section we talk about the program.

III.1 The Data and How to Use It

The data was made available by Nesbitt Burns, a company owned by the Bank of Montreal. Whenever there is a transaction realized at the stock market, the information given by Nesbitt is the time at which the transaction has been realized, (up to the nearest minute), the volume, and the price of the stock at the exchange. We disregard the volume here while keeping in mind that it could be a good variable to add to the analysis of the stock. So, for example, if a stock is traded at 9:30 a.m., then at 9:39 a.m., and then at 9:40 a.m., we will consider these times and the price of the stock at those particular times.

What if there is more than one transaction during one minute? (and this happens a lot, sometimes 9 or 10 transactions during one minute...) Then we only consider the value of the stock at the end of that minute. So we obviously have an approximating error here, but I think that it is of minimal importance. The idea is to have a cloud of points (Fig. III.1), each representing the price of the stock at a certain minute.

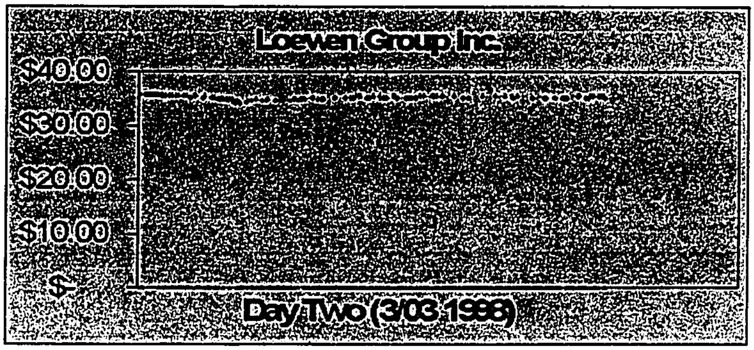

FIGURE III.1: One day of transaction = 390 minutes.

If there is no transaction during a minute, the default value of the price is 0$. This is a somewhat critical move. Why do we do it? Three reasons. The first reason is that sometimes, the price of the stock will fluctuate for a few consecutive minutes, and then stop for another few minutes. But a few minutes later, when the buyers and sellers start transacting again, the price of the stock may not be the same as it was when they stopped a few minutes ago. So there is no use in putting a dot at every non-transaction minute.

Let us clarify this with an example. Say the stock is at $35.00 at 9:35 a.m., and that there are no transactions between 9:35 a.m. and 10:00 a.m. Well, at 10:00, the stock may very well be listed at $35.25. The reasons and consequences of these abrupt changes will not be taken into account, as long as we keep in mind that whatever happened in the past should continue happening in the future (talking about the abrupt changes). So joining a line of dots from 9:35 to 10:00 is irrelevant since we would be forced to either have an abrupt change or 'smooth-out'the line of dots, and thus estimating too much, giving us too many errors.

The second reason why we give the value of 0$ to non-transaction minutes. By doing so, the cloud of dots is less dense, giving a lower fractal dimension. Also, it is easier to differentiate, just by looking at the graph, when the stock is more 'active', and when it is less 'active'.

The third reason is actually a compliant consequence coming from the fact that in the post-experiment analysis (Chapter IV) we compare the slope of the regression line with the fractal dimension of the cloud of dots. If we were to keep all the dots (transaction and non-transaction minutes), the slope of the regression line would be represented by a number closer to 0. Since we will be looking for a relationship between the fractal dimension and the slope of the linear regression, it is best that the two be close in range... More on that in Chapter IV.

As an example, let us consider the second day of transaction for the company Loewen Group Inc. (of the two companies analyzed). We can see that the data never goes above \$40, and never steps below \$30 (Fig. III.1). In the box-counting experiment, the boxes will cover the cloud of dots in the rectangle of height 40 - 30 = 10 (Fig. III.2).

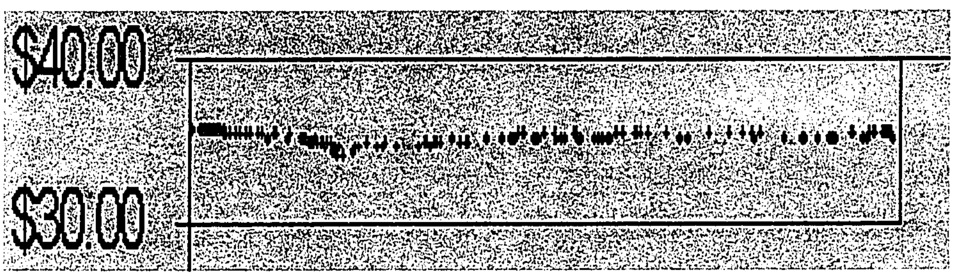

FIGURE III.2: Zoom in of Fig. III.1.

BOXES = RECTANGLES!

As we can see in figure III.3, the boxes are not squares, but the algorithm still goes through each and every one of them, and reports if there is at least one y-value in it. Whether the box is a square, a circle, a triangle, or anything regular and easily duplicable, the theorem will still apply.

Considering one day of transaction, for the first step of the algorithm the big unique covering box will have a length of 390 minutes and a height of \$10 (i.e., length = 1 day, height = 10). For the next step, we will have four smaller boxes, each of length 195 minutes, and height \$5. Now for the next step, the new (16) smaller boxes will have height \$2.50 (Fig. III.3), but the length should be of 98.5 minutes. As we will see in section III.2, the data fed to the program is only read at integer minutes. So we have to take the integer part of the length of the boxes, i.e, here, 98 minutes instead of 98.5.

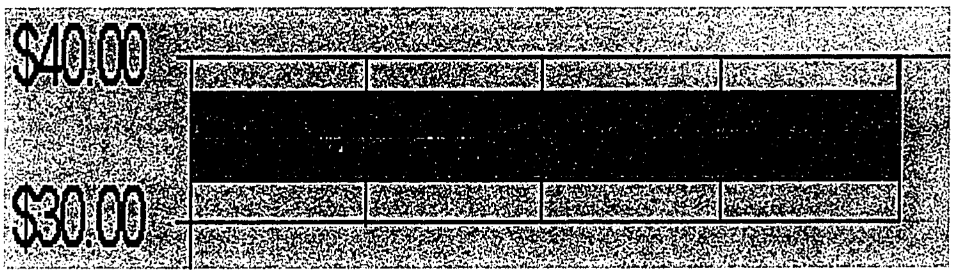

FIGURE III.3: We need 8 (out of 16) covering boxes after 2 subdivisions.

Actually in the experiment, the program analyses the data and finds the maximum and minimum values of the price of the stock for the period considered. Next, if, say the maximum value is \$38.90, and the minimum value is \$32.12, then instead of \$40.00, the program would use \$38.91, and instead of \$30.00, the program would use \$32.11. So in figures III.1,2 and 3, the original rectangle (and its subsequent subdivisions) is overly represented.

III.2 The Setup for the Program

MINIMAL BOX-SIZE

Let us keep in mind that a dot represents (at least) one transaction during one minute. The data is obviously not continuous, but discrete and finite. By looking at the graph, we can see that at some point in the algorithm, boxes will be small enough for each dot to be covered by a box (i.e. no two dots are in the same box). When this happens, there is no use to continue and so, we stop the algorithm. In fact, it is even too late to stop the algorithm at that time. (The determination of when the algorithm has to stop is discussed below.)

Let $E - S$ (End – Start) be the total number of minutes elapsed during the whole period (the period over which we want to estimate the fractal dimension). So in our example (Fig. III.1), one day of transaction has 6.5 hours, and therefore, if we consider one whole day of transaction, $E - S = 390$ minutes.

The height of the boxes is not as important as the width (the length). In the algorithm, the y-values are real numbers. So the loop that checks whether a certain row contains a dot or not, simply verifies with a $\leq$ sign for the base line of the row, and a $<$ sign for the upper line of the row. So having a height of, say \$1.25, is quite alright.

In the algorithm, the width (in minutes) of the smaller and smaller boxes is important to know as it can tell us when to stop. Why is this true? Once the length of a column is of smaller width than the smallest interval of time between any two transactions, only one dot can occur in the whole column. So we must stop dividing the boxes. And in fact, according to other similar box-counting experiments (for example, see [V]), the length of time (width) of the smallest boxes should not be lower than the **average time between transactions**. Let us denote the length of the smallest boxes by s (in minutes), and let A represent the smallest interval of time, and B, the average interval of time between any two transactions. Then we will want to have to following condition on s:

$$A \leq B \leq s$$

Obviously, the average time is always greater than the minimal time. So we will want the algorithm to stop before the boxes get smaller than the average time, and thus certainly before they get smaller than the minimal time.

Now, suppose we know what is the average amount of time between any two transactions (thanks to a subalgorithm in the program that calculates this average and lets the user know what it is before the algorithm starts). Then, we can calculate exactly when the boxes will be small enough to stop the algorithm. For example, if the average time between transactions is 4.5 minutes, we can ask the program to stop the algorithm when the length of the boxes attains a low of 6 minutes... In fact, the optimal s was chosen to be 6 minutes for the experiment.

Note that the width of the boxes (rectangles) after k subdivisions is proportional to the length of the period considered. So the width of the boxes (in minutes), that we will denote $w(k)$, after k subdivisions is $w(k) = \frac{E-S}{2^k}$, where $E - S$ is in minutes, and where k goes from $0, 1, 2, 3, \ldots$ to M (M is defined below).

This leads us to the determination of M. M is the number of steps the algorithm goes through. If $M = 4$, then the original 390 minutes x \$10.00 box is subdivided 4 times, and so, the smallest boxes (after 4 subdivisions, we have 256 boxes) will have length $w(4) = 390/2^4 = 24.625$ minutes (or rather, 24 minutes). An optimally chosen M will be very important to get accurate results in the least-squares approximation of the log-log plot. As s represents the length of the smallest of our covering rectangles (in minutes), it will be of no surprise to see that M is a function of s...

M represents the sufficient and otpimal number of divisions that can be made in the Box-Counting Theorem. M is chosen in the following way: Before the length (in minutes) of the boxes is smaller than the average interval of time between two transactions, the algorithm has to stop. So, assuming we know what s is, then M satisfies the inequality

$$(E - S)/2^M < s,$$

where $E - S$ is the length of the period considered. If we solve for M, we get

$$M > \log(E - S/s)/\log(2).$$

Taking M to be the integer part +1 of the right hand side of the above inequality, i.e., $M = \mathrm{INT}(\log(E - S)/s)/\log(2)) + 1$ will the optimal choice.

The program used in the experiment is listed in the appendix. The computer divides the imaginary rectangle on the $\mathcal{Z}$ x $\Re$ plane, where $\mathcal{Z}$ represents the integer axis, into smaller and smaller boxes, and counts the number of data points in each one. The program is written in Liberty Basic, and should run on any PC. A fast computer is needed in order to obtain quick results...

The program is self-contained in the sense that all the data points are already in it. We cover one week for each of the two companies. The program does not have to read its data from a file or a disk. To save (a lot of) space in the appendix, only the first and last fifteen minutes of transaction for the week are listed in the program.

CHAPTER IV: Post-Experiment Analysis

In this chapter, we analyze the results of the experiment. The first section establishes what was expected from the experiment. The second section considers the results and compares them with the hypothesis.

IV.1 What Pattern Do We Want to Look For?

Our first intuitive hypothesis was to expect an increase in the fractal dimension when the data became more volatile, and a decrease when the data stabilized. This is a rather logical idea when considering the definition of the fractal dimension.

So how can one calculate, or estimate the volatility of the data? Well, another intuition was to think that the slope of the regression line would go through changes that would be proportional to the changes in the volatility. Now this is not as safe an idea as the one in the preceeding paragraph. It is quite possible, however unlikely, that two regression lines be equal (or at least have the same slope), one approximating data that is spread out and volatile, and the other approximating very condensed and almost linear data. We have to keep this in mind throughout the analysis.

IV.2 Results

The two graphs shown in each of the Figures IV.1,2,3 and 4 express respectively the results of the fractal dimension and the results of the absolute value of the slope of the regression line. A sample of the output of the program is listed in the appendix. The program ran a total of thirty times. We can see in the graphs showing the results of the experiment (Figures IV.1-4) that the slope of the regression line and the fractal dimension go through similar changes as time goes on.

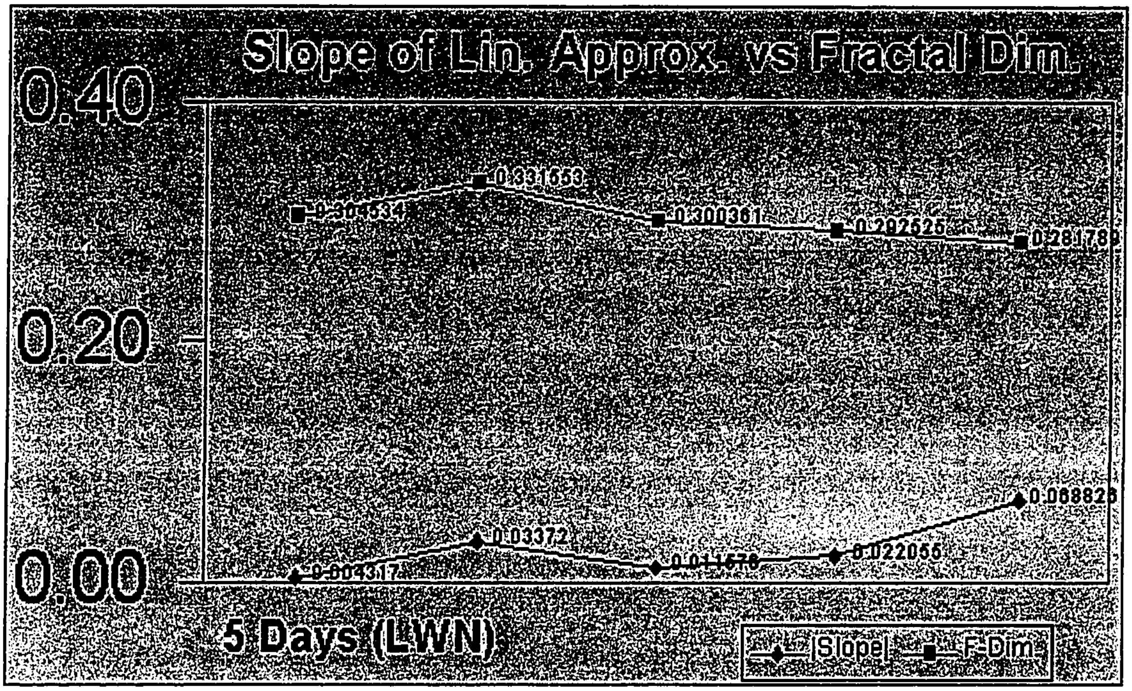

FIGURE IV.1: Results for one week, in five periods. Company: Loewen Group Inc.

Unfortunately, we could only consider the absolute value of the slope of the linear regression. Whether the slope is positive or negative is not represented by any visible sign in the fluctuations of the fractal dimension. This is not in contradiction with the above statement which says that the volatility and the change come together. But so far, no conclusion could be made with regards to the fractal dimension versus the sign of the slope of the regression line.

It is important to say a word concerning the end of the week for Loewen Group Inc. (Fig IV.1). On Friday, especially in the afternoon, not too many transactions were taking place. Consequentely, the data are very scarse. Intuitively, this should result in a lower fractal dimension (and it does). But the slope of the regression line that approximates the same data may very well go wild...

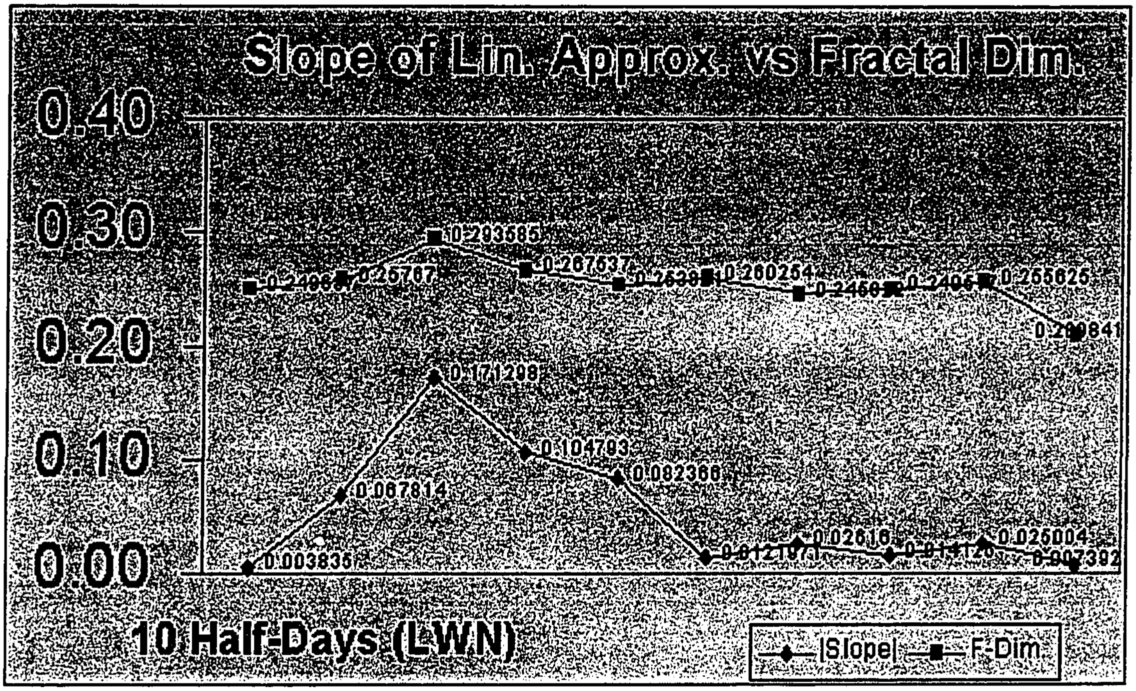

FIGURE IV.2: Results for one week, in ten periods. Company: Loewen Group Inc.

In Figure IV.2, the idea that there is a lack of data for Friday afternoon is more easily seen. When approximating a whole day (Fig. IV.1, Friday), the regression line has a rather large slope, but when considering the day in two parts (Fig. IV.2, Friday), the two slopes (and especially the one representing Friday afternoon) are much smaller. This confirms the idea that the transactions were too rare that Friday.

Although it was of less importance, there was also a period where less transactions were held. If we look at what happened on Wednesday (p.m.) and Thursday (a.m.) for Loewen Group Inc. (Fig. IV.1 & IV.2), we can see that the two graphs diverge. A quick check in the data shows that this particular period was also a slow one.

The next company is Bell International.

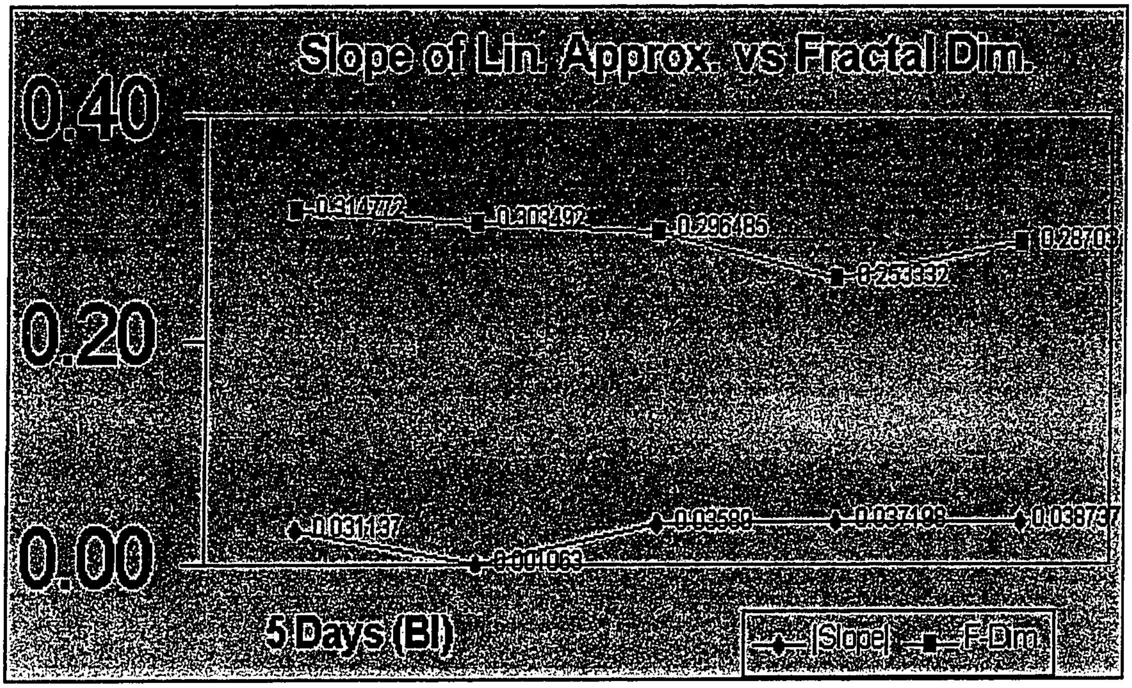

FIGURE IV.3: Results for one week, in five periods. Company: Bell International

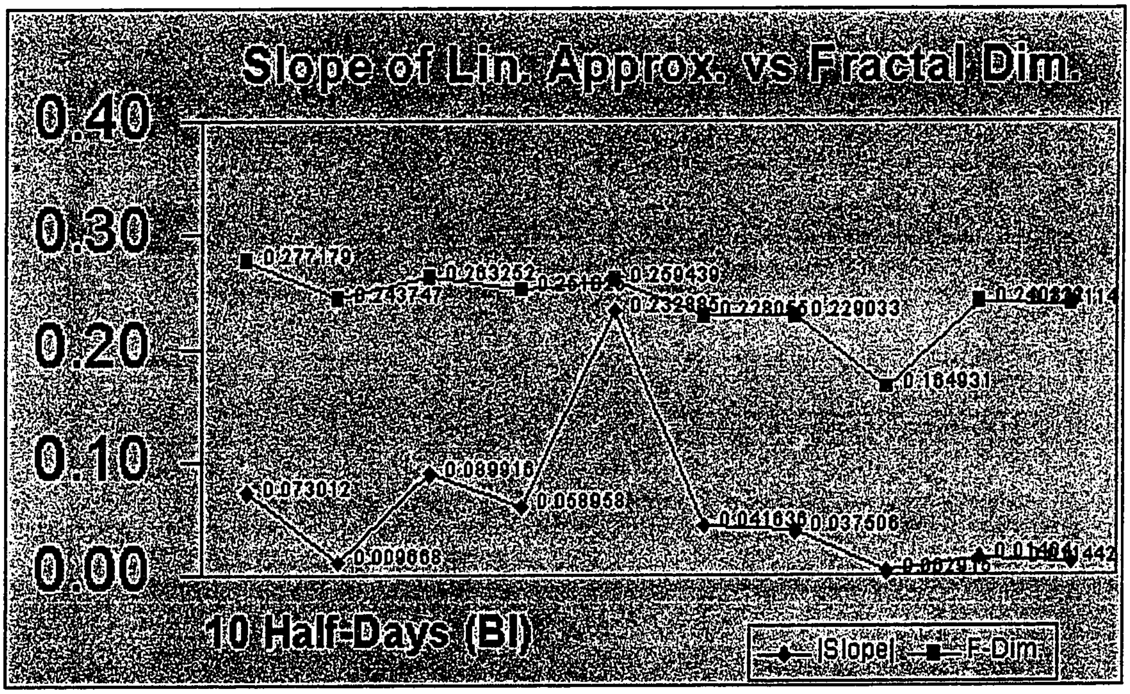

FIGURE IV.4: Results for one week, in ten periods. Company: Bell International

Our first comment concerns the number of transactions held (the volume) during the week of data we have for Bell International. The data representing these transactions is much more dense than the data for Loewen Group Inc. The cohesion seems stronger between the two graphs in Figures IV.3 & IV.4.

A logical statement would say that the more data you have, better are your chances of obtaining an accurate approximation... The hypothesis and the conclusion of this statement are both verified in this experiment.

As mentioned in Chapter I, it is important to specify what range of lengths of covering boxes was used. Since the same range was used for all of the thirty times the results were compiled, let us state it here. The length of the biggest covering box was either 390 minutes, (for one-day periods) or 195 minutes (for half-day periods). The length of the smallest covering boxes, s, was 6 minutes.

A possible future use of a new and improved program could be to help investors who already know what they are buying or selling. More specifically, since the comparison made with the fractal dimension concerns the absolute value of the slope of the regression line, the possiblity of predictions are those of movement, and not of the direction of the movement. In this case, the program would help them to decide **when** to buy or sell. Even so, we would have to find out that the fractal dimension changes slightly before the slope of the regression line does.

But how long before the optimum transaction time could the analysis confirm the intuition? One day? One hour? Five minutes...? Good questions that should and will be investigated in the future but are out of the scope of our actual study.

CHAPTER V: Conclusion

The notion of Fractal Geometry is a very interesting and inviting subject. In the past twenty years, many experiments have been conducted on various subjects. The fractal dimension and fractal interpolation are only two of the many tools used in these experiments and applications. Fractal growth, Brownian motion, countless studies on boundaries, are just a few examples of the ever-growing math branch classified as Fractal Geometry.

Looking back at this text, we first notice all the qualificative words like 'remarkable', 'beautiful', 'surprisingly neat', etc... And I am convinced that in the future, these words will always stick to the subject of Fractal Geometry.

In the first two chapters of the text, we covered the theoretical parts related to the fractal dimension and fractal interpolation functions. In this latter part, the way our theorems expressed the fractal dimension of the fractal functions is quite remarkable. And although the experiment in Chapter III only requires the notions seen in Chapter I, the numerous applications of fractal interpolation functions make it a very inviting subject. I would say that the topic of fractal functions should be seen soon after the basics of the fractal dimension.

For a first experiment with an invented program, we could say that the graphical analysis of the two stocks went very well. We cannot be completely satisfied with the program and should concentrate on what can be improved. The most important upgrade would be that the choice of the length of the smallest covering boxes (s) be done automatically. After analyzing the data, the program should be able to evaluate the optimal value for s.

We considered two companies with different volumes. By volume, we mean the number of stocks exchanged during a certain period. For the whole week, the company Loewen Group Inc. had a relatively low volume and Bell International had an average volume.

If we concerned ourselves only with the fractal dimension (and not the slope of the regression line), then the volume would have been a biasing factor of less importance. But I think that we should look in the other direction and at some point, one could find a way to deal with companies having a low volume. This alone would be a good reason to add the volume as a specific variable in the analysis. Perhaps we could consider the price of the stock as a function of its volume and not as a function of time. Certainly in this way, companies with a low volume and companies with a high volume could be compared equally by taking two different time intervals for each...

A conclusion is one which says that there is a tangible cohesion between the fractal dimension of a cloud of points and the absolute value of the slope of the regression line that approximates the same cloud of points. Future enquiries in this direction could be to represent the strength of the cohesion by a number or on a scale which has a spectrum ranging from no cohesion, to a very strong cohesion.

REFERENCES

Note that a '[ ]'refers to a specific topic from the text.

[B] Barnsley, M.F., *Fractals Everywhere*, Academic Press, 1993.

Barnsley, M.F., Elton, J., Hardin, D., Massopust, P., *Hidden Variable Fractal Interpolation Functions*, SIAM, Vol 20. No. 5 pp.1218-1242, September 1989.

Bowers, C.S., *On Teaching and Learning the Concept of Fractal*, Master's Thesis, Concordia University Montreal, 1991.

[F] Falconer, K.J., *Fractal Geometry, Mathematical Foundations and Applications*, John Wiley & Sons, 1990.

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

[FR] Freund, J.E., *Mathematical Statistics*, Prentice-Hall, 1992.

Hutchinson, J.E., *Fractals and Self-Similarity*, Indiana University Mathematics Journal, Vol. 30, No. 5, p.713-747, 1981.

Kaye, B., *A Random Walk Through Fractal Dimensions*, New York: VCH, 1994.

Lu, N., *Fractal Imaging*, Academic Press, 1997.

Mandelbrot, B.B., *The Fractal Geometry of Nature*, W.H. Freeman and Co., New York, 1982.

Mandelbrot, B.B., *Fractals: Form, Chance and Dimension*, W.H. Freeman and Co., San Fransico, 1977.

Mishra, S.K., *Dimensions of Attractors*, Master's Thesis, Concordia University Montreal, 1990.

Peitgen, H.O., Jurgens, H. and Saupe, D., *Chaos and Fractals, New Frontier of Science*, Springer-Verlag, New York, 1992.

Peters, E.E., *Chaos and Order in the Capital Markets*, John Wiley & Sons, 1996.

Royden, H.L., *Real Analysis*, Macmillan, 1988.

Saether, G., Bendiksen, K., Muller, J., Froland, E., *The Fractal Dimension of Oil-Water Interfaces in Channel Flows*, Institute of Mathematics and its Applications Conference Series, New Series Number 43, p.371-378, Clarendon Press, Oxford, 1993.

Sandogdhar, F., *Connectedness of the Attractor of an Iterated Function System*, Master's Thesis, Concordia University Montreal, 1995.

Sutherland, W.A., *Introduction to Metric and Topological Spaces*, Clarendon Press, Oxford, 1975.

Tricot, C., *Curves and Fractal Dimension*, Springer-Verlag, New York, 1995.

Vassilicos, J.C., Demos, A., Tata, F., *No Evidence of Chaos But Some Evidence of Multifractals in the Foreign Exchange and the Stock Markets*

APPENDIX

This is the listing for the Program.

```
DIM f(2000)

'THE DATA

f(0) = 35.5
f(1) = 0
f(2) = 0
f(3) = 0
f(4) = 0
f(5) = 0
f(6) = 0
f(7) = 0
f(8) = 0
f(9) = 35.3
f(10) = 0
f(11) = 0
f(12) = 0
f(13) = 35.5
f(14) = 35.35
f(1936) = 0
f(1937) = 0
f(1938) = 36.35
f(1939) = 36.35
f(1940) = 0
f(1941) = 36.3
f(1942) = 36.3
f(1943) = 0
f(1944) = 0
f(1945) = 0
f(1946) = 0
f(1947) = 36.3
f(1948) = 0
f(1949) = 36.3
f(1950) = 0

LET L = 33.99       'Low is 34.00$
LET H = 36.51       'High is 36.50$
Let z = H-L
print "At what minute do you want "
print "the experiment to Start"
input S
print "At what minute do you want "
print "the experiment to End"
input E

LET T = 5*390
for i = S to E
if f(i) = 0 then
let r = r+1
```

```
end if
next i
Let a = (E-S) - r
Print "The average time between points "
print "of data is ";(E-S)/a;" minutes."
print
'[chooses]
print "Enter the approximate length (s) of the"
print "smallest covering rectangle (in minutes)"
input s

LET M = int(log((E-S)/s)/log(2))

print
print "With this value of s, the smallest rectangle"
print "will be of length ";int((E-S)/(2^M));" minutes."
print
'print "Do you want to change s? (No)"
'input o
'if o ="Y" then
'goto [chooses]
'end if
print
PRINT "RESULTS OF THE BOX-COUNTING THEOREM."
PRINT
PRINT "n        N(A)      ln(N(A))       ln((2^n)/z)"
PRINT "---------------------------------------------"

FOR n = 0 TO 2
    LET N = 0
    FOR k = 1 TO 2 ^ n
    FOR j = 1 TO 2 ^ n
        LET c = 0
'THE "BOX-DIVIDING" LOOP
        FOR i = int((j-1)*(E-S)/2^n)+S TO int(j*(E-S)/(2^n))+S
            IF ((k - 1) / 2 ^ n)*(H - L) + L <= f(i) AND f(i) < (k / 2 ^ n)*(H -
L) + L THEN
            c = c + 1
            ELSE c = c
            END IF
        NEXT i
        IF c >= 1 THEN
        N = N + 1
        else N = N
        END IF
    NEXT j
    NEXT k
    IF n = 0 THEN
    PRINT n; "        "; N; "         "; LOG(N); "             "; LOG((2^n)/z)
    ELSE
    PRINT n; "        "; N; "         "; LOG(N); "      "; LOG((2^n)/z)
    END IF
    PRINT
    LET k(n) = LOG(N)
NEXT n

FOR n = 3 TO M
```

```
     LET N = 0
     FOR k = 1 TO 2 ^ n + 1
     FOR j = 1 TO 2 ^ n + 1
         LET c = 0
'THE "BOX-DIVIDING" LOOP
         FOR i = int((j-1)*(E-S)/2^n)+S TO int(j*(E-S)/(2^n))+S
             IF ((k - 1) / 2 ^ n)*(H - L) + L <= f(i) AND f(i) < (k / 2 ^ n)*(H -
L) + L THEN
             c = c + 1
             ELSE c = c
             END IF
         NEXT i
         IF c >= 1 THEN
         N = N + 1
         else N = N
         END IF
     NEXT j
     NEXT k
     IF n = 0 THEN
     PRINT n; "        "; N; "         "; LOG(N); "            "; LOG((2^n)/z)
     ELSE
     PRINT n; "        "; N; "         "; LOG(N); "     "; LOG((2^n)/z)
     END IF
     PRINT
     LET k(n) = LOG(N)
NEXT n
'END OF THE MAIN ALGORITHM
'LEAST SQUARES APPROXIMATION ALGORITHM
Print "What is the first point to be"
Print "considered in the L-S Method (n=?)"
input f
FOR i = f TO M
    sumy = sumy + k(i)
    sumxy = sumxy + (LOG((2^i)/z)+10) * k(i)
    sumx = sumx + LOG((2^i)/z)+10
    sumxx = sumxx + (LOG((2^i)/z)+10)^2
NEXT i
Sxx = sumxx - (1 / M) * (sumx) ^ 2
Sxy = sumxy - (1 / M) * (sumx) * (sumy)
dimension = (Sxy) / (Sxx)
PRINT
PRINT "Fractal Dimension ...... "; dimension
'PRINT "(for the "; T/60; " hours of transaction time)"
Print
FOR i = S TO E
    sumy = sumy + f(i)
    sumxy = sumxy + i*f(i)
    sumx = sumx + i
    sumxx = sumxx + i^2
NEXT i
Sxx = sumxx - (1 / (E-S)) * (sumx) ^ 2
Sxy = sumxy - (1 / (E-S)) * (sumx) * (sumy)
slope = (Sxy) / (Sxx)
PRINT
PRINT "Slope of the Regression ...... "; slope
PRINT "(for the "; (E-S)/60; " hours of transaction time)"
END
```

This is the output of the program for the first day of the week for the company Bell International.

```
At what minute do you want
the experiment to Start
?1
At what minute do you want
the experiment to End
?390
The average time between points
of data is 4.1827957 minutes.

Enter the approximate length (s) of the
smallest covering rectangle (in minutes)
?5

With this value of s, the smallest rectangle
will be of length 6 minutes.


RESULTS OF THE BOX-COUNTING THEOREM.
```

n	N(A)	ln(N(A))	ln((2^n)/z)
0	1	0.0	-1.25846099
1	4	1.38629436	-0.56531381
2	8	2.07944154	0.12783337
3	13	2.56494936	0.82098055
4	27	3.29583687	1.51412773
5	53	3.97029191	2.20727491
6	72	4.27666612	2.90042209

```
What is the first point to be
considered in the L-S Method (n=?)
?2

The fractal dimension is... 0.30432428


Slope of the Regression ...... -0.84007081e-2
(for the 6.5 hours of transaction time)
```