Topology, Measure Theory, and Analysis on Fractals

Justin Furlotte

December 15, 2019

Contents

1 Introduction

Fractals are sets that exhibit *self-similarity*. In this thesis, we will discuss the construction of fractals with "iterated function systems" $\{F_i\}_{i=1}^n$ using the contraction mapping theorem and explore some of their properties. It is perhaps surprising at first that the closed interval $[a, b]$ is a fractal in this sense, so that the entire discussion of this article can be interpreted as giving a different viewpoint on the classical (Euclidean) case. We also give a parametrization of fractals via a space of "strings".

In section 3, we discuss several measures on both a fractal itself and its associated string space, specifically, a natural self-similar measure which is related to the iterated function system $\{F_i\}_{i=1}^n$. This self-similar measure is the one most commonly used when developing the theory of analysis on fractals. The s-dimensional Hausdorff measure is also introduced, as it is a very common measure for dealing with fractals in the literature. We also discuss a certain measure on the associated string space that is related to the Hausdorff measure.

In section 4, we investigate one of the most commonly used notions of dimension for fractals, the Hausdorff dimension, which is related to the Hausdorff measure. An important detail is that this dimension, which is known to coincide with other notions of dimension in standard Euclidean cases of $\mathbb{R}^n$ [2], actually allows for non-integer values. For example, as we shall see, the Cantor set has Hausdorff dimension $\log 2/\log 3$.

Finally, in section 5, we introduce the more recent development of Laplace operators on a certain class of fractals. We will even be able to solve Laplace's equation with arbitrary boundary values, given a sensible notion of "boundary" of a fractal.

2 Fractals and Self-Similarity

This section follows [2], [3], and [7].

2.1 Iterated Function Systems

Definition 2.1.1 (Similarity). *A function between metric spaces $f : S \to T$ is called a* similarity *with* ratio $r > 0$ *if*

$$\rho_T(f(x), f(y)) = r\rho_S(x, y)$$

for all $x, y \in S$*.*

Definition 2.1.2 (Contraction)**.** *A function between metric spaces* $f : S \to T$ *is called a* contraction *if there exists some* $0 \leq r < 1$ *such that*

$$\rho_T(f(x), f(y)) \leq r\rho_S(x, y)$$

for all $x, y \in S$*.*

Lemma 2.1.3. *Contractions and similarities are continuous.*

Proof. Let $f : S \to T$ be a contraction with ratio $0 \leq r < 1$, and let $\epsilon > 0$. Set $\delta = \epsilon/r$. Then if $\rho_S(x, y) < \delta$, we have

$$\rho_T(f(x), f(y)) \leq r\rho_S(x, y) < r\delta = \epsilon$$

Exactly the same argument works for similarities. □

The following theorem, called the *contraction mapping theorem*, is an important result and will play a central role in our definition of fractals.

Theorem 2.1.4 (contraction mapping theorem)**.** *Let* (S, ρ) *be a complete metric space, and let* $f : S \to S$ *be a contraction. There exists a unique* $x^* \in S$*, called the* fixed point *of* f*, such that* $f(x^*) = x^*$*. Furthermore, for each* $x_0 \in S$*, the sequence defined by* $x_{n+1} = f(x_n)$ *converges to* x^**.*

Proof. By induction, we have $\rho(x_{j+1}, x_j) \leq r^j\rho(x_1, x_0)$. Let $m < n$ and let $\epsilon > 0$. By applying the triangle inequality and the formula for a finite geometric series, we have

$$\begin{aligned}
\rho(x_m, x_n) &\leq \sum_{j=m}^{n-1} \rho(x_{j+1}, x_j) \\
&\leq \sum_{j=m}^{n-1} r^j \rho(x_1, x_0) \\
&= \rho(x_1, x_0)\frac{r^m - r^n}{1 - r} \\
&= \rho(x_1, x_0)\frac{r^m(1 - r^{n-m})}{r - 1} \\
&\leq \frac{\rho(x_1, x_0)r^m}{1 - r}
\end{aligned}$$

Since $r \in [0,1)$ we can take N large enough that $(\rho(x_1, x_0)r^N)/(1-r) < \epsilon$, which gives us the desired result: if $n, m > N$ then $\rho(x_m, x_n) < \epsilon$. By completeness of (S, ρ), the series x_n converges to some element $x^* \in S$. By lemma 2.1.3, f is continuous, so we also have $f(x_n) \to f(x^*)$. But, by definition, $x_{n+1} = f(x_n)$, so the two limits must be equal, i.e.

$$x_n \to f(x^*) = x^*$$

So the limit x^* is a fixed point of f. To show uniqueness, suppose there were another such fixed point $y^* \neq x^*$. Then, since x^* and y^* are fixed points, we have $\rho(x^*, y^*) = \rho(f(x^*), f(y^*))$. But since f is a contraction, we also have $\rho(f(x^*), f(y^*)) \leq r\rho(x^*, y^*)$. Now $0 \leq r < 1$, so this can only be possible if $\rho(f(x^*), f(y^*)) = \rho(x^*, y^*) = 0$, hence $x^* = y^*$. $\square$

We will use theorem 2.1.4 to define a fractal set. Specifically, given a finite set of functions $\{F_i\}_{i=1}^n$ which all are simultaneously similarities and contractions, we will later see that a fractal set K is the unique set satisfying $K = \cup_i F_i(K)$. This set of functions is of fundamental importance to the study of fractals, and has a special name.

Definition 2.1.5 (Iterated function system)**.** *A finite set $\{F_i\}_{i=1}^n$ of contractions which are also similarities is called an* iterated function system*.*

Given $r > 0$, the *open r-neighbourhood* of a set A in a metric space is defined to be:

$$N_r(A) := \{y \mid \rho(x, y) < r \text{ for some } x \in A\}$$

Also, given a metric space S, we shall denote the collection of nonempty compact subsets of S by $\mathbb{H}(S)$.

Definition 2.1.6 (Hausdorff metric)**.** *Given a metric space (S, ρ), the* Hausdorff metric *is the function $D : \mathbb{H}(S) \times \mathbb{H}(S) \to \mathbb{R}_{\geq 0}$ defined by*

$$D(A, B) := \inf\{r > 0 \mid A \subseteq N_r(B) \text{ and } B \subseteq N_r(A)\}$$

The fact that D is a metric on $\mathbb{H}(S)$ is fairly easy to show. It is clear from the definition that D is nonnegative and symmetric; the fact that $D(A, B) = 0$ if and only if $A = B$ also follows from the definition. The fact that D is finite follows from the fact that D acts on compact sets, and compact sets are bounded.

Only the triangle inequality needs a bit of attention. Let $A, B, C \in \mathbb{H}(S)$ and let $\epsilon > 0$. Then if $a \in A$, there exists $b \in B$ such that $\rho(a,b) < D(A,B) + \epsilon$. Similarly, there exists $c \in C$ such that $\rho(b,c) < D(B,C) + \epsilon$.

Notice that this implies that $A \subseteq N_p(C)$ and similarly $C \subseteq N_p(A)$, where $p = \rho(a,b) + \rho(b,c) = D(A,B) + D(B,C) + 2\epsilon$. Therefore, by definition,

$$D(A,C) \leq p = D(A,B) + D(B,C) + 2\epsilon$$

But this holds for all $\epsilon > 0$, hence $D(A,C) \leq D(A,B) + D(B,C)$ as was claimed. $\square$

Lemma 2.1.7. *Suppose that (S, ρ) is a complete metric space. Then $(\mathbb{H}(S), D)$ is also a complete metric space.*

Proof. Let $(A_n) \subseteq \mathbb{H}(S)$ be Cauchy. We define the candidate limit to be

$$A = \{x \mid \exists x_k \to x \text{ such that } \forall k,\, x_k \in A_k\}$$

There exists N such that if $n, m > N$, then $D(A_n, A_m) < \epsilon/2$. We claim that if $n \geq N$, then $D(A_n, A) < \epsilon$. To see this, fix $n \geq N$ and consider the following.

If $x \in A$, then there is some sequence (x_k) converging to x, hence there exists k_0 such that $\rho(x, x_{k_0}) < \epsilon/2$. Also, $k \geq N$ implies that there exists some $y \in A_n$ with $\rho(x_k, y) < \epsilon/2$, since $D(A_k, A_n) < \epsilon/2$. So $\rho(x,y) \leq \rho(y, x_k) + \rho(x_k, x) < \epsilon$. Since $x \in A$ was arbitrary, $A \subseteq N_\epsilon(A_n)$.

Now if $y \in A_n$, choose $k_1 < k_2 < \ldots$ so that $k_1 = n$, and $D(A_{k_j}, A_m) < \epsilon/2^j$ for all $m \geq k_j$. Notice that by this definition, there could be large gaps between any two successive elements k_j and k_{j+1}. Define the following sequence: set $y_k \in A_k$ arbitrarily if $k < n$, let $y_n = y$, and if y_{k_j} is defined with $k_j \leq k \leq k_{j+1}$ then choose $y_k \in A_k$ such that $\rho(y_{k_j}, y_k) < \epsilon/2^j$. Then (y_k) is Cauchy, hence convergent, say to x. Then $x \in A$ and $\rho(x,y) = \lim_{k\to\infty} \rho(y_k, y) < \epsilon$, so $y \in N_\epsilon(A)$, hence $A_n \subseteq N_\epsilon(A)$. Thus $D(A_n, A) < \epsilon$, so $A_n \to A$ in the Hausdorff metric. $\square$

We are finally ready to define a fractal.

Theorem 2.1.8. *Let $\{F_i\}_{i=0}^n$ be an iterated function system on a metric space (S, ρ). Then the* fractal *or* attractor *set of the iterated function system is the unique compact set K satisfying*

$$K = \bigcup_{i=1}^{n} F_i(K)$$

Furthermore, given any nonempty compact set $V_0 \subseteq S$*,the sequence*

$$V_{m+1} = \bigcup_{i=1}^{n} F_i(V_m)$$

converges to $K = \lim_{m\to\infty} V_m$ *in the Hausdorff metric.*

Proof. Notice that each $F_i : \mathbb{H}(S) \to \mathbb{H}(S)$ is a contraction between complete metric spaces, by lemma 2.1.7. Define

$$F(A) := \bigcup_{i=1}^{n} F_i(A).$$

Now, by lemma 2.1.3, each F_i is continuous. Since the continuous image of a compact set is compact, $F_i(A)$ is compact for any compact set A. Also recall that the finite union of nonempty compact sets is a nonempty compact set, so if A is compact then so is $F(A) = \cup_i F_i(A)$. So F is still a map from $\mathbb{H}(S)$ to $\mathbb{H}(S)$.

Next we show F is a contraction (in the Hausdorff metric). To see this, set $r = \max\{r_i\}$, for r_i the similarity ratio of F_i. Then $r < 1$ since each $r_i < 1$. Let $A, B \in \mathbb{H}(S)$; we wish to show that $D(F(A), F(B)) \leq rD(A, B)$.

Let $q > D(A, B)$. If $x \in F(A)$, then $x = F_i(x')$ for some i, for some $x' \in A$. Since $q > D(A, B)$, there exists $y' \in B$ such that $\rho(x', y') < q$. Thus $y = F_i(y') \in F(B)$ satisfies $\rho(x, y) = r_i\rho(x', y') < rq$.

Since this holds for all $x \in F(A)$, we have $F(A) \subseteq N_{rq}(F(B))$. Similarly, $F(B) \subseteq N_{rq}(F(A))$. So $D(F(A), F(B)) < rq$. Since this holds for all $q > D(A, B)$, we have the desired inequality

$$D(F(A), F(B)) \leq rD(A, B).$$

Hence $F : \mathbb{H}(S) \to \mathbb{H}(S)$ is a contraction between complete metric spaces, so by the contraction mapping theorem (2.1.4), F has a unique fixed point $K \in \mathbb{H}(S)$, so K is nonempty and compact, and the sequence defined by $V_{m+1} = F(V_m)$ converges to K for any $V_0 \in \mathbb{H}(S)$. □

The function $F : \mathbb{H}(S) \to \mathbb{H}(S)$ defined by $F(A) := \cup_{i=1}^{n} F_i(A)$ as in theorem 2.1.8 is sometimes called a *Hutchinson operator*.

Later (definition 5.2.1), we will describe *post-critically finite* (pcf) fractals. Essentially, a fractal that is pcf admits a notion of a finite boundary. Often, the boundary is exactly what one would expect. For instance, the boundary of the interval $[a, b]$ is simply the set of points $\{a, b\}$. The boundary of the Sierpinski gasket (example 2.1.12) is the set of vertices $\{q_0, q_1, q_2\}$ of the equilateral triangle from which it is constructed.

It is well-known that if we take the initial set V_0 in the sequence $V_{m+1} = \cup_i F_i(V_m)$ to be the boundary of a pcf fractal K, then

$$K = \overline{\bigcup_{m \in \mathbb{N}} V_m}.$$

We can also introduce the notion of an m-cell, which is a construction which helps us understand the self-similar nature of fractals. It is a generalization of the formula $K = \cup_i F_i(K)$.

Definition 2.1.9. *Let $\{F_i\}_{i=1}^{n}$ be an iterated function system. Let $w = w_1 w_2 \dots w_m$ be a "word" (i.e. string) of indices of length $|w| = m$, where each $w_j \in \{1, \dots, n\}$. Then*

$$K = \bigcup_{|w|=m} F_w(K),$$

where $F_w := F_{w_1} \circ \cdots \circ F_{w_m}$. This is called a level m decomposition *of K, and each $F_w(K)$ is called an m-*cell, *or* cell of level m.

It is not hard to prove the following formula. Let K be the fractal associated to an iterated function system $\{F_i\}_{i=1}^{n}$. Then

$$\bigcup_{|w|=2} F_w(K) = \bigcup_{j=1}^{n} \bigcup_{i=1}^{n} F_j \circ F_i(K) = \bigcup_{j=1}^{n} F_j(\cup_i F_i(K)) = \bigcup_{j=1}^{n} F_j(K) = K.$$

This argument easily generalizes by induction on m to the general case $|w| = m$. We thus have four descriptions of a fractal set associated with an iterated function system:

$$K = \bigcup_{i=1}^{n} F_i(K) = \lim_{m \to \infty} V_m = \overline{\bigcup_{m \in \mathbb{N}} V_m} = \bigcup_{|w|=m} F_w(K),$$

where again, $V_{m+1} = \cup_{i=1}^{n} F_i(V_m)$. In the limit definition, $V_0 \in \mathbb{H}(S)$ is any nonempty compact set of the metric space S. In the definition $K = \overline{\bigcup_{m \in \mathbb{N}} V_m}$, V_0 is the boundary.

Now that the definition of a fractal has been established, we can give some concrete examples of iterated function systems that give rise to well-known fractal sets. Some of these examples show how clean the iterated function system formulation of fractal sets is in comparison with other constructions.

Example 2.1.10 (Unit interval)**.** *It is perhaps surprising at first that this entire discussion can be applied to the unit interval, which is a self-similar set. Let $F_0(x) = \frac{1}{2}x$ and $F_1(x) = \frac{1}{2}(x+1)$. Notice that these are contractive similarities with ratio* 1/2*. The attractor of this iterated function system is the unit interval $I = [0,1]$:*

$$F_0([0,1]) \cup F_1([0,1]) = [0,1/2] \cup [1/2,1] = [0,1].$$

Its m-cells $F_w(K)$, $|w| = m$, are the subintervals of $[0,1]$ *whose boundary points consist of the dyadic points $k/2^m$. For example, with $m = 2$, the possible words of length m are* 00*,* 01*,* 10*, and* 11*, so*

$$\begin{aligned} F_0(F_0(I)) \cup F_0(F_1(I)) \cup F_1(F_0(I)) \cup F_1(F_1(I)) \\ &= [0,1/4] \cup [1/4,1/2] \cup [1/2,3/4] \cup [3/4,1] \\ &= I \end{aligned}$$

Example 2.1.11 (Cantor Set)**.** *The Cantor set is typically constructed as follows. Let $C_0 = [0,1]$ be the unit interval, and let $C_1 = [0,1/3] \cup [2/3,1]$ be the set constructed by removing the open "middle third" of C_0. Let*

$$C_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$$

be the set constructed by removing the open "middle thirds" of the intervals in C_1. In general, let C_{n+1} be the set resulting from the removal of the open middle third interval from each of the intervals in C_n, and define the Cantor set as

$$C = \bigcap_{n=0}^{\infty} C_n$$

The first few steps of this construction of the Cantor set are shown in figure 1*. The Cantor set is the set of points left after repeating this process infinitely many times.*

There is another construction of C via iterated function systems. Let $F_0(x) = \frac{1}{3}x$ and $F_1(x) = \frac{x+2}{3}$, which define contractions on the compact set $[0, 1]$. Then the attractor of the iterated function system $\{F_0, F_1\}$ is the Cantor set, $C = F_0(C) \cup F_1(C)$.

The see this, suppose that $x \in C$; then $x \in C_k$ for all k, so $x \in C_1$. In particular, either x is in $[0, 1/3]$ or $[2/3, 1]$. Suppose that it's in $[2/3, 1]$; the other case is handled similarly.

Let k be arbitrary. We have that $x \in C_{k+1} = F_0(C_k) \cup F_1(C_k)$. But $F_0(C_k) \subset F_0([0, 1]) = [0, 1/3]$ so x must be in $F_1(C_k)$. In other words, $3x - 2 \in C_k$. Since k was arbitrary, $3x - 2 \in C = \cap_k C_k$, which implies that $x \in F_1(C)$. This proves that $C \subseteq F_0(C) \cup F_1(C)$.

Next, let $x \in F_0(C) \cup F_1(C)$; then either $x \in F_0(C)$ or $x \in F_1(C)$. Suppose that $x \in F_1(C)$. Again, the other case is handled similarly. Then $3x - 2 \in C = \cap_k C_k$, so x must be in $F_1(C_k) \subseteq C_{k+1}$ for all k. So $x \in C_{k+1}$ for all k, and thus $x \in C = \cap_k C_k$, which shows that $F_0(C) \cup F_1(C) \subseteq C$.

Figure 1: The Cantor set.

Example 2.1.12 (Sierpinski Gasket)**.** *A typical construction of the* Sierpinski triangle *is as follows. Start with an equilateral triangle (including its interior); call this set SG_0.*

Remove an "upside-down" triangle in the center, creating three smaller copies of the original triangle (one on top, one on the bottom left, and one on the bottom right). Call this set SG_1.

Next, remove the middle triangle from each of the three triangles in SG_1; call this SG_2. Keep repeating this process and define the Sierpinski gasket to be $SG = \cap_{n \in \mathbb{N}} SG_n$.

There is another construction involving iterated function systems. Let $\{F_i\}_{i=0}^{2}$ be the iterated function system defined by $F_i(x) = \frac{1}{2}(x - q_i) + q_i$, for

$q_i \in \mathbb{R}^2$ the vertices of an equilateral triangle. Then the attractor of $\{F_i\}_{i=0}^2$ is SG. Its 1-cells $F_w(K)$ for $|w| = 1$ are the "next" three copies of SG; the one at the top, the one at the bottom left, and the one at the bottom right. Its 2-cells are the "next" three copies from a given 1-cell, etc.

The Sierpinski Gasket is shown in figure 2.

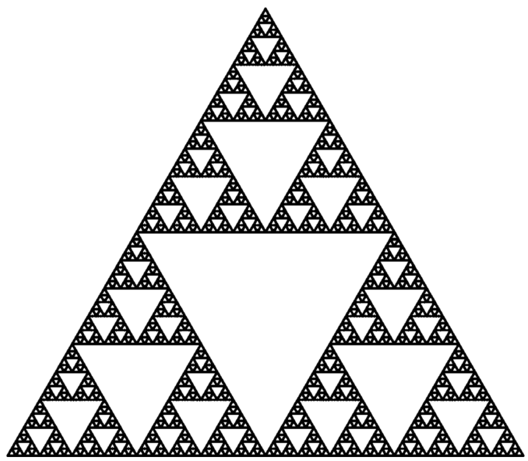


Figure 2: The Sierpinski gasket, or Sierpinski triangle.

Example 2.1.13 (Koch curve)**.** *The Koch curve (see figure 3) is the fractal set associated with the following iterated function system:*

$$F_0(\mathbf{x}) = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix} \mathbf{x}$$

$$F_1(\mathbf{x}) = \begin{pmatrix} 1/6 & -\sqrt{3}/6 \\ \sqrt{3}/6 & 1/6 \end{pmatrix} \mathbf{x} + \begin{pmatrix} 1/3 \\ 0 \end{pmatrix}$$

$$F_2(\mathbf{x}) = \begin{pmatrix} 1/6 & \sqrt{3}/6 \\ -\sqrt{3}/6 & 1/6 \end{pmatrix} \mathbf{x} + \begin{pmatrix} 1/2 \\ \sqrt{3}/6 \end{pmatrix}$$

$$F_3(\mathbf{x}) = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix} \mathbf{x} + \begin{pmatrix} 2/3 \\ 0 \end{pmatrix}$$

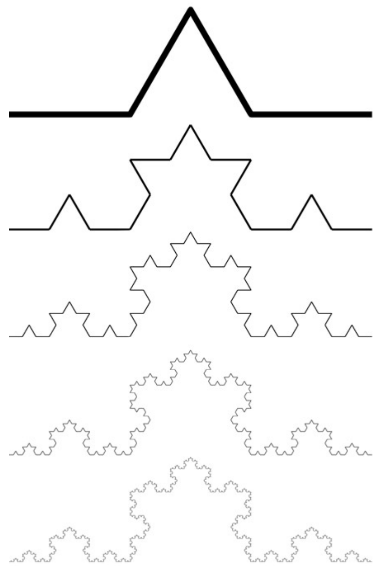


Figure 3: The first few steps $V_m = \cup_i F_i(V_{m-1})$ in the construction of the Koch curve, taking the interval as V_0.

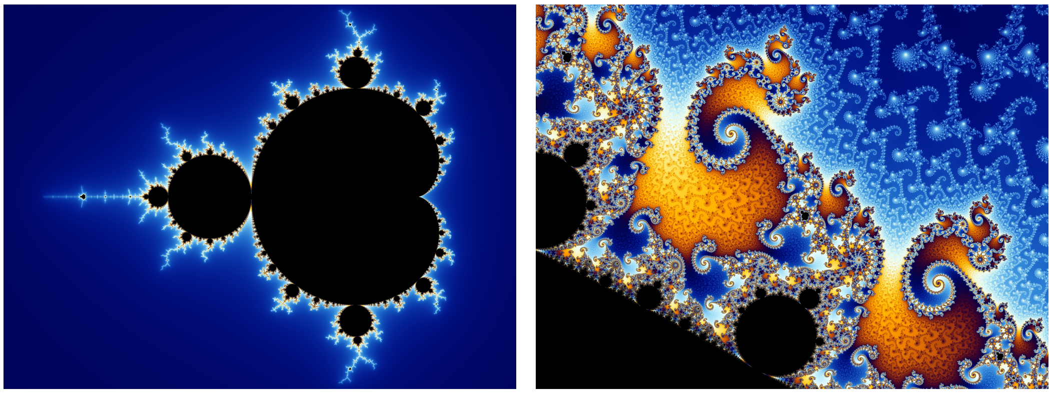
Example 2.1.14 (Mandelbrot Set)**.** *The Mandelbrot set J is the set of all complex numbers c such that the image of* 0 *remains bounded under arbitrary iterations of the function $f_c(z) = z^2 + c$:*

$$J = \{c \in \mathbb{C} \mid f_c^n(0) < \infty \ \forall n \in \mathbb{N}\}.$$

Here f_c^k denotes k iterations of f_c, i.e. $f_c \circ \cdots \circ f_c(z)$. For such a simple example, the Mandelbrot set is famously beautiful, and is shown in figure 4*.*

These pictures are generated by colouring the Mandelbrot set itself black and assigning arbitrary colours to the points $c \in \mathbb{C}$ which cause 0 *to diverge under iterations of f_c, with the colour depending on how quickly* 0 *diverges.*

The Mandelbrot set does not have a known iterated function system, although most mathematicians would classify the Mandelbrot set as a fractal. This is unfortunately the nature of the study of self-similar sets; there is no universally agreed upon definition of a fractal.



(a) The Mandelbrot set.

(b) Zooming in along the boundary of the Mandelbrot set.

Figure 4: The Mandelbrot set is the set of points $c \in \mathbb{C}$ which allow 0 to stay bounded under iteration of $f_c(z) = z^2 + c$.

Next, we introduce the notion of a *similarity-value* or *similarity dimension*. We have chosen to call this the similarity-value because, while it can sometimes be related to a type of fractal dimension, this is not the case in general.

Definition 2.1.15. *Let $\{F_i\}_{i=1}^n$ be an iterated function system. The* similarity value *of $\{F_i\}_{i=1}^n$ is the unique positive number s satisfying $\sum_i r_i^s = 1$, where r_j is the ratio of the similarity F_j.*

Existence and uniqueness of s is not hard to see. Notice that by definition, $r_i \in (0, 1)$ for all i. So, the function

$$\theta(s) := \sum_{i=1}^{n} r_i^s$$

has the property that $\theta(0) \geq 1$ and $\lim_{s\to\infty} \theta(s) = 0$. Furthermore, θ is continuous, so that by the intermediate value theorem, there exists some $s \geq 0$ with $\theta(s) = 1$. In fact,

$$\frac{d}{ds}\theta(s) = \sum_{i=1}^{n} r_i^s \log r_i < 0.$$

Hence θ is monotone decreasing, so s is unique. Another way of thinking about the similarity value is as follows. Consider, for example, a square.

Scaling the side length by a factor $r = 1/2$ creates a square with $1/4 = (1/2)^2$ of the original area. This is reflected in the equation $1 = 4(1/2)^2 = \sum_{i=1}^{4}(1/2)^2$, which implies that the similarity value (or "similarity dimension") of the square is 2. Indeed, the square K can be described as the attractor of four similarities F_i, each scaling the side length by $1/2$ and shifting the resulting smaller square to one of the four corners of the original square, so that $K = \cup_i F_i(K)$.

For a cube, scaling the side length by $1/2$ creates a cube with $1/8 = (1/2)^3$ the original volume, so that the similarity value of the cube is (unsurprisingly) 3, since $1 = 8(1/2)^3 = \sum_{i=1}^{8}(1/2)^3$.

What about something more exotic, like the Sierpinski gasket? Scaling its side by $1/2$ (as the functions in example 2.1.12 do) gives a copy of SG with $1/3$ the original area, giving $1/3 = (1/2)^s$, or $1 = \sum_{i=1}^{3} r_i^s = 3(1/2)^s$, which implies that $s = \log 3/\log 2$.

2.2 String Models

There is another very useful way of describing fractal sets. We begin by defining a string model related to K, and introducing a natural metric on it. Later, we shall also equip it with a measure that has a special relationship to a natural measure on K. We will begin with some simple definitions, and end with an important theorem relating string models to fractals.

We denote an *alphabet* consisting of an arbitrary number of symbols by E. For example, $E = \{0, 1\}$ is an alphabet with two symbols, 0 and 1. A *string* α is simply a concatenation of elements of E, and its length $|\alpha|$ is the number of symbols in the string. The empty string is denoted by Λ.

We will use the shorthand notation $E^n = \prod_{i=1}^{n} E$ to represent the set of all possible strings of length n comprised of letters from the alphabet E. We will denote the set of all possible finite strings by $E^* = \bigcup_{n=1}^{\infty} E^n$, and we will denote the set of infinite strings of symbols by E^ω.

For example, $\alpha = 011010100 \in E^9$ is a string comprised of nine symbols from the alphabet $E = \{0, 1\}$, with length $|\alpha| = 9$. The notation $\beta \upharpoonright n$ represents the first n letters of the string β. For example, $\alpha \upharpoonright 3 = 011$.

Given two strings σ and τ, we denote their concatenation by $\sigma\tau$. For instance, if $\sigma = 011$ and $\tau = 0101$, then $\sigma\tau = 0110101$. We say that a *child* of a string β is any string that can be created by concatenating one additional symbol from E to β. For instance, the children of σ are 0110 and 0111; similarly, σ is said to be the *parent* of both 0110 and 0111. We say that a string is an *ancestor* of another string if there is a chain of parents between them. For example, 01 is an ancestor of 011010, since there is a chain of parents between them:

$$01 \to 011 \to 0110 \to 01101 \to 011010.$$

If $\alpha \in E^*$ is a finite string, we define

$$[\alpha] = \{\sigma \in E^\omega \mid \exists \tau \in E^\omega \text{ such that } \sigma = \alpha\tau\}$$

to be the set of all strings in E^ω that begin with α (i.e. the set of all strings which have α as an ancestor).

There is a relationship between string spaces and fractals. Let $\{F_i\}_{i=1}^n$ be an iterated function system with similarity ratios $\{r_i\}_{i=1}^n$. Since there are n functions in the iterated function system, we will use an alphabet E with n symbols to represent it, say $E = \{1, \dots, n\}$.

Given an alphabet E, a metric for the space E^ω is defined as follows. Let $w_\Lambda = 1$, and using the similarity ratios r_i, recursively define $w_{\alpha e} = w_\alpha r_e$ for $\alpha \in E^*$, and $e \in E$. In other words,

$$w_\alpha = \prod_e r_e,$$

where the product is taken over all symbols e that make up the string α, allowing for repeating of an index. For example, $w_{011} = r_0 r_1 r_1$.

We define a metric ρ on E^ω by $\rho(\sigma, \tau) = w_\alpha$, where α is the *longest common prefix* of σ and τ, which means that α is the longest string for which both σ and τ are in $[\alpha]$. So,

$$\operatorname{diam}[\alpha] = \sup\{\rho(\sigma, \tau) \mid \sigma, \tau \in [\alpha]\} = w_\alpha.$$

It is not hard to show ρ is a metric on E^ω. Let $\sigma, \tau \in E^\omega$. By assumption, each r_j in the ratio list associated with the iterated function system is positive, so $\rho(\sigma, \tau) = \prod r_e \geq 0$.

Suppose that $\rho(\sigma, \tau) = 0$. Then $\prod r_e = 0$, but $0 < r_j < 1$ for all j. Hence there must be infinitely many ratios r_j in the product $\prod r_e$, hence the longest common prefix α is infinite in length, which implies that $\sigma = \tau$. Similarly, if $\sigma = \tau$ then the longest common prefix $\alpha = \sigma = \tau$ is infinite in length, so $\rho(\sigma, \tau) = \prod r_e = 0$, hence $\rho(\sigma, \tau) = 0$ if and only if $\sigma = \tau$.

Symmetry is easy, as the pair (σ, τ) clearly has the same longest common prefix as the pair (τ, σ).

Finally, let σ, τ, ξ be infinite strings. We wish to show ρ satisfies the triangle inequality. Suppose that

- the longest common prefix between σ and τ is α;
- the longest common prefix between τ and ξ is β;
- the longest common prefix between σ and ξ is γ.

Notice that there are three possible cases:

1. if $\alpha = \beta$ then clearly $\gamma = \alpha = \beta$, so $w_\alpha = w_\beta = w_\gamma$;
2. if α is an ancestor of β, then it must be that $\gamma = \alpha$, which implies that $w_\alpha \leq w_\beta = w_\gamma$;
3. if β is an ancestor of α, then it must be that $\gamma = \beta$, which implies that $w_\alpha = w_\gamma \leq w_\beta$.

In any case,

$$\rho(\sigma, \tau) = w_\alpha \leq \max\{w_\beta, w_\gamma\} \leq w_\beta + w_\gamma = \rho(\sigma, \xi) + \rho(\xi, \tau)$$

So ρ is a metric (indeed, an ultrametric) on E^ω. The metric space (E^ω, ρ) will turn out to be very useful in describing fractals.

Proposition 2.2.1. *The collection $\mathcal{B} = \{[\alpha] \mid \alpha \in E^*\}$ is a countable basis for the metric topology of E^ω.*

Proof. Note that although each $[\alpha]$ is uncountable (indeed, it is well known that even the set of all infinite sequences of 0s and 1s is uncountable), the collection $\mathcal{B} = \{[\alpha] \mid \alpha \in E^*\}$ itself is countable. The fact that this is countable is given by the fact that $\mathcal{B}$ is a collection of *finite* combinations of the letters $\{1, \ldots, n\}$. To see that this generates the metric topology is a bit trickier. Let $\mathcal{B}' = \{B_\rho(x, \epsilon) \mid x \in E^\omega,\ \epsilon > 0\}$ denote the usual basis for metric topology on E^ω, where $B_\rho(x, \epsilon) = \{y \in E^\omega \mid \rho(x, y) < \epsilon\}$ is the usual open ball. Let $\mathcal{T}$ and $\mathcal{T}'$ be the topologies generated by $\mathcal{B}$ and $\mathcal{B}'$, respectively.

First, we'll show that $\mathcal{T} \subseteq \mathcal{T}'$. Let $x \in B_\rho(y, \epsilon)$. Then $\rho(x, y) = w_\alpha < \epsilon$, where α the longest common prefix of x and y. In other words, we can write

$x = \alpha x'$ and $y = \alpha y'$. But notice that $w_x = w_{\alpha x'} < w_\alpha < \epsilon$, so choosing $[x]$ gives

$$x \in [x] \subset B_\rho(y, \epsilon)$$

Hence $\mathcal{T} \subseteq \mathcal{T}'$.

Next we'll show that $\mathcal{T}' \subset \mathcal{T}$. Let $x \in [\alpha]$ for some finite string $\alpha \in E^*$. Then choose

$$\epsilon = \frac{w_\alpha}{2},$$

so that $B_\rho(x, \epsilon) = \{y \in E^\omega \mid \rho(x, y) < \epsilon = (w_\alpha/2)\}$, hence

$$x \in B_\rho(x, \epsilon) \subset [\alpha],$$

as desired. Hence $\mathcal{T} = \mathcal{T}'$ and so $\mathcal{B}$ is a countable basis for $(E^\omega, \rho_{1/2})$. □

It is a general fact of topology that proposition 2.2.1 actually also implies that E^ω is separable and Lindelöf [5].

There is an important set of functions on E^ω which will essentially be the "string version" of an iterated function system. For each symbol $e \in E$, we can define a *shift function* $\theta_e : E^\omega \to E^\omega$ by

$$\theta_e(\sigma) = e\sigma.$$

In other words, θ_e just appends the symbol e to the beginning of σ.

Proposition 2.2.2. *Let $\{F_i\}_{i=1}^n$ be an iterated function system with ratio list $\{r_i\}_{i=1}^n$. Let $E = \{1, \ldots, n\}$ be an n-symbol alphabet. Then (E^ω, ρ) is a complete metric space, and $\{\theta_e\}_{e \in E}$ is an iterated function system on E^ω realizing the ratio list $\{r_i\}_{i=1}^n$.*

Proof. Let (σ_n) be a Cauchy sequence in E^ω, let $r_{\min} = \min\{r_1, \ldots, r_n\}$, and let $r_{\max} = \max\{r_1, \ldots, r_n\}$. Since (σ_n) is Cauchy, we have that for all $k \in \mathbb{N}$, there exists $N_k \in \mathbb{N}$ such that

$$\rho(\sigma_n, \sigma_m) < r_{\min}^k$$

for all $m, n \geq N_k$. In particular, setting $n = N_k$, this implies that $\sigma_{N_k} \upharpoonright k = \sigma_m \upharpoonright k$ for all $m \geq N_k$. We will define a candidate limit τ by letting the k^{th} letter of τ be the k^{th} letter of σ_{N_k}. Then $\tau \upharpoonright k = \sigma_{N_k} \upharpoonright k$ for all k. Let $\epsilon > 0$, and choose k large enough so that $r_{\max}^k < \epsilon$. Then

$$\rho(\sigma_m, \tau) \le w_k < r_{\max}^k < \epsilon$$

for all $m > N_k$. Therefore $\sigma_n \to \tau$, so (E^ω, ρ) is complete. Now to show that $\{\theta_e\}_{e\in E}$ is an iterated function system, let $\sigma, \tau \in E^\omega$. Suppose that the longest common prefix between these two strings is α (in other words, $\alpha \in E^*$ is the longest string such that $\sigma, \tau \in [\alpha]$). Then

$$\rho(\theta_e(\sigma), \theta_e(\tau)) = \rho(e\sigma, e\tau) = w_{e\alpha} = r_e w_\alpha = r_e \rho(\sigma, \tau).$$

In other words, θ_e is a similarity with ratio $0 < r_e < 1$, so $\{\theta_e\}_{e\in E}$ is an iterated function system realizing the desired ratio list. □

The next theorem shows another important correspondence between string models and fractal sets. Before proving it, we will need one final definition. Given a metric ρ, the *uniform metric* ρ_u is the metric on the space of bounded functions $f : S \to S$ defined by

$$\rho_u(f_1, f_2) := \sup_x \{\rho(f_1(x), f_2(x)) \mid x \in S)\}.$$

The fact that ρ_u is a metric is easy to see from its definition, as it is defined in terms of a metric ρ.

Theorem 2.2.3. *Let S be a complete metric space and let $\{F_e\}_{e\in E}$ be an iterated function system represented by a string space E^ω, with attractor set K. Then there exists a unique, continuous surjection $h : E^\omega \to S$ with*

$$h(e\sigma) = F_e(h(\sigma))$$

for all $\sigma \in E^\omega$ and for all $e \in E$. Furthermore,

$$h(E^\omega) = K.$$

Proof. Let g_k be the sequence of functions defined as follows. Choose any $a \in S$ and let $g_0(\sigma) = a$ for all $\sigma \in E^\omega$. Then define g_{n+1} by the formula $g_{n+1}(e\sigma) = F_e(g_k(\sigma))$. Notice that g_0 is continuous, and each g_m is also continuous as it is the composition of g_0 with the continuous contractions F_e.

We will show that (g_k) converges uniformly to the desired h. Let $r = \max\{r_1, \dots, r_n\}$. Given arbitrary $\sigma \in E^\omega$,

$$\begin{aligned}\rho(g_{k+1}(e\sigma), g_k(e\sigma)) &= \rho(F_e(g_k(\sigma)), F_e(g_{k-1}(\sigma)) \\ &= r_e\rho(g_k(\sigma), g_{k-1}(\sigma)) \\ &\leq r\rho_u(g_k, g_{k-1}),\end{aligned}$$

where ρ_u is the uniform metric. So, by induction, $\rho(g_{k+1}(e\sigma), g_k(e\sigma)) \leq r^k\rho_u(g_0, g_1)$. This is true for any σ, so $\rho_u(g_{k+1}, g_k) \leq r^k\rho_u(g_0, g_1)$. Hence by the triangle inequality, for any $m > k$ we have

$$\begin{aligned}\rho_u(g_m, g_k) &\leq \sum_{j=k}^{m-1} \rho_u(g_{j+1}, g_j) \\ &\leq \sum_{j=k}^{m-1} r^j\rho_u(g_1, g_0) \\ &\leq \sum_{j=k}^{\infty} r^j\rho_u(g_1, g_0)\end{aligned}$$

Now, E^ω is compact (see [2], exercises 2.6.1(3) and 2.6.6(1)), so the continuous functions g_0 and g_1 are bounded on E^ω, hence $\rho_u(g_1, g_0)$ is finite. Furthermore, $\sum_{j=k}^{\infty} r^j\rho_u(g_1, g_0)$ is (the tail of) a geometric series, which is known to be convergent. Hence $\sum_{j=k}^{\infty} r^j\rho_u(g_1, g_0)$ must go to 0 as $k \to \infty$. Since $\rho_u(g_m, g_k) \geq 0$ but tends to zero as $k \to \infty$, we must have that (g_m) is Cauchy. Since E^ω is complete, the space of continuous functions in the uniform metric $C(E^\omega)$ is complete, so (g_m) is uniformly convergent, say to the continuous function $g_m \to h$.

Finally, we show h has the desired properties. Notice that

$$g_{k+1}(E^\omega) = \bigcup_{e\in E} F_e(g_k(E^\omega)).$$

But, this sequence in k is exactly the sequence defined in theorem 2.1.8, so it converges to the attractor set of the iterated function system

$$\lim_{k\to\infty} g_k(E^\omega) = h(E^\omega) = K,$$

which, of course, implies surjectivity of h. Finally, we show uniqueness. Suppose that h and h' are two functions satisfying $h(e\sigma) = F_e(h(\sigma))$ and $h'(e\sigma) = F_e(h'(\sigma))$. Then

$$\rho(h(e\sigma), h'(e\sigma)) = \rho(F_e(h(\sigma)), F_e(h'(\sigma))) \leq r\rho(h(\sigma), h'(\sigma))$$

for all strings $\sigma \in E^\omega$, and for all $e \in E$. Therefore,

$$\rho_u(h, h') \leq r\rho_u(h, h'),$$

which implies that $\rho_u(h, h') = 0$, and hence that $h = h'$. □

The identity $h(e\sigma) = F_e(h(\sigma))$ of the previous theorem can be rewritten in terms of the shift function:

$$h \circ \theta_e = F_e \circ h$$

We can consider h, called the *addressing function*, as a sort of parametrization of the fractal. In practical terms, $h : E^\omega \to K$ is given by the formula

$$h(e_1 e_2 e_3 \dots) = \lim_{n\to\infty} F_{e_1} \circ F_{e_2} \circ \dots \circ F_{e_n}(x)$$

for x some arbitrary point in the same metric space as K and each $e_j \in E$, or, equivalently, by

$$h(\alpha) = \bigcap_{m\geq 1} F_{\alpha\restriction m}(K).$$

Now is a good time to work through a concrete example.

Example 2.2.4 (Cantor set)**.** *It is well-known that a number $x \in [0,1]$ belongs to the Cantor set C if and only if its base 3 expansion contains only 0s and 2s. To see this, recall the construction from example 2.1.11. We set $C_0 = [0,1]$ and remove the middle third $(1/3, 2/3)$. But this middle third is precisely the set of numbers $y \in [0,1]$ whose first digit after the decimal in the base 3 expansion is 1. Similarly, the second place digit of x is a 1 if and only if x belongs to one of the sets $(1/9, 2/9)$ or $(7/9, 8/9)$, but these are precisely the sets which get removed to construct C_2. An induction argument shows that $x \in C$ if and only if the base 3 expansion of x has no 1s.*

Now, what is the addressing function $h : E^\omega \to \mathbb{R}$ for C? To start, we have two similarities F_0 and F_1, so we will use a two-symbol alphabet, say $E = \{a, b\}$. If $\alpha \in E^\omega$, then $h(\alpha)$ is given as follows: turn all the a's in the string α into 0s, and turn all the b's into 2s, and interpret the result in base 3. For example,

$$h(aabaabaab\dots) = (0.002002002\dots)_3 = \frac{1}{13}$$

Then $h(E^\omega)$ is precisely the Cantor set, as it consists of all possible combinations of 0s and 2s interpreted in base 3. Furthermore, given σ, notice that $h(0\sigma)$ just appends a 0 to the first place after the decimal in the base 3 expansion, which is precisely the effect of $F_0(x) = x/3$ on $h(\sigma)$. Similarly, $h(1\sigma)$ just appends a 2 to the first place after the decimal in the base 3 expansion, which is exactly what F_1 does to $h(\sigma)$. Hence

$$F_0(h(\sigma)) = h(a\sigma),$$

$$F_1(h(\sigma)) = h(b\sigma).$$

So, h is the addressing function for C by theorem 2.2.3.

3 Measure Theory on Fractals

This section follows [2] and [7].

3.1 General Measure Theory

Recall the following definitions:

Definition 3.1.1. *Let $\mathcal{A}$ be a σ-algebra. A* measure *is a function $\mu : \mathcal{A} \to [0, \infty]$ satisfying:*

1. $\mu(\varnothing) = 0$;
2. *if $\{A_n\}_{n=1}^\infty$ are pairwise disjoint sets, then $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^\infty \mu(A_n)$.*

Definition 3.1.2. *Let X be any set, and denote its power set by $\mathcal{P}(X)$. An* outer measure *is a function $\bar{\mu} : \mathcal{P}(X) \to [0, \infty]$ satisfying:*

1. $\bar{\mu}(\varnothing) = 0$;
2. *for all $A, B \subset X$, if $A \subseteq B$, then $\bar{\mu}(A) \le \bar{\mu}(B)$;*
3. *for all $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(X)$, $\bar{\mu}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \le \sum_{n=1}^\infty \bar{\mu}(A_n)$.*

The second property of outer measures is actually also a property of measures, it just is not required as part of the definition. Notice that outer measures are defined on all subsets of a set X, while a measure is only defined on a fixed σ-algebra of subsets of X. As we shall see, there is a way of defining outer measures on any set, and restricting these outer measures so that they become measures. This will then allow us to define integration. Finally, we will extend these ideas to the fractal setting by defining something called a self-similar measure.

First, we will discuss a general theorem about constructing outer measures.

Theorem 3.1.3. *Let X be any set, let $\mathcal{A}$ be a σ-algebra of subsets of X, and let $c : \mathcal{A} \to [0, \infty]$ be any extended real-valued function. Then*

$$\bar{\mu}(B) := \inf_{\mathcal{U}} \left\{ \sum_{A \in \mathcal{U}} c(A) \,\middle|\, \mathcal{U} \text{ is a countable cover for } B \right\}$$

is an outer measure on X satisfying:

1. *$\bar{\mu}(A) \leq c(A)$ for all $A \in \mathcal{A}$;*
2. *if $\bar{\nu}$ is any other outer measure with $\bar{\nu}(A) \leq c(A)$ for all $A \in \mathcal{A}$, then $\bar{\nu}(B) \leq \bar{\mu}(B)$ for all $B \subset X$.*

Proof. First we show $\bar{\mu}$ is an outer measure. Note that the empty cover is a countable cover of the empty set, so $\bar{\mu}(\varnothing) = 0$ since the empty sum is equal to 0. If $A \subset B$, then any cover of B is a cover of A, so we have $\bar{\mu}(A) \leq \bar{\mu}(B)$. Finally, let $(B_n)_{n=1}^{\infty}$ be a countable collection of sets. If $\bar{\mu}(B_k) = \infty$ for some k then the inequality

$$\bar{\mu}\left(\bigcup_{n \in \mathbb{N}} B_n \right) \leq \sum_{n=1}^{\infty} \bar{\mu}(B_n)$$

is trivial, so assume that $\bar{\mu}(B_n) < \infty$ for all n. Let $\epsilon > 0$. Then for each n, choose a cover $\mathcal{U}_n$ such that

$$\bar{\mu}(B_n) + \frac{\epsilon}{2^n} \geq \sum_{A \in \mathcal{U}_n} c(A).$$

Since each $\mathcal{U}_n$ is a cover for B_n, $\mathcal{U} := \cup_n \mathcal{U}_n$ must be a cover for $\cup_n B_n$, so

$$\bar{\mu}\left(\bigcup_{n\in\mathbb{N}} B_n\right) \leq \sum_{A\in\mathcal{U}} c(A) \leq \sum_{n=1}^{\infty}\sum_{A\in\mathcal{U}_n} c(A) \leq \sum_{n=1}^{\infty} \bar{\mu}(B_n) + \epsilon.$$

Since ϵ was arbitrary, we have the desired inequality. Hence $\bar{\mu}$ is an outer measure. Now we can check the two properties claimed in the theorem. For (1), note that $\{B\}$ covers the set B, so $\bar{\mu}(B) \leq \sum_{A\in\{B\}} c(A) = c(B)$.

For (2), let $\mathcal{U}$ be a countable cover of $B \subseteq X$. If $\bar{\nu}$ is any other outer measure satisfying $\bar{\nu}(F) \leq c(F)$ for all $F \subseteq X$, then

$$\bar{\nu}(B) \leq \bar{\nu}\left(\bigcup_{A\in\mathcal{U}} A\right) \leq \sum_{A\in\mathcal{U}} \bar{\nu}(A) \leq \sum_{A\in\mathcal{U}} c(A),$$

so $\bar{\nu}(B)$ is a lower bound for the set

$$\left\{ \sum_{A\in\mathcal{U}} c(A) \,\middle|\, \mathcal{U} \text{ is a countable cover for } B \right\}.$$

But, by definition $\bar{\mu}$ is the greatest lower bound for this set, so $\bar{\nu}(B) \leq \bar{\mu}(B)$. $\square$

Next, we define the Carathéodory measurable sets.

Definition 3.1.4 (Measurable sets)**.** *Let $\bar{\mu}$ be an outer measure on a set X, and let $A \subset X$. Then A is said to be $\bar{\mu}$-measurable if*

$$\bar{\mu}(E) = \bar{\mu}(E \cap A) + \bar{\mu}(E \setminus A)$$

for all $E \subset X$.

The proofs of the following two theorems consist of tedious exercises in set theoretic manipulation and are not very illuminating, and are thus omitted. A proof of the first theorem can be found in [2], theorem 5.2.5. A proof of the second can be found in [2], theorem 5.4.2.

Theorem 3.1.5 (Carathéodory extension theorem)**.** *Let $\bar{\mu}$ be an outer measure on a set X. The collection $\mathcal{A}$ of all $\bar{\mu}$-measurable sets is a σ-algebra on X, and the restriction of $\bar{\mu}$ to this collection is a measure.*

Given an outer measure $\bar{\mu}$, we will denote the measure defined in theorem 3.1.5 simply by μ.

A *metric outer measure* $\bar{\nu}$ is a measure satisfying $\bar{\nu}(A \cup B) = \bar{\nu}(A) + \bar{\nu}(B)$ for any two sets A and B with $\text{dist}(A, B) > 0$. Metric outer measures ensure the measurability of Borel sets.

Theorem 3.1.6. *If $\bar{\nu}$ is a metric outer measure, then the Borel sets are measurable.*

While theorem 3.1.3 does give a nice formula for defining measures from set functions, it unfortunately does not guarantee that the resulting outer measure will be a metric outer measure. For that, we will need the following theorem.

Theorem 3.1.7. *Let $\mathcal{U}$ be a countable family of subsets of a metric space S with the property that for each $x \in S$ and for each $\epsilon > 0$, there exists some $U \in \mathcal{U}$ such that $x \in U$ and $\text{diam} U \leq \epsilon$. Given $\epsilon > 0$, define the subcollection*

$$\mathcal{U}_\epsilon = \{U \in \mathcal{U} \mid \text{diam} U \leq \epsilon.\}$$

Given any set function c, let $\bar{\mu}_\epsilon$ be the associated outer measure defined by theorem 3.1.3 using the collection $\mathcal{U}_\epsilon$, i.e.

$$\bar{\mu}_\epsilon(A) = \inf_{\mathcal{U}_\epsilon} \left\{ \sum_{U \in \mathcal{U}_\epsilon} c(U) \right\}.$$

Then

$$\bar{\mu}(A) := \lim_{\epsilon \to 0} \bar{\mu}_\epsilon(A) = \sup_{\epsilon > 0} \bar{\mu}_\epsilon(A)$$

defines a metric outer measure.

Proof. Let A, B be subsets of S with $\text{dist}(A, B) > 0$. Clearly $\bar{\mu}$ is an outer measure since it was constructed by the method of theorem 3.1.3. Any outer measure satisfies $\bar{\mu}(A \cup B) \leq \bar{\mu}(A) + \bar{\mu}(B)$, so we must show the opposite inequality.

Let $0 < \epsilon < \text{dist}(A, B)$ and let $\mathcal{D}$ be a countable cover of $A \cup B$ by elements of $\mathcal{U}_\epsilon$. Then $\text{diam}\, D \leq \epsilon < \text{dist}(A, B)$ for any D in $\mathcal{D}$, so D intersects at most one of A or B.

Divide $\mathcal{D}$ into two collections: $\mathcal{D}_1$, whose elements intersect A, and $\mathcal{D}_2$, whose elements intersect B. Then

$$\sum_{D\in\mathcal{D}} c(D) = \sum_{D\in\mathcal{D}_1} c(D) + \sum_{D\in\mathcal{D}_2} c(D) \geq \bar{\mu}_\epsilon(A) + \bar{\mu}_\epsilon(B)$$

This inequality holds for any countable cover $\mathcal{D}$ of $A \cup B$, so it still holds when we pass an infimum over $\mathcal{D}$:

$$\inf_{\mathcal{D}} \sum_{D\in\mathcal{D}} c(D) = \bar{\mu}(A \cup B) \geq \bar{\mu}_\epsilon(A) + \bar{\mu}_\epsilon(B).$$

This inequality holds for any $\epsilon > 0$, so it holds when we pass a supremum over ϵ:

$$\bar{\mu}(A \cup B) \geq \sup_{\epsilon>0} \bar{\mu}_\epsilon(A) + \sup_{\epsilon>0} \bar{\mu}_\epsilon(B) = \bar{\mu}(A) + \bar{\mu}(B),$$

which is the desired inequality. Hence $\bar{\mu}$ is a metric outer measure, meaning that Borel sets are measurable (by theorem 3.1.6).

To see the final equality, notice that as ϵ gets smaller, there are fewer possible collections $\mathcal{U}_\epsilon$ over which to take the infimum, and hence $\bar{\mu}_\epsilon$ increases, so we have

$$\lim_{\epsilon\to 0} \bar{\mu}_\epsilon(A) = \sup_{\epsilon>0} \bar{\mu}_\epsilon(A). \qquad \square$$

3.2 Measures for Fractals

Finally we are ready to define three natural (and related) measures for a fractal; two for the fractal itself in a metric space S (namely the standard measure and the Hausdorff measure), and one for the corresponding string space E^ω.

First, we'll discuss the self similar measure. What would be a reasonable list of criteria we would like a fractal measure μ to satisfy? It would make sense that if $A = \cup_{j=1}^N C_j$ is any finite union of cells which intersects only at points, then $\mu(A) = \sum_{j=1}^N \mu(C_j)$.

In addition, a sensible requirement of decomposing m-cells to "the next level" (i.e. $m+1$-cells) is that

$$\mu(F_w(K)) = \sum_i \mu(F_w \circ F_i(K)).$$

These are rather unrestrictive conditions which could lead to many different measures, but the standard measure is perhaps the simplest possible construction:

Definition 3.2.1 (Standard measure). *Let K be a fractal of an iterated function system $\{F_e\}_{e\in E}$, and denote its m-cells by $K_w := F_w(K)$. Given a word $w = w_1 \dots w_m$ from the alphabet E, we will create a set function on the m-cells defined by*

$$c(K_w) = \prod_{i=1}^{m} \mu_{w_i},$$

where the weights μ_e are a set of n positive real numbers satisfying $\sum_{e\in E} \mu_e = 1$.

Then we construct a metric outer measure $\bar{\mu}$ on subsets $C \subset K$ by applying theorem 3.1.7 to the function c. We define the self-similar measure *μ by restricting $\bar{\mu}$ to the σ-algebra of measurable sets.*

The weights of the self similar measure *can be any positive real numbers whose sum is 1, but the* standard measure *is the simplest possible special case of the self similar measure, defined by taking all the weights μ_e as equal. In other words, if $\{F_e\}_{e\in E}$ is an iterated function system with a corresponding n-symbol alphabet E, then*

$$\mu_e = \frac{1}{n} \quad \forall e \in E$$

For ease of notation, given a word $w = w_1 \dots w_m$ of length m, we will write $\mu_w := \prod_{i=1}^{m} \mu_{w_i}$, so that the measure of a cell is simply $\mu(F_w(K)) = \mu_w$.

Once we can construct measures, we can also define integration. We will only be concerned with integrating continuous functions on fractals, so we can restrict ourselves to Riemann integration, rather than the more complicated theory of Lebesgue integration.

Definition 3.2.2 (Integral). *Let μ be a measure, and K be a fractal. If $f \in C(K)$, then we can define the integral*

$$\int_K f d\mu := \lim_{m\to\infty} \sum_{|w|=m} f(x_w)\mu\left(F_w(K)\right),$$

where, for each w, x_w is any element of $K_w = F_w(K)$.

The reason that x_w can be any element of K_w is that continuity of f implies uniform continuity of f since any fractal is compact. Uniform continuity means that given $\epsilon > 0$, we can find a $\delta > 0$ such that $\rho(x, y) < \delta$

only if $\rho(f(x), f(y)) < \epsilon/\mu(K)$. Since each F_j is a contraction and $K = \cup_{|w|=m} F_w(K)$, we can find a collection of m-cells K_i which cover K and have mesh less than δ. Thus

$$\begin{aligned}\rho\left(\sum_{i=1}^{N} f(x_i)\mu(K_i), \sum_{i=1}^{N} f(y_i)\mu(K_i)\right) &\leq \sum_{i=1}^{N} \mu(K_i)\rho\left(f(x_i), f(y_i)\right) \\ &< \sum_{i=1}^{N} \frac{\epsilon\mu(K_i)}{\mu(K)} \\ &= \epsilon.\end{aligned}$$

But $x_i, y_i \in A_i$ were arbitrary, which shows that the choice is immaterial. This is analogous to taking a smaller and smaller Δx_i in the usual Riemann integral

$$\int_X f d\lambda = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i')\Delta x_i,$$

where, as is well known, x_i' can any element of Δx_i.

Proposition 3.2.3. *Let $\{F_e\}_{e\in E}$ be an iterated function system, and let K be its attractor (fractal). If μ is the standard measure on K, then the following* self-similar identity for measure *holds:*

$$\mu(A) = \sum_{e\in E} \mu_e \mu(F_e^{-1}(A)).$$

Moreover, if f is a continuous function on K, the following self-similar identity for integration *also holds:*

$$\int_K f d\mu = \sum_{e\in E} \mu_e \int_K f \circ F_e d\mu$$

Proof. For the measure identity, notice that $\mu(F_e(A)) = \mu_e\mu(A)$. Furthermore, we can split A into a cell decomposition:

$$A = \bigcup_{e\in E} (A \cap F_e(K))$$

Invoking additivity of the measure on these disjoint sets and a using bit of manipulation, we have

$$\begin{aligned}\mu(A) &= \mu(\cup_e(A \cap F_e(K))) \\ &= \sum_{e\in E} \mu(A \cap F_e(K)) \\ &= \sum_{e\in E} \mu(F_e(F_e^{-1}(A \cap F_e(K)))) \\ &= \sum_{e\in E} \mu_e \mu(F_e^{-1}(A)),\end{aligned}$$

as desired. For the proof of the integration identity, notice that by definition $\mu(F_eF_wK) = \mu_e\mu(F_wK)$, so

$$\begin{aligned}\sum_{e\in E} \mu_e \int_K f \circ F_e d\mu &= \sum_{e\in E} \mu_e \lim_{m\to\infty} \sum_{|w|=m} \mu(F_wK) f \circ F_e(x_w) \\ &= \lim_{m\to\infty} \sum_{|w|=m} \sum_{e\in E} \mu(F_eF_wK) f(x_{ew}) \\ &= \lim_{m\to\infty} \sum_{|w|=m+1} \mu(F_wK) f(x_w) \\ &= \int_K f d\mu,\end{aligned}$$

where x_{ew} is any point in F_eF_wK. □

Using precisely the same proof technique, a simple induction argument will actually generalize the self-similar formulas, which reduce to proposition 3.2.3 in the case $m = 1$:

$$\mu(A) = \sum_{|w|=m} \mu_w \mu(F_w^{-1}A),$$

$$\int_K f d\mu = \sum_{|w|=m} \mu_w \int_K f \circ F_w d\mu,$$

where, again, F_w is a shorthand for $F_{w_1} \circ \cdots \circ F_{w_m}$.

There are two other useful measures when discussing fractals: the Hausdorff measure (to be discussed in section 4), and the string measure.

Definition 3.2.4 (String measure). *Let r_i be the ratios of an iterated function system. Recall that $\mathcal{B} = \{[\alpha] \mid \alpha \in E^*\}$ is a countable basis for the metric topology of E^ω (proposition 2.2.1). Recursively define*

$$r(\Lambda) = 1,$$
$$r(e\alpha) = r_e r(\alpha),$$

where, since ρ is the string metric, $r(\alpha) = \operatorname{diam}[\alpha]$. The string measure *$\mathcal{M}$ on E^ω is the measure built from theorem 3.1.7, taking $c([\alpha]) = (\operatorname{diam}[\alpha])^s$ as the set function, where s is the similarity value.*

Note the similarity between the string metric and the string measure. The motivation for this definition will become apparent when we compare it to the Hausdorff dimension.

4 Fractal Dimension

There are several notions of "dimension" one can use when discussing fractals: covering dimension, packing dimension, similarity dimension, etc. In this section we introduce one of the most popular: the Hausdorff dimension. We also give a condition under which computing this dimension is relatively easy, called the Moran open set condition. This section follows [2].

4.1 Hausdorff Dimension

Recall theorem 3.1.7, which states that any given set function $c : \mathcal{P}(X) \to [0, \infty]$ can be made into a metric outer measure on $\mathcal{P}(X)$ via the formula

$$\bar{\mu}(B) := \sup_{\epsilon>0} \inf_{\mathcal{U}_\epsilon} \sum_{A\in\mathcal{U}_\epsilon} c(A),$$

where the infimum is over countable covers $\mathcal{U}_\epsilon$ for B satisfying $\operatorname{diam} U \leq \epsilon$ for all $U \in \mathcal{U}_\epsilon$. The restriction of $\bar{\mu}$ to the σ-algebra of measurable sets is a metric measure. We will use this process to define the *s-dimensional Hausdorff outer measure* $\overline{\mathcal{H}}^s$ as the outer measure associated with the set function $c_s(A) := (\operatorname{diam} A)^s$. The restriction of $\overline{\mathcal{H}}^s$ to the measurable sets is the *s-dimensional Hausdorff measure $\mathcal{H}^s$*.

More explicitly, if a set A is measurable, then define

$$\mathcal{H}^s_\epsilon(A) := \inf_{\mathcal{B}_\epsilon} \sum_{B\in\mathcal{B}_\epsilon} (\operatorname{diam} B)^s,$$

where the infimum is taken over all countable covers $\mathcal{B}_\epsilon$ of A, with the property that $\operatorname{diam} B \leq \epsilon$ for all $B \in \mathcal{B}_\epsilon$. Notice that as ϵ gets smaller, $\mathcal{H}^s_\epsilon(A)$ gets larger (since there are fewer open covers to choose from, hence the infimum becomes larger). Thus

$$\mathcal{H}^s(A) = \sup_{\epsilon>0} \mathcal{H}^s_\epsilon(A) = \lim_{\epsilon\to 0} \mathcal{H}^s_\epsilon(A)$$

We can even write the more explicit formula for the s-dimensional Hausdorff measure,

$$\mathcal{H}^s(A) = \lim_{\epsilon\to 0} \inf_{\mathcal{B}_\epsilon} \sum_{B\in\mathcal{B}_\epsilon} \operatorname{diam}(B)^s$$

Proposition 4.1.1. *Let A be a Borel set. There is a critical value s_0 at which $\mathcal{H}^s(A) = \infty$ for all $s < s_0$, and $\mathcal{H}^s(A) = 0$ for all $s > s_0$.*

Proof. Let $s < t$ be positive real numbers and let A be any Borel set. If $\mathcal{H}^s(A) < \infty$, then

$$\begin{aligned}
\mathcal{H}^t(A) &= \lim_{\epsilon\to 0} \overline{\mathcal{H}^t_\epsilon}(A) \\
&\leq \lim_{\epsilon\to 0} \epsilon^{t-s} \overline{\mathcal{H}^s_\epsilon}(A) \\
&= 0^{t-s} \mathcal{H}^s(A) \\
&= 0.
\end{aligned}$$

The statement that if $\mathcal{H}^t(A) > 0$ then $\mathcal{H}^s(A) = \infty$ is the contrapositive of what we've just shown. This shows the existence of the critical value s_0. □

Definition 4.1.2 (Hausdorff Dimension)**.** *The* Hausdorff dimension *of a measurable set A is the unique value s_0, such that $\mathcal{H}^s(A) = \infty$ for all $s < s_0$, and $\mathcal{H}^s(A) = 0$ for all $s > s_0$. It is denoted by* $\dim_H(A)$*.*

Why would we define the Hausdorff dimension this way? While it may seem convoluted at first glance, the Hausdorff dimension actually preserves some familiar intuitive notions of what we'd expect dimension to be.

For example, consider a line, which ought to have dimension 1. What if we tried to assess its "amount" of dimension 2? Well, we would hope that this quantity is zero; the line has dimension 1, so dimension 2 is too large in helping to describe the line. The same can be said of dimension 3, 4, and so on. On the other hand, it would take infinitely many 0 dimensional objects (points) to create the line.

This is the intuition behind the critical value of the Hausdorff measure. It "picks out" the dimension of a set A by saying that its measure in terms of dimension *greater* than the dimension of A is zero, while its measure in terms of dimension *less* than the dimension of A is infinite.

A key feature of the Hausdorff measure is that $\dim_H$ is not restricted to integer values. It allows for interpolation between dimensions, and as we shall see, fractals often have non-integer Hausdorff dimension.

Now we introduce a relationship between the Hausdorff measure and the measure we created for strings in section 3.

Lemma 4.1.3. *Let $A \subset E^\omega$. The there is an $\alpha \in E^*$ such that $A \subseteq [\alpha]$ and* $\mathrm{diam} A = \mathrm{diam}[\alpha]$.

Proof. Let α be the longest common prefix of all the strings in A. Clearly $A \subseteq [\alpha]$, thus $\mathrm{diam}\, A \leq \mathrm{diam}[\alpha]$.

Let $\sigma \in A$. Then $\sigma \upharpoonright (|\alpha| + 1)$ is not a common prefix for all strings in A, so $\exists \tau \in A$ such that $\tau \upharpoonright (|\alpha| + 1) \neq \sigma \upharpoonright (|\alpha| + 1)$. Since α is the longest common prefix for $\{\sigma, \tau\}$, we have $\rho(\sigma, \tau) = w_\alpha = \mathrm{diam}[\alpha]$, so $\mathrm{diam}\, A \geq w_\alpha = \mathrm{diam}[\alpha]$. ☐

Theorem 4.1.4. *Let $\{F_i\}_{i=1}^n$ be any iterated function system, and let s be the similarity value of the ratios $\{r_i\}_{i=1}^n$. Then $\mathcal{H}^s = \mathcal{M}$ on the space E^ω, where $\mathcal{M}$ is the string measure.*

Proof. Suppose $\mathrm{diam}\, A > 0$. By lemma 4.1.3, there exists α such that $A \subseteq [\alpha]$ and $\mathrm{diam}\, A = \mathrm{diam}[\alpha]$.

So $\mathcal{M}(A)$ has the property that $\mathcal{M}(A) \leq \mathcal{M}([\alpha]) = (\mathrm{diam}[\alpha])^s = (\mathrm{diam}\, A)^s$. But by theorem 3.1.3, $\overline{\mathcal{H}^s_\epsilon}$ is an upper bound of all such measures, so $\overline{\mathcal{H}^s_\epsilon} \geq \overline{\mathcal{M}}$. This holds for all $\epsilon > 0$, so $\overline{\mathcal{H}^s} \geq \overline{\mathcal{M}}$.

Now let $\alpha \in E^*$, $\epsilon > 0$. There exists an integer N large enough so that $r_{\max}^N < \epsilon$, $N > |\alpha|$, so $r(\beta) < \epsilon\ \forall \beta \in E^N$. $[\alpha]$ is the disjoint union of all $[\beta]$ with $\beta \geq \alpha$ and $|\beta| = N$, so

$$\overline{\mathcal{H}^s_\epsilon}([\alpha]) \leq \sum_\beta (\mathrm{diam}[\beta])^s = \sum_\beta \mathcal{M}([\beta]) = \mathcal{M}([\alpha]),$$

where the sum is over all such β. This holds for all $\epsilon > 0$, so $\overline{\mathcal{H}^s} \leq \overline{\mathcal{M}}$, hence $\mathcal{H}^s = \mathcal{M}$. $\square$

Theorem 4.1.5 (Moran's open set condition)**.** *Let $\{F_i\}$ be an iterated function system giving rise to a fractal K. The following condition is called* Moran's open set condition*:*

There exists an open set U, called a Moran open set*, such that*

1. *$F_i(U) \cap F_j(U) = \varnothing$ for all $i \neq j$;*
2. *$F_i(U) \subseteq U$ for all i.*

If Moran's open set condition is satisfied, then $\dim_H(K)$ *is equal to the similarity value.*

A proof of theorem 4.1.5 can be found, for example, in [2] section 6.5.

Example 4.1.6 (Hausdorff dimension of the Cantor set)**.** *The interval* $(0,1)$ *is a Moran open set for C. Recall the iterated function system for C is $F_0(x) = x/3$ and $F_1(x) = (x+2)/3$, so*

$$F_0((0,1)) = (0,1/3) \subset (0,1), \qquad F_1((0,1)) = (2/3,1) \subset (0,1),$$

and

$$F_0((0,1)) \cap F_1((0,1)) = (0,1/3) \cap (2/3,1) = \varnothing.$$

Hence the similarity value for C is its Hausdorff dimension by theorem 4.1.5. The ratios of the similarities F_0 and F_1 are $r_0 = r_1 = 1/3$, so

$$1 = \sum_i r_i^s = \left(\frac{1}{3}\right)^s + \left(\frac{1}{3}\right)^s = \frac{2}{3^s}$$

Which has solution $s = \dim_H(C) = \log 2/\log 3$.

Example 4.1.7 (Hausdorff dimension of the Sierpinski gasket)**.** *Recall example 2.1.12; the interior U of the first triangle, SG_0, is a Moran open set for SG. Indeed, $F_i(U)$ just returns one three smaller triangle interiors (specifically, the top, bottom right, or bottom left triangle interior, depending on i), so $F_i(U) \subset U$ for all i and $F_i(U) \cap F_j(U) = \varnothing$ for any $i \neq j$ since U is the interior, i.e. does not include its boundary and hence there is an empty* *intersection, since the triangles themselves only intersect at a single point along the boundary.*

The ratios of the similarities $\{F_i\}_{i=0}^2$ *are* $r_0 = r_1 = r_2 = 1/2$*, so the similarity value is*

$$1 = \sum_i r_i^s = \left(\frac{1}{2}\right)^s + \left(\frac{1}{2}\right)^s + \left(\frac{1}{2}\right)^s = \frac{3}{2^s},$$

which has solution $s = \dim_H(SG) = \log 3/\log 2$*.*

5 Dirichlet Forms and Laplacians

We will close with a section on some more modern developments on fractals, specifically, the construction of Laplacians and related differential equations on them. There are three related methods to do this: one by constructing a "Dirichlet form" (which is related to the Laplacian), another by taking a limit of "finite graph Laplacians", and yet another by a probabilistic approach. We shall adopt the first two methods here. This section follows [1], [3], [6], and [7].

5.1 Manifold Case

We begin with a brief summary of Laplacians on manifolds. This will give the motivation for how we extend the Laplacian to the fractal setting.

Let M be a compact and connected Riemannian manifold with Riemannian metric g, and let $p \in M$. We denote the tangent space of M at point p by T_pM. We denote the dual of the tangent space, consisting linear functionals on T_pM, by T_p^*M.

Define $\alpha_g : T_pM \to T_p^*M$ by $\alpha_g(u)v = g(u, v)$. Then α_g is an isomorphism between the tangent space and its dual.

Let $p \in M$. We define the gradient ∇ at p of a function f on M to be the unique function ∇f satisfying the equation

$$d_pf(X) = g(\nabla f(p), X)$$

for all $X \in T_pM$, where $d_pf(X)$ is directional derivative of f in the direction of X evaluated at point p, i.e.

$$d_p f(X) = \frac{d}{dt}(f \circ c)\bigg|_{t=0},$$

where $c : (\epsilon, \epsilon) \to M$ is any smooth, parametrised curve satisfying $c(0) = p$ and $\dot{c}(0) = X$.

This definition of gradient allows us to define the divergence $\nabla\cdot$ of a vector field V as the negative adjoint of the gradient with respect to the $L^2(M)$ inner product:

$$\langle V, \nabla f\rangle_{L^2(M)} = \langle (\nabla)^* V, f\rangle_{L^2(M)} = -\langle \nabla \cdot V, f\rangle_{L^2(M)}$$

Finally, we define the *Laplace operator* or *Laplacian* Δ to be the operator satisfying the formula

$$\Delta f := -\nabla^*\nabla f = \nabla \cdot (\nabla f)$$

Next, we prove some basic properties about the Laplacian.

Lemma 5.1.1. *Let V be an inner product space on a field $\mathbb{F}$ and let T be any linear operator. Then its adjoint T^* is also linear with respect to the inner product.*

Proof. Let $x, y \in V$ and let $\alpha, \beta \in \mathbb{F}$. Using linearity in the first argument of the inner product, along with the fact that $(T^*)^* = T$, we have that

$$\begin{aligned}
\langle T^*(\alpha x + \beta y), z\rangle &= \langle \alpha x + \beta y, (T^*)^* z\rangle \\
&= \alpha\langle x, Tz\rangle + \beta\langle y, Tz\rangle \\
&= \alpha\langle T^* x, z\rangle + \beta\langle T^* y, z\rangle \\
&= \langle \alpha T^* x + \beta T^* y, z\rangle
\end{aligned}$$

□

Proposition 5.1.2. *The Laplace operator is linear.*

Proof. $\nabla : C^\infty(M) \to \mathfrak{X}(M)$ is linear, so its adjoint $\nabla^* : \mathfrak{X}(M) \to C^\infty(M)$ is linear as well, where $\mathfrak{X}(M)$ is the set of smooth vector fields with L^2 inner product. Hence $\Delta := -\nabla^*\nabla : C^\infty(M) \to C^\infty(M)$ is linear as well. □

Proposition 5.1.3. *The Laplace operator is symmetric, i.e.*

$$\langle f, \Delta f\rangle = \langle \Delta f, f\rangle$$

Proof. By definition, we have

$$\begin{aligned}
\langle f, \Delta f\rangle &= \langle (\nabla\cdot)^* f, \nabla f\rangle \\
&= -\langle \nabla f, \nabla f\rangle \\
&= -\langle (\nabla)^* \nabla f, f\rangle \\
&= \langle \nabla \cdot \nabla f, f\rangle \\
&= \langle \Delta f, f\rangle
\end{aligned}$$

□

Proposition 5.1.4. *The Laplace operator is positive-definite, i.e.*

$$\langle f, \Delta f\rangle \leq 0$$

Proof. By positive-definiteness of the inner product (i.e. $\langle x, x\rangle \geq 0$), we have

$$\langle f, \Delta f\rangle = \langle (\nabla\cdot)^* f, \nabla f\rangle = -\langle \nabla f, \nabla f\rangle \leq 0$$

□

Proposition 5.1.5. $\Delta u = 0$ *if and only if* u *is a constant function on* M*.*

Proof. Let $\Delta u = 0$. Then

$$\int_M u\Delta u = 0 = -\int_M g(\nabla u, \nabla u),$$

so that $\nabla u = 0$. Now, let $p, q \in M$. M is compact and connected, so let $\gamma : [0,1] \to M$ be a smooth curve such that $\gamma(0) = p$ and $\gamma(1) = q$. By the fundamental theorem of calculus,

$$\begin{aligned}
u(p) - u(q) &= \int_0^1 \frac{d}{ds}(u \circ \gamma)(s) ds \\
&= \int_0^1 d_{\gamma(s)} u(\dot{\gamma}(s)) ds \\
&= \int_0^1 g(\nabla f(\gamma(s)), \dot{\gamma}(s)) ds \\
&= 0
\end{aligned}$$

Since this is true for any p, q we conclude that u is constant.

Next, suppose that u is constant. Then $d_p u(X) = 0 = g(\nabla u, X)$ for all $X \in T_pM$ so we have $\nabla u = 0$, and hence by definition $\Delta u = 0$.

□

Propositions 5.1.2 through 5.1.5 are, in fact, the properties we later shall use as *the definition* of a more general Laplacian on fractals.

Next, we define the Dirichlet form. Let u, v be functions on the manifold M and define the Dicihlet form to be

$$\mathcal{E}(u,v) := \int_M g(\nabla u, \nabla v) d\text{vol}$$

Notice that by taking the adjoint of the gradient, it has the following relationship with the Laplacian:

$$\mathcal{E}(u,v) = -\langle u, \Delta v\rangle_{L^2(M)}$$

This actually gives rise to an alternate way of defining the Laplacian, and is the method we will use for constructing Laplacians on fractals. An important fact that we shall see later is that a function $\tilde{u}$ that minimizes the Dirichlet form

$$\mathcal{E}(\tilde{u},\tilde{u}) = \int_M g(\nabla \tilde{u}, \nabla \tilde{u}) d\text{vol}$$

is actually a solution of *Laplace's equation*, $\Delta \tilde{u} = 0$.

5.2 Fractal Case

There is a well-developed theory of analysis on a certain class of fractals, called post-critically finite (pcf) fractals.

Definition 5.2.1 (Post-critically finite fractal)**.** *Let $\{F_i\}_{i=1}^n$ be an iterated function system with associated string space E^ω and addressing function $h : E^\omega \to K$. The* critical set *is*

$$\mathcal{C} = h^{-1}\left(\bigcup_{i\neq j}[F_i(K)\cap F_j(K)]\right)$$

The post-critical set *is*

$$\mathcal{P} = \bigcup_{n\in\mathbb{N}} \xi^{(n)}(\mathcal{C})$$

where $\xi(w_1w_2w_3w_4\dots) = w_2w_3w_4\dots$ is called the deletion map*. The fractal K is said to be* post-critically finite *if $\mathcal{P}$ is a finite set.*

Definition 5.2.2 (Boundary)**.** *$V_0 := h(\mathcal{P})$ is called the* boundary *of the fractal.*

Post-critical finiteness essentially ensures that fractals have finite boundary and finite intersection of m-cells, which is the foundation for the entire theory of fractal analysis. For the remainder of this section, is will be assumed that fractals are pcf.

Example 5.2.3 (Boundary of the interval)**.** *Let $I = [0, 1]$, which is a fractal with iterated function system $F_0(x) = x/2$, $F_1(x) = (x + 1)/2$. The fixed points of each F_0 is 0, and the fixed point of F_1 is 1. Notice that $F_0(I) \cap F_1(I) = \{1/2\}$, so the critical set is*

$$\mathcal{C} = h^{-1}(\{1/2\}) = \{0\bar{1}, 1\bar{0}\}$$

Since $h(0\bar{1}) = F_0 \circ F_1 \circ F_1 \circ \ldots (x) = F_0(1) = 1/2$, and similarly $h(1\bar{0}) = F_1 \circ F_0 \circ F_0 \circ \ldots (x) = F_1(0) = 1/2$. The post critical set is therefore

$$\mathcal{P} = \bigcup_{n \geq 1} \xi^{(n)}(\{0\bar{1}, 1\bar{0}\}) = \{\bar{1}, \bar{0}\},$$

which is finite, so $[0, 1]$ is a pcf fractal. But $\mathcal{P}$ is precisely the set of points which h maps to the fixed points 0 and 1, so the boundary of I is (unsurprisingly)

$$V_0 = h(\mathcal{P}) = \{0, 1\}$$

Example 5.2.4 (Boundary of the Sierpinski Gasket)**.** *Let SG be the Sierpinski gasket, with iterated function system $F_i(x) = 1/2(x - q_i) + q_i$ for q_i the vertices of an equilateral triangle. The fixed point of each F_j is q_j, hence $h(\bar{j}) = F_j \circ F_j \circ F_j \circ \ldots (x) = q_j$, where $\bar{j} := jjj\ldots$.*

The intersection $F_1(SG) \cap F_2(SG)$ contains one point which we call p_3, and similarly $F_2(SG) \cap F_3(SG) = \{p_1\}$, and $F_1(SG) \cap F_3(SG) = \{p_2\}$.

Now, one can check that $h^{-1}(\{p_1\}) = \{2\bar{3}, 3\bar{2}\}$, and similarly $h^{-1}(\{p_2\}) = \{1\bar{3}, 3\bar{1}\}$ and $h^{-1}(\{p_3\}) = \{1\bar{2}, 2\bar{1}\}$.

So the critical set is $\mathcal{C} = h^{-1}(\{p_1, p_2, p_3\}) = \{2\bar{3}, 3\bar{2}, 1\bar{3}, 3\bar{1}, 1\bar{2}, 2\bar{1}\}$, and the post critical set is therefore

$$\mathcal{P} = \bigcup_{n \geq 1} \xi^{(n)}(\{2\bar{3}, 3\bar{2}, 1\bar{3}, 3\bar{1}, 1\bar{2}, 2\bar{1}\}) = \{\bar{3}, \bar{2}, \bar{1}\},$$

but this is precisely the set of points which map to the fixed points q_i, so the boundary of SG is simply

$$V_0 = h(\mathcal{P}) = \{q_1, q_2, q_3\}.$$

We will be using both the interval and the Sierpinski gasket as a running example throughout this section.

Let V be any set, and let $\ell(V) := \{f \mid f : V \to \mathbb{R}\}$. If V is finite, then it is understood to be equipped with the inner product

$$\langle u, v\rangle = \sum_{p \in V} u(p)v(p)$$

We begin by defining the Laplacian and Dirichlet form on finite sets.

Definition 5.2.5 (Finite Laplacian)**.** *Let V be a finite set. A Laplacian is a linear operator $\Delta : \ell(V) \to \ell(V)$ which satisfies the following properties for any $u \in \ell(V)$:*

1. *$\langle u, \Delta u\rangle = \langle \Delta u, u\rangle$ (symmetry)*
2. *$\langle u, \Delta u\rangle \leq 0$ (negative semi-definite)*
3. *$\Delta u = 0$ if and only if u is a constant function on V*
4. *if $p \neq q$, then $\Delta_{pq} \geq 0$, where $\Delta_{pq} := \langle \chi_p, \Delta \chi_q\rangle = (\Delta \chi_q)(p)$ and χ_q is the characteristic function of the set $\{q\}$.*

Notice that these are precisely the properties we proved about the Laplace operator in the Riemannian setting (propositions 5.1.2 through 5.1.5); here, we take these properties as the definition of the Laplacian.

The collection of Laplacians on V is denoted by $\mathcal{L}(V)$, and the collection of linear operators satisfying only the first three properties is denoted $\widetilde{\mathcal{L}}(V)$.

Lemma 5.2.6. *Let V be finite, and let $H : \ell(V) \to \ell(V)$ be a linear operator. Define $H_{pq} = (H\chi_q)(p)$. For any $p \in V$, $u \in \ell(V)$, we can write*

$$(Hu)(p) = \sum_{q \in V} H_{pq} u(q)$$

Proof.
$$\begin{aligned}\sum_{q\in V} H_{pq}u(q) &= \sum_{q\in V}(H\chi_q)(p)(u(q))\\ &= \sum_{q\in V}\begin{cases} H(p) & q=p \\ 0 & q\neq p\end{cases} u(q)\\ &= \sum_{q\in V}\begin{cases}(H(p))(u(p)) & q=p\\ 0 & q\neq p\end{cases}\\ &= (Hu)(p)\end{aligned}$$

□

Definition 5.2.7 (Finite Dirichlet Form)**.** *Let V be a finite set. A Dirichlet form is a bilinear map $\mathcal{E}:\ell(V)\times\ell(V)\to\mathbb{R}$ satisfying the following properties for any $(u,v)\in\ell(V)\times\ell(V)$:*

1. $\mathcal{E}(u,v)=\mathcal{E}(v,u)$ *(symmetry)*
2. $\mathcal{E}(u,u)\geq 0$ *(positive semi-definite)*
3. $\mathcal{E}(u,u)=0$ *if and only if u is a constant function on V*
4. $\mathcal{E}(u^*,u^*)\leq\mathcal{E}(u,u)$*, where* $u^*=\min\{\max\{0,u\},1\}$*. (Markov property)*

The collection of Dirichlet forms on V is denoted by $\mathcal{DF}(V)$, and the collection of bilinear forms satisfying only the first three properties is denoted by $\widetilde{\mathcal{DF}}(V)$.

There is an important correspondence between Laplacians and Dirichlet forms which was alluded to in section 5.1. In fact, to any symmetric linear operator $H:\ell(V)\to\ell(V)$, we can associate a Dirichlet form by $\mathcal{E}_H(u,v):=-\langle u,Hv\rangle$.

Theorem 5.2.8. *Let V be a finite set, and let $\pi:\widetilde{\mathcal{L}}(V)\to\widetilde{\mathcal{DF}}(V)$ be defined by $\pi(H)(u,v):=\mathcal{E}_H(u,v)$. Then π is a bijection and $\mathcal{DF}(V)=\pi(\mathcal{L}(V))$.*

Proof. π is surjective, since given $\mathcal{E}_H$, we can reverse engineer H using matrix coefficients. Choose coefficients H_{pq} so that

$$-\sum_{p\in V}u(p)\left(\sum_{q\in V}H_{pq}v(q)\right)=\mathcal{E}_H(u,v),$$

and define $H : \ell(V) \to \ell(V)$ by $Hu(p) := \sum_{q \in V} H_{pq} u(q)$. It is also injective, since if $H_1 \neq H_2$ are two different operators in $\widetilde{\mathcal{L}}(V)$, then there is a function v such that $H_1 v \neq H_2 v$, hence $-\langle u, H_1 v\rangle = \mathcal{E}_{H_1}(u,v) \neq \mathcal{E}_{H_2}(u,v) = -\langle u, H_2 v\rangle$. So π is bijective. However, we still need to show that $\pi(H)$ satisfies the necessary properties to be in $\widetilde{\mathcal{DF}}(V)$ if and only if H satisfies the necessary properties to be in $\widetilde{\mathcal{L}}(V)$.

First we show $\pi(\widetilde{\mathcal{L}}(V)) = \widetilde{\mathcal{DF}}(V)$. Let $\Delta \in \widetilde{\mathcal{L}}(V)$. We check that $\pi(\Delta) : \ell(V) \times \ell(V) \to \mathbb{R}$ given by $\pi(\Delta)(u,u) = \mathcal{E}_\Delta(u,u)$ is in $\widetilde{\mathcal{DF}}$.

Symmetry of $\mathcal{E}_\Delta$ is given by symmetry of Δ itself. By nonpositive definiteness of Δ, we have that

$$\mathcal{E}_\Delta(u,u) = -\langle u, \Delta u\rangle \geq 0$$

Next, if $u \in \ell(V)$ is constant, then we have $\Delta u = 0$, hence

$$\mathcal{E}_\Delta(u,u) = -\langle u, \Delta u\rangle = -\langle u, 0\rangle = 0$$

Thus $\pi(\widetilde{\mathcal{L}}(V)) \subseteq \widetilde{\mathcal{DF}}(V)$. Next, we show that $\widetilde{\mathcal{DF}}(V) \subseteq \pi(\widetilde{\mathcal{L}}(V))$. Let $\mathcal{E} \in \widetilde{\mathcal{DF}}(V)$. The associated linear operator is the operator Δ satisfying $\mathcal{E}(u,u) = -\langle u, \Delta u\rangle$.

Symmetry of $\mathcal{E}$ implies symmetry of Δ. $\mathcal{E}$ is nonnegative-definite, so $\mathcal{E}(u,u) = -\langle u, \Delta u\rangle \geq 0$ implies that $\langle u, \Delta u\rangle \leq 0$, which is precisely the statement that Δ is nonpositive-definite. Let u be constant. Then $\mathcal{E}(u,u) = 0 = -\langle u, \Delta u\rangle$, which implies that $\Delta u = 0$; similarly, if $\Delta u = 0$ then $\mathcal{E} = 0 = -\langle u, \Delta u\rangle$ which implies u must be constant (since $\mathcal{E} \in \widetilde{\mathcal{DF}}$), and so $\widetilde{\mathcal{DF}}(V) = \pi(\widetilde{\mathcal{L}}(V))$.

Next we will show that $\pi(\mathcal{L}(V)) \subseteq \mathcal{DF}(V)$. Suppose that $\Delta \in \mathcal{L}(V)$ and consider $\pi(\Delta) = \mathcal{E}_\Delta$. Let $u^* = \min\{\max\{0,u\},1\}$, and notice that

$$\begin{aligned}
\mathcal{E}_\Delta(u,u) &= -\sum_{p \in V}\sum_{q \in V} u(p)\Delta_{pq}u(q) \\
&= \frac{1}{2}\sum_{p \in V}\sum_{q \in V} \Delta_{pq}\left(u(p) - u(q)\right)^2
\end{aligned}$$

We wish to show that $\mathcal{E}_\Delta(u,u) \geq \mathcal{E}_\Delta(u^*,u^*)$ $\forall u \in \ell(V)$. Let u be arbitrary; there are several possible cases to consider. For instance, suppose that $u(q) \in (0,1)$ and $u(p) \leq 0$. Then $\left(u^*(p) - u^*(q)\right)^2 = u(q)^2 < \left(u(p) - u(q)\right)^2$.

A similar argument works for all other cases. In every case, we get the inequality

$$(u^*(p) - u^*(q))^2 \leq (u(p) - u(q))^2\,,$$

and since $\Delta_{pq} \geq 0$ for all $p \neq q \in V$, we conclude that all the terms in the summation are positive (or 0 if $p = q$), hence $\mathcal{E}_\Delta(u^*, u^*) \leq \mathcal{E}_\Delta(u, u)$ for each u in $\ell(V)$, as desired. So $\pi(\mathcal{L}(V)) \subseteq \mathcal{DF}(V)$.

Next we want to show that $\mathcal{E}_H \in \mathcal{DF}(V)$ implies that $H \in \mathcal{L}(V)$, so suppose that $\mathcal{E}_H \in \mathcal{DF}(V)$. Suppose for contradiction that $H \in \widetilde{\mathcal{L}}(V) \backslash \mathcal{L}(V)$. Then $\exists p \neq q$ with $H_{pq} < 0$. Without loss of generality, set $H_{pq} = -1$. Denote $u(p) = x$, $u(q) = y$, and $u(a) = z$ for all $a \neq p, q$. Then

$$\mathcal{E}_H(u, u) = \alpha(x - z)^2 + \beta(y - z)^2 - (x - y)^2.$$

Now $\mathcal{E}_H$ is nonnegative-definite, so we can guarantee that $\alpha, \beta > 0$. Consider the case $x = 1$, $z = 0$. Then $\mathcal{E}_H(u, u) = \alpha - 1 + 2y + (\beta - 1)y^2$, and $\mathcal{E}_H(u^*, u^*) = \alpha - 1$.

We can consider a case where $y < 0$, and $|y|$ is small enough to make $\mathcal{E}_H(u, u) \leq \mathcal{E}_H(u^*, u^*)$, which implies that $\mathcal{E}_H \notin \mathcal{DF}(V)$, which is the desired contradiction. Hence $\mathcal{DF}(V) = \pi(\mathcal{L}(V))$.

□

In particular, this means that given a Dirichlet form we can uniquely find an associated Laplacian (and vice-versa). This is an extremely important result because, as we shall see, functions which minimize $\mathcal{E}$ are solutions to Laplace's equation, and hence problems involving Laplace operators can be reformulated as problems involving Dirichlet forms.

We have defined Dirichlet forms on finite sets, but now we need to define them on arbitrary measure spaces. The reason for introducing Dirichlet forms is that we can use them to talk about Laplace's equation on simply a measure space, rather than explicitly defining a Laplacian on a manifold.

Definition 5.2.9 (Dirichlet Form)**.** *Let (X, μ) be a measure space. Let $\mathcal{D}$ be a dense subspace of $L^2(X, \mu)$. A Dirichlet form is a bilinear map $\mathcal{E} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ satisfying the following properties for any $(u, v) \in \mathcal{D} \times \mathcal{D}$.*

1. $\mathcal{E}(u, v) = \mathcal{E}(v, u)$ *(symmetry)*
2. $\mathcal{E}(u, u) \geq 0$ *(nonnegative-definite)*

3. $\mathcal{E}(u^*, u^*) \leq \mathcal{E}(u, u)$*, where* $u^* = \min\{\max\{0, u\}, 1\}$*. (Markov property)*

4. $\mathcal{D} \subset L^2(X, \mu)$ *is a Hilbert space when equipped with the inner product* $\langle u, v \rangle_{\mathcal{D}} := \langle u, v \rangle_{L^2(X,\mu)} + \mathcal{E}(u, v)$.

Definition 5.2.10 (Restriction)**.** *Let $\mathcal{E}$ be a Dirichlet form on a set V (finite or infinite), and let U be a proper subset of V. We define the restriction*

$$\mathcal{E}(f, f)|_U := \inf\{\mathcal{E}(g, g) \mid g|_U = f\}$$

Using this, we can define a notion of compatible sequences of Dirichlet forms $\mathcal{E}_n$ on the graph approximations V_n as defined in section 2.

Definition 5.2.11 (Compatibility)**.** *Let $(V_n, \mathcal{E}_n)$ be a sequence of Dirichlet forms $\mathcal{E}_n$ defined on an increasing sequence of sets V_n. The sequence is said to be* compatible *if the following restriction property holds for all n:*

$$\mathcal{E}_{n+1}|_{V_n} = \mathcal{E}_n$$

Now we shall introduce a Dirichlet form defined on any proper subset of a fractal K.

Proposition 5.2.12. *Let V_n be an increasing sequence of sets converging to a dense subset $V^* := \cup_n V_n$ of a set V. Let $\mathcal{E}_n \in \mathcal{DF}(V_n)$ $\forall n \in \mathbb{N}$, and suppose that $(V_n, \mathcal{E}_n)$ is a compatible sequence. Then we define*

$$\mathcal{E}(u, v) := \lim_{n\to\infty} \mathcal{E}_n(u|_{V_n}, v|_{V_n}),$$

which exists (but may be infinite). Then $\mathcal{E}$ is a Dirichlet form on V^*.

Proof. Since $\mathcal{E}_n \in \mathcal{DF}$ for all n, we have by symmetry of $\mathcal{E}_n$ that

$$\mathcal{E}(u, v) = \lim_{n\to\infty} \mathcal{E}_n(u, v) = \lim_{n\to\infty} \mathcal{E}_n(v, u) = \mathcal{E}(v, u),$$

so $\mathcal{E}$ is symmetric. Also $\mathcal{E}_n(u, u) \geq 0$ for all $u \in \ell(V)$ and for all $n \in \mathbb{N}$, so, in the limit, $\mathcal{E}(u, u) \geq 0$. It is clear from the definition that $\mathcal{E}_m(\alpha u, v) = \alpha\mathcal{E}_m(u, v)$ and $\mathcal{E}_m(u + v, w) = \mathcal{E}_m(u, w) + \mathcal{E}_m(v, w)$; since these hold for each m, they hold for the limit $\mathcal{E}$ as well. Thus $\mathcal{E}$ is an inner product on V^*.

Next, notice that V^* is dense in $\overline{V^*}$ (in the fractal setting, V^* will be dense in the fractal $K = \overline{V^*}$, which is a measure space).

Now we wish to show $\mathcal{E}(u,u) = 0$ if and only if u is constant. First suppose u is constant; then since $\mathcal{E}_n$ is a Dirichlet form for each n, we have $\mathcal{E}_n(u,u) = 0$ for all n and hence the limit $\mathcal{E}(u,u) = 0$ is also zero. dom

Next let $\mathcal{E}(u,u) = 0$. Notice that by the definition of the restriction and the fact that $(V_n, \mathcal{E}_n)$ is compatible,

$$\mathcal{E}_{n+1}|_{V_n}(u,u) = \mathcal{E}_n(u,u) = \inf\{\mathcal{E}_{n+1}(g,g) \mid g|_{V_n} = u,\}$$

so we have that $\mathcal{E}_n(u,u)$ is an increasing sequence (which proves that $\mathcal{E}$ is well-defined, but may be infinite). Now, suppose by contradiction that u is not constant; then each $\mathcal{E}_n(u,u)$ is strictly greater than 0 and the sequence is increasing, hence the limit $\mathcal{E}(u,u)$ is strictly greater than 0. This contradicts our assumption that $\mathcal{E}(u,u) = 0$, hence u must be constant.

Next, let $u^* = \min\{\max\{0,u\},1\}$ and let $u \in \ell(V)$ be any other real valued function on V. Since $\mathcal{E}_n \in \mathcal{DF}$, we have $\mathcal{E}_n(u^*,u^*) \leq \mathcal{E}_n(u,u)$ for all n. So $\mathcal{E}(u^*,u^*) = \lim_{n\to\infty} \mathcal{E}_n(u^*,u^*) \leq \lim_{n\to\infty} \mathcal{E}_n(u,u) = \mathcal{E}(u,u)$, as desired.

Finally, we will need to show completeness in the norm induced by the inner product. In fact, we will have to show that this holds for functions in $\operatorname{dom}\mathcal{E} := \{f : V^* \to \mathbb{R} \mid \mathcal{E}(f,f) < \infty$ modulo constant functions.

We can identify $\operatorname{dom}\mathcal{E}/$constant functions with the space $\mathcal{E}' := \{u \in \operatorname{dom}\mathcal{E} \mid u(q_0) = 0\}$.

Let (u_n) be a sequence in $\mathcal{E}'$ such that

$$\mathcal{E}(u_n - u_k) \to 0$$

As $n,m \to \infty$. Fix m. We have $0 \leq \mathcal{E}_m(u_n - u_k) \leq \mathcal{E}(u_n - u_k)$, so $\mathcal{E}_m(u_n - u_k) \to 0$ also. It then follows from the definition of $\mathcal{E}_m$ that $\lim_{k\to\infty} u_k(x)$ exists for all $x \in V_m$, so define $u \in \ell(V^*)$ by

$$u(x) := \lim_{k\to\infty} u_k(x)$$

This implies that $\mathcal{E}_m(u_n - u) = \lim_{k\to\infty} \mathcal{E}_m(u_n - u_k)$. But since (u_n) is Cauchy, we can make $\mathcal{E}_m(u_n - u_k)$ arbitrarily small for any m, so we can make $\mathcal{E}_m(u_n - u)$ arbitrarily small for any m. Hence $\mathcal{E}(u_n - u) \to 0$, as desired. Therefore $\mathcal{E} : V^* \times V^* \to \mathbb{R}$ is a Dirichlet form. $\square$

We define the domain of $\mathcal{E}$ to be the set

$$\operatorname{dom}\mathcal{E} := \{f : V^* \to \mathbb{R} \mid \mathcal{E}(f,f) < \infty\}$$

Recall that the sequence $V_n \to V^*$ converging to a dense subset using V_0 as the boundary is exactly the scenario we had in the fractal case, from section 2. Thus all of the preceding discussion applies to fractals sets (and their approximating graphs) as well.

We've built up a theory of Dirichlet forms, but now it's time to actually construct one. We will start by defining the following bilinear "energy form" on approximating set V_m of a fractal:

$$E_m(u,v) = \sum_{x \sim_m y} [u(x) - u(y)]\,[v(x) - v(y)]$$

Recall $x \sim_m y$ whenever x and y have an edge joining them in the m^{th} level graph approximation V_m. There is also the associated quadratic form

$$E_m(u) := E_m(u,u) = \sum_{x \sim_m y} [u(x) - u(y)]^2\,,$$

which satisfies the following polarization identity with the bilinear form:

$$\mathcal{E}_m(u,v) = \frac{1}{4}[\mathcal{E}_m(u+v) - \mathcal{E}_m(u-v)].$$

The reason to define E_m this way will become apparent soon. Let's first check that $E_m(u,v)$ is a finite Dirichlet form on V_m. It is clear from the definition that E_m is symmetric and nonnegative-definite, and it is also easy to see from the formula that $E_m(u) = 0$ if and only if u is constant. Finally, if we replace u by $u^* = \min\{1, \max\{u, 0\}\}$, then each of the terms $[u(x) - u(y)]^2$ can only decrease or stay the same, which can be easily verified but is somewhat tedious.

Suppose $u(x), u(y) < 0$; then $u^*(x) = u^*(y) = 0$, so $[u^*(x) - u^*(y)]^2 = 0 < [u(x) - u(y)]^2$.

If $0 < u(x) < 1$ and $u(y) < 0$, then $u^*(x) = u(x)$ and $u^*(y) = 0$, so $[u^*(x) - u^*(y)]^2 = [u^*(x)]^2 < [u(x) - u(y)]^2$ since $u(y)$ is strictly negative.

If $0 < u(x), u(y) < 1$ then we simply have equality, $[u^*(x) - u^*(y)]^2 = [u(x) - u(y)]^2$.

If $0 < u(x) < 1$ and $u(y) > 1$, then $u^*(x) = u(x)$ and $u^*(y) = 1$, so $[u^*(x) - u^*(y)]^2 = [u(x) - 1]^2 < [u(x) - u(y)]^2$.

If $u(x), u(y) > 1$ then $u^*(x) = u^*(y) = 1$ so $[u^*(x) - u^*(y)]^2 = 0 \leq [u(x) - u(y)]^2$.

Finally if $u(x) > 1$ and $u(y) < 0$, then $u^*(x) = 1$ and $u^*(y) = 0$ so $[u^*(x) - u^*(y)]^2 = 1 < [u(x) - u(y)]^2$.

Symmetry takes care of all the other cases. So E_m satisfies the Markov property, and is therefore a finite Dirichlet form.

Definition 5.2.13 (Harmonic extension)**.** *Let $u \in \ell(V_m)$ be a function, where V_m is an approximating set for the fractal K. A* harmonic extension of u *to V_{m+1} is a function $\tilde{u} \in \ell(V_{m+1})$ which minimizes the Dirichlet form in the sense that*

$$\tilde{u}|_{V_m} = u$$

and

$$E_m(\tilde{u}) \leq E_m(f)$$

for any other extension f of u.

Often, the Dirichlet form will need to be *renormalized* at each step, in the sense there are coefficients c_{xy}, such that, given $E_m(u) = \sum_{x \sim_m y}(u(x) - u(y))^2$, we have

$$E_{m+1}(\tilde{u}) = \sum_{x \sim_m y} c_{xy}(u(x) - u(y))^2.$$

In fact, we will concentrate on cases where the coefficients c_{xy} are equal for all x, y, so that $rE_{m+1}(\tilde{u}) = E_m(u)$ for some $r \in \mathbb{R}$. Notice that we write $r = 1/c_{xy}$ for all x and y. In this case, we write the renormalized Dirichlet form as

$$\mathcal{E}_m(u) := \frac{1}{r} E_m(u).$$

Then, if $\tilde{u}$ is a harmonic extension, then $\mathcal{E}_m(u) = \mathcal{E}_{m+1}(\tilde{u})$ for all m. The unit interval and Sierpinski gasket, for example, are both cases where all c_{xy} are equal.

Example 5.2.14 (Interval)**.** *Let $I = [0, 1]$ be the unit interval, which has the iterated function system $F_1 = x/2$, $F_2 = (x + 1)/2$. Then $V_0 = \{0, 1\}$ (see example 5.2.3) and $V_1 = \{0, 1/2, 1\}$. So given some function u defined on V_0, the energy form takes the value $E_0(u) = [u(1) - u(0)]^2$.*

How would we go about defining the harmonic extension of u to V_1? We need to find $\tilde{u}$ that minimizes

$$E_1(\tilde{u}) = [\tilde{u}(1) - \tilde{u}(1/2)]^2 + [\tilde{u}(1/2) - \tilde{u}(0)]^2.$$

The only undefined quantity is $\tilde{u}(1/2)$*, and some calculus shows that the minimizing value is* $\tilde{u}(1/2) = 1/2[u(1) + u(0)]$*. Substituting this in, we actually find that*

$$E_1(\tilde{u}) = \frac{1}{2} E_0(u).$$

Similarly, at the m^{th} *level graph approximation* $V_m = \left\{\frac{k}{2^m}\right\}_{k=0}^{2^m}$*, the only "new" terms in the* $(m+1)^{th}$ *Dirichlet form are those involving odd values of* k*. Given any odd* k*, say* $k = 2j+1$*, the term* $\tilde{u}\left(\frac{2j+1}{2^{m+1}}\right)$ *only appears twice in the sum for* E_{m+1} *since any point in* V_{m+1} *only has two neighbouring points to which it is connected. Specifically,* $\tilde{u}\left(\frac{2j+1}{2^{m+1}}\right)$ *appears in the terms*

$$\left[u\left(\frac{2j+2}{2^{m+1}}\right) - \tilde{u}\left(\frac{2j+1}{2^{m+1}}\right)\right]^2 + \left[\tilde{u}\left(\frac{2j+1}{2^{m+1}}\right) - u\left(\frac{2j}{2^{m+1}}\right)\right]^2.$$

Again some simple calculus shows that this is minimized by setting the new value to be the average between its two neighbours

$$\tilde{u}\left(\frac{2j+1}{2^{m+1}}\right) = \frac{1}{2}\left[u\left(\frac{2j+2}{2^{m+1}}\right) + u\left(\frac{2j}{2^{m+1}}\right)\right].$$

Notice that by this extension algorithm, we are essentially constructing a linear function with the desired values at the boundary, which is precisely the solution we'd expect from solving the 1-dimensional Laplace equation. Again, a computation yields

$$E_{m+1}(\tilde{u}) = \frac{1}{2} E_m$$

This is true for any m*, so the renormalization constant for the Dirichlet forms on the unit interval is* $r = 1/2$*.*

The Sierpinski gasket is another case where the coefficients c_{xy} are all equal.

Example 5.2.15 (Sierpinski Gasket)**.** *Recall that the boundary of SG is the set* $V_0 = \{q_0, q_1 q_2\}$ *of vertices of the equilateral triangle from which SG* *is constructed. Suppose that $u \in \ell(V_0)$ is defined so that $u(q_0) = 1$ and $u(q_1) = u(q_2) = 0$. Then*

$$E_0(u) = \sum_{x \sim_0 y} (u(x) - u(y))^2 = (1-0)^2 + (0-1)^2 + (0-0)^2 = 2.$$

How do we find the harmonic extension $\tilde{u}$ on V_1? Certainly we will have $\tilde{u}(F_0(q_1)) = \tilde{u}(F_0(q_2)) := x$ at the two points in V_1 closest to q_0 and $\tilde{u}(F_1(q_2)) = y$ at the point in V_1 farthest from q_0 (see figure 5).

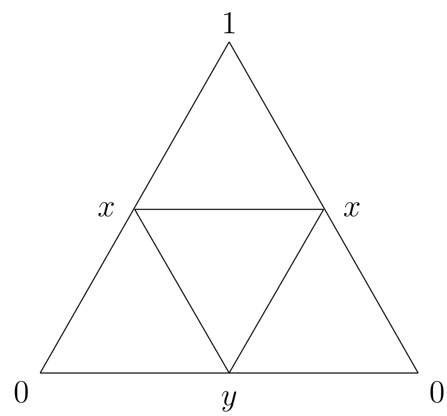


Figure 5: The set V_1 and the value of $\tilde{u}$ on these points.

The energy form on V_1 using the extended function $\tilde{u}$ is

$$E_1(\tilde{u}) = 2(x-1)^2 + 2x^2 + 2y^2 + 2(x-y)^2.$$

Setting the derivatives of E_1 equal to zero, we get two linear equations to solve:

$$4x = 1 + x + y$$
$$4y = 2x$$

By inspection, a solution is $x = 2/5$, $y = 1/5$. By symmetry, exactly the same result would occur if we had set the value 1 at any of the other boundary points. Furthermore, E_1 is a quadratic form, so taking its derivative gives a *set of linear minimizing equations. If the values of u on V_0 are a, b, c then $\tilde{u}$ must satisfy the following rule:*

$$\tilde{u}(z) = \frac{2}{5}a + \frac{2}{5}b + \frac{1}{5}c,$$

where z is the point shown in figure 6. Of course, symmetry takes care of the points x and y as well.

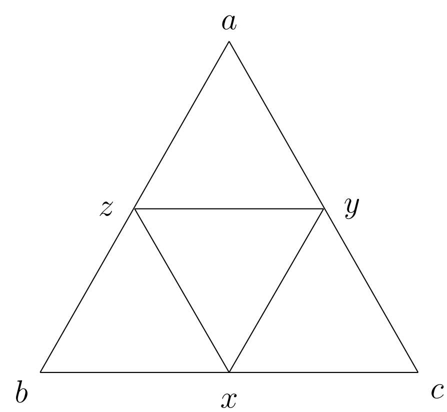


Figure 6: The set V_1 with $\tilde{u}$ taking arbitrary values on these points.

It requires considerably more work, but one can show (see [7] section 1.3) that this rule actually holds locally at each m-cell, with $r = 3/5$, so that $E_{m+1}(\tilde{u}) = (5/3)E_m(u)$ for each m.

Example 5.2.14 provides us with a hint as to why we defined the Dirichlet form as $E_m(u) = \sum_{x \sim_m y}[u(x) - u(y)]^2$. The renormalized Dirichlet form can be written explicitly in the simple case of the interval:

$$\begin{aligned}\mathcal{E}_m(u) &= 2^m \sum_{k=0}^{2^m-1} \left[u\left(\frac{k+1}{2^m}\right) - u\left(\frac{k}{2^m}\right)\right]^2 \\ &= \frac{1}{2^m} \sum_{k=0}^{2^m-1} \left(\frac{u\left(\frac{k+1}{2^m}\right) - u\left(\frac{k}{2^m}\right)}{1/2^m} \right)^2.\end{aligned}$$

Assuming u is continuous and differentiable, we can apply the mean value theorem to obtain points x_k such that

$$\mathcal{E}_m(u) = \sum_{k=0}^{2^m-1} \left(\frac{du}{dx}(x_k)\right)^2 \left(\frac{1}{2^m}\right).$$

Passing the limit, we see that this is a Riemann sum converging to

$$\mathcal{E}(u) = \lim_{m\to\infty} \mathcal{E}_m(u) = \int_0^1 \left(\frac{du}{dx}\right)^2 dx,$$

which is the classical Dirichlet form on the Euclidean interval. This is the motivation for defining the finite Dirichlet form the way we did.

It is therefore natural to define

$$\mathcal{E}(u) := \lim_{m\to\infty} \mathcal{E}_m(u),$$

which satisfies the compatibility property of 5.2.11, and hence converges to a Dirichlet form by 5.2.12. With the Dirichlet form defined on fractal sets, we can define an associated Laplacian.

Definition 5.2.16 (Fractal Laplacian)**.** *Let K be a fractal set and let $f : K \to \mathbb{R}$ be a function in $\mathrm{dom}\mathcal{E} \cap L^2(\mu)$. The Laplacian of f, with respect to a measure μ, is the continuous function $\Delta_\mu f$ satisfying*

$$\mathcal{E}(f, v) = -\langle v, \Delta_\mu f\rangle_{L^2(\mu)} = -\int_K v\Delta_\mu f d\mu$$

for any v that vanishes along the boundary V_0 of K.

The fact that we require v to vanish on the boundary is analogous to requiring that g vanish on the boundary in the integration by parts formula

$$\int_a^b \left(\frac{dg}{dx}\right)^2 dx = g\frac{dg}{dx}\bigg|_a^b - \int_a^b g\frac{d^2g}{dx^2}dx = -\langle g, \Delta g\rangle_{L^2(\lambda)},$$

relating the classical Dirichlet form $\int_a^b \left(\frac{dg}{dx}\right)^2 dx$ to the classical Laplacian. Theorem 5.2.8 guarantees that Δ_μ exists and is uniquely defined by this equation. Notice that the Laplace operator depends on our choice of measure.

Lemma 5.2.17. *Let u, v be defined on V_m. If $\tilde{u}$ is the harmonic extension of u to V_{m+1} and v' is any extension (not necessarily harmonic) of v, then $\mathcal{E}_{m+1}(\tilde{u}, v') = \mathcal{E}_m(u, v)$.*

Proof. By the polarization identity,

$$\mathcal{E}_m(u,v) = \frac{1}{4}[\mathcal{E}_m(u+v) - \mathcal{E}_m(u-v)].$$

We have $\mathcal{E}_{m+1}(\tilde{u},\tilde{v}) = \mathcal{E}_m(u,v)$. Consider the function $v' - \tilde{v}$. By definition, we have the following:

$$E_{m+1}(\tilde{u}, v' - \tilde{v}) = \sum_{x \sim_{m+1} y} [\tilde{u}(x) - \tilde{u}(y)][(v'(x) - \tilde{v}(x)) - (v'(y) - \tilde{v}(y))].$$

Notice that any term in this sum containing $x \in V_m$ is zero, since the two extensions of v must agree on this set. Thus the only terms in the sum which might contribute anything are those for which $x \in V_{m+1} \setminus V_m$. Hence the only remaining terms must be of the form

$$\sum_{x \sim_m y} \alpha[\tilde{u}(x) - u(y)],$$

where $\alpha = [(v'(x) - \tilde{v}(x)) - (v'(y) - \tilde{v}(y))]$. These also sum to zero, because for $x \in V_{m+1} \setminus V_m$, $\tilde{u}(x)$ is just the average of the values of the neighbouring points in V_{m+1}.

Thus $\mathcal{E}_{m+1}(\tilde{u}, v' - \tilde{v}) = 0$. This implies that $v' - \tilde{v}$ is constant on V_{m+1}, which means $v' = \tilde{v} + p$ for some $p \in \mathbb{R}$. By the definition of $\mathcal{E}_m$, the extra constant p will "cancel out", giving the desired result

$$\mathcal{E}_{m+1}(\tilde{u}, v') = \mathcal{E}_{m+1}(\tilde{u}, \tilde{v}) = \mathcal{E}_m(u,v). \qquad \square$$

Definition 5.2.18 (Harmonic function)**.** *Let h take any finite value on the boundary V_0. Given the increasing sequence of approximating sets $V_m \to V^*$, the extended function $\tilde{h} \in \ell(V^*)$ is said to be* harmonic *if it minimizes $\mathcal{E}_m$ for all m.*

Not only does the Dirichlet form allow us to define the Laplace operator, but the two also have a special relationship in terms of harmonic functions. Next comes the theorem that shows why Dirichlet forms are so useful:

Theorem 5.2.19. *A function $h \in \operatorname{dom} \mathcal{E}$ is harmonic (i.e. minimizes the Dirichlet forms $\mathcal{E}_m$ for all m) if and only if h is a solution to Laplace's equation $\Delta_\mu h = 0$.*

Proof. Let h be harmonic; notice that $\mathcal{E}_0(h,v) = 0$ because v vanishes on the boundary V_0. By lemma 5.2.17, $\mathcal{E}(h,v) = \mathcal{E}_0(h,v)$, thus

$$-\int_K v\Delta_\mu h d\mu = \mathcal{E}(h,v) = \mathcal{E}_0(h,v) = 0.$$

But since this holds for any v, it must be that $\Delta_\mu h = 0$. Now suppose that $\Delta_\mu h = 0$; then $\mathcal{E}(h,h) = \int_K (\Delta_\mu h)^2 d\mu = 0$. But the sequence $\mathcal{E}_m \to \mathcal{E}$ satisfies

$$0 \leq \mathcal{E}_m(h,h) \leq \mathcal{E}_{m+1}(h,h) \leq \mathcal{E}(h,h) = 0 \quad \text{for all } m.$$

Hence $\mathcal{E}_m(h,h) = 0$ for all m, i.e. h is harmonic. □

This allows us to solve the fractal differential equation $\Delta_\mu f = 0$, with boundary conditions given by the values of f on the boundary V_0, by finding the harmonic function associated with the repeated harmonic extension of $f|_{V_0}$ from V_0 to V_1, to V_2, etc. Of course, the challenge then becomes finding a suitable harmonic extension "algorithm"; even the simple case of the interval took a bit of work in example 5.2.14. The Sierpinski gasket in example 5.2.15 was significantly harder.

Furthermore, although the resulting extended function is only defined on $V^* = \cup_m V_m$, recall that $\overline{V^*} = K$. Let $u \in \operatorname{dom} \mathcal{E}$; we'll show that u is uniformly continuous, hence it can be extended to a continuous function on K naturally by setting $u(x) = \lim u(x_n)$ for a sequence (x_n) in V^* converging to $x \in K$.

If $x \sim_m y$, then

$$r^{-m}[u(x) - u(y)]^2 \leq \mathcal{E}_m(u) \leq \mathcal{E}(u),$$

since $r^{-m}[u(x) - u(y)]^2$ is an individual term in the sum of positive terms which make up $\mathcal{E}_m(u)$. This implies that $|u(x) - u(y)| \leq r^{m/2}\mathcal{E}(u)^{1/2}$. Now consider any "chain" of elements $x_m, \ldots, x_{m+k}$ which are related in the sense that for any $\ell > m$, $x_\ell \in V_\ell$ and $x_\ell \sim_{\ell+1} x_{\ell+1}$. Then

$$
\begin{aligned}
|u(x_m) - u(x_{m+k})| &\le \sum_{i=1}^{k} |u(x_m) - u(x_{m+i})| \\
&\le \sum_{i=1}^{k} r^{(m+\ell)/2} \mathcal{E}(u)^{1/2} \\
&= r^{m/2} \mathcal{E}(u)^{1/2} \sum_{i=1}^{k} r^{i/2} \\
&= \frac{r^{m/2}}{1 - r^{1/2}} \mathcal{E}(u)^{1/2},
\end{aligned}
$$

which can be made arbitrarily small by taking large enough values of m, since $0 < r < 1$. So $\operatorname{dom} \mathcal{E} \subset C(K)$.

Finite Dirichlet forms also satisfy the self-similar relation $\mathcal{E}_{m+1}(u) = \sum_i r^{-1} \mathcal{E}_m(u \circ F_i)$ for any m by definition, hence the limit satisfies this as well:

$$
\mathcal{E}(u) = \sum_i r^{-1} \mathcal{E}(u \circ F_i)
$$

Definition 5.2.20 (Tent function)**.** *Let K be a fractal. Fix m, and consider the usual approximating set V_m. Let $S(V_m)$ denote the set of continuous functions u such that $u \circ F_w$ is harmonic for all $|w| = m$.*

The tent function *is defined to be the continuous function $\psi_x^{(m)} \in S(V_m)$ which satisfies*

$$
\psi_x^{(m)}(y) = \delta_{xy}
$$

for all y in V_m, where x is any point in $V_m \setminus V_0$ and δ_{xy} is the Dirac delta.

The reason these are called tent functions is that in the case of the interval $K = [a, b]$, $\psi_x^{(m)}$ is shaped like a triangle, or "tent", on the sets V_m, which are just dyadic points $\ell/2^m$ in $[a, b]$ (see example 5.2.14).

Our definition of the Laplacian up until this point is fairly useless if we just want to know the value of $\Delta_\mu u(x)$ at a single point x. The following proposition takes care of this issue by introducing a "graph Laplacian" Δ_m on the set V_m, which converges (after some rescaling) to Δ_μ.

Proposition 5.2.21. *Suppose that a Laplacian of u exists. The following pointwise formula for the Laplacian holds at each $x \in V^* \setminus V_0$:*

$$\Delta_\mu u(x) = \lim_{m\to\infty} r^{-m} \left(\int_K \psi_x^{(m)} d\mu \right)^{-1} \Delta_m u(x),$$

where

$$\Delta_m u(x) := \sum_{y \sim_m x} (u(y) - u(x)) \quad \textit{for all } x \in V_m \setminus V_0$$

and $\psi_x^{(m)}(y) := \delta_{xy}$ is the tent function at x.

Proof. Given u, notice that

$$\begin{aligned}
\mathcal{E}_m(u, \psi_x^{(m)}) &= r^{-m} \sum_{x \sim_m y} [u(x) - u(y)][\psi_x^{(m)}(x) - \psi_x^{(m)}(y)] \\
&= r^{-m} \sum_{x \sim_m y} [u(x) - u(y)][1 - 0] \\
&= r^{-m} \Delta_m u(x).
\end{aligned}$$

But by definition,

$$\mathcal{E}(u, \psi_x^{(m)}) = \int_K \psi_x^{(m)} \Delta_\mu u d\mu.$$

Hence

$$r^{-m} \Delta_m u(x) = \int_K \psi_x^{(m)} \Delta_\mu u d\mu.$$

Dividing both sides by $\int_K \psi_x^{(m)} d\mu$, we get

$$r^{-m} \Delta_m u(x) \left(\int_K \psi_x^{(m)} d\mu \right)^{-1} = \frac{\int_K \psi_x^{(m)} \Delta_\mu u d\mu}{\int_K \psi_x^{(m)} d\mu}$$

for all m. But $\Delta_\mu u$ is continuous, and clealy $\lim_{m\to\infty} \psi_x^{(m)}$ is the Dirac delta δ_{xy} on V^*, so

$$\frac{\int_K \Delta_\mu u \psi_x^{(m)} d\mu}{\int_K \psi_x^{(m)} d\mu} \to \Delta_\mu u.$$

Hence

$$\lim_{m\to\infty} r^{-m}\left(\int_K \psi_x^{(m)} d\mu\right)^{-1} \Delta_m u(x) = \Delta_\mu u(x). \qquad \square$$

6 Conclusion

Fractal sets exhibit interesting self-similar structure. For instance, the fractal itself satisfies the self-similar formula

$$K = \bigcup_{|w|=m} F_w(K)$$

for any m. It can also be described, however, as a limit of approximating sets

$$K = \lim_{m\to\infty} V_m,$$

where $V_{m+1} = \cup_{i=1}^n F_i(V_m)$, and V_0 is any nonempty compact set. These are both consequences of the contraction mapping theorem. A fractal constructed from an iterated function system $\{F_i\}_{i=1}^n$ has an associated string space E^ω, where E is a set of n symbols. This complete, compact ultrametric space is related to the fractal by the addressing function h, via the formula

$$h \circ \theta_e = F_e \circ h.$$

The function h defines a natural "parametrization" of K, in the sense that $h(E^\omega) = K$. The natural measure μ on K also satisfies a self similar formula

$$\mu(A) = \sum_{|w|=m} \mu_w \mu(F_w^{-1} A),$$

which leads to another self-similar formula for the integral,

$$\int_K f d\mu = \sum_{|w|=m} \mu_w \int_K f \circ F_w d\mu.$$

The Hausdorff measure and the associated Hausdorff dimension are also useful in describing fractals. This notion of dimension allows for non-integer values; for instance, the Cantor set has Hausdorff dimension $\dim_H(C) = \log 2/\log 3$, and the Sierpinski gasket has Hausdorff dimension $\dim_H(SG) = \log 3/\log 2$. Moran's open set condition gives conditions that allow for a much simpler calculation of the Hausdorff dimension, using the similarity value.

With an understanding of the basic structure of fractals, one can a Laplacian on a certain class of fractals, called pcf fractals. This Laplacian is constructed indirectly via a Dirichlet form, but it can in fact be described locally as well. These generalized Dirichlet forms and Laplacians both reduce to the classical Dirichlet form and Laplacian in the Euclidean case.

References

[1] Gregory Derfel, Peter J. Grabner, and Fritz Vogl: *Laplace Operators on Fractals and Related Functional Equations* J. Phys. A: Math. Gen. 2012.

[2] Gerald Edgar: *Measure, Topology, and Fractal Geometry* second edition, Springer, 2008.

[3] Jun Kigami: *Analysis on Fractals* first edition, Cambridge University Press, 2001.

[4] Kasso Okoudjou: *Analysis on fractals: An introduction* `https://www.nist.gov/system/files/documents/2017/04/24/anafracnist.pdf` ACMD Seminar NIST, Gaithersburg, MD, 2017.

[5] James R. Munkres: *Topology* second edition, Pearson Eduction Inc. 2018.

[6] Anders Pelander: *A Study of Smooth Functions and Differential Equations on Fractals* Department of Mathematics, Uppsala University, 2007.

[7] Robert S. Strichartz: *Differential Equations on Fractals: a Tutorial* Princeton University Press, 2006.

[8] Robert Meyers, Robert S. Strichartz, and Alexander Teplyaev: *Dirichlet Forms on the Sierpinski Gasket* Pacific Journal of Mathematics, Vol 217 No. 1, pg 149-174, 2004.