An Introduction to Fractals

Gabriel Rémond-Tiedrez

2020/21

Contents

1 Introduction

How would you measure the length of the United Kingdom's coastline? Would you take into account all the nooks and crannies? If not, then how would you choose which ones to consider and which ones not to? These questions seem to imply that there is a right answer but there really isn't. Benoit B. Mandelbrot tried to find a solution to this, but instead he created a whole new field of mathematics by focusing on the structure of the object. He did not focus on the general shape, but on how this shape was constructed.

Through iterations of certain functions, we can create shapes with a structure that resembles that of a coastline: jagged and arbitrarily small if we look closely enough, like the Koch snowflake [2][pp. 19-20]. Then the study of such objects will show us how to deal with rough and irregular shapes. Coastlines are far from being the only such object we find in nature: trees, lightning, real snowflakes or clouds are other examples.

Before looking at such functions we must look at their building blocks: contracting functions, and for that we need a crash course in metric spaces.

2 Metric Spaces

First, we need an understanding of distance and how some functions interact with it.

Definition 2.1. A *metric space* is a pair (M, d), where M is a set and $d : M \times M \to [0, \infty)$ is a distance function. Furthermore, d satisfies the following properties:

1. **Non-Degeneracy :** $d(x, y) = 0$ if and only if $x = y$ for all $x, y \in M$,
2. **Symmetry :** $d(x, y) = d(y, x)$ for all $x, y \in M$,
3. **Triangle Inequality :** $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in M$.

The added properties enforced on d are to make it a *metric*, then (M, d) has some structure we can work with.

Example 2.1. A distance function that satisfies all these properties is the usual Euclidian metric in $\mathbb{R}^n$, $d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$.

Definition 2.2. For a set $A \subset M$, the *diameter* of A is the largest possible distance between any 2 of its points. We write

$$\text{diam } A = \sup\{d(x, y) : x, y \in A\}.$$

If this supremum is finite, we say A is bounded, otherwise A is unbounded.

Example 2.2. Let us look at the unit circle $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. We can check that the maximum distance between $z, w \in S^1$ occurs when $z = -w$, then $d(z, w) = 2$ so diam $A = 2$ and A is bounded.

Definition 2.3. A sequence of points [1] $x_n \in M$ is called

1. *Cauchy* if $d(x_n, x_m) \to 0$ as $n, m \to \infty$,
2. *convergent* to $x \in M$ if $d(x_n, x) \to 0$ as $n \to \infty$.

A notion we need to recall is that of sets closed under limits, i.e. $B \subset M$ is closed if for any sequence $b_n \in B$ such that $b_n \to b$ for some $b \in M$, we have $b \in B$.

Using the Triangle Inequality 2.1.3, we can show that every convergent sequence is Cauchy, but now arises the problem that in some metric spaces, certain sequences are Cauchy but do not converge to anything in the space.

Example 2.3. The typical example for such a metric space is the rationals $\mathbb{Q}$, where the sequence $x_n = \frac{\lfloor n\sqrt{2} \rfloor}{n}$ converges to $\sqrt{2} \notin \mathbb{Q}$. Note that our distance function here is $d(x, y) = |x - y|$.

[1] We say points but they could be anything from sets or functions to graphs. "Points" are just the objects in M we defined the distance function for.

Definition 2.4. A metric space (M, d) is *complete* if every Cauchy sequence is convergent.

Let us now look at a class of functions defined by how they interact with distance.

Definition 2.5. A function $f : M \to M$ is called *contracting* or a *contraction* if there exists $\lambda \in (0, 1)$ such that

$$d(f(x), f(y)) \leq \lambda \cdot d(x, y) \text{ for all } x, y \in M.$$

Such a scaling factor λ is definitely not unique, because for any λ we can consider $\frac{\lambda+1}{2} \in (0, 1)$ and it will satisfy the equation as well.

However, some functions not only change distances in the metric space, but do so more consistently.

Definition 2.6. A function $f : M \to M$ is called a *similarity* if there exists $\mu > 0$ such that

$$d(f(x), f(y)) = \mu \cdot d(x, y) \text{ for all } x, y \in M.$$

We call μ the *ratio* of f. In contrast with contracting functions, this ratio is unique and intrinsic to the structure of f.

If we take $\mu \in (0, 1)$ our function is contracting as well, and we can take $\lambda \in [\mu, 1)$.

Example 2.4. Consider the metric space $(\mathbb{R}, d)$ with $d(x, y) = |x - y|$, just as in the rationals. Let f be defined by $f(x) = \frac{1}{2}|x|$. If $xy \geq 0, d(f(x), f(y)) = \frac{1}{2}d(x, y)$ so we would think that this function is a similarity with ratio $\frac{1}{2}$. But if $xy < 0$, for example take $x = -1, y = 1$, then $d(f(-1), f(1)) = d(\frac{1}{2}, \frac{1}{2}) = 0$ which definitely does not respect the apparent ratio. In this case, f is a contraction but not a similarity.

These contractions have quite nice properties, one of them being that they are Lipschitz continuous, and the far more important one being that they have a unique fixed point. This is a crucial notion to understand as it will later on prove the existence of fractal sets, and the proof will show how to find such sets.

Theorem 2.1. *Let f be a contraction on the complete metric space (M, d). Then there exists a unique point $x \in M$ such that $f(x) = x$. We call x the fixed point of f.*

Proof. Pick any point $x_0 \in M$ and construct a sequence of points x_n by defining $x_{n+1} = f(x_n)$. Now let $n, m \in \mathbb{N}$ be integers. Without loss of generality, we can say that $n \geq m$ and write $n = m + k$ for some $k \geq 0$.

$$\begin{aligned}
d(x_n, x_m) &= d(x_{m+k}, x_m) \\
&= d(f(x_{m+k-1}), f(x_{m-1})) \quad \text{now use } f\text{'s contracting property} \\
&\leq \lambda \cdot d(x_{m+k-1}, x_{m-1}) \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad \text{repeat } m-1 \text{ more times} \\
&\leq \lambda^m \cdot d(x_k, x_0)
\end{aligned}$$

Recall that since d is a metric, by definition it satisfies the Triangle Inequality 2.1.3. We can then bound $d(x_k, x_0)$ from above by writing :

$$\begin{aligned}
d(x_k, x_0) &\leq d(x_k, x_{k-1}) + d(x_{k-1}, x_0) \\
&\leq d(x_k, x_{k-1}) + d(x_{k-1}, x_{k-2}) + d(x_{k-2}, x_0) \\
&\vdots \\
&\leq \sum_{i=0}^{k-1} d(x_{i+1}, x_i) \quad \text{now using } f\text{'s contracting property again} \\
&\leq \sum_{i=0}^{k-1} \lambda^i \cdot d(x_1, x_0) = d(x_1, x_0) \cdot \sum_{i=0}^{k-1} \lambda^i = d(x_1, x_0) \cdot \frac{1-\lambda^k}{1-\lambda}.
\end{aligned}$$

Since $\lambda \in (0,1), 1 - \lambda^k \leq 1$ so altogether we obtain

$$d(x_n, x_m) \leq \lambda^m \cdot d(x_k, x_0) \leq \lambda^m \cdot d(x_1, x_0) \cdot \frac{1-\lambda^k}{1-\lambda} \leq \lambda^m \cdot \frac{d(x_1, x_0)}{1-\lambda}.$$

So letting $n, m \to \infty$ we see that $d(x_n, x_m) \to 0$, so x_n is Cauchy. Now recall that we imposed on M to be complete so x_n converges, say to x. As we said f is Lipschitz continuous so $f(x) = f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x$. We therefore have that x is a fixed point of our contraction f.

Now suppose y is another fixed point of f, then $d(x, y) = d(f(x), f(y)) \leq \lambda \cdot d(x, y)$. Since $d(x, y) \geq 0$ and $\lambda < 1$, we must have that $d(x, y) = 0$ so by Non-Degeneracy, $x = y$, we have uniqueness.

$\square$

Another name for such a fixed point is an *attractor* because we can imagine it pulling in every point towards itself, as long as we iterate through f enough times.

3 Hyperspaces and the Hausdorff metric

In order to find fractals, we must first define a metric between sets, then we will see that contractions on specific subsets lead to what we want.

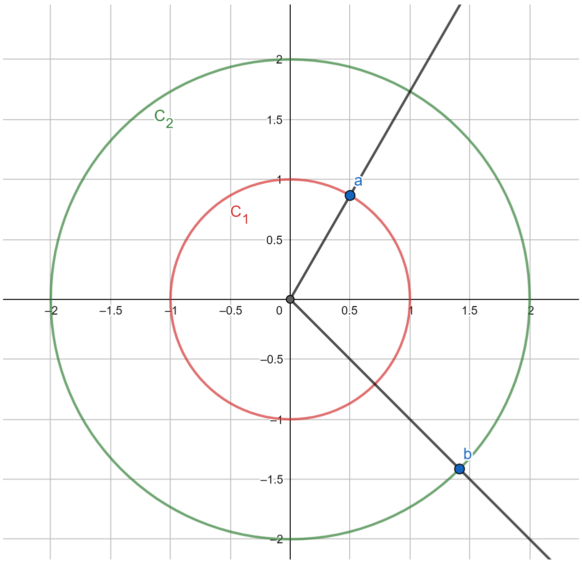


Figure 1: Illustration of Example 3.1

Definition 3.1. Let (M, d) be a metric space with $A, B \subset M$. Then the *Hausdorff distance* between A and B is

$$d_{\mathbb{H}}(A, B) = \inf\{\delta \geq 0 : A \subset \mathcal{N}_\delta(B), B \subset \mathcal{N}_\delta(A)\},$$

where $\mathcal{N}_\delta(A) = \{x \in M :$ there exists $a \in A$ such that $d(x, a) \leq \delta\}$ is the δ-neighbourhood of A. In other words, if $d_{\mathbb{H}}(A, B) = \delta$, every point in A has a point in B that is at distance at most δ, and vice versa.

Example 3.1. Let $C_1 = \{x \in \mathbb{R}^2 : |x| = 1\}$ and $C_2 = \{x \in \mathbb{R}^2 : |x| = 2\}$.

We can see that for any $a \in C_1$, the closest C_2 ever gets is at distance 1. Similarly, for any $b \in C_2$, there always is a point in C_1 at distance 1 but none closer. To see this clearly, we can draw a radius passing through a or through b and obtain a point on the other circle. Therefore $d_{\mathbb{H}}(C_1, C_2) = 1$.

Note that we called this function $d_{\mathbb{H}}$ the Hausdorff distance, not metric. This is because we need to be careful about which sets we plug into the function. Otherwise we might get weird results, as shown in [2, p. 71]. Let us go through some examples in $\mathbb{R}$:

- $d_{\mathbb{H}}(\mathbb{R}, \{0\}) = \infty$. Note that here it is not a limit that goes to infinity but a distance equal to infinity, something that does not make sense with our definition of a metric.

- Let A be any open subset of $\mathbb{R}$ and $\bar{A}$ be its closure. Then $d_{\mathbb{H}}(A, \bar{A}) = 0$ even though the 2 sets are not equal, causing problems for the Non-Degeneracy part of a metric (2.1.1). This problem arises because $d_{\mathbb{H}}$ is defined as an infimum, i.e. $\bar{A} \not\subseteq \mathcal{N}_0(A)$, but $\bar{A} \subset \mathcal{N}_\delta(A)$ for any $\delta > 0$.

- $d_{\mathbb{H}}(\varnothing, \{0\}) = \infty$ as well, because of properties of the infimum.

Now knowing what happens when we do otherwise, we only consider Hausdorff distances between non-empty closed and bounded sets. In fact, we will be more careful and restrict it to compact sets. From now on, whenever we talk about a metric space, we assume that closed and bounded means compact. Note that this is the case in $\mathbb{R}^n$, which is where most of our examples will be.

Definition 3.2. For a metric space (M, d) we define the *hyperspace* of M to be $\mathbb{H}(M) = \{A \subset M : \text{A is non-empty and compact}\}$.

Theorem 3.1. *For a metric space (M, d), the hyperspace $\mathbb{H}(M)$ paired with Hausdorff distance $d_{\mathbb{H}}$ is a metric space.*

Proof. We want to show that $d_{\mathbb{H}}$ is a metric on $\mathbb{H}(M)$.

1. Non-Degeneracy: suppose $A, B \in \mathbb{H}(M)$ and that $d_{\mathbb{H}}(A, B) = 0$. By definition this means $A \subset \mathcal{N}_0(B) = B$ and $B \subset \mathcal{N}_0(A) = A$, so $A = B$.

   Since we are working with sets in $\mathbb{H}(M)$, the infimum in the definition of $d_{\mathbb{H}}$ is a minimum, i.e. it is attained. Using the inclusions with 0 therefore makes sense.

   Showing that $d_{\mathbb{H}}(A, A) = 0$ uses the same argument, just backwards.

2. Symmetry: clearly $d_{\mathbb{H}}(A, B) = d_{\mathbb{H}}(B, A)$.

3. Triangle Inequality: let $A, B, C \in \mathbb{H}(M)$. Define $\delta_1 = d_{\mathbb{H}}(A, B)$ and $\delta_2 = d_{\mathbb{H}}(B, C)$. By definition of Hausdorff distance, $B \subset \mathcal{N}_{\delta_1}(A)$, therefore $\mathcal{N}_{\delta_2}(B) \subset \mathcal{N}_{\delta_2}(\mathcal{N}_{\delta_1}(A))$. Observe that the set on the right hand side is contained within $\mathcal{N}_{\delta_1+\delta_2}(A)$. Again by definition of Hausdorff distance, we know that $C \subset \mathcal{N}_{\delta_2}(B)$. Putting the chain of inclusions together we obtain that $C \subset \mathcal{N}_{\delta_1+\delta_2}(A)$. By the exact same argument, we see that $A \subset \mathcal{N}_{\delta_1+\delta_2}(C)$, so using the property that infimum is a lower bound, $d_{\mathbb{H}}(A, C) \leq \delta_1 + \delta_2$. We therefore have the triangle inequality.

□

Note that by restricting to non-empty bounded sets, we always have finite distance. Consider this as a sanity check that we have a "proper" metric.

Theorem 3.2. *If the metric space (M, d) is complete, then so is $(\mathbb{H}(M), d_{\mathbb{H}})$.*

Proof. In order to prove this, we need to show that any Cauchy sequence in $\mathbb{H}(M)$ is convergent to some point [2] in $\mathbb{H}(M)$.

[2] A point in $\mathbb{H}(M)$ is a non-empty compact subset of M.

Let A_n be a Cauchy sequence in $\mathbb{H}(M)$ and let

$$A = \{a \in M : \text{there exists a sequence } a_k \in A_k \text{ such that } a_k \to a\}.$$

Observe that the sequences we construct take one point a_k in each A_k, and if a sequence converges, then its limit will be in A. We will show that $A_n \to A$.

First, pick $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $n, m \geq N$ implies $d_{\mathbb{H}}(A_n, A_m) \leq \frac{1}{2}\varepsilon$. Now choose and fix n such that $n \geq N$. We want to show that $A_n \subset \mathcal{N}_\varepsilon(A)$ and vice versa, therefore showing that $d(A_n, A) < \varepsilon$.

Pick any $a \in A$, then by construction there exists a sequence $x_k \in A_k$ such that $x_k \to a$. For k large enough, both $d(x_k, a) \leq \frac{1}{2}\varepsilon$ and $d_{\mathbb{H}}(A_k, A_n) \leq \frac{1}{2}\varepsilon$ are satisfied. Therefore, there exists $x \in A_n$ such that $d(x_k, x) \leq \frac{1}{2}\varepsilon$, which means $d(x, a) \leq d(x, x_k) + d(x_k, a) \leq \varepsilon$, so $A \subset \mathcal{N}_\varepsilon(A_n)$.

Now pick any $x \in A_n$. Then choose integers $k_1 < k_2 < \dots$ such that $k_1 = n$ and $d_{\mathbb{H}}(A_{k_j}, A_m) \leq 2^{-j}\varepsilon$ for all $m \geq k_j$. We know these k_j's exist because the A_n's are Cauchy. From here, define a sequence $x_k \in A_k$ in the following way:

- if $k < n$, choose any point in A_k for x_k, it makes no difference,
- if $k = n$, choose $x_n = x$,
- if $k_j < k \leq k_{j+1}$, choose x_k to satisfy $d(x_{k_j}, x_k) \leq 2^{-j}\varepsilon$. These points exist because of how we chose our k_j's.

Then this sequence x_k is Cauchy and because M is complete, x_k converges, say to a. By construction $a \in A$, and we can upper bound $d(x, a) = d(x, \lim_{k\to\infty} x_k) = \lim_{k\to\infty} d(x, x_k) \leq \varepsilon$. We have shown that $A_n \subset \mathcal{N}_\varepsilon(A)$. By choosing $\varepsilon = 1$ we have also proved that A is non-empty.

Putting both inclusions together, we have that $d_{\mathbb{H}}(A_n, A) \leq \varepsilon$, so letting ε go to 0, we see that A_n converges to A.

Recall that for $\mathbb{H}(M)$ to be complete, we need $A \in \mathbb{H}(M)$, i.e. we still have to show A is closed and bounded, therefore compact in our case.

Let us show that A is bounded. Fix $\varepsilon = 1$ then let N be such that $d_{\mathbb{H}}(A_N, A) < 1$. This means exactly that $A \subset \mathcal{N}_1(A_N)$. Observe that since A_N is bounded, so is $\mathcal{N}_1(A_N)$ therefore A must be bounded as well.

Now we show that A is closed. Let $x_n \in A$ be a sequence that converges to some point $x \in M$ such that $d(x_n, x) \leq \frac{1}{n}$. Now let $y_n \in A_n$ be a point such that $d(y_n, x_n) \leq d_{\mathbb{H}}(A_n, A) + \frac{1}{n}$. Then $d(y_n, x) \leq d(y_n, x_n) + d(x_n, x) \leq d_{\mathbb{H}}(A_n, A) + \frac{2}{n}$. The right hand side goes to 0, so $y_n \to x$. Therefore, by definition of A, $x \in A$ so A is closed.

□

The only property of metric spaces that we have not touched yet for hyperspaces are contracting functions, so let us see what they look like.

Definition 3.3. Given a list $f_1, f_2, \ldots, f_n$ of continuous functions in (M, d), an *iterated function system* is the function $F : \mathbb{H}(M) \to \mathbb{H}(M)$ such that for $X \in \mathbb{H}(M)$,

$$F(X) = \bigcup_{i=1}^{n} f_i(X).$$

We will shorten it to i.f.s. for future references. Let us be careful with the nature of X in this equation. On the left-hand side it is a point in $\mathbb{H}(M)$ and on the right-hand side it is a subset of M. We make the difference here because the f_i's are defined on M, not for points in $\mathbb{H}(M)$.

We should also be worried about if F actually maps $\mathbb{H}(M)$ to itself. Recall we assumed closed and bounded meant compact, therefore since continuous images of compact sets are compact, F does have co-domain $\mathbb{H}(M)$.

An i.f.s. is a function that takes $X \subset M$, copies it n times, applies each f_i to each copy and outputs the union of all the images.

Theorem 3.3. *Given a list of contracting functions $f_1, f_2, \ldots, f_n$ on M, the i.f.s. F obtained from the f_i's is contracting in $\mathbb{H}(M)$.*

Proof. Let $X, Y \in \mathbb{H}(M)$ with $d_{\mathbb{H}}(X, Y) = \delta$. We want to show that $d_{\mathbb{H}}(F(X), F(Y)) \leq \lambda\delta$ for some $\lambda \in (0, 1)$.

Now let z be any point in $F(X)$, then $z = f_i(x)$ for some $i \in \{1, 2, \ldots, n\}$ and some $x \in X$. From the definition of Hausdorff distance, there exists $y \in Y$ such that $d(x, y) \leq \delta$. Then apply f_i to obtain that $d(f_i(x), f_i(y)) \leq \lambda_i\delta$, where λ_i is a scaling factor for f_i. In fact, if we let $\lambda = \max\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, we can see that $d(z, f_i(y)) \leq \lambda\delta$. Since this holds for any $z \in F(X)$, we have that $F(X) \subset \mathcal{N}_{\lambda\delta}(F(Y))$. By the same argument we get the inclusion $F(Y) \subset \mathcal{N}_{\lambda\delta}(F(X))$, which means that

$$d_{\mathbb{H}}(F(X), F(Y)) \leq \lambda\delta = \lambda d_{\mathbb{H}}(X, Y),$$

i.e. F is contracting in $\mathbb{H}(M)$.

□

Note that we could have $n = 1$ or have n copies of the same function in an i.f.s. but that would mean we need to study contracting functions one at a time, which we have already done. So we should look at $n \geq 2$ with all functions distinct. Moreover, i.f.s.'s with $n \geq 2$ have fixed points that start getting interesting.

Definition 3.4. For a given contracting i.f.s. F, we call its fixed point X a *homogeneous self-similar fractal set.*

Using the machinery we built, we know that such objects exist.

Let us think about what these fractals really are. F is a function that takes a set, copies it n times, shrinks the copies down in a certain way and maybe moves the smaller images around and rotates them. X is a set that is invariant to all those changes, a set that has a structure such that it remains unchanged by F. We say "structure" and not "shape" because the shape matters very little, but the structure of X is what makes it stand out for F.

In most cases we will work with, the functions considered will not only be contractions, but will be similarities. This will make dealing with i.f.s.'s easier because we can focus on the unique ratio of each function instead of just their contracting property.

Example 3.2. Let our metric space be $(\mathbb{R}^2, d)$, where d is the usual Euclidian metric (example 2.1). We will make the Sierpinski carpet so we will need 8 functions, which seems like a lot to deal with, but focus on the structure and the functions will be self-explanatory.

Let

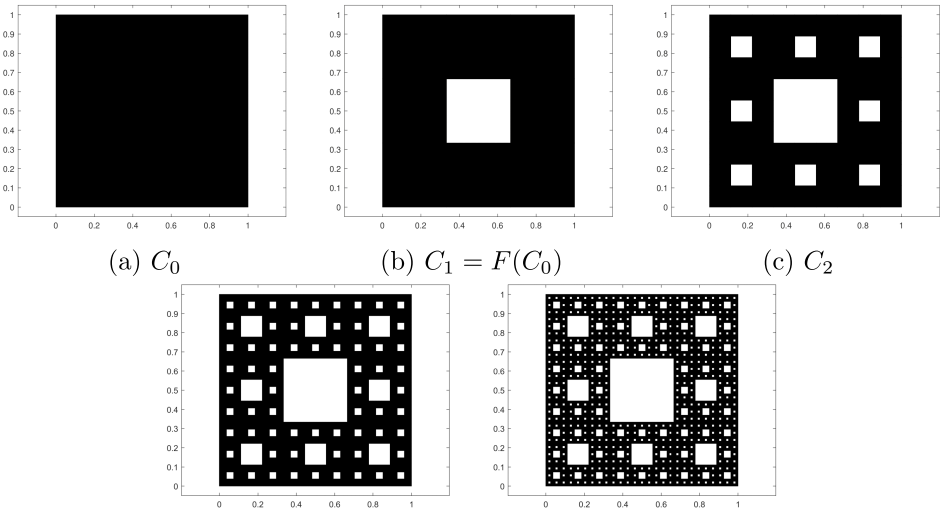
$$
\begin{array}{lll}
f_7(x,y) = \frac{1}{3}(x, y+2), & f_6(x,y) = \frac{1}{3}(x+1, y+2), & f_5(x,y) = \frac{1}{3}(x+2, y+2), \\[2ex]
f_8(x,y) = \frac{1}{3}(x, y+1), & & f_4(x,y) = \frac{1}{3}(x+2, y+1), \\[2ex]
f_1(x,y) = \frac{1}{3}(x, y), & f_2(x,y) = \frac{1}{3}(x+1, y), & f_3(x,y) = \frac{1}{3}(x+2, y).
\end{array}
$$

Letting F be the i.f.s. corresponding to this list, what does F do exactly? It copies the input 8 times, shrinks it down by a factor of a third and translates the output to various points. Define the sequence $C_n \in \mathbb{H}(M)$ such that $C_{n+1} = F(C_n)$ with $C_0 = [0,1]^2$ being a unit square. Because all the f_i's are contractions, the i.f.s. is contracting by Theorem 3.3. We can then apply the fixed point Theorem 2.1 and the hyperspace completeness Theorem 3.2 to see that the limit of C_n should be the fixed point of F, a self-similar fractal. Let us see what a couple of iterations look like (figure 2). Then we call the limit of this sequence the Sierpinski Carpet, a set of which we now know the rough shape and that we will explore in further detail in a later section of this essay.

Definition 3.5. For a given i.f.s. F such that $F(X) = \bigcup_{i=1}^{n} f_i(X)$ for some similarities f_i, the *ratio list* of F is defined to be $(\mu_1, \mu_2, \ldots, \mu_n)$, where μ_i is the ratio of f_i.

In example 3.2, the i.f.s. for the Sierpinski carpet has ratio list $(\frac{1}{3}, \ldots, \frac{1}{3})$ with the same ratio 8 times. But an i.f.s. could very well have different ratios in the ratio list.

We then call ratio lists that have $\mu_i < 1$ for all i *contracting.*



(a) C_0 (b) $C_1 = F(C_0)$ (c) C_2

(d) C_3 (e) C_4

Figure 2: Sierpinski Carpet approximations

4 Dimensions

Definition 4.1. The *similarity value* of a contracting ratio list $(\mu_1, \ldots, \mu_n)$ is the unique positive number s such that $\sum_{i=1}^{n} \mu_i^s = 1$.

A very slick proof that s exists and is unique uses the Intermediate Value Theorem [2, pp. 117-118].

An intuition behind this is that s is an indicator of how the union of each shrunk down image adds up back to the fractal.

Definition 4.2. If a set X is the fixed point of an i.f.s. with ratio list having similarity value s, then we say s is the *similarity dimension* of X.

Example 4.1. Suppose we have an i.f.s. F_1 with ratio list $(\frac{1}{n}, \ldots, \frac{1}{n})$, with the same ratio n times, then the fixed point X_1 of F_1 would have similarity dimension s_1, where s_1 solves

$$n\left(\frac{1}{n}\right)^{s_1} = 1,$$

which we can see only works if $s_1 = 1$, so X_1 has similarity dimension 1. I invite you to think about an i.f.s. that would have such a ratio list and what its fixed point could be.

Now consider F_2 as being the i.f.s. for the Sierpinski carpet, as seen in Example 3.2. Then F_2 has similarity value s_2, where s_2 solves

$$8\left(\frac{1}{3}\right)^{s_2} = 1,$$

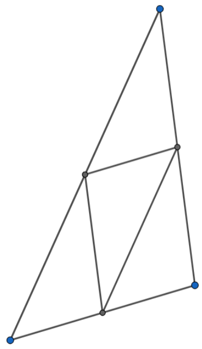


Figure 3: Construction to calculate similarity dimension of a triangle

which gives $s_2 = \frac{\log 8}{\log 3}$, approximately 1.89279.

Can we compute the similarity dimension of a triangle? What we need to do is find an i.f.s. that has this triangle as its fixed point. We can do this if we cut the triangle in 4 equal smaller triangles, each of which is congruent to the original one but smaller by a factor of $\frac{1}{2}$ (figure 3). Then the i.f.s. would have ratio list $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and from this we can see that the similarity dimension of any triangle is 2.

Another shape we could expect to have dimension 2 is a disk. But in trying to compute the similarity dimension of a disk, we need to replace it with smaller copies of itself, which is impossible. We therefore cannot compute the similarity dimension of a disk, an elementary shape.

This motivates another attempt to find a dimension that encapsulates the fractal nature of some sets.

Let X be some object lying in a metric space M. We want to calculate the similarity dimension of X but it has no self-similarity whatsoever. We therefore want to find some sort of dimension that does not require an i.f.s. . Suppose we want to check if s is the dimension of X, whatever that means.

The first attempt will be to try and cover X with sets. By cover we mean a collection of sets $\{C_i\}$ such that $X \subset \bigcup_i C_i$. Then we enforce on these sets to satisfy diam $C_i \leq \delta$. From this, calculate $\inf \sum_i (\text{diam } C_i)^s$, where the infimum is taken over all covers of X.

By minimizing this sum, we try to limit the amount of space that $\bigcup_i C_i \setminus X$ takes, i.e. we want the cover to fit X as much as possible without it exceeding the boundary of X. A question we could ask ourselves is why we take diameters to the s-th power. Think about a simple 2-dimensional plane. To calculate the area of a square of sidelength c, we raise c to the 2^{nd} power. It seems to make sense to raise distances to the power of the dimension.

Now we let δ go to 0 to have the most precise cover we can. Note that now, if we increase s, we decrease the sum because we raise numbers tending to 0 to a higher power, making them even smaller. Similarly if we decrease s, we increase the sum. Then a problem that happens is that the sum will start misbehaving: it will very easily blow up to infinity or collapse to 0 if we take s too small or too large respectively [3][p. 7]. We can analyse this carefully to find the threshold where the sum goes from infinite to null [2][p. 168]. Then for this unique value s_0, the sum will be non-zero and finite. We interpret this as meaning s_0 is the right dimension to take when attempting to measure X. We call s_0 the *Hausdorff dimension* of X [3].

The second option we can consider is trying to pack X with sets instead. By packing we mean a collection of sets $\{P_i\}$ such that $\bigcup_i P_i \subset X$. Again, we restrict these packings to those that satisfy diam $P_i \leq \delta$ for some $\delta > 0$. The other difference with Hausdorff dimension is that we calculate $\sup \sum_i (\text{diam } P_i)^s$.

The reason we maximize the sum this time is because we want the packing to fill X as much as we can, so we must try to limit the amount of space that $X \setminus \bigcup_i P_i$ takes. By doing this we fill X with the most precision possible and try to leave no space left in X that isn't also in a P_i.

Again make δ go to 0 and observe how the sum misbehaves depending on s. Exactly in the same way, there will be a unique s_0 such that the sum is non-zero and finite. We call this number the *packing dimension* of X.

If X is nice enough, the similarity, Hausdorff and packing dimensions are one and the same, hinting at the fact that this s_0 is intrinsic to the object rather than dependent on the method used to calculate it.

The one downside of the Hausdorff and packing dimension is calculating them. Since we take suprema, infima and limits, they seem tough to compute. However, it is possible to approximate the Hausdorff dimension using a grid system.

Suppose we want to calculate the approximate Hausdorff dimension of an object $X \subset \mathbb{R}^n$. We first place it in an n-dimensional grid filled with hypercubes with sidelength δ. Let $N(\delta)$ be the number of hypercubes that X touches, then we look at how $N(\delta)$ changes when δ varies, most importantly when δ goes to 0. Then just by collecting some data and putting it in a graph, we can easily calculate what we call the *box-counting* dimension [4] .

The main idea is that instead of considering the most efficient covering, we prioritise hypercubes because of how simple they are to work with. Note that the hypercubes that X touches form a cover of X, so this process agrees with how we calculate the Hausdorff dimension.

In [5], Mahieu evaluates the box-counting dimension of the UK coastline to be 1.26,and figure 4 shows what the construction looks like.

[3]There are some objects that do not have such a value, in which case we say that the dimension is 0 or infinite.

[4]This is similar to how Kirillov defines Hausdorff dimension in [4][pp. 15-16], up to changing hypercubes to balls of radius δ.

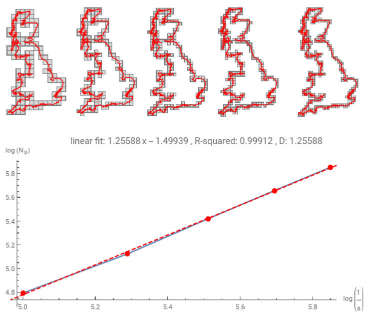


Figure 4: Box-counting dimension of the UK coastline

5 Two Examples

5.1 The Sierpinski Carpet

Recall from example 3.2 that the Sierpinski Carpet is defined as the limit of the sequence C_n, where C_0 is a unit square.

Now suppose that we take a different starting set, let us start with a rectangle instead of a unit square. Furthermore, let us move this rectangle away from the origin. Let $R_0 = [5, 7] \times [9, 12]$ be our first term of the sequence R_n, where $R_{n+1} = F(R_n)$.

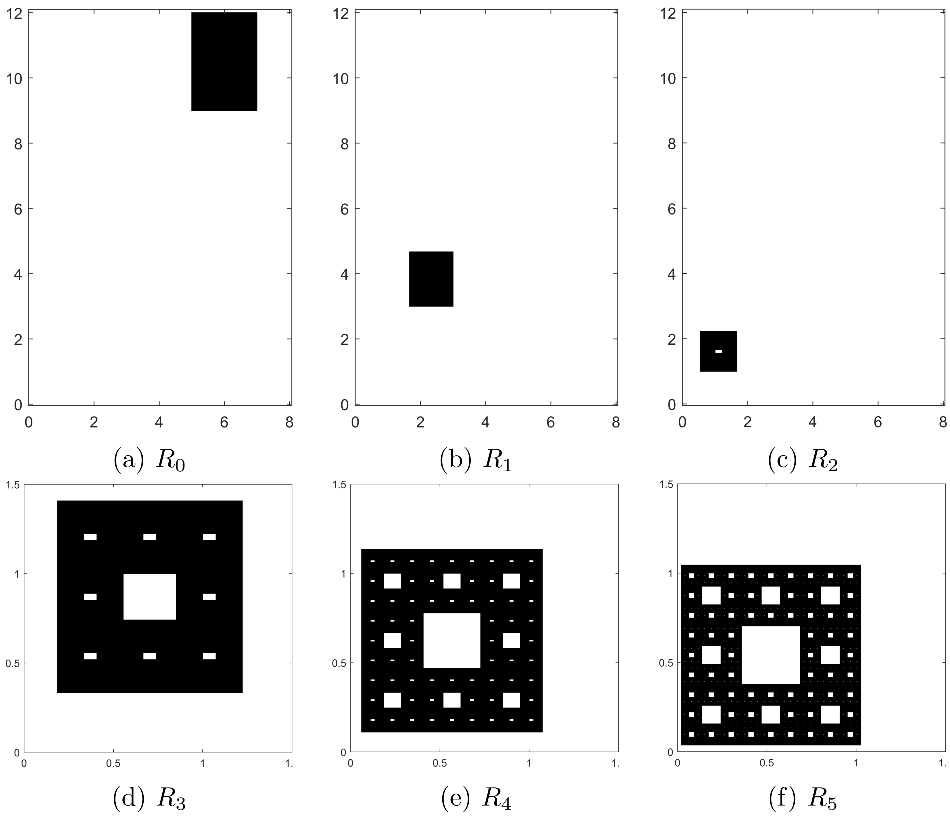
In figures 5a, 5b and 5c the axes are identical, which helps see what the i.f.s. F really does initially, i.e. shrinks the rectangle down and moves it towards the origin. Quickly enough, 3 iterations in, we already see the Sierpinski Carpet structure appear. The more iterations we do, the more it looks like the Sierpinski Carpet and the closer it gets to where it should be – in the $[0, 1]^2$ square.

Recall theorem 2.1, we showed that when iterating a contracting function at any point infinitely many times, it would lead to the function's fixed point. We observe that this holds for i.f.s.'s just as much and that the term *attractor* still makes sense: the Sierpinski Carpet pulls in any – non-empty, closed and bounded – set to itself and forces its structure onto them.

Another starting point we could take is an equilateral triangle, which gives something that looks exactly like the Sierpinski Carpet after just 3 iterations – figure 6.

Since the first term of the sequence does not influence the resulting limit, let us take the easiest option and start with the unit square C_0 as in example 3.2. We could interpret the i.f.s. acting on C_0 as punching precise holes out of it, then on C_1 and C_2 and so on. Let us calculate the area of the Sierpinski Carpet, that is, the area of C_0 minus all the area that we remove when iterating.

Note that C_1 is composed of 8 shrunk down copies of C_0, that is, 8 squares with sidelength $\frac{1}{3}$. Similarly, C_2 is composed of 8 small C_1's, so 8^2



(a) R_0 (b) R_1 (c) R_2

(d) R_3 (e) R_4 (f) R_5

Figure 5: First iterations of the Sierpinski Carpet i.f.s. on a rectangle

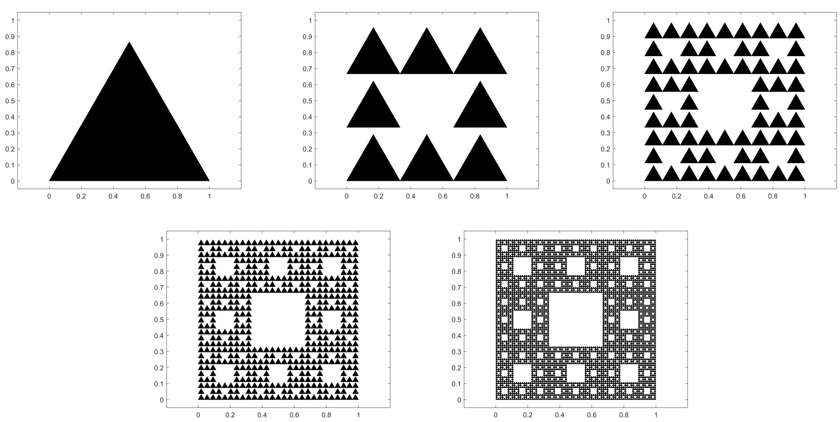


Figure 6: Sierpinski Carpet starting with an equilateral triangle

squares with sidelength $\frac{1}{3^2}$. We can keep going to see that C_n is composed of 8^n small squares with sidelength $\frac{1}{3^n}$.

Let d_n be the area we "remove" when putting C_n through F. We see that $d_0 = \frac{1}{3^2}$ and from there that $d_1 = 8 \cdot \frac{1}{3^4}$. In general, we have $d_n = 8^n \cdot \left(\frac{1}{3^{n+1}}\right)^2 = \frac{1}{9} \cdot \left(\frac{8}{9}\right)^n$. Then the area of the Sierpinski Carpet is the area of C_0 minus all the d_n's: Area $= 1 - \sum_{n=0}^{\infty} d_n = 1 - \frac{1}{9} \cdot \frac{1}{1-\frac{8}{9}} = 1 - 1 = 0$.

So the Sierpinski Carpet has no area. Another interesting property is that it has infinite perimeter, which we will not bother with here as it is very similar to the area calculation.

5.2 A Fractal Tree

A perfect example of fractal geometry in nature is a tree. Let us decompose the structure: start with a trunk and add some branches to it, each of which has smaller branches, which again have even smaller branches. Of course, the branches do stop at some point for real trees, but using our knowledge of fractals, let us make a model of a mathematically perfect tree. [5]

Let us start with a straight line going from $x_0 = (0,0)$ to $y = (0,1)$. This will be our trunk and we will call this T_0. We now need to choose the points where the branches will grow out of, let us call them $x_1, x_2, \dots$ with x_1 being the first from the bottom and the others going up after that. Pick a parameter $r \in (0,1)$. Then let $x_1 = x_0 + r \cdot (y - x_0)$. Note that since $x_0 = (0,0)$, we could not take it into account, but writing it that way makes it clearer what we want to do, i.e. we start at x_0 and go distance $r \cdot \|y - x_0\|$ in direction $y - x_0$. Then we define $x_{i+1} = x_i + r \cdot (y - x_i)$, just like x_1.

Think of this as ignoring the bottom part of the tree and just considering it from x_i upwards, then doing the same construction we did to get x_1, by replacing x_0 by x_i. Figure 7 shows the first 5 x_i's with $r = 0.3$.

Now that we know where our branches start from, we need to think about how we want our branches to look. We choose for them to be straight, in fact we actually want them to be exactly like the trunk, only smaller. Let us call b_i the branch growing from x_i. We need to have that branch at a certain angle with the trunk, call it θ. In fact, we make the branches alternate sides, i.e. b_1 will be on the right and b_2 on the left, and so on.

What about their length? An option could be to take b_i to have length $\|x_i - x_{i-1}\|$, i.e. the same length as the part of the trunk we removed. However, after some experimenting we observe that the branches get quite small if we do this, so we add a final parameter α: the ratio between the length of b_i and $\|x_i - x_{i-1}\|$. We can then make α larger than 1 and have longer branches.

Figure 8 shows how the parameters influence the shape of our tree.

[5] The approach we will be using here is very similar to the one Bonenberg started with in [1].

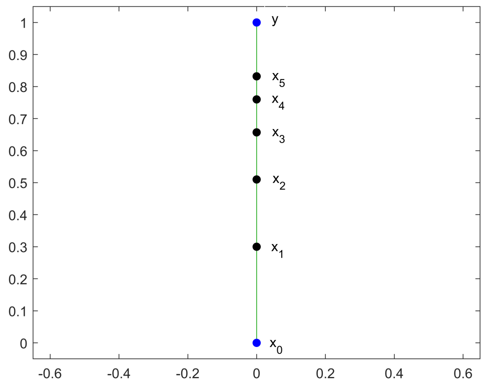
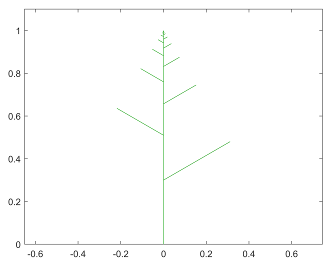
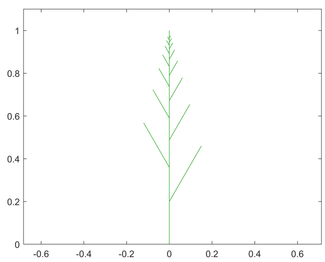


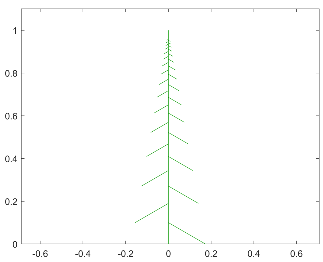
Figure 7: T_0 with the first 5 x_i's and $r = 0.3$



(a) $r = 0.3$, $\theta = \frac{\pi}{3}$ and $\alpha = 1.2$

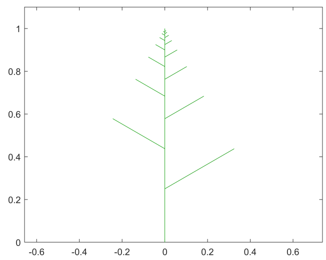


(b) $r = 0.2$, $\theta = \frac{\pi}{6}$ and $\alpha = 1.5$

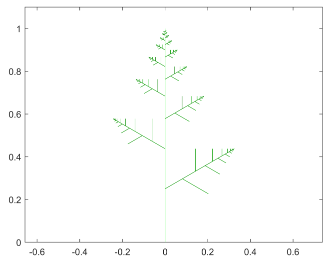


(c) $r = 0.1$, $\theta = \frac{2\pi}{3}$ and $\alpha = 1.2$

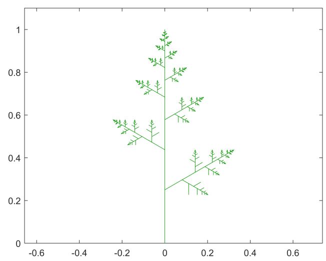
Figure 8: Branch construction for 3 different triplets of parameters



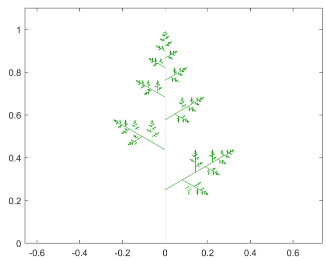
(a) T_1



(b) T_2



(c) T_3



(d) T_4

Figure 9: First iterations for $r = 0.25$, $\theta = \frac{\pi}{3}$ and $\alpha = 1.5$

Recall that to have a self-similar fractal, we need a contracting i.f.s. and the fractal will be its fixed point. Here, a contracting function in the i.f.s. should take the initial tree – just the trunk – contract, rotate and translate it to output a smaller version of itself, i.e. a branch. Call f_i the function that takes T_0 to branch b_i [6].

Let F be the i.f.s. composed by the f_i's, and just like with the Sierpinski Carpet, define a sequence T_n by $T_{n+1} = F(T_n)$. Figure 9 shows the first 4 iterations we obtain for some triplet of parameters.

Now we quite naturally want to know what the dimension of this tree is. We will calculate the similarity dimension and use the nice self-similarity properties the tree has. We note that θ has no importance in this calculation, but rather we need to focus on r and α.

A bit of algebra shows that the ratio of f_i is $\mu_i = \alpha r(1-r)^{i-1}$, so in order to find the similarity dimension of the tree we need s_0 such that $\phi(s_0) = 1$

[6] Since our x_i's go on forever, the b_i's do too, therefore so do the functions f_i's. This could be a problem because we have not yet had to deal with i.f.s.'s that have infinitely many functions composing them. For the figures we only calculate a given number of branches, but we can easily imagine the repeating pattern if we zoomed in at the top of the tree.

for $\phi(s)$ defined as follows:

$$\begin{aligned}\phi(s) =& \sum_{i=1}^{\infty}(\alpha r(1-r)^{i-1})^s \\ =&(\alpha r)^s \sum_{i=0}^{\infty}((1-r)^s)^i \\ =&(\alpha r)^s \frac{1}{1-(1-r)^s}\end{aligned}$$

We want $\phi(s_0) = 1$ so we require $(\alpha r)^{s_0} = 1-(1-r)^{s_0}$, a relatively simple equation, which we can prove has a solution, given $\alpha r < 1$ [7]. However finding a general solution is tricky.

This is disappointing but it comes to show how careful one needs to be when dealing with fractals. Note that for any given pair α and r, we can approximate the dimension using the Intermediate Value Theorem. That way we can estimate the similarity dimension of the tree in Figure 9 to be 1.23.

6 Conclusion

Let us go back to the questions we asked at the start. Could we now measure a given coastline? No, but we now understand why this was a trick question to begin with. Just like the Sierpinski Carpet has infinite perimeter when we consider every tiny side, we can evaluate a coastline's length to be infinite.

The reason this happens is because coastlines are rough, irregular and crinkly shapes. One thing that we can do is measure just how rough they are, using our new dimensions.

Let us focus on box-counting dimension. We can evaluate the dimension of the UK and Norwegian coastlines to be 1.25 and 1.52 respectively. Simply put, a higher dimension means more nooks and crannies, i.e. more roughness and irregularities.

The tree we made is an example of all the possibilities fractals contain. We can simulate trees in 3 dimensions [1]. We can simulate a real-looking mountain with all the irregularities we want, as long as we iterate the right object through the right function [7]. We can analyse the fractal nature of organs and look for imperfections out of the ordinary in order to prevent diseases [6]. We can use approximations of space-filling curves like the Hilbert Curve in antennas to catch all sorts of frequencies [8]. The list goes on.

[7] If $\alpha r \geq 1$, it would mean the first branch would be bigger than or have the same size as the tree, meaning the i.f.s. is not contracting.

References

[1] Karol Bonenberg. Real time design of fractal trees using oppenheimer algorithm, Mar 2010. URL https://github.com/bonenberg/Recursive-Oppenheimer-Tree.

[2] Gerald Edgar. *Measure, Topology, and Fractal Geometry.* Springer Science & Business Media, 2nd edition, 2008. ISBN 978-0-387-74748-4.

[3] Kenneth J. Falconer. *The Geometry of Fractal Sets*, volume 85. Cambridge University Press, 1986. ISBN 0-521-33705-4.

[4] Aleksandr Aleksandrovich Kirillov. *A Tale of Two Fractals.* Springer, 2013. ISBN 978-0-8176-8381-8.

[5] Erik Mahieu. Box-counting the dimension of coastlines, Jan 2014. URL https://demonstrations.wolfram.com/BoxCountingTheDimensionOfCoastlines/.

[6] Hannah V Meyer, Timothy JW Dawes, Marta Serrani, Wenjia Bai, Paweł Tokarczuk, Jiashen Cai, Antonio de Marvao, Albert Henry, R Thomas Lumbers, Jakob Gierten, et al. Genetic and functional insights into the fractal structure of the heart. *Nature*, 584(7822):589–594, 2020.

[7] Mohsen Sharifi, Fatemeh Hashemi Golpaygani, and Mehdi Esmaeli. A new fractal-based approach for 3d visualization of mountains in vrml standard. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 100–105, 2004.

[8] Douglas H Werner and Suman Ganguly. An overview of fractal antenna engineering research. *IEEE Antennas and propagation Magazine*, 45(1): 38–57, 2003.