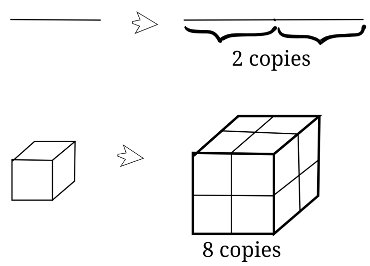
Final Paper: the Hausdorff Measure

source: Stein and Shakarchi, Chapter 7

1 Motivation

In this paper, we will introduce the concept of the Hausdoff measure. The Hausdorff measure uses a volumetric concept similar to the natural Lebesgue measure, but does so in a way that formalizes the notion of dimension. Specifically, we have an innate idea of positive integer dimensions - a line has one dimension, a square has two dimensions, a cube has three dimensions, and so on. Intuitively, we can think of this dimension as a description of the way a set scales - if we have a n-dimensional set, then we expect that if we scale each point in the set by λ, we create λ^n disjoint copies of the original set. For example, if we have a line, scaling the line by a factor of 2 will create 2 disjoint copies of the line, while scaling a cube by a factor of 2 will create 8 disjoint copies.



With this definition of measure, we can define objects to be n-dimensional where n is not a natural number, but instead a nonnegative real number, which gives us further tools to analyze sets which are not interesting in integer dimensions.

Specifically, we will find that sets like the Cantor set, which have measure 0 in one dimension, have positive finite measure in a fractional dimension. With the intuitive definition of measure above, we can see that if we take each point of the Cantor set to triple its original location, then after the first step we are creating Cantor sets in $[0, 1]$ and $[2, 3]$, so scaling the Cantor set by a factor of 3 creates two copies of the set, and by our intuitive definition above, this means that if n is the dimension of the Cantor set, $3^n = 2$, so $n = \log_3 2$. We will formalize this notion later in the paper.

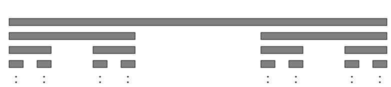


Figure 1: the original Cantor set

Figure 2: the Cantor set scaled by a factor of three

Moreover, we can use the Hausdorff measure to analyze fractals. We know that many fractals are uninteresting in the Lebesgue measure - for example, the Von Koch curve which is pictured in Figure 1 and which we will formally define later, has infinitely many wrinkles, so that it has infinite measure when considered using the one-dimensional Lebesgue measure (one can imagine stretching it out over the real line, and see that it would stretch out infinitely since each step of the fractal creation adds length to the curve). However, it can be covered by arbitrarily thin rectangles in two dimensions, and so its two-dimensional Lebesgue measure is zero. Thus, the Lebesgue measure, or analysis in integer dimensions, does not give us much interesting information about the Von Koch curve, but using the Hausdorff measure, we will find that it actually has positive finite measure in $\log_2 3$ dimensions.

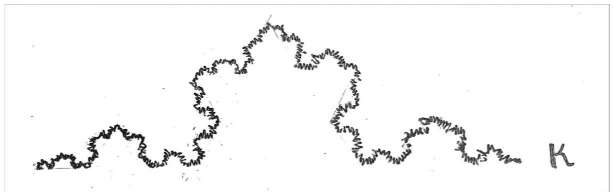


Figure 3: the Von Koch curve

2 Introduction

2.1 Defining the Hausdorff Measure

The Hausdorff measure is not defined as a singular measure function, but rather as a series of measures $m_\alpha(E)$, where α is the dimension and can be any positive real number. In an intuitive sense, $m_\alpha(E)$ is computed by looking at the minimal covering of E using α-dimensional sets.
Moreover, this means that $m_\alpha(E)$ should be zero when α is bigger than the dimension of E and $m_\alpha(E)$ should be infinite when α is smaller than the dimension of E; for example, we can see that when we cover a unit square with cubes, it can be covered by arbitrarily thin cubes, so it should have zero measure in three dimensions, and we can see that it is significantly "bigger" than any finite line, so it should have infinite measure in one dimension. However, we can see that a unit square should have finite measure in two dimensions, because it can be covered by squares whose total volume isn't too small.

Definition 2.1. More formally, we define the **exterior α-dimensional Hausdorff measure** of sets $E \in \mathbb{R}^d$ to be

$$m_\alpha^*(E) = \lim_{\delta\to 0} \inf \left\{ \sum_k (\operatorname{diam} F_k)^\alpha : E \subset \bigcup_{k=1}^\infty F_k, \operatorname{diam} F_k \leq \delta \text{ for all } k \right\},$$

where $\operatorname{diam} S$ is the **diameter** of the set S, or $\sup \big\{|x-y| : x, y \in S\big\}$.

That is, for each set $E \in \mathbb{R}^d$, the exterior α-dimensional Hausdorff measure of E is computed by taking, for each δ, possible countable coverings of E by sets whose diameter is at most δ, and then computing $\sum_k (\operatorname{diam} F_k)^\alpha$, which makes sense intuitively if we consider $(\operatorname{diam} F_k)^\alpha$ to be approximately the α-dimensional mass of F_k. As usual, we take the infimum over all possible coverings, but then we also take the limit as $\delta \to 0$. We will see soon that the requirement that the sets are arbitrarily small, which is not necessary in the definition of the Lebesgue measure, is necessary here for additivity for positively disjoint sets to hold in the Hausdorff measure.

Since we want the exterior measure to exist for all E, we must check that this limit always exists; we can see that the quantity

$$\mathcal{H}_\alpha^\delta(E) = \inf \left\{ \sum_k (\operatorname{diam} F_k)^\alpha : E \subset \bigcup_{k=1}^\infty F_k, \operatorname{diam} F_k \leq \delta \text{ for all } k \right\}$$

increases as δ decreases, so

$$m_\alpha^*(E) = \lim_{\delta\to 0} \mathcal{H}_\alpha^\delta(E)$$

always exists, though it may sometimes be infinite.

We can see also that using $(\operatorname{diam} F_k)^\alpha$ as our α-dimensional mass of F_k makes sense with the intuitive definition of dimension we discussed in the motivation; since scaling a set F_k by a factor of λ would scale its diameter by a factor of λ, $(\operatorname{diam} F_k)^\alpha$ would be scaled by a factor of $(\lambda)^\alpha$, which is what we wanted.

2.2 Properties of the Hausdorff Measure

We will begin by showing that the exterior α-dimensional Hausdorff measure is a metric Carathéodory exterior measure, by showing that it has the relevant properties. This is an application of Theorem 1.2 from Chapter 6 of the textbook, which is explained in Appendix A.

Property 1 (Monotonicity)**.** If $E_1 \subset E_2$ then $m_\alpha^*(E_1) \leq m_\alpha^*(E_2)$.

This follows directly from the fact that we are using the infimum of coverings of the sets, and any covering of E_2 will also be a covering of E_1.

Property 2 (Sub-additivity)**.** For any countable family $\{E_j\}$ of sets in $\mathbb{R}^d$, $m_\alpha^*\left(\bigcup_{j=1}^\infty E_j\right) \leq \sum_{j=1}^\infty m_\alpha^*(E_j)$.

Proof. This is very similar to the proof of sub-additivity for the exterior Lebesgue measure; for any δ and any $\epsilon > 0$, we know by the definition of infimum that for each j we can pick a covering $\{F_{j,k}\}_{k=1}^\infty$ of E_j such that for all k, $\operatorname{diam} F_k \leq \delta$ and $\sum_k (\operatorname{diam} F_{j,k})^\alpha < m_\alpha^*(E_j) + \epsilon/2^j$. Then, $\{F_{j,k}\}$ is a covering of $\bigcup_j E_j$, where $\operatorname{diam} F_{j,k} \leq \delta$ for each j, k, so

$$\mathcal{H}_\alpha^\delta\left(\bigcup_j E_j\right) \leq \sum_{j,k} (\operatorname{diam} F_{j,k})^\alpha \leq \sum_j (m_\alpha^*(E_j) + \epsilon/2^j) \leq \epsilon + \sum_j m_\alpha^*(E_j).$$

Since ϵ is arbitrary, we get that $\mathcal{H}_\alpha^\delta \leq \sum_j m_\alpha^*(E_j)$, and since this is true for every δ, we can take the limit as $\delta \to 0$ to get that $m_\alpha^*\left(\bigcup_{j=1}^\infty E_j\right) \leq \sum_{j=1}^\infty m_\alpha^*(E_j)$. □

Property 3 (Additivity for Positively Separated Sets)**.** If $d(E_1, E_2) > 0$, then $m_\alpha^*(E_1 \cup E_2) = m_\alpha^*(E_1) + m_\alpha^*(E_2)$.

Proof. We can see that $m_\alpha^*(E_1 \cup E_2) \leq m_\alpha^*(E_1) + m_\alpha^*(E_2)$ follows directly from countable sub-additivity, so we just need to prove the other direction.

To prove that $m_\alpha^*(E_1 \cup E_2) \geq m_\alpha^*(E_1) + m_\alpha^*(E_2)$, we can see that for any $\delta > 0$ where $\delta < d(E_1, E_2)$, for every covering $\{F_k\}$ of $E_1 \cup E_2$ where each $\operatorname{diam} F_k \leq \delta$, we can define $\{F'_k\} \supset E_1$ and $\{F''_k\} \supset E_2$ by

$$F'_k = F_k \cap E_1, \qquad F''_k = F_k \cap E_2$$

for each k. Since $\operatorname{diam} F_k < d(E_1, E_2)$, for each k one of F'_k, F''_k must be the empty set, so

$$(\operatorname{diam} F'_k)^\alpha + (\operatorname{diam} F''_k)^\alpha \leq (\operatorname{diam} F_k)^\alpha,$$

and therefore

$$\sum_k (\operatorname{diam} F'_k)^\alpha + \sum_k (\operatorname{diam} F''_k)^\alpha \leq \sum_k (\operatorname{diam} F_k)^\alpha.$$

Since this is true for every such covering F_k, and $\operatorname{diam} F'_k \leq \delta$ and $\operatorname{diam} F''_k \leq \delta$ for each k, we get that

$$\mathcal{H}_\alpha^\delta(E_1) + \mathcal{H}_\alpha^\delta(E_2) \leq \mathcal{H}_\alpha^\delta(E_1 \cup E_2),$$

and taking the limit as δ approaches 0 gives us that $m_\alpha^*(E_1 \cup E_2) \geq m_\alpha^*(E_1) + m_\alpha^*(E_2)$. □

From Property 1, Property 2, and the fact that the empty set has measure zero, we can see that the exterior α-dimensional Hausdorff measure is a Carathéodory exterior measure, and from Property 3 it is a metric Carthéodory exterior measure, so it is a countably additive measure when restricted to the Borel sets. From this point onwards, we shall discuss only Borel sets, and write $m_\alpha(E)$ instead of $m_\alpha^*(E)$.

Definition 2.2. For Borel sets $E \in \mathbb{R}^d$, the measure $m_\alpha(E)$ is called the **α-dimensional Hausdorff measure** of E.

Property 4 (Countable Additivity)**.** If $\{E_j\}$ are countably many disjoint Borel sets, then

$$m_\alpha\left(\bigcup_j E_j\right) = \sum_j m_\alpha(E_j).$$

(This is simply a restatement of what we said above.)

Our next property discusses how the Hausdorff measure behaves under rotations, translations, and scaling. This should make sense, because we based the Hausdorff measure on the intuitive notion of how an α-dimensional object should scale.

Property 5. The Hausdorff measure is invariant under rotations and translations;

$$m_\alpha(E+h) = m_\alpha(E) \qquad \text{for } h \in \mathbb{R}^d$$

and

$$m_\alpha(rE) = m_\alpha(E)$$

where r is a rotation. Moreover, it scales as follows:

$$m_\alpha(\lambda E) = \lambda^\alpha m_\alpha(E) \qquad \text{for } \lambda > 0.$$

Intuitively, as the Hausdorff measure is dependent on diameters of sets, the first two properties follow from the fact that the diameter of a set is invariant under translation and rotation. Moreover, the last property follows from the fact that $\operatorname{diam}(\lambda S) = \lambda \operatorname{diam}(S)$ for all sets S, and then we take the diameter to the α power when computing the Hausdorff measure. A formalization of this idea is left as an exercise to the reader.

Finally, the remaining properties of the Hausdorff measure help us understand how the measure of a set varies over different α, allowing us to define the Hausdorff dimension.

Property 6. The quantity $m_0(E)$ counts the number of points in E, while $m_1(E) = m_{\mathcal{L}}(E)$, where $m_{\mathcal{L}}(E)$ is the Lebesgue measure on $\mathbb{R}$, for all Borel sets $E \in \mathbb{R}$.

We can see that $m_0(E)$ is $\lim_{\delta\to 0} \mathcal{H}_0^\delta(E)$, where $\mathcal{H}_0^\delta$ counts the number of sets of diameter at most δ it takes to cover E. As δ approaches 0, we need a separate set for each point in E, so $m_0(E)$ counts the number of points in E.

For the one-dimensional measure, we will show that we can express any covering of E by intervals I as a covering of E by sets F_n whose total diameter is the same as the total length of the intervals. Specifically, we can split each interval I into smaller intervals, of length at most δ, and since the diameter of an interval is equal to its length, the sum of these lengths must equal the sum of the lengths of the intervals. In the reverse direction, for any covering of E by sets F_n, we can construct a covering of E by intervals I by noting that we can cover each F_n by an interval whose length is exactly the diameter of F_n. Thus, since we can translate coverings back and forth between the two measures, the Lebesgue measure of a Borel set in $\mathbb{R}$ must equal its one-dimensional Hausdorff measure.

Property 7. If E is a Borel subset of $\mathbb{R}^d$, then $c_d m_d(E) = m_{\mathcal{L}}(E)$, for some constant c_d that depends only on the dimension d.

The constant c_d that we use is exactly $m_{\mathcal{L}}(B)/(\text{diam}(B))^d$, where B is the d-dimensional unit ball. Note that this ratio is the same for all d-dimensional balls, so this states that the ratio of the d-dimensional Lebesgue measure to the d-dimensional Hausdorff measure is the ratio of the volume of a ball to the d^{th} power of its diameter, which makes sense because when computing the Hausdorff measure, we use the d^{th} power of the diameter as an approximation for the size of a set.

Proof. First, we will show that $c_d m_d(E) \leq m_{\mathcal{L}}(E)$. We know that for every $\delta, \epsilon > 0$, we can find a covering $\{B_j\}$ of E such that every B_j is a ball with diameter less than δ and $\sum_j m_{\mathcal{L}}(B_j) \leq m_{\mathcal{L}}(E) + \epsilon$. Then, we have that

$$\mathcal{H}_d^\delta \leq \sum_j (\text{diam}\, B_j)^d = \frac{1}{c_d} \sum_j m_{\mathcal{L}}(B_j) \leq \frac{1}{c_d} m_{\mathcal{L}}(E) + \frac{1}{c_d}\epsilon.$$

Taking the limit as δ approaches 0 gives us that $m_d(E) \leq \frac{1}{c_d} m_{\mathcal{L}}(E) + \frac{1}{c_d}\epsilon$, and since ϵ is arbitrary, we get that $c_d m_d(E) \leq m_{\mathcal{L}}(E)$.

Then, we will show the opposite direction. This requires the **isodiametric inequality**, which states that for any diameter, the set with that diameter which has the largest volume is a ball. We will use this without proof.

Specifically, by the infimum in the definition of the Hausdorff measure, for any $\delta, \epsilon > 0$, we can find a covering $\{F_n\}$ of E such that for every n, $\text{diam}\, F_n \leq \delta$ and $\sum_n (\text{diam}\, F_n)^d \leq \mathcal{H}_d^\delta(E) + \epsilon$. Then, we have that by countable subadditivity,

$$m_{\mathcal{L}}(E) \leq m_{\mathcal{L}}\left(\bigcup_n F_n\right) \leq \sum_n m_{\mathcal{L}}(F_n),$$

and by the isodiametric inequality, $m_{\mathcal{L}}(F_n) \leq m_{\mathcal{L}}(B_n)$ where B_n is a ball with the same diameter as F_n. Then,

$$m_{\mathcal{L}}(E) \leq \sum_n m_{\mathcal{L}}(B_n) = \sum_n c_d (\text{diam}\, F_n)^d \leq c_d \mathcal{H}_d^\delta(E) + c_d \epsilon.$$

Taking the limit as delta and epsilon go to 0 gives us $m_{\mathcal{L}}(E) \leq c_d m_d(E)$.

Thus, $m_{\mathcal{L}}(E) = c_d m_d(E)$. □

Thus, for sets embedded in $\mathbb{R}^d$, the d-dimensional Haussdorf measure of a set is proportional to its Lebesgue measure, so the two measures behave similarly under nice conditions.

Now, we can prove the final property that leads us to the definition of the Hausdorff dimension:

Property 8. If $m_\alpha^*(E) < \infty$ and $\beta > \alpha$ then $m_\beta^*(E) = 0$, and if $m_\alpha^*(E) > 0$ and $\beta < \alpha$, then $m_\beta^*(E) = \infty$.

Proof. For the first statement, note that for any $\delta > 0$, for any set F such that $\text{diam}\, F \leq \delta$, then $(\text{diam}\, F)^\beta = (\text{diam}\, F)^{\beta-\alpha} (\text{diam}\, F)^\alpha \leq \delta^{\beta-\alpha} (\text{diam}\, F)^\delta$. But this implies that

$$\mathcal{H}_\beta^\delta(E) \leq \delta^{\beta-\alpha} \mathcal{H}_\alpha^\delta(E) \leq \delta^{\beta-\alpha} m_\alpha^*(E).$$

Then, taking the limit as δ approaches 0 gives us $m_\beta^*(E) \leq 0$, so $m_\beta^*(E) = 0$.

The second statement is just the contrapositive of the first. □

This means that for every set E, there is at most one α such that $m_\alpha(E)$ is both positive and finite.

To gain an intuition for how the α-dimensional Hausdorff measure works in different dimensions, we will consider a few examples.

Example 2.3. First, if I is a finite line segment in $\mathbb{R}^d$ then $m_1(I)$ is positive and finite.

Example 2.4. Similarly, if Q is a k-cube; that is, the product of k intervals and $d-k$ points, then $m_k(Q)$ is positive and finite.

Example 2.5. If E is a set in $\mathbb{R}^d$, then $m_\alpha(E) = 0$ for all $\alpha > d$.

3 Hausdorff Dimension

3.1 Defining the Hausdorff Dimension

The properties above intuitively lead to the notion of the Hausdorff dimension of a set E being the unique α for which E has β-dimensional Hausdorff measure 0 for all $\beta > \alpha$ and β-dimensional Hausdorff measure ∞ for all $\beta < \alpha$.

Definition 3.1. The **Hausdorff dimension** of a set E is the unique α such that

$$m_\beta(E) = \begin{cases} 0 & \text{if } \beta > \alpha \\ \infty & \text{if } \beta < \alpha. \end{cases}$$

We write this as $\alpha = \dim E$.

Note that this is not necessarily saying that $m_\alpha(E)$ is positive or finite, and it may be the case that E has Hausdorff dimension α but $m_\alpha(E)$ is 0 or ∞. If we know that $m_\alpha(E)$ is positive and finite, and E is bounded, we say that E has **strict Hausdorff dimension** α.

Corollary 3.2. If $0 < m_\alpha(E) < \infty$ for some α, then α must be the strict Hausdorff dimension of E.

This follows directly from Property 8.

Definition 3.3. If E is a set of fractional dimension, we call it a **fractal**.

There is no easy general method for computing the Hausdorff measure of a set. It is sometimes clear what the dimension is for nice sets, as in Example 2.4, but for more interesting sets E we have to prove that its dimension is α by bounding $m_\alpha(E)$ above and below, so that it is clear that the dimension of E is α, but not necessarily what its α-dimensional Hausdorff measure actually is.

3.2 Computing the Hausdorff Dimension

We will begin by computing the Hausdorff dimension of the Cantor set.

3.2.1 The Cantor Set

Recall that the Cantor set is the set $\mathcal{C}$ that was the set remaining after we start with the interval $[0, 1]$ and then repeatedly remove the middle third of each interval. For a formal definition of the Cantor set, refer to Appendix B.

Figure 4: the process of creating the Cantor set

We said in the motivation for this paper that if we think of dimension as a sort of scaling factor, then it would make sense for the Cantor set to have a dimension of $\log_3 2$, because scaling each point by a factor of three would result in two disjoint copies of the Cantor set, one in $[0,1]$ and one in $[2,3]$. We will now see formally that this is indeed the Hausdorff dimension of the Cantor set.

Theorem 3.4. The Cantor set $\mathcal{C}$ has $\dim \mathcal{C} = \frac{\log 2}{\log 3}$.

For this section, we will define $\alpha_c = \frac{\log 2}{\log 3}$, for ease of notation. Then, we will show that $\dim \mathcal{C} = \alpha_c$ by showing that the α_c-dimensional Hausdorff measure of $\mathcal{C}$ is positive and finite. In order to show that it is positive, however, we use the Cantor-Lebesgue function, which is defined in Appendix B, to map $\mathcal{C}$ onto $[0,1]$, and use a property of the mapping based on the fact that it satisfies a Lipschitz condition. We will define these properties first.

Definition 3.5. A function f defined on a subset E of $\mathbb{R}^d$ satisfies a **Lipschitz condition with exponent** γ if there exists some $M > 0$ such that

$$\big|f(x) - f(y)\big| \leq M|x-y|^\gamma \qquad \text{for all } x, y \in E.$$

Intuitively, this is a way of saying that a function does not jump around too much over a narrow space. We can use this to place a bound on how much the measure of the image of a set can differ from its preimage under the mapping.

Lemma 3.6. Suppose a function f defined on a compact set E satisfies a Lipschitz condition with exponent γ. Then,

1. $m_\beta(f(E)) \leq M^\beta m_\alpha(E)$ if $\beta = \alpha/\gamma$
2. $\dim f(E) \leq \frac{1}{\gamma} \dim E$,

where $M > 0$ is the value such that $\big|f(x) - f(y)\big| \leq M|x-y|^\gamma$ for all $x, y \in E$.

Proof. For any $\epsilon, \delta > 0$, let $\{F_n\}$ be a collection of sets that cover E such that for each n, $\operatorname{diam} F_n \leq \delta$ and $\sum_n (\operatorname{diam} F_n)^\alpha < \mathcal{H}_\delta^\alpha + \epsilon$. Then, $\big\{f(E \cap F_n)\big\}$ is a collection of sets that cover $f(E)$, and we can see that for each n, $f(E \cap F_n)$ has diameter at most $M(\operatorname{diam} F_n)^\gamma$. But this means that

$$\mathcal{H}_{M\delta}^{\alpha/\gamma}(f(E)) \leq \sum_n (\operatorname{diam} F_n)^{\alpha/\gamma} \leq M^\beta \sum_n (\operatorname{diam} F_n)^\alpha \leq M^\beta \mathcal{H}_\delta^\alpha + M^\beta \epsilon.$$

Taking the limit as δ and ϵ go to zero, we get that $m_\beta(f(E)) \leq M^\beta m_\alpha(E)$. Then, the second part follows directly from the first. ☐

Lemma 3.7. The Cantor-Lebesgue function F on $\mathcal{C}$ satisfies a Lipschitz condition with exponent $\gamma = \alpha_c$.

Proof. We defined the Cantor-Lebesgue function F as the limit of a sequence of functions $\{F_n\}$ where F_n increased by at most 2^{-n} on all intervals of length 3^{-n}. Thus, F_n increases by at most $\big(3/2\big)^n|x-y|$ on an interval $[x, y]$, and since we also know that $\left|F(x) - F_n(x)\right| \leq 2^{-n}$ for all $x \in [0,1]$, we get that

$$\begin{aligned}\left|F(x) - F(y)\right| &\leq \left|F(x) - F_n(x)\right| + \left|F_n(x) - F_n(y)\right| + \left|F_n(y) - F(y)\right| \\ &\leq \left(\frac{3}{2}\right)^n |x - y| + \frac{2}{2^n}\end{aligned}$$

for all $x, y \in [0,1]$. But since this is true for all n, we can pick an n depending on x and y to make this inequality nice.

Specifically, we can always pick n such that $1 \leq 3^n|x-y| \leq 3$ by taking n to be $\lceil \log|x-y| \rceil$. Then, we get that

$$\left|F(x) - F(y)\right| \leq \frac{5}{2^n}$$

, and since $\alpha_c = \log_3 2$, this means that $\left|F(x) - F(y)\right| \leq 5(3^{-n})^{\alpha_c}$. But since $3^n|x-y| \geq 1$, we can see that $|x - y| \geq 3^{-n}$, and

$$\left|F(x) - F(y)\right| \leq 5|x - y|^{\alpha_c}.$$

Thus, F satisfies a Lipschitz condition with exponent $\gamma = \alpha_c$. □

Now, we have enough background to prove Theorem 3.4.
We will prove first that $m_{\alpha_c}(\mathcal{C}) \leq 1$ and then that $m_{\alpha_c}(\mathcal{C}) > 0$. This shows that $m_{\alpha_c}(\mathcal{C})$ is positive and finite, which means α_c must be the dimension of $\mathcal{C}$ from Corollary 3.2.

For the first part, note that we defined $\mathcal{C}$ as $\bigcap_n C_n$, where each C_n has 2^n intervals of length 3^{-n} and covers $\mathcal{C}$. Then, for each $\delta > 0$, we can simply pick N such that $3^{-N} < \delta$. This makes C_N a valid covering of $\mathcal{C}$, and we can see that

$$\mathcal{H}^{\delta}_{\alpha_c}(\mathcal{C}) \leq 2^n (3^{-n})^{\alpha_c} = 1$$

since $(3^{-n})^{\alpha_c} = 2^{-n}$ by definition. Taking the limit as δ goes to 0 gives us that $m_{\alpha_c}(\mathcal{C}) \leq 1$.

For the second part, we can use Lemma 3.7 to apply Lemma 3.6 to the Cantor-Lebesgue function, giving us

$$m_1([0,1]) \leq M m_{\alpha_c}(\mathcal{C}).$$

But we know that $m_1([0,1]) = 1$, so $m_{\alpha_c}(\mathcal{C}) \geq 1/M$ (since M is nonzero), and this means that $m_{\alpha_c}(\mathcal{C}) > 0$.

Thus, $m_{\alpha_c}(\mathcal{C})$ is positive and finite, so $\dim \mathcal{C} = \log_3 2$. □

So now we have seen one example of computing the dimension of a fractional-dimension set. We can consider another, similar example - the Sierpinski triangle.

3.2.2 The Sierpinski Triangle

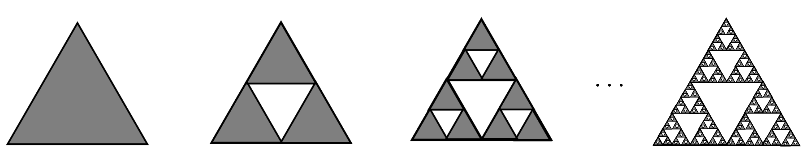


Figure 5: the steps to create the Sierpinski triangle

The Sierpinski triangle is a Cantor-like set that is constructed in the plane. We begin with a closed equilateral triangle S_0 with unit-length sides. Then, the first generation S_1 is defined to be S_0 with the middle open equilateral triangle removed, as pictured, so that S_1 is the union of three smaller equilateral triangles. Then, in each generation S_{n+1}, we take the equilateral triangles in S_n and remove the middle equilateral triangle as we did in the creation of S_1, so that generation S_n has 3^n closed equilateral triangles, each with side length 2^{-n}. We can see moreover that each S_n is a compact set, and $S_{n+1} \subset S_n$ for all $n \geq 0$.

Then, in a similar fashion to the Cantor set, we define the Sierpinski triangle as follows.

Definition 3.8. The **Sierpinski triangle** $\mathcal{S}$ is the compact set defined to be the intersection of each generation in its creation, or

$$\mathcal{S} = \bigcap_n S_n.$$

We can see here that similarly to the Cantor set, scaling each point in the Sierpinski triangle by 2, so that S_0 now has a side length of 2, would create three disjoint copies of our original triangle, so we might expect that the Sierpinski triangle has a dimension of $\log_2 3$. Indeed, this is what we will now formally prove.

Theorem 3.9. The Sierpinksi triangle $\mathcal{S}$ has strict Hausdorff dimension $\log_2 3$.

We will define $\alpha_s = \log_2 3$ for ease of notation.

We will again prove this by showing that $m_{\alpha_s}(\mathcal{S})$ is positive and finite.

We can show that $m_{\alpha_s}(\mathcal{S}) \leq 1$ in a very similar procedure to that of the Cantor set. We know that for any $\delta > 0$, we can find N such that $2^{-N} < \delta$. Then, we can see that S_N is a covering of $\mathcal{S}$ that contains 3^N triangles, and since the diameter of an equilateral triangle is equal to its side length, the diameter of each of these triangles is $2^{-N} < \delta$. Thus,

$$\mathcal{H}^{\delta}_{\alpha_s}(\mathcal{S}) \leq 3^N (2^{-N})^{\alpha_s} = 1,$$

since $(2^{-N})^{\alpha_s} = 3^N$. Taking the limit as δ approaches 0, we get that $m_{\alpha_s}(\mathcal{S}) \leq 1$, so the α_s-dimensional Hausdorff measure of $\mathcal{S}$ is finite.

The other direction, showing that $m_{\alpha_s}(\mathcal{S}) > 0$, is a bit more complicated. First, we will call the lower-left vertex of each triangle the shiny vertex of that triangle - this is important because the shiny vertices in any S_n will also be part of $\mathcal{S}$. We can see that since there are 3^n triangles in S_n, S_n must have 3^n shiny vertices. Then, we will claim that there is a positive constant c such that for any $\delta > 0$, for any covering $\{F_n\}$ of $\mathcal{S}$ such that $\operatorname{diam} F_n \leq \delta/2$ for each n,

$$\sum_n (\operatorname{diam} F_n)^{\alpha_s} \geq c.$$

Then, since every set F_n is contained in a ball B_n with diameter at most $2 \operatorname{diam} F_n$, it is enough to show that for any covering $\{B_n\}$ of $\mathcal{S}$ by balls B_n such that $\operatorname{diam} B_n \leq \delta$,

$$\sum_n (\operatorname{diam} B_n)^{\alpha_s} \geq c,$$

because this means that every covering $\{F_n\}$ has $\sum_n (\operatorname{diam} F_n)^{\alpha_s} \geq c/2^{\alpha_s}$, which is still a constant. We will prove this now, picking c to be c''/c', where $c' = 9\pi/4$, or the area of a circle with diameter 3, and $c'' = \sqrt{3}/4$, or the area of an equilateral triangle with side length 1, though the exact values of c' and c'' are unimportant.

For any such covering of $\mathcal{S}$ by balls, we can choose k such that

$$2^{-k} \leq \min_n \operatorname{diam} B_n < 2^{-k+1},$$

and consider the shiny points in the k^{th} generation that our balls cover.

Lemma 3.10. Suppose B is a ball in our covering such that

$$2^{-\ell} \leq \operatorname{diam} B \leq 2^{-\ell+1} \qquad \text{for some } \ell \leq k.$$

Then, B covers at most $3^{k-\ell}/c$ shiny vertices from the k^{th} generation.

Proof. Let B^* be the ball with the same center as B but three times its diameter. Then, for any k^{th} generation triangle $\triangle_k$ whose shiny vertex is in B, we can define $\triangle'_\ell$ to be the ℓ^{th} generation triangle containing $\triangle_k$, and since $\operatorname{diam} B \geq 2^{-\ell}$, we can see that $\triangle'_\ell$ must be fully contained in B^*.

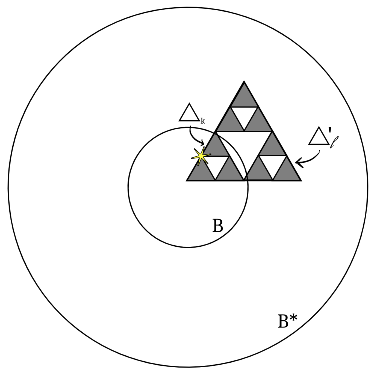


Figure 6: for a shiny vertex in B, the relevant triangles $\triangle_k$ and $\triangle'_\ell$

But the area of each triangle in the ℓ^{th} generation is $4^{-\ell}c''$ and the area of B^* is $4^{-\ell}c'$, so there are at most $1/c$ triangles in the ℓ^{th} generation that are fully contained in B^*. Each one contains $3^{k-\ell}$ triangles of the k^{th} generation so there can be at most $3^{k-\ell}/c$ shiny points of the k^{th} generation covered by B. ☐

Now, we can use this lemma to get back to proving that $\sum_n (\operatorname{diam} B_n)^{\alpha_s} \geq c$. For each natural number ℓ, we can define N_ℓ to be the number of balls B in $\{B_n\}$ such that $2^{-\ell} \leq \operatorname{diam} B < 2^{-\ell+1}$. Then, we can see that

$$\sum_n (\operatorname{diam} B_n)^{\alpha_s} \geq \sum_\ell N_\ell (2^{-\ell})^{\alpha_s} = \sum_\ell N_\ell 3^{-\ell}.$$

But since the number of shiny vertices of the k^{th} generation covered by the balls is at most $\sum_\ell N_\ell 3^{k-\ell}/c$, and we need all 3^k shiny vertices to be covered, we get that $\sum_\ell N_\ell 3^{k-\ell}/c \geq 3^k$, so $\sum_\ell N_\ell 3^{-\ell} \geq c$. Thus, using our above inequality, $\sum_n (\operatorname{diam} B_n)^{\alpha_s} \geq c$, and since this is true for every possible covering of $\mathcal{S}$, $m_{\alpha_s}(\mathcal{S}) \geq c$.

Thus, $m_{\alpha_s}(\mathcal{S})$ is positive and finite, so $\dim S = \alpha_s$. ☐

We used very similar steps in proving the Hausdorff dimension of these two sets. We will generalize these steps in the next section, and then use the generalization to compute the dimension of one more (slightly different) fractal.

3.3 Self-Similarity

What we used to intuit, and in a more indirect way, to compute the Hausdorff dimension of the Sierpinski triangle and the Cantor set was the fact that both could be expressed as the union of scaled down versions of themselves. For example, $\mathcal{C} \cap [0, 1/3]$ is a scaled-down version of the Cantor set, $\mathcal{C} \cap [1/3, 2/3] = \emptyset$, and $\mathcal{C} \cap [2/3, 1]$ is also a scaled-down version of the Cantor set. We can also express $\mathcal{C}$ as the union of four scaled-down versions of itself in intervals of length 1/9, and so on. Similarly, the Sierpinski triangle is the union of three scaled-down versions of itself, one in each of the three triangles in S_1, or nine scaled-down versions of itself, one in each of the nine triangles in S_2, and so on.

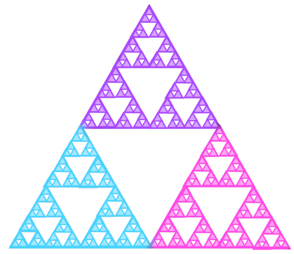


Figure 7: the three scaled-down versions of the Sierpinksi triangle, in different colors

In this section, we will formalize the notion of self-similarity, and use it to prove some more general ideas about the Hausdorff dimension of a fractal.

Definition 3.11. A **similarity** with **ratio** $r > 0$ is a mapping $S : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\big|S(x) - S(y)\big| = r|x - y|\,.$$

That is, a similarity is a mapping that preserves the relative position of the points in the original set. It is left as an exercise to the reader to show that every self-similarity can be expressed as the composition of a translation, rotation, and dilation by r.

Definition 3.12. We say that a set $F \in \mathbb{R}^d$ is **self-similar** if there exist finitely many similarities $S_1, \dots S_m$ with the same ratio r such that

$$F = S_1(F) \cup S_2(F) \cup \cdots \cup S_m(F).$$

This makes sense with how we were describing the Cantor set and Sierpinski triangle - the sets are self-similar because they are equal to a finite number of scaled copies of themselves.

Example 3.13. Formally, we can see that $\mathcal{C}$ is self-similar because it can be expressed as $S_1(\mathcal{C}) \cup S_2(\mathcal{C})$, where

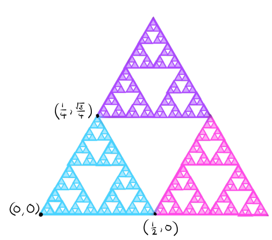
$$S_1(x) = \frac{x}{3} \qquad \text{and} \qquad S_2(x) = \frac{x}{3} + \frac{2}{3}.$$

We can see that both these similarities have $r = 1/3$.

Example 3.14. The Sierpinski triangle $\mathcal{S}$ is self-similar because it can be expressed as $S_1(\mathcal{S}) \cup S_2(\mathcal{S}) \cup S_3(\mathcal{S})$, where

$$S_1(x) = \frac{x}{2}, \qquad S_2(x) = \frac{x}{2} + \left(\frac{1}{4}, \frac{\sqrt{3}}{4}\right), \qquad S_3(x) = \frac{x}{2} + \left(\frac{1}{2}, 0\right).$$

Each of these similarities has ratio 1/2.



Another example, which is very similar to the first two, is the **Cantor dust** $\mathcal{D}$, another two-dimensional version of the Cantor set. This version can be defined for any $0 < \mu < \frac{1}{2}$, and it is constructed iteratively as follows. First, we begin with the unit square D_0. Then, to construct D_1, remove everything except the four squares in the corner of side length μ. We repeat this step iteratively, so that to construct D_2 we look at each smaller square and remove everything except the corner squares of side length μ^2, and D_3 has 64 squares of side length μ^3, and so on. This gives us a sequence of compact sets $D_0 \supset D_1 \supset D_2 \supset \cdots$, and as one might expect, we define the Cantor dust to be

$$\mathcal{D} = \bigcap_n D_n.$$

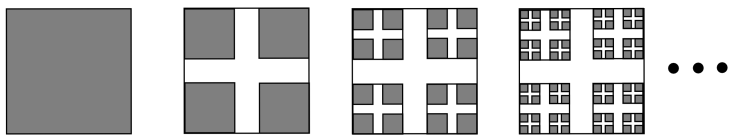


Figure 8: the first few steps in creating the Cantor dust

Example 3.15. The Cantor dust $\mathcal{D}$ is also a self-similarity; it can be expressed as $S_1(\mathcal{D}) \cup S_2(\mathcal{D}) \cup S_3(\mathcal{D}) \cup S_4(\mathcal{D})$, where we define the similarities

$$S_1(x) = \mu x, \qquad S_2(x) = \mu x + (0, 1-\mu), \qquad S_3(x) = \mu x + (1-\mu, 0), \qquad S_4(x) = \mu x + (1-\mu, 1-\mu).$$

Each of these similarities has ratio μ.

First, we will prove that given any set of similarities, assuming the ratio is decreasing the size of its input, we can find a set that is self-similar using those similarities.

Theorem 3.16. Suppose $S_1, S_2, \cdots S_m$ are m similarities, each with the same ratio $0 < r < 1$. Then, there exists a unique non-empty compact set F such that

$$F = S_1(F) \cup S_2(F) \cup \cdots \cup S_m(F).$$

We prove this by essentially taking a large ball B and iteratively applying the similarities to the ball. Because the ratio is less than 1, the ball contracts as we do this, so repeatedly applying the similarities converges to such a set F.

We begin with a lemma that allows us to find such a relevant ball.

Lemma 3.17. There exists a closed ball B so that $S_j(B) \subset B$ for all $1 \leq j \leq m$.

Proof. We note that for any x, we can express

$$\left|S_j(x)\right| \leq \left|S_j(x) - S_j(0)\right| + \left|S_j(0)\right|,$$

and by definition of a similarity, this means that

$$\left|S_j(x)\right| \leq r|x| + \left|S_j(0)\right|.$$

Then, we can see that for any x within a ball B of radius $R_j \geq \left|S_j(0)\right|/(1-r)$ centered at the origin, the above inequality tells us that $S_j(x)$ is also at most R_j away from the origin, so $S_j(B) \subset B$. But then, we can define B to be a closed ball centered at the origin with radius $\max_j \left|S_j(0)\right|/(1-r)$, $S_j(B) \subset B$ for all j, which is exactly what we wanted. ☐

Then, for ease of notation, we will define $\overline{S}(A)$ to be $S_1(A) \cup \cdots \cup S_m(A)$ for any set A. Note that while each S_j is a mapping of points, $\overline{S}$ is a mapping from sets to sets.

Moreover, to better understand the contraction of sets that we are creating, we will introduce the notion of Hausdorff distance between sets.

Definition 3.18. We define the **Hausdorff distance** between compact sets A and B to be

$$\text{dist}(A, B) = \inf \left\{ \delta : A^\delta \supset B \text{ and } B^\delta \supset A \right\},$$

where A^δ is the set, for any $\delta > 0$,

$$A^\delta = \left\{ x : \text{dist}(x, A) < \delta \right\}.$$

Intuitively, we can think of the Hausdorff distance as how far we need to expand one set to reach the furthest end of the other. We will leave it to the reader to verify that the Hausdorff distance satisfies the three properties of a valid distance.

Similarities interact with distance in the way one might expect.

Lemma 3.19. If $S_1, \ldots, S_m$ are similarities with ratio r, then for any compact sets A, B,

$$\text{dist}(\overline{S}(A), \overline{S}(B)) \leq r \,\text{dist}(A, B).$$

Proof. We can define $D_{A,B}$ to be the set $\left\{\delta : A^\delta \supset B \text{ and } B^\delta \supset A\right\}$. Then, we can see that for any $\delta \in D_{A,B}$, for any $x \in B$, $d(x, A) < \delta$, where $d(x, A)$ is the standard Euclidean distance. But then by definition of a similarity, for any $x \in \overline{S}(B)$, $d(x, \overline{S}(A)) < r\delta$, so $\overline{S}(B) \subset \overline{S}(A)^{r\delta}$ and we can similarly show that $\overline{S}(A) \subset \overline{S}(B)^{r\delta}$. But since this is true for all such δ, we get that $D_{\overline{S}(A),\overline{S}(B)} \supset \left\{r\delta : \delta \in D_{A,B}\right\}$, so by taking the infimum we can see that $\text{dist}(\overline{S}(A), \overline{S}(B)) \leq r \,\text{dist}(A, B)$. $\square$

Then, we can use these two lemmas to prove Theorem 3.16.

We begin by constructing such an F. First, pick the ball B as defined in Lemma 3.17. We will define the sequence of sets $F_0 = B$ and $F_n = \overline{S}(F_{n-1})$. We can see that since B is compact and nonempty, so is each F_n. Moreover, since $\overline{S}(B) \subset B$, we have that for each n, $F_n \subset F_{n-1}$. Then, we can let

$$F = \bigcap_{n=1}^{\infty} F_n,$$

and clearly F is also compact and nonempty. Moreover, $\overline{S}(F) = \bigcap_{n=2}^{\infty} F_n = F$, so F has the property we wanted.

To see that F is unique, we can see that for any other compact set G such that $\overline{S}(G) = G$, Lemma 3.19 tells us that $\text{dist}(\overline{S}(F), \overline{S}(G)) = \text{dist}(F, G) \leq r \,\text{dist}(F, G)$. Since $r < 1$, this implies that $\text{dist}(F, G) = 0$ and therefore $F = G$. $\square$

Now, for a self-similar set F, if the similarities $S_1(F), \ldots S_m(F)$ do not overlap too much, we can compute in general the Hausdorff dimension of F.

For disjoint $S_1(F), \ldots, S_m(F)$, this dimension is not difficult to find. We can simply use countable additivity of the Hausdorff measure to see that

$$m_\alpha(F) = \sum_{j=1}^{m} m_\alpha(S_j(F)).$$

But since every S_j scales F by a factor of r, it also scales the diameter of each set in a cover of F by a factor of r, so $m_\alpha(S_j(F)) = r^\alpha m_\alpha(F)$, so

$$m_\alpha(F) = mr^\alpha m_\alpha(F).$$

But this means when $m_\alpha(F)$ is positive and finite, we must have $1 = mr^\alpha$, so $\alpha = (\log m)/(\log 1/r)$. Note that this just implies that *if* there is a strict Hausdorff dimension of F, it has this value, but it may be the case that F has no strict Hausdorff dimension.

We will show that this is true in a slightly more general case, which we will call separated similarities.

Definition 3.20. A set of similarities $S_1, \ldots S_m$ are **separated** if we can find a bounded open set $\mathcal{O}$ such that

$$\mathcal{O} \supset S_1(\mathcal{O}) \cup S_2(\mathcal{O}) \cup \cdots \cup S_m(\mathcal{O})$$

and the sets $S_j(\mathcal{O})$ are disjoint.

Example 3.21. The similarities S_1, S_2 as defined for the Cantor set are separated, because if we take $\mathcal{O} = (0,1)$ we can see that $S_1(\mathcal{O}) = (0, 1/3)$ and $S_2(\mathcal{O}) = (2/3, 1)$, so $\mathcal{O} \supset S_1(\mathcal{O}) \cup S_2(\mathcal{O})$ and the two sets are disjoint.

Now, we can get to our big theorem about the Hausdorff dimension of self-similar sets.

Theorem 3.22. If $S_1, S_2, \ldots, S_m$ are separated similarities with ratio $0 < r < 1$, then their self-similar set $F \subset \mathbb{R}^d$ has strict Hausdorff dimension $(\log m)/(\log 1/r)$.

We will prove this theorem using a very similar strategy to the proof of the Hausdorff dimension of the Sierpinski triangle. We will define $\alpha = (\log m)/(\log 1/r)$ for ease of notation. Then, we will first prove that $m_\alpha(F)$ is finite, and then prove that it is positive.

To prove that it is finite, we consider the ball B as defined in Lemma 3.17. We know that F_i, as defined in the proof of Theorem 3.16 is the union of m^i balls of diameter cr^i, where c is the diamter of B. Thus, for any $\delta > 0$, we can find a k such that F_i is a covering of F by balls of diameter less than δ. Thus,

$$\mathcal{H}_\alpha^\delta(F) \leq m^i (cr^i)^\alpha \leq c^\alpha,$$

since $mr^\alpha = 1$ by definition. Then since this is true for all $\delta > 0$, we can take the limit as δ approaches 0 to get that $m_\alpha(F) \leq c^\alpha$, which means F has finite α-dimensional Hausdorff measure.

To prove that it is positive, we begin by picking a point $\overline{x}$ in F. Then, we define the shiny vertices of the i^{th} generation (or the shiny vertices of F_i) to be the m^i points

$$S_{n_1} \circ \cdots \circ S_{n_i}(\overline{x}) \qquad 1 \leq n_1 \leq m, \ldots 1 \leq n_i \leq m,$$

which are the m^i points that $\overline{x}$ maps to in the i^{th} generation. We will label each such vertex $(n_1, \ldots, n_i)$ and note that we don't care if multiple shiny vertices map to the same point.

We also take an open set $\mathcal{O}$ that has the property from the definition of separated similarities. We will similarly define the shiny open sets of the i^{th} generation to be sets $\mathcal{O}$ maps to in the i^{th} generation, or

$$S_{n_1} \circ \cdots \circ S_{n_i}(\mathcal{O}) \qquad 1 \leq n_1 \leq m, \ldots 1 \leq n_i \leq m,$$

which are again labeled $(n_1, \ldots n_i)$.

Then, since the shiny open sets of the first generation are disjoint by definition, we can inductively see that the shiny open sets of the i^{th} generation are disjoint for all i. Moreover, again by the separation property, for any $i \geq \ell$, each shiny open set of the ℓ^{th} generation contains $m^{i-\ell}$ shiny open sets of the i^{th} generation.

Then, for any shiny vertex v of the i^{th} generation, let $\mathcal{O}(v)$ be the shiny open set with the same label as v. Then, we can see that by definition of a similarity,

$$d(v, \mathcal{O}(v)) = r^i d(\overline{x}, \mathcal{O}), \tag{1}$$

and

$$\operatorname{diam}(\mathcal{O}(v)) = r^i \operatorname{diam}(\mathcal{O}), \tag{2}$$

where what matters is that $d(\overline{x}, \mathcal{O})$ and $\operatorname{diam}(\mathcal{O})$ are constants with relation to v and i.

Similarly to the proof for the Sierpinski triangle, we want to prove that there is some constant c such that for $\delta > 0$ and every covering $\{B_j\}$ of F by balls B_j with radius less than δ,

$$\sum_j (\operatorname{diam} B_j)^\alpha \geq c.$$

For any such covering $\{B_j\}$, we look at the ball with the minimum diameter and define k to be the integer such that

$$r^k \leq \min_j \operatorname{diam} B_j \leq r^{k-1}.$$

Then, we prove a lemma very similar to Lemma 3.10.

Lemma 3.23. For any ball B in our covering, define $\ell \geq k$ to be the integer such that

$$r^\ell \leq \operatorname{diam} B \leq r^{\ell-1}.$$

Then B contains at most $c_1 m^{k-\ell}$ shiny vertices of the k^{th} generation, where c_1.

Proof. Note by Equation 1 and Equation 2, there exists some constant c' such that if we define B^* to be the dilation of B by a factor of c', then for any shiny vertex v contained in B, B^* contains $\mathcal{O}(v)$ and moreover B^* contains the shiny open set from the ℓ^{th} generation that contains $\mathcal{O}(v)$.

Then, B^* has volume at most $r^{d\ell}$ times c'^d times the volume of a unit ball, and each shiny open set in the ℓ^{th} generation has volume r^{dk} times the volume of $\mathcal{O}$, which means that at most c_1 shiny open sets of the ℓ^{th} generation can fit in B^*, where c_1 is dependent only on $\overline{x}$ and $\mathcal{O}$. But this means that at most $c_1 m^{k-\ell}$ shiny open sets of the k^{th} generation fit in B^*, so at most $c_1 m^{k-\ell}$ shiny vertices of the k^{th} generation fit in B. □

Then, we can use this lemma to return to proving that

$$\sum_j (\operatorname{diam} B_j)^\alpha \geq c.$$

Specifically, let N_ℓ be the number of balls B in our covering such that $r^\ell \leq \operatorname{diam} B \leq r^{\ell-1}$. Then, we can see that

$$\sum_j (\operatorname{diam} B_j)^\alpha \geq \sum_\ell N_\ell (r^\ell)^\alpha.$$

But we know that the total number of shiny vertices of the k^{th} covered by these balls is at most $\sum_\ell N_\ell c_1 m^{k-\ell}$, and since the balls must cover all m^k shiny vertices of the k^{th} generation, we get that

$$\sum_\ell N_\ell c_1 m^{k-\ell} \geq m^k$$

$$\sum_\ell N_\ell m^{-\ell} \geq 1/c_1.$$

Then, taking $c = 1/c_1$, and since by our definition of α, $(r^\ell)^\alpha = m^{-\ell}$,

$$\sum_j (\operatorname{diam} B_j)^\alpha \geq \sum_\ell N_\ell m^{-\ell} \geq c.$$

Thus, $m_\alpha(F) \geq c$, so the α-dimensional Hausdorff measure of F is also positive, and the strict Hausdorff dimension of F is α. □

This is a really cool general theorem about the Hausdorff dimension of a fractal, and to show this we will use it to compute the Hausdorff dimension of a fractal that is less similar to the Cantor set.

3.3.1 The Von Koch Curve

The Von Koch curve is a curve again defined iteratively. It is better described in pictures, where each K_n is a curve embedded in $\mathbb{R}^2$, and each K_n adds a "bump" in the middle of each line segment in K_{n-1}, so that K_n is the curve consisting of 4^n line segments, each of length 3^{-n}.

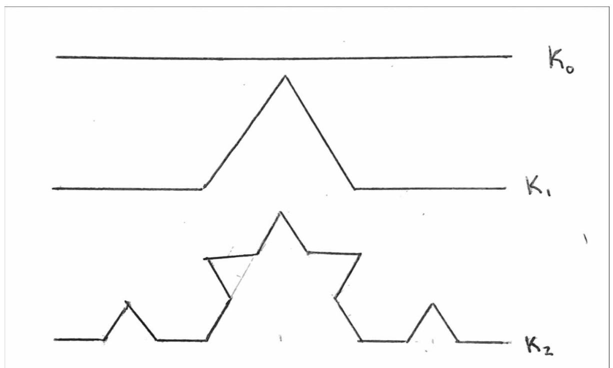


Figure 9: the first steps in creating the Von Koch curve

Definition 3.24. We define the **Von Koch curve** $\mathcal{K}$ to be the limit $\lim_{n\to\infty} K_n$, or the limit as we keep adding bumps to the line.

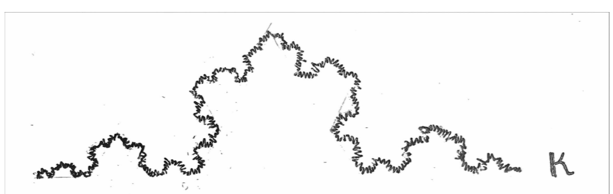


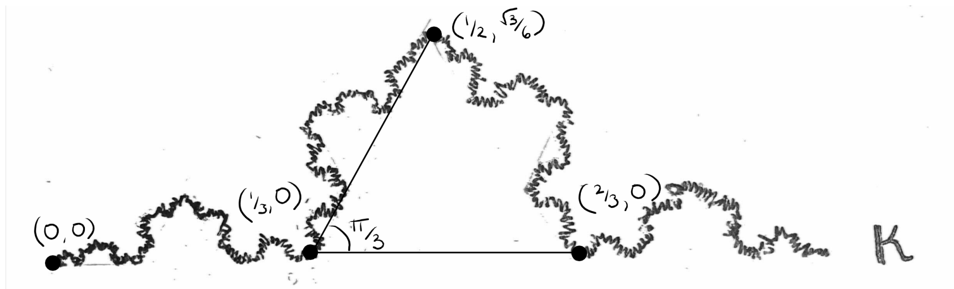
Figure 10: the Von Koch curve

This is a very informal definition of $\mathcal{K}$, but we can define it more formally as a self-similarity.

Definition 3.25. We define the **Von Koch curve** to be the unique self-similarity in $\mathbb{R}^2$ for the similarities

$$S_1(x) = \frac{1}{3}(x)$$
$$S_2(x) = \frac{1}{3}\rho(x) + \left(\tfrac{1}{3}, 0\right)$$
$$S_3(x) = \frac{1}{3}\rho^{-1}(x) + \left(\tfrac{1}{2}, \tfrac{\sqrt{3}}{6}\right)$$
$$S_4(x) = \frac{1}{3}(x) + \left(\tfrac{2}{3}, 0\right),$$

where each of these similarities has ratio $r = \frac{1}{3}$ and ρ is defined to be a rotation about the origin by $\pi/3$.



From what we learned in the previous subsection, we can now easily prove the following theorem.

Theorem 3.26. The Von Koch curve has strict Hausdorff dimension $\log_3 4$.

Proof. First, we can see that these similarities are separated by taking the open set $\mathcal{O}$ shown in the following picture. Showing that $\mathcal{O}$ follows the properties we want for separated similarities is a matter of algebra, so it is left as an exercise to the reader.

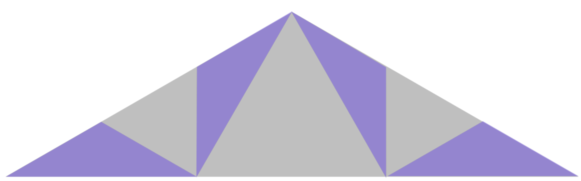


Figure 11: our set $\mathcal{O}$ is the full grey triangle, and the similarities are applied to $\mathcal{O}$ in purple

Now, we can cite Theorem 3.22 directly to see that since $m = 4$ and $r = \frac{1}{3}$, the Von Koch curve has strict Hausdorff dimension $\log_4 3$. □

This is the conclusion of our discussion on the Hausdorff dimension, but we will turn to look at more interesting things we can do with self-similar sets.

4 Space-Filling Curves

We will now see that fractals and self-similarities are useful also in studying the Lebesgue measure. Specifically, we look at a type of fractal that has a Hausdorff dimension of exactly two, but which is constructed out of a line, allowing us to make mappings from sets with finite nonzero one-dimensional Lebesgue measure to sets with finite nonzero two-dimensional Lebesgue measure. The family of fractals that involve drawing a singular line to fill a unit square are called space-filling curves, and they tend to look like the sort of doodles one might make if they are trying to fill time as well as space.

In this section, we will look at the Peano curve specifically. The Peano curve looks like the following, and we will define it formally later in the section.

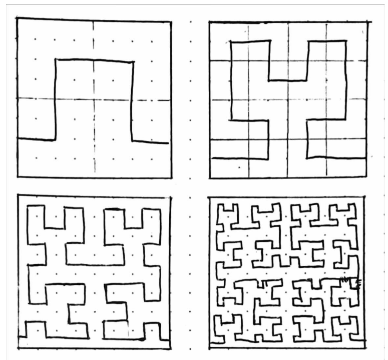


Figure 12: the initial steps in creating the Peano curve

We can see that this is a self-similarity - though I will not formalize this, we can see that the Peano curve is equal to four copies of itself, each scaled by 1/4, which means it is not difficult to use the previous section to see that it has a strict Hausdorff dimension of 2. However, we can prove something even more specific about the Peano curve and similar space-filling curves.

We will define the Peano curve as the image of a mapping $\mathcal{P} : [0,1] \to [0,1] \times [0,1]$, so that we can analyze it mathematically. What we find is that the Peano curve fills the entire unit square, and moreover that besides some measure-zero subset of $[0,1]$, the Peano curve maps any subset of the unit interval to a subset of the unit square in a way that preserves measure.

In this section, we will use m_1 to denote one-dimensional Lebesgue measure and m_2 to denote two-dimensional Lebesgue measure. Then, formally, what we will prove is that

Theorem 4.1. There exists a mapping $\mathcal{P} : [0,1] \to [0,1] \times [0,1]$ such that:

1. $\mathcal{P}$ is both continuous and surjective.
2. $\mathcal{P}$ satisfies a Lipschitz condition of exponent 1/2, so

$$\left|\mathcal{P}(x) - \mathcal{P}(y)\right| = |x-y|^{1/2}.$$

3. For any interval $[a,b] \subset [0,1]$, $\mathcal{P}([a,b])$ is a compact subset of the unit square such that $m_2(\mathcal{P}([a,b])) = b-a$.

The last item of this theorem implies the following corollary.

Corollary 4.2. There are measure-zero subsets $Z_1 \subset [0,1]$ and $Z_2 \subset [0,1] \times [0,1]$ such that $\mathcal{P}$ is bijective from

$$[0,1] \setminus Z_1 \text{ to } [0,1] \times [0,1] \setminus Z_2$$

and for any subset E of $[0,1] \setminus Z_1$, E is measurable if and only if $\mathcal{P}(E)$ is measurable, and

$$m_2(\mathcal{P}(E)) = m_1(E).$$

We can note that it is impossible for such a mapping to be bijective on the whole interval $[0,1]$; if a continuous and surjective mapping $F : [0,1] \to [0,1] \times [0,1]$ were also bijective then it would have an inverse G that would also be continuous and injective. But then for any points $a, b \in [0,1] \times [0,1]$, we can pick any two curves from a to b and see that the preimage of those two curves must intersect at some point besides $G(a)$ and $G(b)$, so G is not injective.

To define this mapping and prove this theorem, we first need to understand ways of dividing up the unit interval and unit square, so that we can map one to the other.

4.1 Quartic Intervals and Dyadic Squares

We begin by defining the quartic intervals and dyadic squares, and listing properties of the two that are useful for mapping one to the other.

Definition 4.3. The **quartic intervals** are the intervals constructed by repeatedly dividing the unit interval into fourths. Specifically, the k^{th} generation quartic intervals are all intervals of the form

$$\left[\tfrac{\ell}{4^k}, \tfrac{\ell+1}{4^k}\right] \qquad \text{for integer } \ell \text{ with } 0 \leq \ell < 4^k.$$

Definition 4.4. A **chain** of quartic intervals is a sequence of intervals

$$I_1 \supset I_2 \supset I_3 \supset \cdots$$

where for each k, I_k is a k^{th} generation quartic interval.

Quartic intervals have a few relevant properties.

Proposition 4.5.

1. For any chain of quartic intervals $\{I_k\}$, there exists a unique $t \in [0,1]$ such that $\{t\} = \cap_k I_k$.
2. For any $t \in [0,1]$ there exists a chain of quartic intervals $\{I_k\}$ such that $\{t\} = \cap_k I_k$.
3. The set of all $t \in [0,1]$ such that this chain is not unique has measure 0.

The proof of these properties is left as an exercise to the reader; it is very similar to some of the proofs in the homework problem about the Cantor-Lebesgue function.

Then, we can naturally represent each chain $\{I_k\}$ as a string "$.a_1a_2a_3\ldots$" where each $a_k = 0, 1, 2$, or 3, so that the point t such that $\{t\} = \cap_k I_k$ is exactly

$$t = \sum_k a_k 4^{-k}.$$

Note that the points t that don't have a unique representation are the ones for which after a certain k, all a_k's are 3, or after a certain k, all a_k's are 0.

We will define dyadic squares in a very similar way.

Definition 4.6. The **dyadic squares** are the squares constructed by repeatedly bisecting the sides of the unit square. Specifically, there are 4^k dyadic squares in the k^{th} generation, each with side length $(1/2)^k$.

Definition 4.7. A **chain** of dyadic squares is a sequence of squares

$$S_1 \supset S_2 \supset S_3 \supset \cdots,$$

where for each k, S_k is a k^{th} generation dyadic square.

Proposition 4.8.

1. For any chain $\{S_k\}$ of dyadic squares, there exists a unique point $x \in [0,1]\times[0,1]$ such that $\{x\} = \bigcap_k S_k$.
2. For any $x \in [0,1] \times [0,1]$, there exists a chain $\{S_k\}$ of dyadic squares such that $\{x\} = \bigcap_k S_k$.
3. The set of $x \in [0,1] \times [0,1]$ such that this chain is not unique has measure zero.

These proofs follow the exact format as for the quartic intervals, so they are left to the reader.

Moreover, as in the quartic intervals, each chain $\{S_k\}$ can be encoded as "$.b_1b_2b_3\ldots$" where each $b_k = 0, 1, 2$, or 3, and then the x such that $x = \bigcap_k S_k$ is exactly

$$x = \sum_k \bar{b}_k 2^{-k},$$

where

$$\bar{b}_k = \begin{cases} (0,0) & \text{if } b_k = 0 \\ (0,1) & \text{if } b_k = 1 \\ (1,0) & \text{if } b_k = 2 \\ (1,1) & \text{if } b_k = 3. \end{cases}$$

4.2 Dyadic Correspondence

Now, we will look at ways to map the quartic intervals to the dyadic squares. We will find that mappings that have specific properties fulfill most of the results of Theorem 4.1 and then we will define the Peano mapping as a specific mapping that has this property.

Definition 4.9. A **dyadic correspondence** is a mapping ϕ from quartic intervals to dyadic squares such that

1. The mapping ϕ is bijective.
2. If I is a quartic interval of the k^{th} generation, then $\phi(I)$ is a dyadic interval of the k^{th} generation.
3. If $I \subset J$ then $\phi(I) \subset \phi(J)$.

Definition 4.10. Given a dyadic correspondence ϕ, the **induced mapping** ϕ^* is a mapping from $[0,1] \to [0,1] \times [0,1]$ defined as follows: for $t \in [0,1]$, if $\{t\} = \bigcap_k I_k$ then

$$\phi^*(t) = \bigcap_k \phi(I_k),$$

which we can do since $\big\{\phi(I_k)\big\}$ is a chain of dyadic squares.

Our induced mapping ϕ^* is well-defined almost everywhere; it is well-defined when t corresponds to a unique chain $\{I_k\}$.

Moreover, we can see that whenever I is a quartic interval of the k^{th} generation, $\phi^*(I) = \phi(I)$ and $m_1(I) = 4^{-k} = m_2(\phi^*(I))$.

Now we can prove our big theorem about dyadic correspondences.

Theorem 4.11. For any dyadic correspondence ϕ, there exist measure-zero sets Z_1 and Z_2 such that

1. The induced mapping ϕ^* is a bijection from $[0,1] \setminus Z_1$ to $[0,1] \times [0,1] \setminus Z_2$.
2. For any subset E of $[0,1] \setminus Z_1$, E is measurable if and only if $\phi^*(E)$ is measurable, and $m_1(E) = m_2(\phi^*(E))$.

To prove the first part, we will define $\mathcal{N}_1$ to be the collection of chains of quartic intervals that correspond to a t with a non-unique chain, and $\mathcal{N}_2$ to be the collection of chains of dyadic squares that correspond to an x with a non-unique chain.

Then, since ϕ is a bijection from quartic intervals to dyadic squares, it is also a bijection from $\mathcal{N}_1 \cup \phi^{-1}(\mathcal{N}_2)$ to $\phi(\mathcal{N}_1) \cup \mathcal{N}_2$, and therefore it is a bijection from $(\mathcal{N}_1 \cup \phi^{-1}(\mathcal{N}_2))^{\mathrm{c}}$ to $(\phi(\mathcal{N}_1) \cup \mathcal{N}_2)^{\mathrm{c}}$. Then, we define Z_1 to be the subset of $[0,1]$ that corresponds to the chains in $\mathcal{N}_1 \cup \phi^{-1}(\mathcal{N}_2)$ and Z_2 to be the subset of $\phi(\mathcal{N}_1) \cup \mathcal{N}_2$.

Clearly, then ϕ^* is well-defined and a bijection from $[0,1] \setminus Z_1$ to $[0,1] \times [0,1] \setminus Z_2$. We now need to prove that Z_1 and Z_2 have measure zero, and to do so we use the following two lemmas.

Lemma 4.12. For a given sequence $\{f_k\}$ where each $f_k = 0, 1, 2,$ or 3, the set

$$E_0 = \left\{ x : x = \sum_k a_k 4^{-k} \text{ and there exists } r > 0 \text{ such that } a_k \neq f_k \text{ for all } k \geq r \right\}$$

has measure 0.

Proof. We can see that for any given $r > 0$ the set $\{x : a_r \neq f_r\}$ has measure $3/4$, so the set

$$\{x : a_r \neq f_r \text{ and } a_{r+1} \neq f_{r+1}\}$$

has measure $(3/4)^2$, and inductively we get that

$$m(\{x : a_k \neq f_k \text{ for all } k \geq r\}) = \lim_{n\to\infty} (3/4)^n = 0.$$

Then, since E_0 is the countable union of such sets, $m(E_0) = 0$ as well. □

We will not prove the following lemma, but it follows the same proof as above.

Lemma 4.13. For a given sequence $\{f_k\}$ where each $f_k = 0, 1, 2,$ or 3, the set

$$E_0 = \left\{ x : x = \sum_k \bar{b}_k 2^{-k} \text{ and there exists } r > 0 \text{ such that } b_k \neq f_k \text{ for all } k \geq r \right\}$$

has measure 0.

Then, since all elements of $\mathcal{N}_1$ correspond to strings where all $a_k = 0$ for sufficiently large k or all $a_k = 3$ for sufficiently large k, we can apply Lemma 4.12 to the sequence $\{1\}$ to see that the set of points corresponding to chains in $\mathcal{N}_1$ is a measure-zero set.

Similarly, if we consider the strings $\mathcal{N}_2$ can correspond to, we find that applying Lemma 4.13 to the sequence $1, 2, 1, 2, \ldots$ tells us that the set of points corresponding to chains in $\mathcal{N}_2$ has measure zero. We can similarly show that $\phi(\mathcal{N}_1)$ and $\phi^{-1}(\mathcal{N}_2)$ have measure zero. Thus, we have shown the first part of Theorem 4.11.

Now, we will show that on the interval where ϕ^* is defined, it is measure-preserving. First, we know from lecture that any open set $\mathcal{O}$ in the unit interval can be expressed as a countable union of closed intervals $\bigcup_n I_n$ such that the closed intervals have disjoint interiors. Moreover, inspection of the proof of this theorem shows us that furthermore, we can force each I_n to be quartic intervals. Similarly, we can show that any open set in the unit square can be expressed as the countable union of dyadic squares whose interiors are disjoint.

Then, for any measure-zero set $E \subset [0,1] \setminus Z_1$, for any $\epsilon > 0$ we can cover E by quartic intervals $\{I_j\}$ such that $\sum_j m_1(I_j) < \epsilon$. Then, since $\phi^*(E) \subset \bigcup_j \phi^*(I_j)$,

$$m_2(\phi^*(E)) \leq \sum_j m_2(\phi^*(I_j)) = \sum_j m_1(I_j) < \epsilon,$$

so taking the limit as ϵ approaches 0, we get that $m_2(\phi^*(E)) = 0$ so $\phi^*(E)$ is measurable. By a similar argument, we can see that $(\phi^*)^{-1}$ maps measure zero sets in $[0,1]\times[0,1]\setminus Z_2$ to measure zero sets in $[0,1]\setminus Z_1$.

But then since we know we can express any open set as the countable union of quartic intervals, we can use the same argument to see that for any open set $\mathcal{O} \subset [0,1]$, $\phi^*(\mathcal{O}\setminus Z_1)$ is measurable and $m_2(\phi^*(\mathcal{O}\setminus Z_1)) = m_1(\mathcal{O})$. Thus, ϕ^* is measure-preserving for G_δ subsets of $[0,1]$, and since every measurable set differs from a G_δ set by a measure zero set, we can combine the previous two paragraphs to see that for any measurable $E \subset [0,1]\setminus Z_1$, $\phi^*(E)$ is measurable and $m_2(\phi^*(E)) = m_1(E)$. A similar argument holds for $(\phi^*)^{-1}$, and so this mapping is measure-preserving. □

Now that we have proved the big theorem for dyadic mappings, we will look at the Peano mapping as a specific type of dyadic mapping.

4.3 Constructing the Peano Mapping

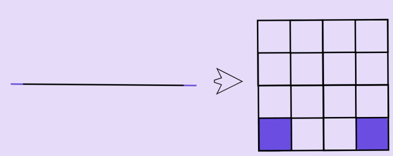
We will find that the Peano mapping is the unique way of tracing along the dyadic squares so that the k^{th} generation path travels in the same direction as the $(k-1)^{\text{th}}$ generation path and we only travel between dyadic squares that share a side. We will now formalize this definition.

Definition 4.14. We say that two quartic intervals of the same generation are **adjacent** if they share a point, and two dyadic squares of the same generation are **adjacent** if they share a side.

Now, we will construct the dyadic correspondence associated with the Peano mapping.

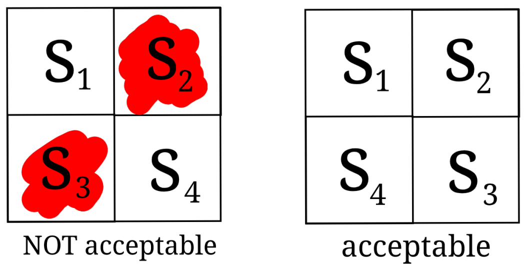
Lemma 4.15. There is a unique dyadic correspondence ϕ_P such that

1. If I and J are adjacent quartic intervals of the same generation, then $\phi_P(I)$ and $\phi_P(J)$ are adjacent dyadic squares of the same generation.
2. In each generation, if I is the left-most quartic interval of that generation then $\phi_P(I)$ is the bottom-left dyadic square of that generation, and if J is the right-most quartic interval of that generation then $\phi_P(J)$ is the bottom-right quartic interval of that generation.



To prove this, we need the following definition.

Definition 4.16. For any square S and its four sub-squares, an acceptable **traverse** of S is an ordering S_1, S_2, S_3, S_4 of the four sub-squares so that if one sub-square follows another in the ordering, the two sub-squares are adjacent.



We can see that if we color the subsquares like a checkerboard so that S_1 is white, then S_4 must be black, and our traverse alternates black and white squares.

Moreover, for any square S, if we are given a starting sub-square S_1 and an edge σ of S so that we have to end at a sub-square touching σ, there is only one valid traverse of S that fits this criteria.

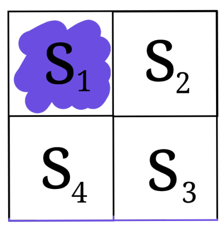
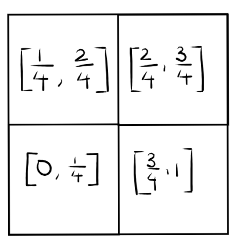


Figure 13: an acceptable traverse given a start square and an end edge

Now, we can go back to proving our lemma. First, it is clear that ϕ_P is defined for the first generation: it must be the mapping below.



Then, we can see that if ϕ_P is defined for all generations up to generation k, we can inductively define it for generation $k+1$. This is done as follows. First, number the k^{th} generation quartic intervals in increasing order as $I_1, I_2, \ldots, I_{4^k}$ and number the k^{th} generation dyadic squares as $S_j = \phi_P(I_j)$ for each $1 \le j \le 4^k$.

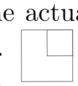
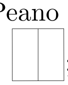
Then, we can divide I_1 into the four $(k+1)^{\text{th}}$ generation quartic intervals $I_{1,1}, \ldots I_{1,4}$ in increasing order, and then define the mapping $S_{1,j} = \phi_P(I_{1,j})$ where $S_{1,j}$ is the unique valid traverse of the four subsquares S_1 such that $S_{1,1}$ is the lower-left dyadic square of the unit square, and the traverse ends at the edge shared between S_1 and S_2. This is possible because by the inductive hypothesis S_1 and S_2 are adjacent and S_1 is the lower-left dyadic square.

Then, we similarly divide I_2 into its four quartic sub-intervals and define the mapping $S_{2,j} = \phi_P(I_{2,j})$ where $S_{2,j}$ is the unique traverse of S_2 where $S_{2,1}$ is the unique subsquare adjacent to $S_{1,4}$ and the traverse ends at the edge shared between S_2 and S_3.

We inductively repeat this process for each S_k for $1 \leq k \leq 4^k - 1$, where we can see that we always enter each S_k through a white subsquare and leave through a black subsquare.

Specifically, this means that we will enter S_{4^k} from a white subsquare, so the unique tranverse that ends at the bottom edge of S_{4^k} must leave through the black subsquare touching the bottom edge, which is the bottom-right dyadic square of the unit cube, which is what we needed it to be.

Thus, we have inductively created a valid ϕ_P, and since the tranverses we used were unique, this ϕ_P is unique and we have proved our lemma. □

Then, we can construct the actual Peano curve. For each generation k, the curve $\mathcal{P}_k$ though each dyadic square is some rotation of  or ; specificially, the rotation that connects the previous and next dyadic square in our ordering.

Specifically, for each generation k, we map the center of the interval I_j to the center of the square S_j, where the ordering of the squares is defined as in the previous lemma, and we map the start and end of the interval to the middle of the left edge of the bottom-left square and the middle of the right edge of the bottom right-square, respectively. Then, we play connect-the-dots to define the rest of the map $\mathcal{P}_k$, connecting these points linearly in the given order of the dyadic squares.

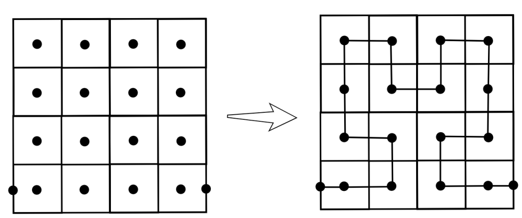


Figure 14: creating $\mathcal{P}_2$ from our second generation of quartic intervals and dyadic squares

Then, since each of these are continuous functions, and for any $t \in [0,1]$, since $\mathcal{P}_k(t)$ and $\mathcal{P}_{k+1}(t)$ are both in the same k^{th} generation dyadic square, we get that

$$\left|\mathcal{P}_k(t) - \mathcal{P}_{k+1}(t)\right| \leq \frac{\sqrt{2}}{2^{-k}},$$

so we can define

$$\mathcal{P}(t) = \lim_{k\to\infty} \mathcal{P}_k(t) = \mathcal{P}_1(t) + \sum_{k=1}^{\infty} \mathcal{P}_{k+1}(t) - \mathcal{P}_k(t),$$

and since this converges absolutely and uniformly, $\mathcal{P}$ exists and is also a continuous function.

Moreover, since $\mathcal{P}_k(t)$ visits every k^{th} generation dyadic square, $\mathcal{P}$ is dense in the unit square, and since it is also continuous, it must be a surjective mapping from $[0,1]$ to $[0,1] \times [0,1]$. This proves part of Theorem 4.1. To prove the measure-preserving part, we will show that $\mathcal{P}$ is exactly ϕ_P^*.

Lemma 4.17. For every $0 \leq t \leq 1$, $\phi_P^*(t) = \mathcal{P}(t)$.

Proof. First, we can see that ϕ_P^* is well-defined for every t. Specifically, we can see that if $t = \bigcap_k I_k$ and $t = \bigcap_k J_k$ for two distinct chains of quartic intervals, then for all sufficiently large k I_k and J_k must be adjacent. But that means that $\phi_P(I_k)$ is adjacent to $\phi_P(J_k)$ for all sufficiently large k, so $\bigcap_k \phi_P(I_k) = \bigcap_k \phi_P(J_k)$.

Then, we can see that by our definition of $\mathcal{P}$, for all t,

$$\bigcap_k \phi_P(I_k) = \lim_k \mathcal{P}_k(t) = \mathcal{P}(t),$$

so the two are equal. □

We can see moreover that this implies $\phi_P(I) = \mathcal{P}(I)$ for any quartic interval I. But then, since any interval (a, b) can be written as $\bigcup_j I_j$ where the I_j are quartic intervals with disjoint interiors, and then $\mathcal{P}(I_j) = \phi_P(I_j)$ must be dyadic squares with disjoint interiors and $\mathcal{P}((a, b)) = \bigcup_j \mathcal{P}(I_j)$, we get that

$$m_2(\mathcal{P}((a,b))) = \sum_j m_2(\mathcal{P}(I_j)) = \sum_j m_1(I_j) = m_1((a,b)) = b - a.$$

This proves the measure-preserving part of Theorem 4.1, and we are done. □

Appendices

A Exterior Measures

The content of this appendix comes from Chapter 6 of Stein and Sakarchi.
As a reminder, we will first review the definition of an exterior measure.

Definition A.1. An exterior measure is a function μ_* defined on a set X that has the following properties:

- The exterior measure of the empty set, $\mu_*(\emptyset)$, is zero.
- (monotonicity) For any sets $E_1 \subset E_2 \subset X$ $\mu_*(E_1) \leq \mu_*(E_2)$.
- (countable sub-additivity) For a countable family of sets $\{E_i\}$ where each $E_i \subset X$,

$$\mu_*\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mu_*(E_i).$$

Moreover, recall that we proved in class the following theorem.

Theorem A.2. For any exterior measure μ_* on a set X, the collection $\mathcal{M}$ of Carathéodory sets on X forms a σ-algebra, and μ_* restricted to $\mathcal{M}$ is a measure.

Then, on sets X that are metric spaces - that is, there is a distance function d defined on the space that fulfills all three properties of a distance - we can define metric exterior measures.

Definition A.3. A **metric exterior measure** is an exterior measure μ_* defined on a metric space X such that

$$\mu_*(A \cup B) = \mu_*(A) + \mu_*(B) \qquad \text{whenever } d(A, B) > 0,$$

where we define the distance between sets A and B to be

$$d(A, B) = \inf \left\{ d(x, y) | x \in A, y \in B \right\}.$$

Using just these properties of a metric exterior measure, we can prove the following theorem.

Theorem A.4. If μ_* is a metric exterior measure on a metric space X, then the Borel sets $\mathcal{B}_X$ are measurable, and therefore μ_* restricted to $\mathcal{B}_X$ is a measure.

Proof. Since we know that measurability is closed under countable unions and intersections, and measure-zero sets are measurable, we simply need to show that closed sets are Carathéodory measurable.

Then, for an arbritrary closed set F and an arbritary set $A \subset X$ such that $\mu_*(A) < \infty$, we split A into subsets A_n for each positive integer n, such that

$$A_n = \left\{ x \in A : d(x, F) \geq \frac{1}{n} \right\}.$$

Then, clearly for each n we have that $A_n \subset A_{n+1}$, and since F is closed, $F^c \cap A = \bigcup_n A_n$.

Then, for each n, since $d(F, A_n) > 0$ and μ_* is a metric exterior measure,

$$\mu_*(A) \geq \mu_*((F \cap A) \cup A_n) = \mu_*(F \cap A) + \mu_*(A_n).$$

We want to take the limit as n tends to infinity to get one side of Carathéodory measurability, but to do that, we need

$$\lim_{n\to\infty} \mu_*(A_n) = \mu_*(F^c \cap A).$$

To do that, we will divide A into "onion rings;" that is, define $B_n = A_{n+1} \cap A_n^c$, so $B_n = \left\{x \in A : \frac{1}{n+1} \leq d(x, F) < \frac{1}{n}\right\}$. Then, the triangle inequality tells us that

$$d(B_{n+1}, A_n) \geq \frac{1}{n(n+1)},$$

so since μ_* is a metric exterior measure,

$$\mu_*(A_{2k+1}) \geq \mu_*(A_{2k-1}) + \mu_*(B_{2k}).$$

But repeating this tells us that

$$\mu_*(A_{2k+1}) \geq \sum_{j=1}^{k} \mu_*(B_{2j}),$$

and similarly

$$\mu_*(A_{2k}) \geq \sum_{j=1}^{k} \mu(B_{2j-1}).$$

But since $\mu_*(A)$ is finite, both sums $\sum_{j=1}^{\infty} \mu(B_{2j-1})$ and $\sum_{j=1}^{\infty} \mu(B_{2j})$ must be convergent.

This tells us that we can take the limit of

$$\mu_*(A_n) \leq \mu_*(F^c \cap A) \leq \mu_*(A_n) + \sum_{j=n+1}^{\infty} B_j$$

to get that $\lim_{n\to\infty} \mu_*(A_n) = \mu_*(F^c \cap A)$.

Then, we can return to the inequality $\mu_*(A) \geq \mu_*(F \cap A) + \mu_*(A_n)$; taking the limit as n goes to infinity gives us $\mu_*(A) \geq \mu_*(F \cap A) + \mu_*(F^c \cap A)$.

The other direction follows directly from countable subadditivity, so all closed sets F are Carathéodory measurable, and therefore all Borel sets are measurable. □

B The Cantor Set

The description of the Cantor set is sourced from the first homework, and the description of the Cantor-Lebesgue function is sourced from Chapter 3 of Stein and Shakarchi.

To construct the Cantor set, we begin with the unit interval $[0, 1]$, which we will call C_0. Then, we define C_1 to be the union of two intervals $\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. We repeat this process indefinitely, so that to construct C_n we remove the middle third of each interval in C_{n-1}, and each set C_n consists of 2^n closed intervals of length 3^{-n}.

Figure 15: the first steps in creating the Cantor set

Definition B.1. Then, we define the **Cantor set** $\mathcal{C}$ to be $\bigcap_n C_n$.

Since this is the intersection of decreasing compact sets, we know that $\mathcal{C}$ is also a compact set and it is nonempty because, for example, the point $\frac{1}{3}$ is in $\mathcal{C}$.

The Cantor set is uncountable but has measure zero; proving this will be left as an exercise to the reader.

We then turn to defining the Cantor-Lebesgue function by first defining functions $F_n : [0, 1] \to [0, 1]$ based on our sets C_n. We want each $F_n(x)$ to be linearly increasing on C_n and constant on C_n^c. Specifically, we know since C_n is the union of 2^n disjoint closed sets, C_n^c is the union of $2^n - 1$ disjoint open sets, so we define $F_n(x)$ such that if x is in the k^{th} open interval in C_n^c, $F_n(x) = k/2^n$. Moreover, $F_n(0) = 0$ and $F_n(1) = 1$. For example,

$$F_1(x) = \begin{cases} \frac{3}{2}x & \text{if } 0 \leq x \leq \frac{1}{3} \\ \frac{1}{2} & \text{if } \frac{1}{3} < x < \frac{2}{3} \\ \frac{3}{2}x - \frac{2}{3} & \text{if } \frac{2}{3} \leq x \leq 1. \end{cases}$$

Definition B.2. Then, we define the **Cantor-Lebesgue function** to be the function $F : [0, 1] \to [0, 1]$ such that $F(x) = \lim_{n\to\infty} F_n(x)$.

Since $\{F_n\}$ is a sequence of continuous increasing functions and for each x,

$$\left|F_n(x) - F_{n-1}(x)\right| \leq 2^{-n-1},$$

so these functions converge uniformly, we can see that F is continuous and increasing. Moreover, $F(0) = 0$, $F(1) = 1$, and F is constant on every interval in $\mathcal{C}^c$.