CANTOR SETS IN TOPOLOGY, ANALYSIS, AND FINANCIAL MARKETS

ZIQIAN (ALEXA) JIN

ABSTRACT. Cantor sets provide intriguing examples, counterexamples, and illustrations of a variety of concepts in different mathematical fields. This paper explores the applications of Cantor sets – the Cantor ternary set in particular – to the areas of topology, measure theory, analysis and the real world. Several notable topics discussed include homeomorphism, dimensions, Cantor functions, and the fractal market hypothesis.

CONTENTS

1. Introduction 1
1.1. The standard Cantor set 1
1.2. Ternary representation 2
1.3. Generalization of the standard Cantor set 3
2. Topological properties 4
2.1. Basic topological properties 4
2.2. Nowhere dense and totally disconnected 5
2.3. Homeoporphism 5
3. Applications in measure theory 7
3.1. The measure of Cantor sets 7
3.2. Dimension 8
4. Applications in analysis 10
4.1. Cantor functions 10
4.2. Volterra's Function 11
5. Applications in the real world: fractal phenomena 12
5.1. Fractal Geometry in Nature 12
5.2. Mandelbrot and the Fractal Market 13
5.3. The fractal market hypothesis 14
Acknowledgments 15
References 15

1. INTRODUCTION

1.1. **The standard Cantor set.** Discovered in 1874 by Henry John Stephen Smith and introduced by German mathematician Georg Cantor in 1883, the Cantor ternary set, also called the standard Cantor set, C is created by iteratively deleting the open middle third from a set of line segments. One starts by deleting the

Date: DEADLINES: Draft AUGUST 14 and Final version AUGUST 28, 2021.

open middle third $\left(\frac{1}{3}, \frac{2}{3}\right)$ from the unit interval $[0, 1]$, leaving two line segments: $\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. Next, the open middle third of each of these remaining segments is deleted, leaving four line segments: $\left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$. This process is continued ad infinitum, where the nth set is

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right)$$

for $n \geq 1$, and $C_0 = [0, 1]$. The standard Cantor set is formed by the intersection of all the C_n, i.e. $C = \bigcap_{n=0}^{\infty} C_n$.

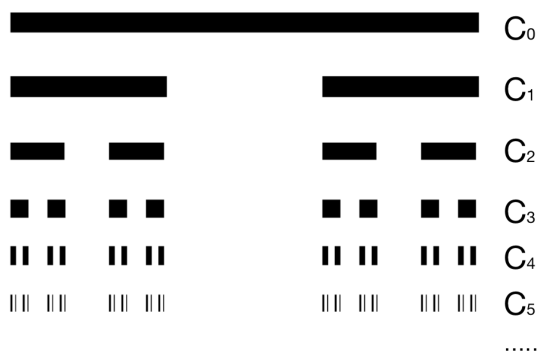


FIGURE 1. Creation of the standard Cantor set

1.2. **Ternary representation.**

Theorem 1. *The standard Cantor set consists of exactly those numbers in [0, 1] that can be written in base 3 without 1's.*

Note that this is sometimes given as an alternative, though perhaps less intuitive, definition of the standard Cantor set.

Proof. We begin by showing that the standard Cantor set only consists of such elements:

Write all numbers from the unit interval in base 3. For instance, the number 0.25 in base 10 is written in base 3 as 0.020202... . If a number can be written using only 0 and 2, consider the latter representation of it. For instance, the number 0.1 in base three is the same as 0.022222... in base three. It is easy to check that .1 is in the standard Cantor set, which is why we consider it represented as .022222... .

We proceed by induction, starting with the base case: those in C_1 are either smaller than 1/3 or bigger than 2/3. Therefore, their first decimal digits are either 0 or 2.

Now, the inductive step: suppose all elements in C_n can be written in base 3 with only 0 and 2 in the first n decimal places. Then since $C_{n+1} = \frac{C_n}{3} \cup \left(\frac{2}{3} + \frac{C_n}{3}\right)$ for $n \geq 1$, we deduce that all elements in C_{n+1} can be written in base 3 with only 0 and 2 on the first $n + 1$ decimal places. Specifically, their 2 to $(n + 1)$ digits are copied from the 1 to n digits of the corresponding elements in C_n, and the first digit is either 0 or 2 depending on whether they belong to the left or right portion of the Cantor set.

Supposing that there is any 1 that appears in the ternary representation of an element in the Cantor set, it would have to appear in a certain place, say the kth digit. This is impossible by the above induction, which gives that there is no digit where a 1 can appear.

Now for the other direction, we want to show by induction that all the numbers in [0, 1] that can be written in base-3 without 1's are in the standard Cantor set.

For the base case, we know that all the numbers in [0, 1] that are written base 3 with either 0 or 2 in the first decimal digit are in C_1, since all the numbers in $[0, 1/3]$ have a 0 as the first digit of their ternary representation (writing $1/3 = .0222...$), and all the numbers in $[2/3, 1]$ have a 2 as the first digit (writing $1 = .222$).

Now, suppose all numbers that can be written in base 3 with only 0 and 2 on the first n decimal places are in C_n. Now, we take out from these elements those that have a 1 in the $n+1$th decimal place. then we see all the remaining numbers are in C_{n+1} $C_{n+1} = \dfrac{C_n}{3} \cup \left(\dfrac{2}{3} + \dfrac{C_n}{3}\right)$ for $n \geq 1$. If a number has no 1 in its ternary, there will be no C_k which it doesn't appear in, so it will be in C, the intersection of all C_k.

□

Theorem 2. *The standard Cantor set is uncountable*

One of the most notable properties of the standard Cantor set is that it contains uncountably many points. This is despite C being very small in that it is the result of an intersection of countably many decreasing sets, a notion we will make more precise later.

Proof. First, consider every element of C in its ternary representation. Supposing the standard Cantor set is countable, then there is a listing of all the elements of C, for instance:

0.200200222...,
0.002202002...,
0.202220002...,
0.002020200...,
...

Now, construct a new element by swapping the nth decimal digit of the nth element in the list from either 0 to 2 or 2 to 0. In this case the new element is 0.0202..., which is different from any existing element in the list. However, by the above theorem, this new number must also be in C, which contradicts that we had listed all elements of C. Therefore, it is impossible to find a complete enumeration of all the elements in the standard Cantor set, so it must be uncountable. Note that this is very similar to the common diagonalization argument which shows that $\mathbb{R}$ is uncountable.

□

1.3. **Generalization of the standard Cantor set.** The word "ternary" in the standard Cantor set meant that the open middle 1/3 of each interval was being removed during each step of the construction. However, we could have selected a different ratio to remove at each step, say 1/2 or 1/4. In fact, we could even vary the ratio we remove in each step, not remove the same amount in each "branch" of the set on a given step, or not remove intervals from the exact center of the previous intervals. Then, as before, we get a decreasing sequence of sets C_k, and define the generalized Cantor set to be their intersection, C. The figure below provides an example of such a sequence of removed intervals. To obtain a set that is similar enough to our prototypical ternary Cantor set, we have one additional restriction: C cannot contain any intervals.

Although Cantor sets can be defined even more generally, we will limit our discussion in this paper to bounded subsets of $\mathbb{R}$.

FIGURE 2. Example of the first four iterations of a generalized Cantor set

Two notable variants of the standard Cantor set are the "fat" Cantor sets and the "thin" Cantor sets, which we will dive into later in the paper.

2. Topological properties

Now, we prove several topological properties of the standard Cantor set. At the end of this section, we will show that all Cantor sets, as we have defined them, are homeomorphic to each other, which implies that all Cantor set possess these topological properties.

2.1. **Basic topological properties.**

Theorem 3. *The standard Cantor set is closed.*

Proof. The Cantor set is an intersection of countably many sets, each of which is a finite union of closed intervals, so closed itself. Thus, the standard Cantor set is the intersection of countably many closed sets, which implies it is closed. □

Theorem 4. *The standard Cantor set is compact.*

Proof. We already have that the standard Cantor set is closed. Furthermore, the set has an upper bound of 1 and a lower bound of 0. Hence, the Cantor set is closed and bounded, and by the Heine-Borel theorem, which states that a subset of $\mathbb{R}$ is compact if it is closed and bounded, it follows that the standard Cantor set is compact. □

Definition 2.1. The set S is perfect if $S = S'$, where S' denotes the set of all limit points of S. [3]

Since a closed set contains all of its limit points, this definition is equivalent to every point of a closed set being a limit point. We already have that the Cantor set is closed, so we only need to show every point of C is a limit point.

Theorem 5. *Every point in the standard Cantor set is a limit point.*

Proof. Given any point x in the Cantor set, there are 2 situations for its ternary representation:

Case 1: There are a finite number of digits. Suppose the number of digits is n. Approach x through the sequence $x + 2/3^{n+1}$, $x + 2/3^{n+2}$, $x + 2/3^{n+3}$,..., where all the terms are also elements in the standard Cantor set. Therefore, x is a limit point of C.

Case 2: There are infinite digits. Suppose the point is written in base-3 as $0.x_1x_2x_3x_4$.... As we have shown, each x_k is either 0 or 2. Now, x through the sequence $0.x_1$, $0.x_1x_2$, $0.x_1x_2x_3$, $0.x_1x_2x_3x_4$,... (where x_k is the kth digit). All the terms in the sequence that converges to x are also elements in the standard Cantor set, since x_k is either 0 or 2. As a result, x is a limit point of the standard Cantor set □

2.2. **Nowhere dense and totally disconnected.**

Definition 2.2. A set is nowhere dense if its closure has empty interior. [3]

This definition expresses the idea that a set is not "tightly clustered" in any location.

Theorem 6. *The standard Cantor set is nowhere dense.*

Proof. We have shown the Cantor set is closed, so its closure is itself. Suppose a subset of the Cantor set is dense, then the subset contains at least one interval, denoted as $A = [a, b]$ where $a \geq 0$ and $b \leq 1$ ($a, b \in R$, $a < b$). From theorem 2.2, we know that written in base 3, both a and b are composed of only 0 and 2, given that the two points are both in the Cantor set. Now, locate the first from left decimal digit that a and b differ from each other and replace it with a 1. For example, if a=0.20202... and b=0.20222... in base, then let c = 0.2021... in base 3 where the remaining digits don't really matter. Since the first 3 digits of a, b, and c are the same, we are able to argue that $a < c < b$ and that c is not in the ternary Cantor set since it contains the digit 1. A contradiction is therefore achieved, and the Cantor set is nowhere dense. □

Definition 2.3 (Totally Disconnected)**.** A totally disconnected space is a topological space that has no non-trivial connected subsets. In other words, the only connected components in any totally disconnected space X are the one-point sets.

Theorem 7. *The standard Cantor set is totally disconnected.*

Proof. This proceeds from what we established in the previous proof. Specifically, any two elements of the standard Cantor set are separated by at least one point not in C. If this is the case, no two distinct points can be part of the same connected component, so the set is totally disconnected. □

2.3. **Homeoporphism.**

Definition 2.4. A homeomorphism is a continuous bijection between topological spaces that has a continuous inverse [3]. Homeomorphism is an important concept in topology, since it expresses a notion of topological equivalence. Thus, two sets which are homeomorphic share many topological properties.

Theorem 8. *All Cantor sets are homeomorphic to each other.*

Proof. Given two Cantor sets C and C' on the unit interval, suppose they are constructed by the intersection of $C_0, C_1, C_2...$ and $C'_0, C'_1, C'_2....$ let f_0 be the linear map bijection from C_0 to C'_0, both of which are entire intervals, sending endpoint to endpoint. $f_0(x)$ is continuous within its domain. Similarly, as shown in figure 3, let f_1 be the combination of linear map from the left interval of C_1 to left interval C'_1, and likewise for the right intervals..., and let f_k analogue for the kth sets. All these maps are continuous, because they are continuous on disjoint closed intervals.

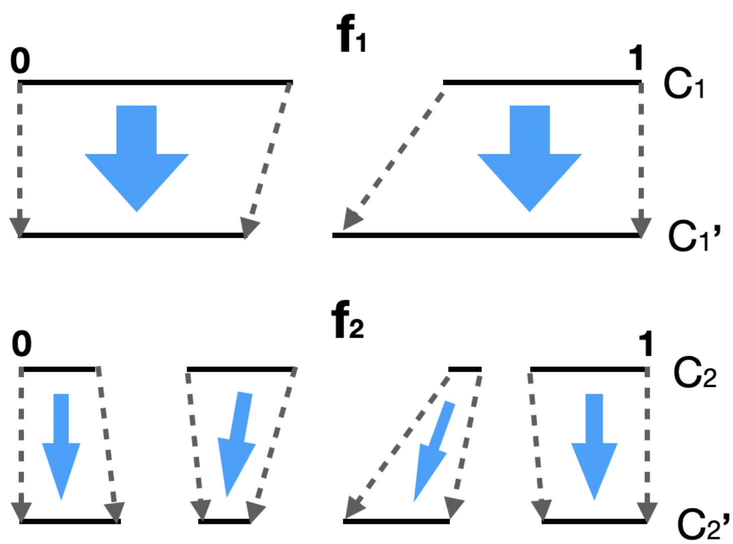


FIGURE 3. Examples of the maps

Now, define g_k as the restriction of f_k that maps only C to C_k. Since the domain C is a subset of all C_k for $k \in \mathbb{N}$, we derive that g_k is continuous for all $k \in \mathbb{N}$. We want to show that these g_k converge uniformly to some map g. The range of g will have to be the intersection of all C'_k, so C'.

Denote M_k as the one among 2^k intervals of C'_k with the maximum length. Since the Cantor sets are always nowhere dense, we deduce that $\lim_{k\to\infty} M_k = 0$. Specifically, if the value of $\lim_{k\to\infty} M_k = 0$ is positive, then a subset of the Cantor set contains at least one interval, contradicting that its closure has empty interior. Now denote N_k as the supremum of $|g_k - g_m|$ on the entire unit interval for any $m > k$. Therefore, given $\epsilon > 0$, there always exists K such that for all $k > K$, $|g_k - g_m| \leq N_k \leq M_k < \epsilon$ for all $x \in C$ and $m > k$. The sequence is Cauchy in the uniform norm, so it uniformly converges to the desired function g.

We have successfully shown that g_n converge uniformly to g. g is, therefore, a continuous map from C to C'. By the same token, we are able to construct a continuous map from C' to C by simply reversing the positions of C and C' and keeping all other aspects of our argument the same.

As a result, in order for there to be a continuous bijection between C and C', we only need to prove that the map from C to C' is bijective. In fact, each element in C and C' can be considered as an infinite sequence of L and R, where L stands for choosing the left interval and R stands for choosing the right interval in a given iteration. A bijective map between a point in C and one in C' can be established if they have identical L/R sequences, but it is easy to see that g_k gives these identical sequences for the first k L/R choices, so g itself gives the desired mapping for all of the infinitely many L/R choices.

Therefore, we have proven that Cantor sets C and C' are homeomorphic to each other. Since C and C' can be any arbitrary Cantor sets, we deduce that all the Cantor sets are homeomorphic to each other. □

3. Applications in measure theory

3.1. **The measure of Cantor sets.**

3.1.1. *Definition of Lebesgue measure.* Now that we have discussed the topological properties of Cantor sets, it is fundamental question also to ask how "big" they are. This idea is trivial for finitely many disjoint intervals – just add up the lengths – yet in the infinite case is somewhat more complicated. The concept of the Lebesgue measure, one particularly useful type of measure in mathematics, is basically the total length of the shortest possible intervals that encapsulate a given subset. A full discussion of this measure is beyond the scope of this paper, but it suffices to note that it gives a more rigorous notion of size to sets.

The Lebesgue measure on $\mathbb{R}$ satisfy the following properties:

1. $m(A) \geq 0$
2. $m(\emptyset) = 0$
3. $m([a, b]) = b - a$
4. It is countably additive. Namely, for all countable collections $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in Σ,

$$m\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m\left(E_k\right)$$

As an immediate consequence of properties 1 and 4, if $A \subseteq B$ then $m(A) \leq m(B)$. It is easy to check that this consequence along with property 3 implies that points have measure 0, and in fact, countable sets also have measure 0. We can also calculate the measure of the standard Cantor set.

3.1.2. *The standard Cantor set.* Since we remove the middle 1/3 of each remaining interval in each iteration, the Lebesgue measure of C_n is $(2/3)^n$ (2^n intervals each of length 3^{-n}). Each C_n contains C, so the measure of C is no larger than that of any C_n. Taking the limit of it as n goes to infinity gives us zero, which is a fairly counterintuitive result: countable sets all have measure zero, but the Cantor set gives an example of a set that is uncountable and also measure zero.

Now, generalizing the standard Cantor set can lead to even more counter-intuitive results. We begin with a theorem.

Theorem 9. *There exists a nowhere dense set with positive measure.*

This theorem can be illustrated by the following category of Cantor sets.

3.1.3. *Fat Cantor sets.* Instead of removing a constant portion of the original set in each iteration, fat Cantor sets are created by removing progressively smaller portions of the original set in each step such that the ratio of what is being removed to the interval it is being removed *from* goes to 0 as n goes to infinity.

Ex: remove the middle $(1/k)^n$ of C_{n-1}, where $k > 3$.

FIGURE 4. Example of a fat Cantor set

Unlike the standard ternary cantor sets, these fat Cantor sets have a positive measure, which is odd because they are nowhere dense and don't contain even one interval. Take the example mentioned earlier that removes the middle intervals of lengths $(1/k)^n$ from C_{n-1}, $k > 3$.

The Lebesgue measure of the removed intervals
$= 1/k + (1/k)^2 * 2 + (1/k)^3 * 4 + ...$
$= 1/2 * (2/k + (2/k)^2 + (2/k)^3 + ...)$
$= 1/2 * 2/k * 1/(1 - 2/k)$
$= (1/k) * (k/k - 2)$
$= 1/(k - 2)$.
Therefore, the Lebesgue measure of the corresponding fat Cantor set is $(k-3)/(k-2)$

An example of the fat Cantor set is the Smith–Volterra–Cantor set (SVC): $k = 4$ in this case, and its Lebesgue measure is $1/2$.

3.2. **Dimension.**

3.2.1. *Definition.* In mathematics, the notion of fractional dimension, an intrinsic property of a set, is an extension of the idea that a line is one-dimensional, a plane is two-dimensional, and space is three-dimensional. First, let us explore one way to approach how the dimensions of, say, a line segment and a rectangle are defined.

A line segment has dimension 1, because as we stretch it to twice its original length, its 'substance'–length–doubles as well. In the case of a rectangle, if we stretch all sides to twice their original scales, its substance–namely the area–quadruples. Taking the logarithm of 4 over 2 gives us 2.

Put in an equation, we can write that

$$S_1/S_2 = S^D$$

where S_1 is the new substance, S_2 is the old substance, S stands for the stretch, and D is the dimension. The dimension is the exponent by which the size changes when scaled by a certain amount.

As in the two examples above, you might expect that only integer dimensions are taken. As will be shown below, however, the dimension of mathematical objects are not necessarily integers and can take on many arbitrary values.

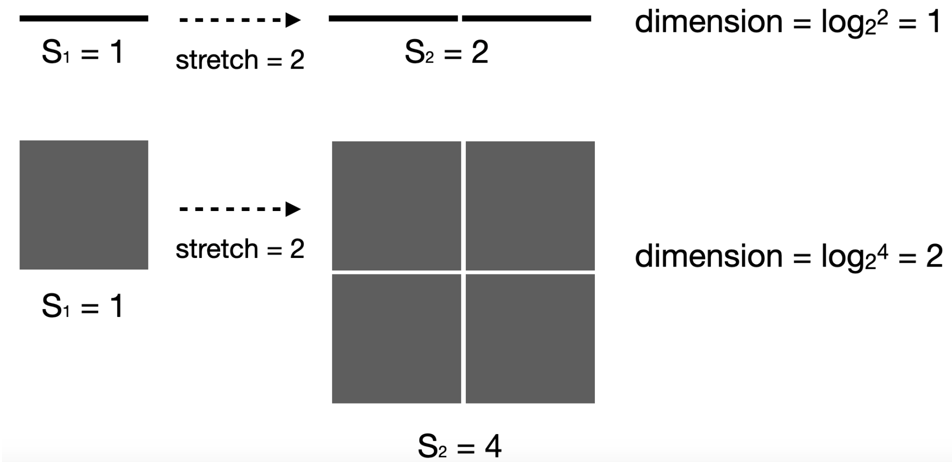


FIGURE 5. dimensions of line segments and squares

3.2.2. *Dimension of Cantor sets.*

Theorem 10. *The dimension of the standard Cantor set is equal to* $log(2)/log(3)$ $= 0.631$.

Proof. Because of its self-similar nature, the left third of the object is the exact replica of it as a whole and is of 1/3 of its original scale. So, the standard Cantor set is just the union of two smaller identical Cantor sets. When you scale by a factor of 3, the 'substance' of the set doubles, because you now have two copies identical to the original. Hence, if the dimension is d, then we may write $3^d = 2$, giving $d = log2/log3$ □

Cantor sets serve as some of the easiest examples of objects of non-integer dimension, and in fact, one can slightly modify the construction of the standard Cantor set to achieve a dimension equal to *any* value in (0, 1).

Theorem 11. *For any* $\alpha \in (0, 1)$, *there is a set of dimension* α

Proof. Let's start with the unit interval as usual. If we take out the middle 1/k of each existing interval in each iteration, then the left $(1 - 1/k)/2$ of the set is the exact replica of it as a whole and is of 1/2 of its original scale. Then:

$$dim = log(2)/log(2/(1 - 1/k)) = log(2)/(log(2) - log(1 - 1/k))$$

The dimension approaches 0 as k approaches 1, and it approaches 1 as k approaches infinity. Since k can be any number greater than 1 in $\mathbb{R}$, and the expression for dimension is continuous, we know that there are sets of any dimension from 0 to 1. □

In fact, we can actually achieve a dimension of 0 using a Cantor set, even though all of the above sets have positive dimension.

Theorem 12. *There exists an uncountable set of dimension zero*

This theorem can be illustrated by another type of generalized Cantor sets called the thin Cantor sets. First let's have the definition.

3.2.3. *The Thin Cantor set.* The thin Cantor set is created, contrary to the fat counterparts, by removing progressively larger portion of the original set in each step. An example would be removing the middle (1-1/n) of C_{n-2} in each iteration. Similar to the standard Cantor set, the measure of thin Cantor sets is zero. Now, we want to show that the dimension of thin Cantor sets is also zero

Proof. Suppose we remove the middle (1-1/n) of C_{n-2}, denoting $M_n = 1 - 1/n$. As n goes to infinity, M_n goes to 1. If we have any Cantor set used in the previous theorem, for n sufficiently large, more of each interval is being deleted in the thin Cantor set after step n than the constant ratio being deleted. Thus, the dimension of a thin Cantor set cannot be larger than the dimension of *any* of the middle 1/k sets, which means the dimension has to be 0, since it is not larger than any number in (0, 1). □

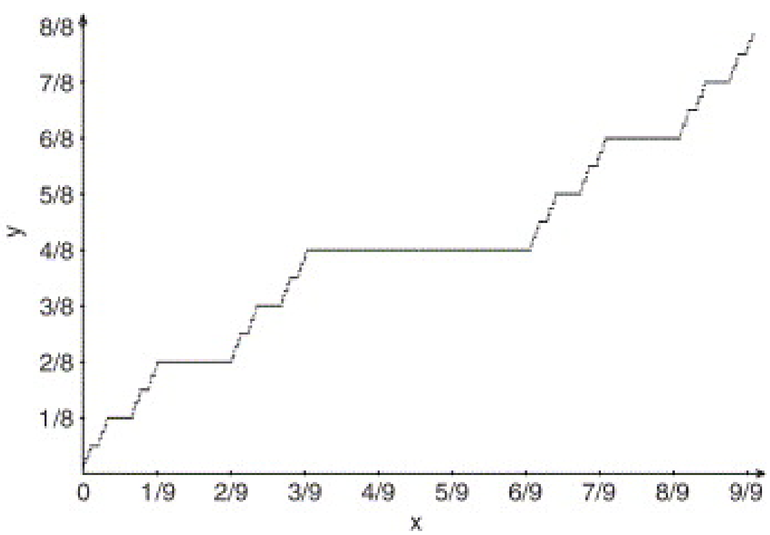
In a similar manner, it can be argued that fat Cantor sets have dimension 1.

4. Applications in analysis

4.1. **Cantor functions.** The Cantor function is an example of a function that is continuous, even uniformly continuous, but fails to satisfy the stronger definition of absolute continuity. It is a notorious counterexample in analysis, because it challenges our intuitions about continuity, derivative, and measure.

The standard Cantor function, or the Cantor ternary function c : [0,1] → [0,1] is defined as follows:

1. Express x in base 3.
2. If x contains a 1, replace every digit strictly after the first 1 by 0.
3. Replace any remaining 2s before the 1 with 1s.
4. Interpret the result in binary.



The result is a ladder-like non-decreasing function that exhibits a point symmetry across (1/2, 1/2).

To find the values of individual points in the domain, for example, let x = 0.25. Then x = 0.020202... in base 3. Since there is no 1 in the ternary representation, replacing all 2s with 1s gives us x' = 0.010101... Now, we convert it back from base 2 to base 10. x' = 1/4 + 1/16 + 1/64 +... = 1/4/(1-1/4) = 1/3

4.1.1. *Continuity.*

Theorem 13. *The Cantor function is uniformly continuous.*

Proof. Given $\epsilon > 0$, suppose n is the smallest integer such that $\epsilon > 1/2^n$. Let $\delta = 1/3^n$. Then given $x_0 \in D = [0, 1]$, we have that for any $x \in (x_0 - \delta, x_o + \delta)$, $|c(x) - c(x_0)| < 1/2^n = \epsilon$. Therefore, the Cantor function is uniformly continuous. □

4.1.2. *Absolutely Continuous.*

Definition 4.1 (Absolute Continuity)**.** Let I be an interval in the real line R. A function f: I to R is absolutely continuous on I if for every positive number ϵ there is a positive number δ such that whenever a finite sequence of pairwise disjoint sub-intervals (x_k, y_k) of I with $x_k < y_k \in I$ satisfies $\sum_k (y_k - x_k) < \delta$ then $\sum_k |f(y_k) - f(x_k)| < \varepsilon$

Proof. Pick $\epsilon < 1$. Since the standard Cantor set has measure zero, for every $\delta > 0$, we can find a collection of intervals (x_k, y_k) that cover the points in the standard Cantor set such that $\sum |x_k - y_k| < \delta$ However, since the Cantor function only changes on the Cantor set, we have $\sum |c(x_k) - c(y_k)| = 1$, contradicting that $\epsilon < 1$. The standard Cantor function is therefore not absolutely continuous. □

4.1.3. *Differentiation.* The Cantor function has zero derivative on C^C (the complement of the Cantor Set on the interval $(0, 1)$) and is not differentiable on the ternary Cantor set

Proof. The first half of the theorem is trivial, since the Cantor function is constant on the open set C^C. The second half can be proved as follows.

given $x \in C$:

Case 1: x has a finite number of digits in base 3. Denote that number as n. Let $h = 2/3^k$ where $k > n$. Then $\lim_{k\to\infty} h = 0$. So $\lim_{h\to 0}(c(h + x) - c(x))/h$ $= \lim_{k\to\infty}(c(h + x) - c(x))/h = (1/2^k)/(2/3^k) = \lim_{k\to\infty} 3^k/2^{k+1} = \infty$. Similarly, it can be proved that the limit does not exist when h is negative.

Case 2: x has an infinite number of digits in base 3. Let y be the first k digits of x, then y is also in C. Then $\lim_{x\to y}(c(x)-c(y))/x-y = \lim_{k\to\infty}(c(x)-c(y))/x-y$ $> \lim_{k\to\infty}(1/2^{k+1})/(2/3^{k+1}) = \lim_{k\to\infty} 3^{k+1}/2^{k+2} = \infty$ □

4.2. **Volterra's Function.** The Cantor function shows us some of the limits of the fundamental theorem of calculus that relates differentiation and integration. Going a step further, through one of the fat Cantor sets, Italian mathematician Vito Volterra (1860-1940) constructed a function that is differentiable with bounded derivative, but whose derivative is not integrable. More details on the construction can be found in [5], but here we provide an overview.

Volterra defined the function $F : [0, 1] \to \mathbb{R}$ as follows, where C is the Smith-Volterra Cantor set defined on page 8:

$$F(x) = \begin{cases} f_{a,b}(x), & \text{if } x \in (a, b) \text{ for some interval}(a, b) \subset [0, 1]\backslash C \\ 0, & \text{if } x \in C \end{cases}$$

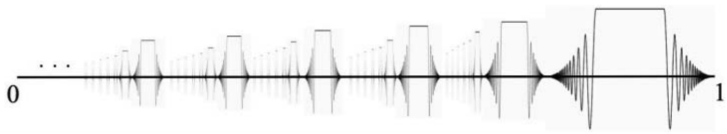


FIGURE 6. Volterra's Function

Here, we would like to pick our endpoints a and b so they match exactly those intervals removed in the construction of the SVC set. Thus, each "hole" in that set has exactly one function placed in it.

Now, we need to specify the function we are filling the gaps with. $f_{a,b} : [a, b] \to \mathbb{R}$ is defined such that:

1. $f_{a,b}(x) = 0$, for $x = a$ or $x = b$
2. $f_{a,b}(x) = (x-a)^2 \sin\left(\frac{1}{x-a}\right)$, for $a < x \le x_1$, where x_1 is the largest number less than or equal to $\frac{a+b}{2}$ for which $(x-a)^2 \sin\left(\frac{1}{x-a}\right)$ has maximum value.
3. $f_{a,b}(x) = (x_1 - a)^2 \sin\left(\frac{1}{x_1-a}\right) = (b-x_2)^2 \sin\left(\frac{1}{b-x_2}\right)$, if $x_1 \le x \le x_2$.
4. $f_{a,b}(x) = (b-x)^2 \sin\left(\frac{1}{b-x}\right)$, if $x_2 \le x < b$, where x_2 is the smallest number greater than or equal to $\frac{a+b}{2}$ for which $(b-x)^2 \sin\left(\frac{1}{b-x}\right)$ has maximum value.

The basic idea about the Volterra function is we use a special property of the function $g(x) = x^2 \sin(1/x)$. Namely, if we require that $g(0) = 0$, this function is differentiable everywhere, but its derivative $2x \sin(1/x) - \cos(1/x)$ is not continuous at 0. With the above construction, we force, near every point in the SVC, the derivative to be discontinuous. For the endpoints of the removed intervals, we basically have a copy of the function g around them, so clearly F' is discontinuous there. There is a slight subtlety here, because not every point in the SVC is an endpoint, but endpoints are dense in Cantor sets, so the function will still fail to be continuous on *all* of the Smith-Volterra Cantor set, if we consider approaching any point through successively closer endpoints.

Thus, we have that F' fails to be continuous on C, a set of positive measure, so by the Riemann-Lebesgue lemma, it is not Riemann integrable. This is why we needed to specifically use a Cantor set of positive measure.

5. Applications in the real world: fractal phenomena

Because of its self-similar nature, the standard Cantor set is the prototype of a fractal. In fact, a well established mathematical branch, fractal geometry is widely applied to study patterns and phenomena in various aspects of our lives, and in this paper, we picked three examples with the closest relationships to the Cantor sets.

5.1. **Fractal Geometry in Nature.** Among the numerous fractal structures observed in nature–spirals, tree branches, snow flakes– Saturn's rings have a special relationship to the Cantor sets.

Note the different sizes in the gaps of Saturn's rings in Figure 7, which look like the intervals removed from a Cantor set. The figure on the right consists of the product of fat Cantor set and a circle. The fat Cantor set has positive measure,

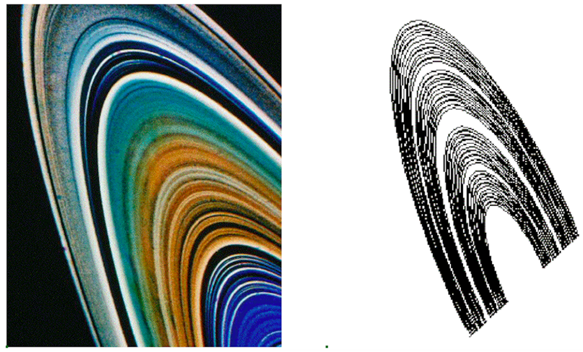


FIGURE 7. Left: Saturn's rings (NASA). Right: a product of a Cantor set and a circle. [1]

so it's product with a circle should have positive area. Thus, the fat Cantor set specifically provides an interesting comparison with Saturn, because if the rings' cross section were a different Cantor set of zero area, the rings would have almost zero area to reflect light and so would be almost invisible.[1]

5.2. **Mandelbrot and the Fractal Market.** Compared to unambiguous self-similar patterns in art and nature, the applications of Cantor sets and fractals to the financial world come in a more subtle way. Mathematician Benoit Mandelbrot (1987) once compared markets to turbulent seas in his "Ten Heresies of Finance," [2] where he argues that :" the very heart of finance is fractal." In discussing the applications of fractals to analyzing markets, he states that the simplest fractals scale the same way in all directions, hence are called self-similar. If the fractals scale in many different ways at different points–the exact reality of the markets ... their mathematical properties become intricate and powerful."

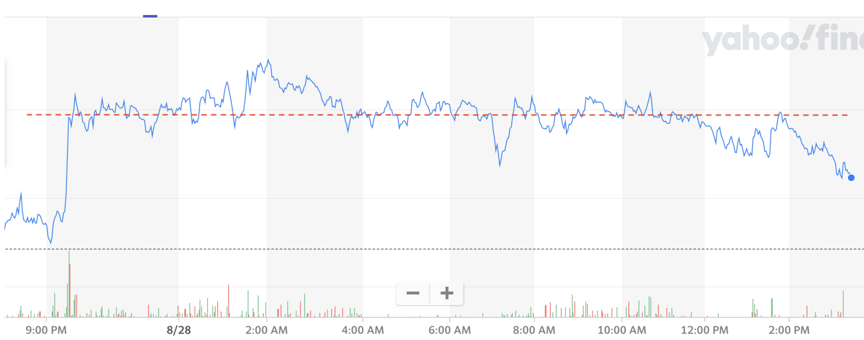


FIGURE 8. Bitcoin Price, for example, in the past 18 hours, with an estimated Hurst Coefficient of around 0.4-0.5

The comparisons with nature lead to idea that financial markets are similar to the behavior of various natural phenomena in the world. The history of shifts from classical to modern views on the market modeling were outlined, visualizing some

FIGURE 9. Turbulent sea

misconduct of classical approach to market modeling and providing examples of utilizing the fractional approach.

5.3. **The fractal market hypothesis.**

5.3.1. *An alternative to EMH.* For decades, the Efficient Market Hypothesis has been the dominant foundation for the modeling of financial markets. It states that stocks always trade at their fair value on exchanges, making it impossible for investors to purchase undervalued stocks or sell stocks for inflated prices. The core idea of the Efficient Market Hypothesis lie in the observation that stock prices exhibit random walks, which can be modeled by something called geometric Brownian motion. Modeling with geometric Brownian motion suggests that the percentage change of a stock price in a given future time interval is completely independent of its previous prices. Furthermore, the distribution of the percentage changes after a given time has passed t should be normally distributed, with variance proportional to t.

The model of geometric Brownian motion is useful, but not perfect. For instance, one can modify it by adding a "drift" term to capture the reality that stock prices tend to increase over time. A more core issue, though, is the idea of fat tails, which reflect the disproportionate influence of rare events on the economy. The reality of fat tails has laid the foundation of a new theory – the "Fractal Market Hypothesis".

One of the central arguments in the fractal marker hypothesis is that the frequency distribution of returns looks the same at different investment horizons, which is the total length of time that an investor expects to hold a security or a portfolio. The longer-term horizons are based more upon fundamental information, and shorter-term investors base their views on more technical information. As long as the market maintains this fractal structure, with no characteristic time scale, the market remains stable because each investment horizon provides liquidity to the others.

As a result, the geometric Brownian motion, as a stochastic process to model stock movements in EMH according to the Black–Scholes model, can be potentially replaced by the *fractional Brownian motion* with a special parameter Hurst coefficient "H". For self-similar time series, H is directly related to fractal dimension, D, where $1 < D < 2$, such that $D = 2 - H$. Increments are independent only when H $= 1/2$, for $H > 1/2$, increments are positively correlated and for $H < 1/2$ they are negatively correlated. [6] The values of the Hurst exponent vary between 0 and 1, with higher values indicating a smoother trend, less volatility, and less roughness.

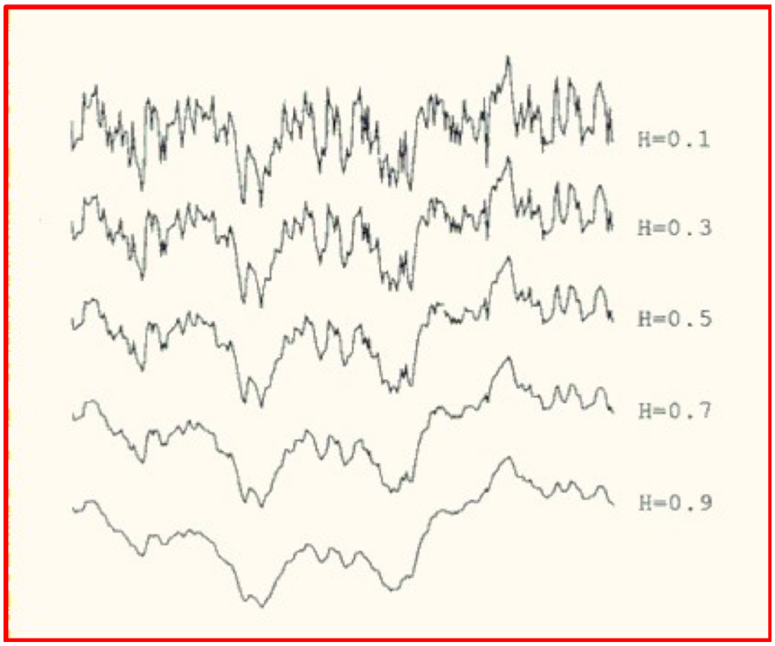


FIGURE 10. Patterns corresponding to different H coefficients

An intriguing point where Cantor sets come into play is when we observe the level set of a one dimensional fractional Brownian motion mentioned above. Assuming in case of figure 8 that the H index of bitcoin price is around 0.5 (0.495 $\pm$ 0.102) according to [7], the chance of the stock price going up or down is close to random. Therefore, any typical level set (e.g. the red line in figure 7) is a closed, perfect set resulting from the properties of Brownian motions (both fractional and geometric) [8]. Furthermore, the level sets will almost surely not contain any intervals, meaning with the above properties that they must be Cantor sets.

Acknowledgments

It is a pleasure to thank my mentor Jake Fiedler for introducing to me the unfamiliar topic of measure theory, discussing and editing my paper, and giving me readings that I found really interesting. I would also like to thank Daniil Rudenko for his engaging lectures on group theory, combinatorics and projective geometry. Lastly, I want to thank Peter May for organizing the REU program - it has been a great experience.

References

[1] https://users.math.yale.edu/public_html/People/frame/Fractals/Extras/Saturn.html
[2] Benoit B. Mandelbrot (Author), Richard L. Hudson. The Misbehavior of Markets A Fractal View of Risk, Ruin, and Reward. Profile Books. 2010.

[3] Walter Rudin, Principles of Mathematical Analysis. McGraw-Hill. 1976.
[4] Edgar E. Peters, Fractal Market Analysis: Applying Chaos Theory to Investment and Economics. Wiley. 2009.
[5] Juan Carlos Ponce-Campuzanoa, Miguel Ángel Maldonado-Aguilar. Vito Volterra's construction of a nonconstant function with a bounded, non-Riemann integrable derivative. Taylor Francis. 2015.
[6] L. DECREUSEFOND and A.S. ÜSTÜNEL. Stochastic Analysis of the Fractional Brownian Motion. 1997.
[7] Mariusz Tarnopolski. Modeling the price of Bitcoin with geometric fractional Brownian motion: a Monte Carlo approach. 2017.
[8] https://www.stat.berkeley.edu/~aldous/205B/bmbook.pdf