The Cantor Set

Aaron Anderson
LA Math Circle

1 Geometric Series

Throughout this handout, we will make use of what's called a geometric series: the infinite sum $\sum_{i=0}^{\infty} x^i$ for some real x. You may already know a formula for this series: $\frac{1}{1-x}$. When does this formula work, and why?

Problem 1

(a) Let's start with a finite version of the geometric series: $\sum_{i=0}^{n-1} x^i$. Prove that this is $\frac{1-x^n}{1-x}$ whenever that number is defined. When is that number not defined?

(b) Now back to the infinite case. Using just algebra, and not worrying about whether sums converge, calculate $(1-x)\sum_{i=0}^{\infty} x^i$ by subtracting $\left(\sum_{i=0}^{\infty} x^i\right) - x\left(\sum_{i=0}^{\infty} x^i\right)$ like you would with a polynomial, by grouping like terms. From this calculation, convince yourself that if the number $\sum_{i=0}^{\infty} x^i$ is defined, it should be $\frac{1}{1-x}$.

(c) Now we show that this formula works exactly when $|x| < 1$. If you are familiar with limits, try to use them in your justifications, but otherwise, it's ok to be vague. When $|x| < 1$, convince yourself that x^n shrinks towards 0 as n gets large enough, and thus for large n, $\frac{1-x^n}{1-x}$ should be close to $\frac{1}{1-x}$.

(d) Now if $x \le -1$ or $x \ge 1$, convince yourself that x^n gets further and further away from 0 when n gets very large, so the finite sums $\sum_{i=0}^{\infty} x^i$ are actually very different from $\frac{1}{1-x}$, so this is not the infinite sum after all.

2 Constructing the Cantor Set

In this section, we will construct a subset $C_\infty \subset [0,1]$, known as the *Cantor set*, which has many interesting and surprising properties. It pops up as a counterexample repeatedly in the mathematical fields of analysis and topology, defying the intuition that we develop from studying more "normal" sets such as intervals.

2.1 Removing Intervals

To define C_∞, we first define $C_n \subset [0,1]$ for each natural number $n = 0, 1, 2, 3, \ldots$. We start with $C_0 = [0,1]$, and iteratively remove pieces. To form C_1, we remove the middle third of the interval $[0,1]$, but not its endpoints: that is, we remove the set $(\frac{1}{3}, \frac{2}{3})$ of points x such that $\frac{1}{3} < x < \frac{2}{3}$.

Problem 2 We can visualize C_0 as a filled-in piece of the number line, with square brackets to denote that the endpoints of the interval are included:

Using square brackets to denote included endpoints or parentheses to denote excluded endpoints, draw the set C_1 here:

Problem 3 Now to define C_2, we take each of the intervals remaining in C_1 ($[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$) and remove the middle third of each of these. Draw the set C_2:

0 1
[ ]

For each subsequent step, if we've defined C_n, it'll be a union of several closed intervals all of the same length, separated from each other. Then to define C_{n+1}, we remove the middle third of each of those intervals. Draw a few more iterations, such as $C_3, C_4, C_5, \ldots$, as many as you feel you need to get the picture:

0 1
[ ]

0 1
[ ]

0 1
[ ]

How long are the intervals in C_n, and how many are there?

Problem 4 Now we have an infinite sequence of sets, which we can use to define C_∞. Let $C_\infty = \bigcap_{n=1}^{\infty} C_n$, that is, C_∞ is the set of points which are in C_n for every number n.

Given your pictures from earlier, draw (an approximate version of) C_∞, the Cantor set:

0 1
[ ]

2.2 Ternary

Another definition of the Cantor set relies on the ternary notation for real numbers. Our standard way of writing a real number in the interval $[0, 1]$ is as a decimal sequence, which we might write generally as $0.a_1a_2a_3\ldots$ where $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is a valid digit less than 10. The actual number that this represents is given by the infinite sum $\sum_{i=1}^{\infty} a_i 10^{-i}$.

We can of course generalize this notation to bases other than 10. You've likely worked with binary numbers in some form, and we can also write any number in $[0, 1]$ as $0.a_1a_2a_3\ldots$ or $\sum_{i=1}^{\infty} a_i 2^{-i}$ where each a_i is in $\{0, 1\}$, which is a binary expansion for that number. In fact, we could use any natural number b greater than 1 as our base, and let our valid set of digits be $\{0, 1, \ldots, b-1\}$, and then evaluate $0.a_1a_2a_3\ldots$ as $\sum_{i=1}^{\infty} a_i b^{-i}$. Today, we're interested in *ternary* numbers, base $b = 3$.

Notational aside: We call the numbers chosen from $\{0, 1, \ldots, b-1\}$ *digits* regardless of which base b we're in, but there are cuter names for digits from specific bases. Thanks to their ubiquity in computers, most people know that binary digits are called bits, but it's also accepted to call ternary digits *trits*, and sometimes it's useful to call the usual decimal digits *dits*, although usually these are just called digits.

Problem 5 Note that we said this notation could be used to write the numbers in $[0, 1]$, and not $[0, 1)$. What is a base b expansion $0.a_1a_2a_3\ldots$ for 1? Why can't we write any number greater than 1 this way?

We can use ternary notations to define the Cantor set a different way: as the set of numbers that can be written as a ternary sequence without using any 1s. In the next two problems, we prove that this defines the same set C_∞ that we defined earlier.

Problem 6 Every ternary sequence consisting only of 0s and 2s (avoiding 1s) is in the Cantor set C_∞ defined above.

Problem 7 Every point in the Cantor set can be written with a ternary sequence consisting of only 0s and 2s. (Note: it may also be possible to write a point in the Cantor set as a ternary sequence that contains a 1. Why isn't this a problem?)

Problem 8 Show that the only time two strings of ternary digits $0.a_1a_2\ldots$ and $0.b_1b_2\ldots$ give rise to the same number is when there is some n such that $a_1 = b_1$, $a_2 = b_2, \ldots, a_{n-1} = b_{n-1}$, but $b_n = a_n + 1$, and for $m > n$, $a_m = 2$, while $b_m = 0$. In other words, the only numbers that have an ambiguous ternary representation are ones for which there is a ternary representation involving only a finite number of non-zero digits. (Bonus: Generalize this for any base.)

3 Measure and Probability

3.1 Measure

As the Cantor set is defined by starting with the unit interval $[0, 1]$ and removing parts, we may wonder how much is left after this process? That is, we're asking what the measure, or length, of the Cantor Set is. The measure of an interval $[a, b]$ or (a, b) is just its length, $b-a$, and the measure of a finite union of intervals which don't overlap is just the sum of the measures of the intervals.

Problem 9 What is the measure of C_1? How about C_n in general (where n is finite)?

Problem 10 To calculate the measure of C_∞, we need to observe one more property of measure. As no interval has negative length, no set has negative measure. Also, if $A \subset B$, then the measure of A is less than or equal to the measure of B. Given this property, what is the measure of C_∞?

3.2 Fat Cantor Sets

At each step in our construction, we removed the middle third of each remaining interval. We can generalize this, instead of removing the middle third, removing some other fraction of each remaining interval at step n. Let us define $D_0 = [0, 1]$, and if D_n is a union of non-adjacent closed intervals, like in the Cantor set construction, we take each interval $[a, b]$, and delete an open interval from the middle with length $r_n(b-a)$. If $r_n = \frac{1}{3}$ for each n, then this is the regular Cantor set construction.

Problem 11 For a general sequence $r_0, r_1, r_2, \ldots$ with $0 < r_n < 1$ for each n, what is the measure of the resulting set $D_\infty = \bigcap_{n=0}^{\infty} D_n$? Can you come up with a sequence $r_0, r_1, r_2, \ldots$ such that this measure is $\frac{1}{2}$? In this case, how different does D_∞ look from C_∞?

3.3 Probability

Measure is closely related to probability. If we choose a point on the unit interval $[0, 1]$ uniformly at random, then there is a 50% chance that the point is on the left half of the interval, in the smaller interval $[0, \frac{1}{2}]$, which has measure $\frac{1}{2} = 50\%$. Similarly, if we have any other subset $A \subset [0, 1]$, then the probability that our randomly-selected point is in A is the measure of A. As we have calculated the measure of the Cantor set, we now also know that this is the probability that a uniformly randomly selected point in $[0, 1]$ belongs to the Cantor set. This probability is 0, even though there are infinitely many points in the Cantor set that we could in theory choose! If this seems strange, perhaps it will make more sense when choosing points in the following way:

Problem 12 Let's pick a point in $[0,1]$ randomly in a different way. To do this, we will pick random numbers from the set $\{0,1,2\}$, each with probability $\frac{1}{3}$, and chain them together into an infinite ternary sequence, which will then correspond to a point in $[0,1]$. Beware that this may not be the same thing as what we were doing before, as there may be more than one ternary sequence resulting in the same point.

If we do this, what is the probability that none of the first n digits of our sequence is a 1? How about the probability that none of the digits at all are 1s?

4 Cardinality

4.1 Intro/Review of Cardinality

Having quantified the size of the Cantor set with the related notions of measure and probability, we can try to compute its size in another sense. The set-theoretic notion of size is *cardinality*, where two sets have the same size if there is a *bijection* between them. That is, sets A and B have the same cardinality if there is a function $f : A \to B$ (mapping elements of A to elements of B) with the following two properties:

- **Injective/One-to-one** If $x, y \in A$ with $x \neq y$, then $f(x) \neq f(y)$, so every element of B has at most one element of A mapping to it.
- **Surjective/Onto** If $y \in B$, then there is some $x \in A$ such that $f(x) = B$, that is, every element of B has at least one element of A mapping to it.

We will denote the cardinality of a set A as $|A|$. Then if $f : A \to B$ satisfies both of those properties and is thus a bijection, $|A| = |B|$. If we only know that there is a one-to-one function, an *injection*, from A to B, then we know that $|A| \leq |B|$, and if we know there is a surjection from A to B, then $|A| \geq |B|$.

This definition works fine for finite sets, where any set with n elements has the same cardinality as the set $\{1, 2, 3, \ldots, n\}$, and is thus said to have cardinality n.

Determining the cardinality of an infinite set is in general much more complicated. But not all infinite sets have the same cardinality - for instance, $\mathbb{N}$, the set of natural numbers, is strictly smaller than the set $\mathbb{R}$ of real numbers, or even $[0,1]$. As a bijection $f : \mathbb{N} \to A$ can be seen as counting the set A, as then the elements of A can be listed as $f(0), f(1), f(2), \ldots$, we call sets that are the same cardinality as $\mathbb{N}$ *countable*, and all other infinite sets that are not countable *uncountable*. All uncountable sets are bigger than all countable sets.

4.2 Endpoints

First, we concern ourself with a particular subset of the Cantor set. Each set C_n in our definition was a union of separate intervals. Let E_n be the set of the endpoints of these intervals, that is, $E_0 = \{0, 1\}$, $E_1 = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$, and so on. Now let $E_\infty = \bigcup_{n=0}^{\infty}$ be the union of all of these sets, the set of all endpoints of any E_n.

Problem 13 Prove that $E_\infty \subset C_\infty$, or that every endpoint makes it into the final Cantor set C_∞.

Problem 14 Given a finite number n, how big is E_n? Is E_∞ countable or uncountable?

Problem 15 Let T be the set of infinite ternary sequences that only contain 0s and 2s. Show that the function f that sends each sequence $a_1, a_2, a_3, \ldots$ in T to the number $0.a_1a_2a_3\ldots$ is a bijection between T and C_∞.

Is T countable or uncountable? How about C_∞?

Problem 16 Show that there is a point in the Cantor Set that is not the endpoint of any interval used in its construction (e.g. not the endpoint of one of the intervals of length 3^{-n} in the set C_n). Can you find any such points? Are any of them rational? If you've found a rational non-endpoint point, how does it fit into your picture from the first section where we construct the Cantor set by removing intervals?

5 The Cantor Function

Here we seek to define a function $g : [0, 1] \to [0, 1]$ in terms of ternary and binary expansions.

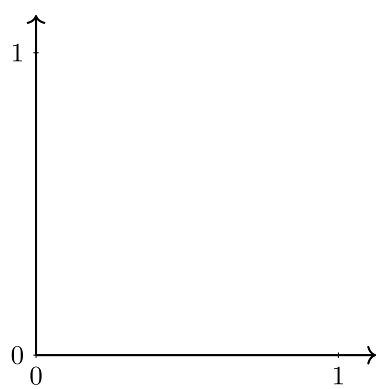
First we define a function f on the Cantor set C_∞. If $x \in C_\infty$ has the ternary expansion $0.a_1a_2\ldots$ (with only 0s and 2s), then we define $f(x)$ to be the *binary* number $0.b_1b_2\ldots$, where $b_i = a_i/2$ for each i.

Problem 17 Prove that this definition is well-defined, in that each $x \in C_\infty$ has a unique ternary expansion with only 0s and 2s, so there isn't ambiguity in choosing which sequence to use to define $f(x)$.

Problem 18 Prove that $f : C_\infty \to [0, 1]$ is (not necessarily strictly) increasing, and is surjective (onto). Note that this means that the measure-0 C_∞ can be stretched around in ways that respect its order to cover the measure-1 $[0, 1]$!

Problem 19 Prove that there is a unique way to extend f to a (not necessarily strictly) increasing function $g : [0, 1] \to [0, 1]$. (By extend, we mean that if $x \in C_\infty$, $g(x) = f(x)$.)

Problem 20 Graph g here:



Hint: because C_∞ has measure 0, it is essentially invisible, so you can focus on graphing the function on the intervals between C_∞.

Depending on the definition of continuous you prefer to use, conclude (possibly just by looking at the graph) that g is continuous. This means that it is possible for a function which is (piecewise) constant on all but a measure 0 set can still change enough to cover actual distance – moving from $g(0) = 0$ to $g(1) = 1$ – without any sudden jumps!