

Cantor's Set and Cantor's Function

Eric Hanson

February 5, 2016

1 Cantor's Set

The Cantor set is an interesting subset of $[0, 1]$, with many properties that help illuminate concepts in analysis. It can often serve as a counter-example or edge-case on which to test ideas, and to construct further unusual objects; one such object is the Cantor function, which we'll define here as well. To me, one of the most immediate clarifications provided by the Cantor set is the idea that looking at an object on different levels yields different notions of size, and these do not have to play nicely with each other. As we will see, the Cantor set is nowhere dense, yet uncountable.

1.1 Construction

We will define a sequence of sets $\{E_n\}_{n=0}^\infty \subseteq [0, 1]$, and define the Cantor set (or "Cantor ternary set") as $K = \bigcap_{n=0}^\infty E_n$.

To start, take

$$E_0 = [0, 1].$$

Divide the interval in three equal pieces and throw away the middle one to define

$$E_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1].$$

Repeat this process by throwing away the middle third of each interval in E_1 to obtain

$$E_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{3}{9}] \cup [\frac{6}{9}, \frac{7}{9}] \cup [\frac{8}{9}, 1].$$

In general, we recursively define E_n in terms E_{n-1} by discarding the middle third of the intervals which constitute E_{n-1} . We may write this definition as

$$E_n = \frac{E_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{E_{n-1}}{3} \right), \quad E_0 = [0, 1]$$

where we are using the notation of multiplying and adding to sets by multiplying and adding to their elements: for a set $A \subset \mathbb{R}$ and $c, d \in \mathbb{R}$,

$$cA + d := \{ca + d : a \in A\}.$$

Contents

1	Cantor's Set	1
1.1	Construction	1
1.2	Properties	2
2	Cantor's function	6



Figure 1: E_0 .



Figure 2: E_1 .



Figure 3: E_2 .

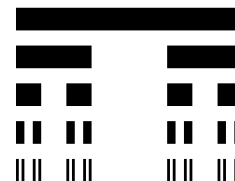


Figure 4: From top to bottom, E_0 to E_4 .

In this way, we create a decreasing sequences of sets

$$E_0 \supset E_1 \supset \dots \supset E_n \supset \dots$$

Note: By construction, E_n is the disjoint union of 2^n intervals, each of length 3^{-n} . This is a key property we will use in our proofs of the properties of the Cantor set.

1.2 Properties

Let

$$K := \bigcap_{n=0}^{\infty} E_n.$$

We have the following properties of K .

Theorem.

1. K is closed.
2. $\text{int } K = \emptyset$.
3. K has no isolated points: that is, $\forall x \in K, \forall \epsilon > 0, \exists y \in K, y \neq x, y \in B(x, \epsilon)$ where $B(x, \epsilon) := (x - \epsilon, x + \epsilon)$ is the open ball of radius ϵ around x .
4. $x \in K \iff x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$ where (a_k) is a sequence with each $a_k \in \{0, 2\}$.
5. K is uncountable.

Remark. Note: a set is called *nowhere dense* if its closure has empty interior. Thus (1) and (2) imply K is nowhere dense.

Proof.

1. E_n is closed for each n so $K = \bigcap_{n=0}^{\infty} E_n$ is closed as well.
2. (By contradiction). Assume $\text{int } K \neq \emptyset$. So $\exists x \in \text{int } K$. Thus, for some $\delta > 0$,

$$B(x, \delta) \subseteq \text{int } K \subseteq K.$$

Choose $n \in \mathbb{N}$ large enough such that $3^{-n} < 2\delta$. We have

$$(x - \delta, x + \delta) \subseteq K \subseteq E_n.$$

But E_n is a disjoint union of intervals of length $3^{-n} < 2\delta$. This is a contradiction.

$A := B$ is notation to mean that we are defining the quantity A to be B .

$\text{int } A = \bigcup_{\substack{O \subseteq A \\ O \text{ is open}}} O$ is the largest open set contained in A .

No isolated points: Arbitrarily close to any point in K lies another point in K .

The number x has a ternary expansion without any 1s.

In contrast, $\mathbb{Q} \cap [0, 1]$ has empty interior, but is dense in $[0, 1]$.

3. Key observation: endpoints of intervals in E_n are not removed in subsequent E_m (for $m > n$). All endpoints, $\{0, 1, \frac{1}{3}, \frac{2}{3}, \dots\} \subset K$.

Let $x \in K$ and $\epsilon > 0$ be given. We choose n large enough so that $3^{-n} < \epsilon$. Then $x \in E_n \implies x \in [a, b]$ for an interval $[a, b] \subset E_n$ which is one of the 2^n intervals making up E_n . Since the length of each interval is 3^{-n} , we have $b - a < \epsilon$. Then we have both

$$|x - b| < \epsilon, \quad |x - a| < \epsilon.$$

Hence, even if $x = a$ or $x = b$, we have found $y \neq x$ so that $y \in K \cap B(x, \epsilon)$.

4. We want to show $x \in K \iff x = \sum_{k=1}^{\infty} \frac{a_k}{3^k}$ for $a_k \in \{0, 2\}$.

First, let us notice and deal with a point of confusion¹. Consider $\frac{1}{3} \in K$. Then $\frac{1}{3} = \sum_{k=1}^{\infty} \frac{a_k}{3^k}$. We could obtain such ternary expansion by $a_1 = 1$ and $a_n = 0$ for $n > 1$, which doesn't fit our criteria. But we can choose other coefficients:

$$\sum_{k=2}^{\infty} \frac{2}{3^k} = 2 \left(\frac{1}{9}\right) \sum_{k=0}^{\infty} \left(\frac{1}{3}\right)^k = \frac{2}{3^2} \frac{1}{1 - (1/3)} = \frac{1}{3}.$$

In other words, ternary expansions are not unique. However, ternary expansions which only contain 0's and 2's are unique. That is the content of the following claim.

Claim 1. If

$$\sum_{k=1}^{\infty} \frac{a_k}{3^k} = \sum_{k=1}^{\infty} \frac{b_k}{3^k}$$

where each $a_k, b_k \in \{0, 2\}$, then $a_k = b_k$ for all $k \in \mathbb{N}$.

Proof. Suppose not; then there is a minimal N such that $a_n \neq b_n$. Then one of a_n or b_n is zero and the other two; wlog take $a_n = 0$, $b_n = 2$. Expanding our sums,

$$\sum_{k=1}^{N-1} \frac{a_k}{3^k} + 0 + \sum_{k=N+1}^{\infty} \frac{a_k}{3^k} = \sum_{k=1}^{N-1} \frac{b_k}{3^k} + \frac{2}{3^N} + \sum_{k=N+1}^{\infty} \frac{b_k}{3^k}.$$

Using $a_k = b_k$ for $k < N$,

$$\sum_{k=N+1}^{\infty} \frac{a_k}{3^k} = \frac{2}{3^N} + \sum_{k=N+1}^{\infty} \frac{b_k}{3^k}. \tag{*}$$

But

$$\sum_{k=N+1}^{\infty} \frac{a_k}{3^k} \leq 2 \sum_{k=N+1}^{\infty} 3^{-k} = 2 \cdot 3^{-N-1} \sum_{k=0}^{\infty} 3^{-k} = 2 \cdot 3^{-N-1} \frac{1}{1 - (1/3)} = 3^{-N}.$$

¹ for Eric

Note we have used the geometric series

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1 - a}$$

which converges absolutely for $|a| < 1$; we will use this several times.

So, returning to (\star) , we have

$$\frac{2}{3^N} + \sum_{k=N+1}^{\infty} \frac{b_k}{3^k} \leq \frac{1}{3^N} \implies \sum_{k=N+1}^{\infty} \frac{b_k}{3^k} \leq -\frac{1}{3^N},$$

which of course is a contradiction, since $b_k \in \{0, 2\}$. \square

Now, let $S = \{\sum_{k=1}^{\infty} \frac{a_k}{3^k} : a_k \in \{0, 2\}\}$.

If $x \in S$, then $x = \sum_{k=1}^{\infty} \frac{a_k}{3^k}$ for some $a_k \in \{0, 2\}$. Let $x_n = \sum_{k=1}^n \frac{a_k}{3^k}$ be the n^{th} partial sum.

Claim 2. For each n , the partial sum x_n is a left-endpoint of an interval of E_n .

Proof. By induction on n . Either $x_1 = 0$ or $\frac{2}{3}$, either of which is a left endpoint of E_1 .

Assume that x_{n-1} is a left endpoint of E_{n-1} . Then

$$[x_{n-1}, x_{n-1} + \frac{1}{3^{n-1}}] \subset E_{n-1}.$$

Then by construction of E_n , we can take out the middle thirds to obtain intervals of E_n :

$$[x_{n-1}, x_{n-1} + \frac{1}{3^n}] \cup [x_{n-1} + \frac{2}{3^n}, x_{n-1} + \underbrace{\frac{1}{3^{n-1}}}_{=\frac{3}{3^n}}] \subset E_n.$$

So $x_n = x_{n-1} + \frac{a_n}{3^n}$ is a left endpoint of an interval in E_n , whether $a_n = 0$ or $a_n = 2$. \square

Now, since each x_n is a left-endpoint of E_n and we don't remove endpoints, we have $x_n \in K$ for all n . But K is closed so contains all its limit points, and we have $x := \lim_{n \rightarrow \infty} x_n \in K$. Thus, $S \subset K$.

To show that $K \subset S$, we will prove the contrapositive

$$x \notin S \implies x \notin K,$$

in order to access properties of S . First notice the following property.

Claim 3. The intervals removed to obtain K are all of the form

$$I_{k,m} = \left(\frac{3k+1}{3^m}, \frac{3k+2}{3^m} \right) : m \in \mathbb{N}, \quad 0 \leq k \leq 3^{m-1} - 1.$$

In fact,

$$[0, 1] \setminus E_n = \bigcup_{m=1}^n \bigcup_{k=0}^{3^{m-1}-1} I_{k,m}.$$

Note that S is the set of numbers which have a ternary expansion without any 1's, not the set of numbers for which every ternary expansion contains no 1's.

Proof. The proof, by induction on n is left as an exercise. \square

Let us proceed to the proof of the contrapositive. Suppose $x \notin S$. Suppose x contains a '1' in its n^{th} digit of its ternary expansion, i.e.

$$x = \sum_{k=1}^{n-1} \frac{a_k}{3^k} + \frac{1}{3^n} + \sum_{k=n+1}^{\infty} \frac{a_k}{3^k}.$$

We will take n to be the first digit which is '1', if there are more than one.

Claim 4. We have that $0 < \sum_{k=n+1}^{\infty} \frac{a_k}{3^k} < \frac{1}{3^n}$.

Proof. If $a_k = 0$ for all $k > n$, then our number written in ternary is of the form $x = 0.\star 1000\cdots$, where \star represents the arbitrary first $n-1$ digits (each in $\{0,2\}$). Then we could choose the expansion $x = 0.\star 02222\cdots$ instead, and avoid having a 1. In other words, $x \in S$. Similarly, if $a_k = 2$ for all $k > n$, then we are writing $x = 0.\star 1222\cdots$, so x admits the expansion $x = 0.\star 2222\cdots$ (using our geometric series again), and hence, $x \in S$. Thus, if $x \notin S$, this tail condition holds. \square

Now, we may rewrite

$$\sum_{k=1}^{n-1} \frac{a_k}{3^k} = \sum_{k=1}^{n-1} \frac{3(3^{n-k-1}a_k)}{3^n} = \frac{3\lambda}{3^n}$$

where $\lambda = \sum_{k=1}^{n-1} 3^{n-k-1}a_k \in \mathbb{N} \cup \{0\}$. With this definition, Claim 4 yields

$$\frac{3\lambda + 1}{3^n} < x < \frac{3\lambda + 2}{3^n}.$$

So $x \in I_{\lambda,n}$ and thus $x \notin K$ by Claim 3, as we wanted.

5. K is uncountable. i.e. $|K| = |[0,1]|$. We already know $K \subset [0,1]$, so we know $|K| \leq |[0,1]|$. Hence, we need to find a surjection from $K \rightarrow [0,1]$ to complete the proof.

We define a function $F : K \rightarrow [0,1]$, such that F takes $x \in K$ and changes all of the 2's in its ternary expansion to 1s and then outputs the number with corresponding binary expansion. We have that

$$K = \left\{ \sum_{k=1}^{\infty} \frac{a_k}{3^k} : a_k \in \{0,2\} \right\}.$$

For $x \in K$, we write $x = \sum_{k=1}^{\infty} \frac{a_k}{3^k}$ where $a_k \in \{0,2\}$. Then we define

$$F(x) = \sum_{k=1}^{\infty} \frac{a_k}{2} \frac{1}{2^k}$$

Since by Claim 1 we have a unique ternary expansion of x that contains no 2's, we have a well-defined function (there is no ambiguity in how to compute $F(x)$ given x).

Let $y \in [0, 1]$. Then y admits a binary expansion: $y = \sum_{k=1}^{\infty} \frac{b_k}{2^k}$ for $b_k \in \{0, 1\}$. Then let $x = \sum_{k=1}^{\infty} \frac{2b_k}{3^k}$, so that $F(x) = \sum_{k=1}^{\infty} \frac{b_k}{2^k} = y$.

So we are done.

Remark. Let's notice that the binary expansion of y does have ambiguity, and thus F is not injective. For example, $\frac{1}{2} = 0.0111 \dots$ in binary, and $\frac{1}{2} = 0.100 \dots$ in binary. Thus $x_1 = 0.022 \dots$ in ternary and $x_2 = 0.200 \dots$ in ternary each have $F(x_1) = F(x_2) = \frac{1}{2}$. But $x_1 = \frac{1}{3}$ and $x_2 = \frac{2}{3}$. This is not a problem; we only needed to find a surjection from K to $[0, 1]$.

Note: the lower bound of F on K is 0 (all $a_k = 0$) and the upper bound is when all the a_k are 1s, as $\sum \frac{1}{2^k} = 1$.

□

2 Cantor's function

We defined $F : K \rightarrow [0, 1]$, and we wish to extend it to the whole interval $[0, 1]$. Let x and y be the left and right endpoints respectively of the same removed interval (x, y) which was removed in step n .

Then, by Claim 3,²

$$x = \sum_{k=1}^n a_k 3^{-k} + 0 \cdot 3^{-n} + \sum_{k=n+1}^{\infty} 2 \cdot 3^{-k},$$

$$y = \sum_{k=1}^n a_k 3^{-k} + 2 \cdot 3^{-n} + \sum_{k=n+1}^{\infty} 0 \cdot 3^{-k}.$$

Then,

$$F(x) = \sum_{k=1}^n \frac{a_k}{2} 2^{-k} + 0 \cdot 2^{-n} + \sum_{k=n+1}^{\infty} 1 \cdot 2^{-k}$$

But $\sum_{k=n+1}^{\infty} 2^{-k} = 2^{-n}$, so

$$= \sum_{k=1}^n \frac{a_k}{2} 2^{-k} + 2^{-n}.$$

$$F(y) = \sum_{k=1}^n \frac{a_k}{2} 2^{-k} + 1 \cdot 2^{-n} + \sum_{k=n+1}^{\infty} 0 \cdot 2^{-k}$$

$$= F(x)$$

So F takes on the same values on the endpoints of the removed intervals. But K together with the removed intervals is $[0, 1]$.

So we may define the extension $C : [0, 1] \rightarrow [0, 1]$ as $C(x) = F(x)$ for $x \in K$, and if $y \in (a, b)$ a removed interval, $C(y) = F(a) = F(b)$. See fig. 5 for a plot of $C(x)$.

We will prove three properties of C , as follows.

² Recall our ternary expansion excludes 1s, so we expand $\frac{1}{3^m} = \sum_{k=m+1}^{\infty} 3^{-k}$.

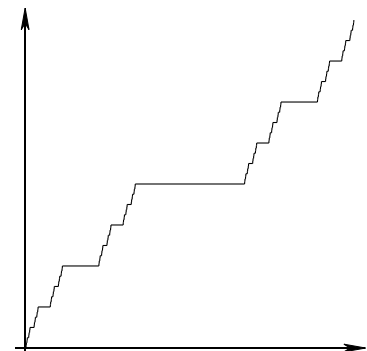


Figure 5: The Cantor function; image taken from <https://commons.wikimedia.org/wiki/File:CantorEscalier.svg>.

Theorem.

1. C is increasing,
2. C is continuous,
3. C has zero derivative on $[0, 1] \setminus K$.

Proof.

1. First we will show that C is increasing on K , and then extend to $[0, 1]$.

Let $x, y \in K$ with $x < y$ and set $x = \sum_{k=1}^{\infty} x_k 3^{-k}$ and $y = \sum_{k=1}^{\infty} y_k 3^{-k}$ their canonical ternary expansions³. If $x < y$ then they lie in the same interval until some n , and at that point y is in a later interval. That is, for some n , $x_n < y_n$, and for all $k < n$, $x_k = y_k$. But if $x_n < y_n$, since both are contained in $\{0, 2\}$, then $x_n = 0$ and $y_n = 2$. So we have

$$\begin{aligned} C(x) &= \sum_{k=1}^{\infty} \frac{x_k}{2} 2^{-k} = \sum_{k=1}^{n-1} \frac{x_k}{2} 2^{-k} + \sum_{k=n+1}^{\infty} \frac{x_k}{2} 2^{-k}, \\ &\leq \sum_{k=1}^{n-1} \frac{y_k}{2} 2^{-k} + 2^{-n}, \\ &\leq \sum_{k=1}^{n-1} \frac{y_k}{2} 2^{-k} + 2^{-n} + \sum_{k=n+1}^{\infty} \frac{y_k}{2} 2^{-k} = C(y). \end{aligned}$$

Now, let $x, y \in [0, 1]$, with $x < y$ and assume $C(x) > C(y)$.

We know then there is some $x', y' \in K$ with $C(x) = C(x')$ and $C(y) = C(y')$, choosing x' to be the left endpoint of the removed interval to which x belongs, and similarly with y' . Then, since $C(x) \neq C(y)$, x' and y' must not correspond to the endpoints of the same interval.

So then since $x < y$, we must have $x' < y'$; see fig. 6. But then since $x', y' \in K$ we must have $C(x) = C(x') \leq C(y') = C(y)$, since we showed C was increasing on K . This is a contradiction. Hence, for $x, y \in K$ with $x < y$, we have $C(x) \leq C(y)$; that is, C is increasing.

2. Next, C is continuous on $[0, 1]$. Clearly, C is continuous at $x \in [0, 1] \setminus K$, because there on a neighborhood of x , C is constant.

So, let $x \in K$. We will first show that if $y \in K$ is close to x , then $C(x)$ is close to $C(y)$. WLOG, take $y > x$ and $y - x < 3^{-m}$, then, choosing $\{x_k\}$ and $\{y_k\}$ as the respective coefficients of x and y 's canonical ternary expansions, we have

$$y - x = \sum_{k=1}^{\infty} (y_k - x_k) 3^{-k} < 3^{-m}$$

³ Meaning expansions with $x_k \in \{0, 2\}$ for all $k \in \mathbb{N}$.

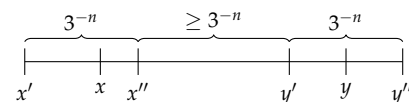


Figure 6: We know x' and y' are endpoints of different removed intervals, and each removed interval is of length 3^{-n} . We set x'' and y'' to denote the other endpoints of those removed intervals, as pictured, for clarity. We see that necessarily $x < y$ if $x' < y'$.

Claim 5. Then we must have for all $k < m$, $x_k = y_k$.

Proof. Suppose not; let $\ell < m$ be the minimal index such that $x_\ell \neq y_\ell$. We have two cases. First, let $x_\ell = 2$, $y_\ell = 0$. Then

$$y - x = 0 + -2 \cdot 3^{-\ell} + \sum_{k=\ell+1}^{\infty} \leq -2 \cdot 3^{-\ell} + 3^{-\ell} < 0$$

which is a contradiction. On the other hand, if $y_\ell = 2$ and $x_\ell = 0$,

$$y - x = 2 \cdot 3^{-\ell} + \sum_{k=\ell+1}^{\infty} \geq 2 \cdot 3^{-\ell} - 3^{-\ell} = 3^{-\ell} > 3^{-m},$$

a contradiction. \square

Using this,

$$C(y) - C(x) = \sum_{k=1}^{\infty} \frac{1}{2} (y_k - x_k) 2^{-k} = \sum_{k=m}^{\infty} \frac{1}{2} (y_k - x_k) 2^{-k} \leq 2^{-m}$$

Clearly then, for any $\epsilon > 0$, we can choose m large enough such that $2^{-m} < \epsilon$.

Now, if $y \notin K$, then we know there exists $y' \in K$ such that $C(y) = C(y')$ and $|x - y'| \leq |x - y|$, by choosing y' to be the endpoint of the removed interval that y resides within; there are two such endpoints, so we choose the one closer to x so that $|x - y'| \leq |x - y|$; see fig. 7. But then for $\epsilon > 0$ we know there is a $\delta > 0$ such that if $|x - y'| < \delta$, then $|C(x) - C(y')| < \epsilon$. But $C(y) = C(y')$, so just take $|x - y| < \delta$.



Figure 7: We choose x' to be the endpoint closer to y than x .

3. For the third property, if we take $x \in [0, 1] \setminus K$, we notice that C is constant on an open interval around x , so has derivative zero:

$$\lim_{t \rightarrow 0} \left| \frac{C(x+t) - C(t)}{t} \right| = \lim_{t \rightarrow 0} \left| \frac{0}{t} \right| = 0.$$

\square