THE CANTOR SET - A BRIEF INTRODUCTION

DYLAN R. NELSON

Abstract. George Cantor (1845-1918) was the originator of much of modern set theory. Among his contributions to mathematics was the notion of the Cantor set, which consists of points along a line segment, and possesses a number of fascinating properties. In the following brief paper we introduce and define the Cantor set, its construction, and basic properties. Several interesting remarks and theorems relating to Cantor sets are then demonstrated, and their connection to real analysis and general topology explored.

1. Preliminaries

The Cantor set has many definitions and many different constructions. Although Cantor originally provided a purely abstract definition, the most accessible is the Cantor "middle-thirds" or ternary set construction. Begin with the closed real interval $[0,1]$ and divide it into three equal open subintervals. Remove the central open interval $I_1 = (\frac{1}{3}, \frac{2}{3})$ such that

$$[0,1] - I_1 = \left[0, \frac{1}{3}\right] \bigcup \left[\frac{2}{3}, 1\right].$$

Next, subdivide each of these two remaining intervals into three equal open subintervals and from each remove the central third. Let I_2 the removed set, then

$$I_2 = \left(\frac{1}{3^2}, \frac{2}{3^2}\right) \bigcup \left(\frac{7}{3^2}, \frac{8}{3^2}\right)$$

and

$$[0,1] - \left(I_1 \bigcup I_2\right) = \left[0, \frac{1}{3^2}\right] \bigcup \left[\frac{2}{3^2}, \frac{3}{3^2}\right] \bigcup \left[\frac{6}{3^2}, \frac{7}{3^2}\right] \bigcup \left[\frac{8}{3^2}, 1\right].$$

We can then subdivide each of the intervals that comprise $[0,1] - (I_1 \bigcup I_2)$ into three subintervals, removing their middle thirds, and continue in the previous manner. The sequence of open sets I_n is then disjoint, and we traditionally define the Cantor set $\mathcal{C}$ as the closed interval with the union of these $I_n's$ subtracted out. That is, $\mathcal{C} = [0,1] - \bigcup I_n$. The formal definition follows:

Definition 1.1. The **Cantor set** $\mathcal{C}$ is defined as $\mathcal{C} = \bigcap_{n=1}^{\infty} I_n$, where I_{n+1} is constructed, as above, by trisecting I_n and removing the middle third, I_0 being the closed real interval $[0,1]$.

Several interesting properties of the Cantor set are immediately apparent. Since it is defined as the set of points not excluded, the "size" of the set can be thought of as the proportion of the interval $[0,1]$ removed. If we add up the contribution from $\frac{2}{3}$ removed n times we find that

$$\sum_{n=0}^{\infty} \frac{2^n}{3^{n+1}} = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + ... = \frac{1}{3}\left(\frac{1}{1-\frac{2}{3}}\right) = 1.$$

Where the geometric sum has its well known solution. As a result, the proportion remaining "in" the Cantor set is $1 - 1 = 0$, and it can contain no intervals of non-zero length. For assume by contradiction that it does contain some interval (a, b). Choose $n \in \mathbb{N}$ such that $\frac{1}{3^n} < b - a$. Since the Cantor set is contained in the finite intersection of closed intervals, all of length less than $(b - a)$, we have that this intersection and so $\mathcal{C}$ cannot contain (a, b).

Theorem 1.2. *The Cantor set is nonempty.*

Proof. Each trisection of I_n to form I_{n+1} leaves exactly two endpoints. For example, removing $\left(\frac{1}{3}, \frac{2}{3}\right)$ from $[0, 1]$ leaves the points $p_0 = \frac{1}{3}$ and $p_1 = \frac{2}{3}$. In fact, since the Cantor set is the infinite intersection of each I_n, $\mathcal{C}$ contains the endpoints of each such subinterval, and is clearly non-empty. In fact, it is infinite. □

Definition 1.3. A subset A of a metric space M is nowhere dense if its closure has an empty interior. That is, if $int(\overline{A}) = \emptyset$.

Remark 1.4. For any subset B of a metric space M, we have that $int(B) = (\overline{B^c})^c$. It is clear then that a subset A is nowhere dense if and only if $(\overline{A})^c$ is dense in its parent space M, i.e. if and only if the interior of $(\overline{A})^c$ is equal to the parent space.

Theorem 1.5. *The Cantor set is closed and nowhere dense.*

Proof. We have already seen that $\mathcal{C}$ is the intersection of closed sets, which implies that $\mathcal{C}$ is itself closed. Furthermore, as previously discussed, the Cantor set contains no intervals of non-zero length, and so $int(\mathcal{C}) = \emptyset$. □

A related idea to that of being nowhere dense is for a metric space to be totally disconnected.

Definition 1.6. A metric space M is totally disconnected if, for any $\epsilon > 0$ and $p \in M$ there exists a clopen subset U of M such that $p \in U \subset M_\epsilon(p)$. That is, there is an arbitrarily small clopen neighborhood centered on every point of M.

With this definition we can prove two more important facts about the Cantor set.

Theorem 1.7. *The Cantor set $\mathcal{C}$ is perfect and totally disconnected.*

Proof. Fix any $\epsilon > 0$ and point $p \in \mathcal{C}$. Let $n \in \mathbb{N}$ be sufficiently large such that $\frac{1}{3^n} < \epsilon$. Then, p is guaranteed to be in one of the intervals (I_n for some $n \in \mathbb{N}$) that make up $\mathcal{C}$, each of length $\frac{1}{3^n}$. The endpoints of the Cantor set in this interval are infinite in number, and all contained in the open interval $(p - \epsilon, p + \epsilon)$. So, p is a cluster point of $\mathcal{C}$, $M_\epsilon(p)$ containing an infinite number of points. And since we are considering any $p \in \mathcal{C}$, $\mathcal{C}$ is perfect. Furthermore, this interval I_n is closed in $\mathbb{R}$ and so in the Cantor set $\mathcal{C}$ as well. Since $I_n^c = \mathcal{C} \backslash I_n$ consists of a countable number of closed intervals, it is itself closed. We can then represent $\mathcal{C}$ as the disjoint union of two clopen sets, $(\mathcal{C} \bigcap I_n)$ and $(\mathcal{C} \bigcap I_n^c)$, the result being that the Cantor set $\mathcal{C}$ is totally disconnected. □

2. Interesting Properties

We have already showed that the Cantor set is nowhere dense. Perhaps the most interesting property is that it is also uncountable. In its construction we remove the same number of points as the number left behind to form the Cantor set, which leads us to this result.

Theorem 2.1. *The Cantor set is uncountable.*

Proof. We demonstrate a surjective function $f : \mathcal{C} \to [0,1]$. As a result, we have that $\#\mathcal{C} \geq \#[0,1]$, i.e., that the cardinality of the Cantor set is at least equal to that of $[0,1]$. However, since $\mathcal{C} \subseteq [0,1]$, its cardinality is also less than or equal to that of $[0,1]$, the conclusion being that the two sets have equal cardinality.

To construct this function f, consider the representation of every $p \in [0,1]$ in base 3 (ternary) notation. That is, every number in $[0,1]$ has a decimal representation involving only the numerals $\{0,1,2\}$. We write $\frac{1}{3}$ as 0.1 and $\frac{2}{3}$ as 0.2, so that I_1, the first middle third removed in our construction of the Cantor set, contains only numbers with base 3 representations of the form 0.1xxx... with "xxx... " stricly between 000... and 222... .

So, after removing the first middle third, the remaining elements of $[0,1]$ are of the form 0.0xxx... or 0.2xxx... with no restriction on the digits after those specified. The second step, similarly, removes numbers of the form 0.01xxx... and 0.21xxx..., leaving numbers with base 3 representations whose first two digits are restricted from being 1. Continuing on to any inductive step n, we see that the n^{th} digit cannot be 1. That is, every $p \in \mathcal{C}$ has a base 3 decimal representation consisting of only the digits $\{0,2\}$.

With this in mind, $f : \mathcal{C} \to [0,1]$ can be defined as taking the number not consisting entirely of the digits $\{0,2\}$ and replacing each occurence of $\{2\}$ by $\{1\}$ in its representation. To show that f is surjective, consider any element $a \in [0,1]$. Represent a in its binary form, and then replace each occurence of a $\{1\}$ digit by a $\{2\}$. This new number, which we label b, satisfies $f(b) = a$. As a result, f is surjective, and by our first comment, we can claim the cardinality of $\mathcal{C}$ equal to that of $[0,1]$, which, as we have already seen, is uncountable. □

This is perhaps the strangest property of the Cantor set. Although it can have no finite intervals, $\mathcal{C}$ contains as many points as its parent interval, an uncountable number. As a consequence of this we can see that, for any $p \in [0,1]$, $\exists q_0, q_1 \in \mathcal{C}$ such that $(q_1 - q_0) = p$. That is, there is a pair of numbers from the Cantor set whose difference is p. If we then consider $(q_0 - q_1)$, the result is that p is only restricted to the larger interval $[-1,1]$.

Theorem 2.2. $\mathcal{C} - \mathcal{C} = [-1,1]$. *Or, equivalently,* $\{p_0 - p_1 \mid p_0, p_1 \in \mathcal{C}\} = [-1,1]$.

Proof. First, we show that any number whose representation in base 3 consists of a finite number of digits, can be expressed as the difference of two elements in the Cantor set. We proceed by induction on L, the length of the base 3 representation. For $L = 1$ there are only three possibilities:

$$0.0 = 0.0 - 0.0,$$

$$0.1 = 0.2 - 0.0222...,$$

$$0.2 = 0.2 - 0.0.$$

For $L = 2$ we add a digit to each of the three expansions of $L = 1$. If the added digit is either of $\{0, 2\}$, the two differenced Cantor set members are easily corrected. For example,

$$0.1 = 0.2 - 0.0222... \implies 0.12 = 0.22 - 0.0222...$$

If, on the other hand, the added digit is $\{1\}$, then either $0.01 = 0.02 - 0.0022...$, $0.11 = 0.21 - 0.0222... = 0.2 - 0.02$, or $0.21 = 0.22 - 0.0022...$. To generalize, pick $p \in [0.1]$ with a base 3 representation of length $L = n$. Then, $p = 0.p_1p_2p_3...p_n$ with $p_n \neq 0$. We construct $q, r \in \mathcal{C}$ such that $p = (q - r)$. If we let $L(p)$ denote the length of the base 3 representation of p, then we will also satisfy $L(q) \leq L(r)$. Furthermore, either $L(r) = \infty$ or $L(r) \leq L(p)$. In the case that $L(r) = \infty$, note that every digit of r after the n^{th} place equals $\{2\}$.

Having already satisfied the $L = 1$ and $L = 2$ cases, we take as assumption the $L = n - 1$ case, i.e. that two members of the Cantor set have already been found such that their difference is $0.p_1p_2p_3...p_{n-1}$. In this case, either

$$0.p_1p_2p_3...p_{n-1} = 0.q_1q_2q_3...q_{n-1} - 0.r_1r_2r_3...r_{n-1} \neq 0,$$

or

$$0.p_1p_2p_3...p_{n-1} = 0.q_1q_2q_3...q_{n-1} - 0.r_1r_2r_3...222...$$

Where $r_i = 2, \forall i > n$. Now, consider the added digit p_n. If $p_n = 0$, we are done. If $p_n = 1$, for the first case previously, set

$$0.p_1p_2p_3...p_{n-1}1 = 0.q_1q_2q_3...q_{n-1}2 - 0.r_1r_2r_3...r_{n-1}0222...$$

If, on the other hand, we are in the second case previously where we have a repeating $\{2\}$ digit, set

$$0.p_1p_2p_3...p_{n-1}1 = 0.q_1q_2q_3...q_{n-1} - 0.r_1r_2r_3...r_{n-1}r_n.$$

If $p_n = 2$,simply set $q_n = 2$ and we have finished with the inductive step on $L = n$.

To finish the proof, we address the case where $p \in [0, 1]$ has an infinite base 3 representation. Then, $p = 0.p_1p_2p_3...$ is a limit of a partial expansion $p_k = 0.p_1p_2p_3...p_k$, $k \in \mathbb{N}$. Then, from our previous discussion, $\exists q_k, r_k \in \mathcal{C}$ such that $p_k = (q_k - r_k)$. We have already proven that the Cantor set is a compact subset of $[0, 1]$, which is equivalent to it being covering compact as well as sequentially compact. Since $\mathcal{C}$ is sequentially compact, every sequence $(a_n)_{n\in\mathbb{N}} \subset \mathcal{C}$ has a convergent subsequence $(a_{n_i})_{i\in\mathbb{N}} \to b \in \mathcal{C}$ for some limit point b. Consider the sequence $(q_n)_{n\in\mathbb{N}}$, which has a convergent sequence $(q_{n_i})_{i\in\mathbb{N}}$. Similarly, the sequence $(r_{n_i})_{i\in\mathbb{N}}$ has a convergent subsequence $(r_{n_{i_j}})_{j\in\mathbb{N}}$. Since each subsequence must converge to the same limit as its parent sequence, we have that the subsequence $(q_{n_{i_j}})_{j\in\mathbb{N}}$ is also convergent. Call their limits r and q, respectively, where $r, q \in \mathcal{C}$. Additionally, $p = (q - r)$, and so the desired representation of p is obtained. □

3. Generalizations

Although our construction of the Cantor set in the first section used the typical "middle-thirds" or ternary rule, we can easily generalize this one-dimensional idea to any length other than $\frac{1}{3}$, excluding of course the degenerate cases of 0 and 1.

3.1. **One-dimensional generalization of the Cantor set.** For this construction, first fix a length $\delta \in (0, \frac{1}{2})$. From $[0, 1]$ remove the open middle interval $I_0$of length $(1 - 2\delta)$ such that

$$[0,1] - I_0 = [0, \delta] \bigcup [1 - \delta, 1]$$

where $I_0 = (\delta, 1 - \delta)$. Next, from $[0, \delta]$ and $[1 - \delta, 1]$, remove the open middle intervals $I_{1,1}$ and $I_{1,2}$, each of length $(\delta - 2\delta^2)$. Call their union I_1, giving

$$I_1 = (\delta^2, \delta - \delta^2) \bigcup (1 - (\delta - \delta^2), 1 - \delta^2)$$

with the length of I_1 equal to $(2\delta - 4\delta^2)$. What remains is then $[0, 1] - (I_0 \bigcup I_1)$ of length $1 - (1 - 4\delta^2) = 4\delta^2$. Proceeding in this manner, we construct sequences of open sets I_n, each the finite union of disjoint open intervals.

Definition 3.1. The **generalized Cantor set** $\mathcal{C}_\delta$ is then defined as $\mathcal{C}_\delta = [0, 1] - \bigcup_{i=0}^{\infty} I_i$.

We can see that setting $\delta = \frac{1}{3}$ returns the standard middle-thirds Cantor set. This form of the Cantor set, for any $\delta \in (0, \frac{1}{2})$, shares the same properties of the ternary Cantor set, namely it is a compact, non-empty, perfect, totally disconnected, and uncountable metric space.

3.2. **Higher dimensional generalizations.** We can extend the one-dimensional case to two dimensions by starting with the unit box $[0, 1] \times [0, 1]$ instead of the unit interval. Simiarly, we can extend to three dimensions by starting with the unit cube and symetrically removing middle-third "cubes" in the standard construction. The result is sometimes called the **Cantor dust**, and the two and three-dimension cases are shown below:

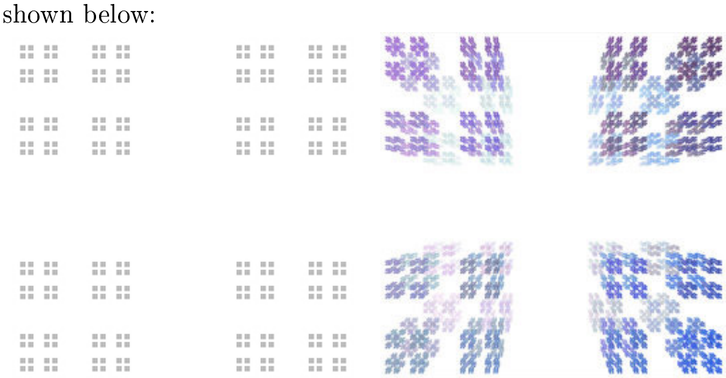

Figure 1. Two and three-dimensional examples of the "Cantor dust."

Each of these images emphasizes the fractal nature of the Cantor set. Firstly, it is self-similar, since "zooming in" on the first iteration in any of the constructions, we can see that it looks identical to the original set. For example, after removing the first middle-third of the unit interval, either of the two remaining thirds is identical to $[0, 1]$, even in cardinality. The defining characteristic of a fractal is that it exhibits the same type of structure at all scales, immediately evident from Figure 1. So, the Cantor set is in some sense one of the most basic examples of a fractal.

Although that is a story for another day, we have already explored several interesting properties of the Cantor set. Perhaps most surprising was the fact that it is uncountable while at the same time being nowhere dense, and containing no finite intervals. It is an interesting mathematical construction with links to several areas of analysis and topology, only the briefest of which we have discussed here.

References

[1] Pugh, C. C., *Real Mathematical Analysis*, (2002).
[2] DiBenedetto, E., *Real Analysis*, (2002).
[3] Aliprantis, C.D., and Burkinshaw, O., *Principles of Real Analysis*, (1981).
[4] Wikipedia contributors, *Cantor set,* http://en.wikipedia.org/wiki/Cantor_set (Accessed 5 April, 2007).

Dylan Nelson, University of California - Berkeley, Berkeley, CA 94704
E-mail address: dnelson@berkeley.edu
URL: http://ugastro.berkeley.edu/~dnelson/