Notes on the Cantor set

1. Definition of the Cantor set

Given a set $A \subset \mathbb{R}$, let $\frac{1}{3}A = \{ \frac{1}{3}x : x \in A \}$; that is, the image of $A$ under the map $x \rightarrow \frac{1}{3}x$. Similarly, $A + \frac{2}{3} = \{ x + \frac{2}{3} : x \in A \}$.

Start with $C_0 = [0, 1]$, the closed unit interval in $\mathbb{R}$.

Let $C_1 = \frac{1}{3}C_0 \cup \left( \frac{1}{3}C_0 + \frac{2}{3} \right) = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$.

Let $C_2 = \frac{1}{3}C_1 \cup \left( \frac{1}{3}C_1 + \frac{2}{3} \right) = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$.

Recursively, let $C_{k+1} = \frac{1}{3}C_k \cup \left( \frac{1}{3}C_k + \frac{2}{3} \right)$.

Note that if for any subset $A \subset [0, 1]$, the sets $\frac{1}{3}A$ and $\frac{1}{3}A + \frac{2}{3}$ are disjoint sets, respectively contained in $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$.

We now see by induction that $C_k$ consists of $2^k$ disjoint, closed intervals of length $3^{-k}$. It is also easy to see that $C_{k+1} \subset C_k$ using induction: if $C_k \subset C_{k-1}$ then

$$C_{k+1} = \frac{1}{3}C_k \cup \left( \frac{1}{3}C_k + \frac{2}{3} \right) \subset \frac{1}{3}C_{k-1} \cup \left( \frac{1}{3}C_{k-1} + \frac{2}{3} \right) = C_k.$$ 

And we have that $C_1 \subset C_0$ to start the induction.

We now let $C = \bigcap_{k=1}^{\infty} C_k$, which is a nonempty compact subset of $[0, 1]$. Also,

$$C = \frac{1}{3}C \cup \left( \frac{1}{3}C + \frac{2}{3} \right).$$

The complement of $C$ is an open subset of $(0, 1)$ since $0 \in C$ and $1 \in C$, so the complement is the countable union of disjoint open intervals. These are the "middle thirds" that you remove to construct $C$, and there are exactly $2^{k-1}$ open intervals of length $3^{-k}$ for $k \geq 1$. Note that

$$\sum_{k=1}^{\infty} 2^{k-1}3^{-k} = \frac{1}{2} \sum_{k=1}^{\infty} \left( \frac{2}{3} \right)^k = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{1 - \frac{1}{3}} = 1$$

so the lengths of those open intervals add up to 1.

2. Ternary expansions

We consider base 3 expansions of numbers $x \in [0, 1]$. We will always take $a_j$ is equal to 0, 1, or 2 in what follows.

Suppose we have a sequence $(a_1, a_2, a_3, \ldots) \in \{0, 1, 2\}^\mathbb{N}$. We associate to this sequence a number

$$\sum_{k=1}^{\infty} \frac{a_k}{3^k}.$$
The sequence \( b_n = \sum_{k=1}^{n} a_k 3^{-k} \) is increasing in \( n \), and bounded above by 
\[
2 \sum_{k=1}^{\infty} \frac{1}{3^k} = \frac{2}{3} \frac{\frac{1}{3}}{1 - \frac{1}{3}} = 1
\]
so the infinite sum makes sense as the least upper bound of the \( b_n \).

We next see that this map from \( \{0, 1, 2\}^\mathbb{N} \to [0,1] \) is onto (but not 1–1).

First we consider \( x \in [0,1) \).

• Suppose \( x \in [0,1) \), and let \( a_1 \) be the largest integer such that \( \frac{a_1}{3} \leq x \). Then since \( 0 \leq x < 1 \) we have \( 0 \leq a_1 \leq 2 \), and

\[
0 \leq x - \frac{a_1}{3} < \frac{1}{3}
\]

Now let \( a_2 \) be the largest integer so \( \frac{a_2}{3^2} \leq x - \frac{a_1}{3} \). Then by (1) we must have \( 0 \leq a_2 \leq 2 \), and

\[
0 \leq x - \frac{a_1}{3} - \frac{a_2}{3^2} < \frac{1}{3^2}
\]

In general, we will take \( a_j \) recursively so that

\[
0 \leq x - \sum_{k=1}^{n} \frac{a_k}{3^k} < \frac{1}{3^n}
\]

Briefly, the \( a_j \) are the largest integers from 0, 1, or 2 so that \( \sum_{k=1}^{n} \frac{a_k}{3^k} \leq x \).

The bound (2) implies that

\[
x = \lim_{n \to \infty} \sum_{k=1}^{n} \frac{a_k}{3^k} = \sum_{k=1}^{\infty} \frac{a_k}{3^k}
\]

This shows the map is onto \([0,1)\). For the number 1, we take \( a_j = 2 \) for all \( j \), and note

\[
\sum_{j=1}^{n} \frac{2}{3^j} = 2 \sum_{k=0}^{n-1} \frac{1}{3^k} = 2 \frac{1 - \left(\frac{1}{3}\right)^n}{1 - \frac{1}{3}} = 1 - \frac{1}{3^n}
\]

Thus \( \sum_{k=1}^{\infty} \frac{2}{3^k} = 1 \).

• We now determine how the map \( \{0,1,2\}^\mathbb{N} \to [0,1] \) can fail to be 1–1. Suppose that we have two sequences \( (a_1, a_2, \ldots) \) and \( (b_1, b_2, \ldots) \) such that

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

Suppose that \( n \) is the first position where \( a_n \neq b_n \), and assume \( a_n > b_n \). We then write

\[
\frac{a_n - b_n}{3^n} + \sum_{k=n+1}^{\infty} \frac{a_k}{3^k} = \sum_{k=n+1}^{\infty} \frac{b_k}{3^k}
\]
Since \[
\frac{a_n - b_n}{3^n} \geq \frac{1}{3^n} \quad \text{and} \quad \sum_{k=n+1}^{\infty} \frac{b_k}{3^k} \leq \sum_{k=n+1}^{\infty} \frac{2}{3^k} = \frac{1}{3^n}
\]
The only way we can have equality is if \(a_n - b_n = 1\), and
\[
a_k = 0 \quad \text{and} \quad b_k = 2 \quad \text{for} \quad k \geq n + 1.
\]
That is, the sequence \(a_k\) has terminal 0’s and \(b_k\) has terminal 2’s, for example
\[
0.1202000000 \cdots = 0.1201222222 \cdots
\]
Other than this kind of case, the ternary expansion of \(x \in [0,1]\) is unique.

Note that the values of \(x\) for which the expansion has two possibilities are precisely those \(x\) of the form
\[
x = \frac{m}{3^n} \quad \text{for some} \quad m \in \{0, \ldots, 3^n - 1\}.
\]

**Observe:** in the above argument, since \(a_n = b_n + 1\) we must have either \(a_n = 1\) or \(b_n = 1\). So different sequences in \(\{0,2\}^\mathbb{N}\) cannot give the same value of \(x\).

### 3. Ternary expansions and the Cantor set

We now claim that the Cantor set consists precisely of numbers of the form
\[
x = \sum_{k=1}^{\infty} \frac{a_k}{3^k}
\]
where each \(a_k\) is either 0 or 2. The map \(\{0,2\}^\mathbb{N} \to C\) is then a bijection by the above observation.

Suppose \(x\) is given by (3). Then
\[
\frac{1}{3}x = \sum_{k=1}^{\infty} \frac{b_k}{3^k} \quad \text{where} \quad b_1 = 0, \quad b_k = a_{k-1} \text{ if } k \geq 2,
\]
\[
\frac{1}{3}x + \frac{1}{3} = \sum_{k=1}^{\infty} \frac{b_k}{3^k} \quad \text{where} \quad b_1 = 2, \quad b_k = a_{k-1} \text{ if } k \geq 2.
\]
Thus, \(x \in C_1\) if and only if it equals a ternary expansion where either \(a_1 = 0\) or \(a_1 = 2\), since it is of the form \(\frac{1}{3}y\) or \(\frac{1}{3}y + \frac{2}{3}\) for some \(y \in [0,1]\). Repeating the argument, we see that \(x \in C_n\) iff it equals a ternary expansion where \(a_k\) is either 0 or 2 for \(1 \leq k \leq n\).

It follows that ternary expansions where every digit is either 0 or 2 belong to \(C_n\) for every \(n\), hence give an element of \(C\). This shows that \(\{0,2\}^\mathbb{N}\) maps into \(C\) under the ternary expansion.

Suppose now that \(x \in C\). If the ternary expansion of \(x\) is unique, the above argument shows that every \(a_k\) is either 0 or 2. If \(x\) has two ternary expansions, let \(n\) be the first digit they differ. Then the above argument
shows that $a_k$ is either 0 or 2 for $1 \leq k < n$, and we know $a_k$ is either identically 0 or identically 2 for $k > n$. And one of the two expansions has $a_n = 1$, the other must have $a_n = 0$ or 2. So $x$ has one expansion where $a_k \neq 1$ for every $k$. This shows that $\{0, 2\}^\mathbb{N}$ maps onto $C$ under the ternary expansion.

4. The Cantor map

We now can construct a map $f : C \to [0, 1]$ that is onto. To do this, expression $x \in C$ uniquely as

$$x = \sum_{k=1}^{\infty} \frac{a_k}{3^k} \text{ where } a_k \in \{0, 2\}.$$ 

Define

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

That is, if $x$ has ternary expansion $(a_1, a_2, \ldots)$, then $f(x)$ has binary expansion $(\frac{1}{2} a_1, \frac{1}{2} a_2, \ldots)$. Since $a_k \in \{0, 2\}$ we have $\frac{1}{2} a_k \in \{0, 1\}$, so the binary expansion gives an element of $[0, 1]$, and the map is onto since every sequence in $\{0, 1\}^\mathbb{N}$ is obtained this way.

The map is not 1–1: for example

$$\frac{1}{3} \to (0, 2, 2, 2, 2, \ldots) \to (0, 1, 1, 1, 1, \ldots) \to \frac{1}{2}$$

and

$$\frac{2}{3} \to (2, 0, 0, 0, 0, \ldots) \to (1, 0, 0, 0, 0, \ldots) \to \frac{1}{2}$$