The Cantor set and the Devil's staircase

Manuela Girotti

MATH 3441 Real Analysis I

In these notes we will prove a few properties of the Cantor set and we will construct a pathological function $\mathscr{D} : [0, 1] \to [0, 1]$ (the "Devil's staircase") that is continuous everywhere, whose derivative is zero almost everywhere, but it somehow magically rises from 0 to 1.

The section about the Devil's staircase will be more of a collection of facts with very few proofs, than a thorough exposition. We will explore the topic more in details in our upcoming course Real Analysis II!

1 The Cantor set $\mathcal{C}$

The (standard) Cantor set is the set $\mathcal{C} \subseteq [0, 1]$ constructed as follows.

We start with the full interval $F_0 = [0, 1]$. Divide F_0 into three equal parts and let I_1 be the open middle third of F_0, that is $I_1 = (\frac{1}{3}, \frac{2}{3})$, and define

$$F_1 = F_0 \setminus I_1 = \left[0, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, 1\right] \ .$$

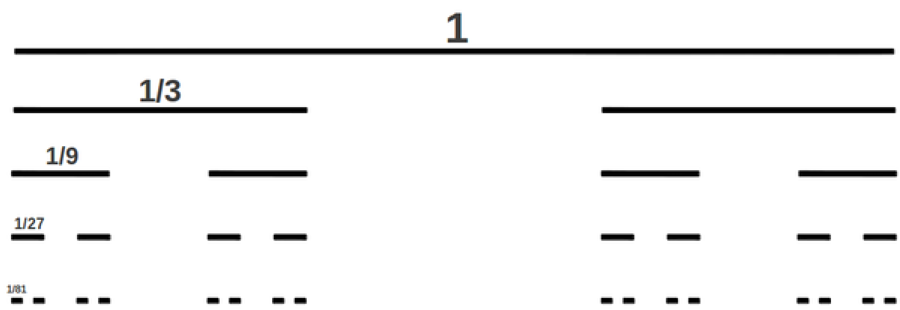
Let I_2 and I_3 be the open middle thirds of the two component intervals of F_1, i.e. $I_2 = (\frac{1}{9}, \frac{2}{9})$ and $I_3 = (\frac{7}{9}, \frac{8}{9})$, and define

$$F_2 = F_1 \setminus (I_2 \cup I_3) = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{3}{9}\right] \cup \left[\frac{6}{9}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \ .$$

Continue the construction iteratively: having constructed the set F_n, which is the disjoint union of 2^n closed intervals each of length $\frac{1}{3^n}$, let $I_{2^n}, I_{2^n+1}, \ldots, I_{2^{n+1}-1}$ be the open middle thirds to of these 2^n component intervals and define

$$F_{n+1} = F_n \setminus (I_{2^n} \cup I_{2^n+1} \cup \ldots \cup I_{2^{n+1}-1}) \ .$$

See the figure below to get an idea of how these sets look like.



Then, the Cantor set is

$$\mathcal{C} = \bigcap_{n=1}^{\infty} F_n.$$

The first straightforward properties of the Cantor set are the following.

Proposition 1. *The Cantor set is closed and nowhere dense.*

Proof. For any $n \in \mathbb{N}$, the set F_n is a *finite* union of closed intervals. Therefore, $\mathcal{C}$ is closed because intersection of a family of closed sets. Notice that this will additionally imply that $\mathcal{C}$ is compact (as $\mathcal{C} \subset [0,1]$).

Now, since $\mathcal{C} = \overline{\mathcal{C}}$, we simply need to prove that $\mathcal{C}$ has empty interior: $\mathcal{C}^\circ = \emptyset$. Assume by contradiction that $\mathcal{C}^\circ \neq \emptyset$, i.e. there exists an interior point $x_0 \in \mathcal{C}$: therefore, $\exists\ \epsilon > 0$ such that $(x_0 - \epsilon, x_0 + \epsilon) \subset \mathcal{C}$. Let $N \in \mathbb{N}$ such that $\frac{1}{3^N} < 2\epsilon$. By construction,

$$\mathcal{C} = \bigcap_{n=1}^{\infty} F_n \subset F_N$$

and F_N is the union of 2^N intervals, each of length $\frac{1}{3^N}$ which is strictly less than the length of the interval $(x_0 - \epsilon, x_0 + \epsilon)$. Thus, the contradiction. $\square$

A little more work is required to prove the following property.

Proposition 2. *The Cantor set is uncountable.*

Let us first consider an equivalent representation of the Cantor set that will help us in the proof. We start by noticing that every $x \in [0,1]$ admits *at most* two representations in base 3 (ternary expansion):

$$x = 0.d_1 d_2 d_3 \ldots = \sum_{j=1}^{\infty} \frac{d_j}{3^j}$$

with $d_j \in \{0,1,2\}$ (the endpoint $x = 1$ has representation $1 = 0.22222\ldots$).

The only cases where $x \in [0,1]$ may admit two equivalent expansions arise when $x \in \mathbb{Q}$ and its denominator is a multiple of 3: for example, $\frac{1}{3} = 0.10000\ldots = 0.02222\ldots$. In these (at most countably many) cases we have two representations for $x = 0.d_1 d_2 d_3 \ldots = 0.c_1 c_2 c_3 \ldots$), however they can be easily identified: let $N := \min\{j \in \mathbb{N} \mid d_j \neq c_j\}$ and (w.l.o.g.) assume $d_N < c_N$; then, necessarily $c_N = d_N + 1$ and $c_{N+1} = c_{N+1} = \ldots = 0$ and $d_{N+1} = d_{N+2} = \ldots = 2$ (in particular, either $c_N = 1$ or $d_N = 1$). Therefore among those two ternary expansion, there is only one which doesn't contain 1's.

Within this setting, the Cantor set can be represented as

$$\mathcal{C} = \{x = 0.d_2 d_2 d_3 \ldots \mid d_j \in \{0,2\}\ \forall\ j \in \mathbb{N}\}\ .$$

Indeed, starting from the first "mid-pinch" I_1 we have that the endpoints

$$\frac{1}{3} = 0.10000\ldots = 0.02222\ldots \qquad \text{and} \qquad \frac{2}{3} = 0.20000\ldots = 0.12222\ldots$$

and any other point $x \in I_1$ has base-3 representation of the form $x = 0.1d_2 d_3 d_4 d_5 \ldots$ where the sequence $d_2 d_3 d_4 d_5 \ldots$ is strictly between $0000\ldots$ and $2222\ldots$.

Therefore, all points $x \in F_1 = [0,1] \setminus I_1$ have base-3 representation of the form $x = 0.0d_2 d_3 d_4 d_5 \ldots$ or $x = 0.2d_2 d_3 d_4 d_5 \ldots$. At the next step, similarly, the "mid-pinch" removes numbers of the form $x = 0.01d_3 d_4 d_5 \ldots$ and $x = 0.21d_3 d_4 d_5 \ldots$, therefore points $x \in F_2 = F_1 \setminus (I_2 \cup I_3)$ have a base-3 representation whose first two digits are restricted from being equal to 1. Continuing on in the construction of the sets F_n, we can see that the points in F_n have a base-3 expansion whose n-th digit is not equal to 1.

We are now ready to prove Proposition 2.

Proof. Assume that $\mathcal{C}$ is countable and collect all of its points in an infinite table:

$$\begin{aligned}
x_1 &= 0.d_1^1 d_2^1 d_3^1 d_4^1 d_5^1 \ldots \\
x_2 &= 0.d_1^2 d_2^2 d_3^2 d_4^2 d_5^2 \ldots \\
x_3 &= 0.d_1^3 d_2^3 d_3^3 d_4^3 d_5^3 \ldots \\
x_1 &= 0.d_1^4 d_2^4 d_3^4 d_4^4 d_5^4 \ldots \\
&\vdots
\end{aligned}$$

By using Cantor diagonalization trick (seen in class), we can easily construct a new point $\bar{x} \in \mathcal{C}$, which has not being accounted for in the table, by considering all the "diagonal" digits d_j^j in the table above and replacing any 0 with 2 and viceversa. Thus, the contradiction. $\square$

2 The Devil's staircase

The formal definition of the Devil's staircase function $\mathscr{D}$ is the following. Recall the ternary representation of a point $x \in [0,1]$:

$$x = \sum_{j=1}^{\infty} \frac{d_j^{(x)}}{3^j} , \qquad \text{with } d_j^{(x)} \in \{0,1,2\} ;$$

denote by N_x the smallest index j such that $d_j^{(x)} = 1$, if it exists, otherwise $N_x = \infty$. Then, the Cantor function is the following

$$\mathscr{D}(x) := \frac{1}{2^{N_x}} + \frac{1}{2} \sum_{j=1}^{N_x - 1} \frac{d_j^{(x)}}{2^j}$$

Remark 3. *Notice that $\mathscr{D}$ is well defined, i.e. its value is independent on the choice of the base-3 expansion in the cases where x admits two of them.*

An alternative way to define such a function (and an easier way to visualize it) is through an iterative construction. We consider the following a sequence of continuous functions $\{f_n\}_{n\in\mathbb{N}} \in F_b([0,1];[0,1]) \cap C^0([0,1])$: let

$$f_1(x) = x$$

and $\forall\ n \in \mathbb{N}$ we have

$$f_{n+1}(x) = \begin{cases} \frac{1}{2} f_n(x) & x \in [0, \frac{1}{3}) \\ \frac{1}{2} & x \in [\frac{1}{3}, \frac{2}{3}] \\ \frac{1}{2} + \frac{1}{2} f_n(3x-2) & x \in (\frac{2}{3}, 1] \end{cases}$$

By induction, it is possible to prove that f_n is indeed a continuous function on $[0,1]$, for every $n \in \mathbb{N}$.

Proposition 4. *The sequence of functions $\{f_n\}$ defined above converges uniformly to the Devil's staircase:*

$$\|f_n - \mathscr{D}\|_\infty \to 0 , \qquad \text{as } n \to \infty .$$

Proof. As a first step, by using again the ternary representation of points in $[0,1]$, we notice that $f_n(x)$ converges *pointwise* to $\mathscr{D}(x)$ (why?). Th uniform convergence follows by noticing that the sequence of functions $\{f_n\}$ is Cauchy.

Indeed, consider the quantity

$$\|f_{n+1} - f_n\|_\infty = \sup_{x\in[0,1]} |f_{n+1}(x) - f_n(x)| = \frac{1}{2} \sup_{x\in[0,1]} |f_n(x) - f_{n-1}(x)| = \frac{1}{2} \|f_n - f_{n-1}\|_\infty , \quad \forall\ n \geq 1;$$

therefore,

$$\|f_{n+1} - f_n\|_\infty \leq \frac{1}{2} \|f_n - f_{n-1}\|_\infty \leq \cdots \leq \frac{1}{2^n} \underbrace{\|f_2 - f_1\|_\infty}_{=\frac{1}{6}} .$$

Recall now the following property (see Homework 6): given a sequence $\{x_n\}$ in a metric space (X,d), if there exists a sequence $\{\gamma_n\}_{n=1}^{\infty} \subset [0,+\infty)$ such that

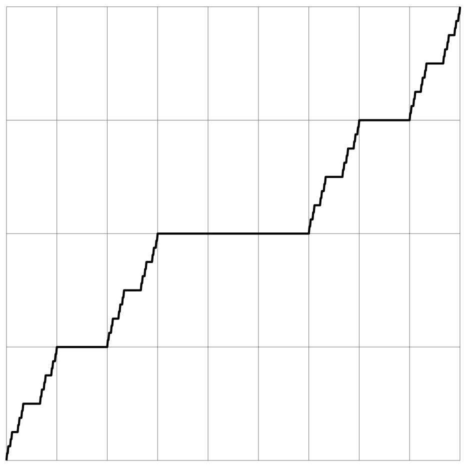


Figure 1: The Devil's staircase $\mathscr{D}$.

$$\text{(i) } d(x_n, x_{n+1}) \leq \gamma_n, \quad \forall\, n \geq 1 \qquad \text{and} \qquad \text{(ii) } \textstyle\sum_{n=1}^{\infty} \gamma_n < \infty\ ,$$

then $\{x_n\}$ is a Cauchy sequence.

In our case we have $\gamma_n = \frac{1}{3 2^{n+1}} \in (0, +\infty\ \forall\, n \geq 1$ and $\sum_n \gamma_n = \frac{1}{3} \sum_n \frac{1}{2^{n+1}} < +\infty$ (geometric series with ration $q = \frac{1}{2} < 1$). Therefore our sequence of functions $\{f_n\} \subset F_b([0,1];[0,1]) \cap C^0([0,1])$ is a Cauchy sequence, therefore it is convergent (recall that $(F_b([0,1];[0,1]), \|\cdot\|_\infty)$ is a Banach space).

$\square$

The Devil's staircase is related to the Cantor set because by construction $\mathscr{D}$ is constant on all the removed intervals from the Cantor set. For example: $\mathscr{D}(x) = \frac{1}{2}$ for $x \in I_1 = (\frac{1}{3}, \frac{2}{3})$, $\mathscr{D}(x) = \frac{1}{4}$ for $x \in I_2 = (\frac{1}{9}, \frac{2}{9})$ and $\mathscr{D}(x) = \frac{3}{4}$ for $x \in I_3 = (\frac{7}{9}, \frac{8}{9})$, and so on.

Further properties are listed (and partly proven) in the Proposition below:

Proposition 5. *The Devil's staircase $\mathscr{D} : [0,1] \to [0,1]$ satisfies the following properties:*

1. *$\mathscr{D}$ is (uniformly) continuous and monotone increasing.*
2. *$\mathscr{D}$ has derivative equal to zero almost everywhere.*
   (the precise formulation is that $\mathscr{D}'(x) = 0\ \forall\ x \in [0,1] \setminus \mathcal{C}$ and $\mathcal{C}$ has measure zero).
3. *The arc length of the graph of $\mathscr{D}$ is equal to 2.*

Proof. (partial – tune in for Real Analysis II for more!

$\mathscr{D}$ is continuous, since uniform limit of continuous functions. It is additionally uniformly continuous because it is defined on the compact set $[0,1]$. Additionally, $\forall\ n \in \mathbb{N}$ we have that $f_n(0) = 0$, $f_n(1) = 1$ and $f_n(x) \leq f_n(y)\ \forall\ x, y \in [0,1]$, $x \leq y$, therefore such properties still hold in the limit: $\mathscr{D}(0) = 0$, $\mathscr{D}(1) = 1$ and $\mathscr{D}(x) \leq \mathscr{D}(y)\ \forall\ x < y$. $\square$