

## Chapter udf

# History and Mythology of Set Theory

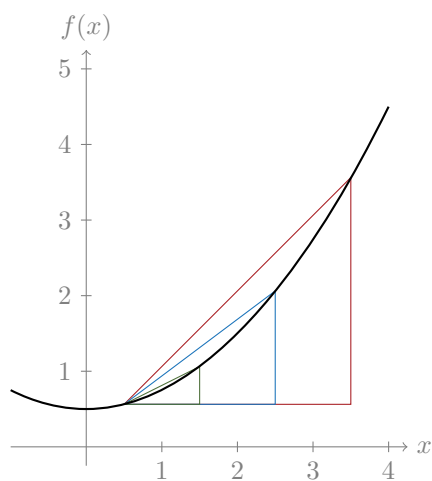
This chapter includes the historical prelude from Tim Button's Open Set Theory text.

### set.1 Infinitesimals and Differentiation

[his:set:infinitesimals:](#)  
[sec](#)

Newton and Leibniz discovered the calculus (independently) at the end of the 17th century. A particularly important application of the calculus was *differentiation*. Roughly speaking, differentiation aims to give a notion of the “rate of change”, or gradient, of a function at a point.

Here is a vivid way to illustrate the idea. Consider the function  $f(x) = x^2/4 + 1/2$ , depicted in black below:



Suppose we want to find the gradient of the function at  $c = 1/2$ . We start by drawing a triangle whose hypotenuse approximates the gradient at that point, perhaps the red triangle above. When  $\beta$  is the base length of our triangle, its height is  $f(1/2 + \beta) - f(1/2)$ , so that the gradient of the hypotenuse is:

$$\frac{f(1/2 + \beta) - f(1/2)}{\beta}.$$

So the gradient of our red triangle, with base length 3, is exactly 1. The hypotenuse of a smaller triangle, the blue triangle with base length 2, gives a better approximation; its gradient is  $3/4$ . A yet smaller triangle, the green triangle with base length 1, gives a yet better approximation; with gradient  $1/2$ .

Ever-smaller triangles give us ever-better approximations. So we might say something like this: the hypotenuse of a triangle with an *infinitesimal* base length gives us the gradient at  $c = 1/2$  itself. In this way, we would obtain a formula for the (first) derivative of the function  $f$  at the point  $c$ :

$$f'(c) = \frac{f(c + \beta) - f(c)}{\beta} \text{ where } \beta \text{ is infinitesimal.}$$

And, roughly, this is what Newton and Leibniz said.

However, since they have said this, we must ask them: what is an *infinitesimal*? A serious dilemma arises. If  $\beta = 0$ , then  $f'$  is ill-defined, for it involves dividing by 0. But if  $\beta > 0$ , then we just get an *approximation* to the gradient, and not the gradient itself.

This is not an anachronistic concern. Here is Berkeley, criticizing Newton's followers:

I admit that signs may be made to denote either any thing or nothing: and consequently that in the original notation  $c + \beta$ ,  $\beta$  might have signified either an increment or nothing. But then which of these soever you make it signify, you must argue consistently with such its signification, and not proceed upon a double meaning: Which to do were a manifest sophism. (Berkeley 1734, §XIII, variables changed to match preceding text)

To defend the infinitesimal calculus against Berkeley, one might reply that the talk of "infinitesimals" is merely figurative. One might say that, so long as we take a *really small* triangle, we will get a *good enough* approximation to the tangent. Berkeley had a reply to this too: whilst that might be good enough for engineering, it undermines the *status* of mathematics, for

we are told that *in rebus mathematicis errores quàm minimi non sunt contemnendi*. [In the case of mathematics, the smallest errors are not to be neglected.] (Berkeley, 1734, §IX)

The italicised passage is a near-verbatim quote from Newton's own *Quadrature of Curves* (1704).

Berkeley’s philosophical objections are deeply incisive. Nevertheless, the calculus was a massively successful enterprise, and mathematicians continued to use it without falling into error.

## set.2 Rigorous Definition of Limits

his:set:limits:sec These days, the standard solution to the foregoing problem is to get rid of the infinitesimals. Here is how.

We saw that, as  $\beta$  gets smaller, we get better approximations of the gradient. Indeed, as  $\beta$  gets arbitrarily close to 0, the value of  $f'(c)$  “tends without limit” to the gradient we want. So, instead of considering what happens *at*  $\beta = 0$ , we need only consider the *trend* of  $f'(c)$  as  $\beta$  approaches 0.

Put like this, the general challenge is to make sense of claims of this shape:

As  $x$  approaches  $c$ ,  $g(x)$  tends without limit to  $\ell$ .

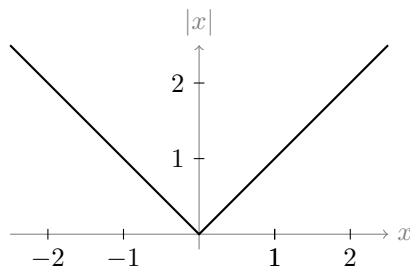
which we can write more compactly as follows:

$$\lim_{x \rightarrow c} g(x) = \ell.$$

In the 19th century, building upon earlier work by Cauchy, Weierstrass offered a perfectly rigorous definition of this expression. The idea is indeed that we can make  $g(x)$  as close as we like to  $\ell$ , by making  $x$  suitably close to  $c$ . More precisely, we stipulate that  $\lim_{x \rightarrow c} g(x) = \ell$  will mean:

$$(\forall \varepsilon > 0)(\exists \delta > 0)\forall x (|x - c| < \delta \rightarrow |g(x) - \ell| < \varepsilon).$$

The vertical bars here indicate absolute magnitude. That is,  $|x| = x$  when  $x \geq 0$ , and  $|x| = -x$  when  $x < 0$ ; you can depict *that* function as follows:



So the definition says roughly this: you can make your “error” less than  $\varepsilon$  (i.e.,  $|g(x) - \ell| < \varepsilon$ ) by choosing arguments which are no more than  $\delta$  away from  $c$  (i.e.,  $|x - c| < \delta$ ).

Having defined the notion of a limit, we can use it to avoid infinitesimals altogether, stipulating that the gradient of  $f$  at  $c$  is given by:

$$f'(c) = \lim_{x \rightarrow 0} \left( \frac{f(c+x) - f(c)}{x} \right) \text{ where a limit exists.}$$

It is important, though, to realise why our definition needs the caveat “where a limit exists”. To take a simple example, consider  $f(x) = |x|$ , whose graph we just saw. Evidently,  $f'(0)$  is ill-defined: if we approach 0 “from the right”, the gradient is always 1; if we approach 0 “from the left”, the gradient is always  $-1$ ; so the limit is undefined. As such, we might add that a function  $f$  is *differentiable* at  $x$  iff such a limit exists.

We have seen how to handle differentiation using the notion of a *limit*. We can use the same notion to define the idea of a *continuous* function. (Bolzano had, in effect, realised this by 1817.) The Cauchy–Weierstrass treatment of continuity is as follows. Roughly: a function  $f$  is continuous (at a point) provided that, if you demand a certain amount of precision concerning the output of the function, you can guarantee this by insisting upon a certain amount of precision concerning the input of the function. More precisely:  $f$  is continuous at  $c$  provided that, as  $x$  tends to zero, the difference between  $f(c + x)$  and  $f(c)$  itself tends to 0. Otherwise put:  $f$  is *continuous* at  $c$  iff  $f(c) = \lim_{x \rightarrow c} f(x)$ .

To go any further would just lead us off into real analysis, when our subject matter is set theory. So now we should pause, and state the moral. During the 19th century, mathematicians learnt how to do without infinitesimals, by invoking a rigorously defined notion of a *limit*.

### set.3 Pathologies

However, the definition of a *limit* turned out to allow for some rather “pathological” constructions.

[his:set:pathology:  
sec](#)

Around the 1830s, Bolzano discovered a function which was *continuous everywhere*, but *differentiable nowhere*. (Unfortunately, Bolzano never published this; the idea was first encountered by mathematicians in 1872, thanks to Weierstrass’s independent discovery of the same idea.)<sup>1</sup> This was, to say the least, rather surprising. It is easy to find functions, such as  $|x|$ , which are continuous everywhere but not differentiable at a particular point. But a function which is continuous everywhere but differentiable *nowhere* is a very different beast. Consider, for a moment, how you might try to draw such a function. To ensure it is continuous, you must be able to draw it without ever removing your pen from the page; but to ensure it is differentiable nowhere, you would have to abruptly change the direction of your pen, constantly.

Further “pathologies” followed. In January 5 1874, Cantor wrote a letter to Dedekind, posing the problem:

Can a surface (say a square including its boundary) be one-to-one correlated to a line (say a straight line including its endpoints) so that to every point of the surface there corresponds a point of the line, and conversely to every point of the line there corresponds a point of the surface?

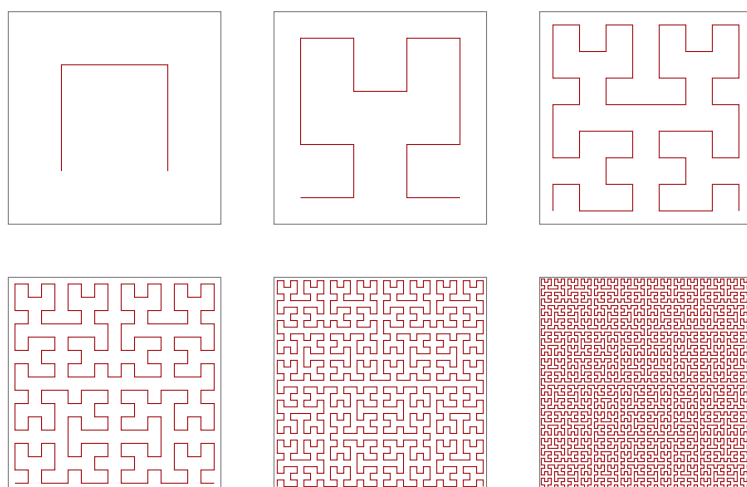
---

<sup>1</sup>The history is documented in extremely thorough footnotes to the Wikipedia article on the Weierstrass function.

It still seems to me at the moment that the answer to this question is very difficult—although here too one is so impelled to say *no* that one would like to hold the proof to be almost superfluous. [Quoted in [Gouvêa 2011](#)]

But, in 1877, Cantor proved that he had been wrong. In fact, a line and a square have exactly the same number of points. He wrote on 29 June 1877 to Dedekind “*je le vois, mais je ne le crois pas*”; that is, “I see it, but I don’t believe it”. In the “received history” of mathematics, this is often taken to indicate just how *literally incredible* these new results were to the mathematicians of the time. (The correspondence is presented in [Gouvêa \(2011\)](#), and we return to it in [section set.4](#). Cantor’s proof is outlined in [section set.5](#).)

Inspired by Cantor’s result, Peano started to consider whether it might be possible to map a line *smoothly* onto a plane. This would be a *curve which fills space*. In 1890, Peano constructed just such a curve. This is truly counter-intuitive: Euclid had defined a line as “breadthless length” (Book I, Definition 2), but Peano had shown that, by curling up a line appropriately, its length can be turned into breadth. In 1891, Hilbert described a slightly more intuitive space-filling curve, together with some pictures illustrating it. The curve is constructed in sequence, and here are the first six stages of the construction:



In the limit—a notion which had, by now, received rigorous definition—the entire square is filled in solid red. And, in passing, Hilbert’s curve is continuous everywhere but differentiable nowhere; intuitively because, in the infinite limit, the function abruptly changes direction at every moment. (We will outline Hilbert’s construction in more detail in [section set.6](#).)

For better or worse, these “pathological” geometric constructions were treated as a reason to doubt appeals to geometric intuition. They became something approaching *propaganda* for a new way of doing mathematics, which would culminate in set theory. In the later myth-building of the subject, it was repeated, often, that these results were both perfectly rigorous and perfectly

shocking. They therefore served a dual purpose: as a warning against relying upon geometric intuition, and as a demonstration of the fertility of new ways of thinking.

## set.4 More Myth than History?

Looking back on these events with more than a century of hindsight, we must be careful not to take these verdicts on trust. The results were certainly novel, exciting, and surprising. But how truly shocking were they? And did they really demonstrate that we should not rely on geometric intuition?

[his:set:mythology:  
sec](#)

On the question of shock, Gouvêa (2011) points out that Cantor’s famous note to Dedekind, “*je le vois, mais je ne le crois pas*” is taken rather out of context. Here is more of that context (quoted from Gouvêa):

Please excuse my zeal for the subject if I make so many demands upon your kindness and patience; the communications which I lately sent you are even for me so unexpected, so new, that I can have no peace of mind until I obtain from you, honoured friend, a decision about their correctness. So long as you have not agreed with me, I can only say: *je le vois, mais je ne le crois pas*.

Cantor knew his result was “so unexpected, so new”. But it is doubtful that he ever found his result *unbelievable*. As Gouvêa points out, he was simply asking Dedekind to check the proof he had offered.

On the question of geometric intuition: Peano published his space-filling curve without including any diagrams. But when Hilbert published his curve, he explained his purpose: he would provide readers with a clear way to understand Peano’s result, if they “help themselves to the following geometric intuition”; whereupon he included a series of *diagrams* just like those provided in [section set.3](#).

More generally: whilst diagrams have fallen rather out of fashion in published proofs, there is no getting round the fact that mathematicians *frequently* use diagrams when proving things. (Roughly put: good mathematicians know when they can rely upon geometric intuition.)

In short: don’t believe the hype; or at least, don’t just take it on trust. For more on this, you could read Giaquinto (2007).

## set.5 Cantor on the Line and the Plane

Some of the circumstances surrounding the proof of Schröder-Bernstein tie in with the history we discussed in [section set.3](#). Recall that, in 1877, Cantor proved that there are exactly as many points on a square as on one of its sides. Here, we will present his (first attempted) proof.

[his:set:cantorplane:  
sec](#)

Let  $L$  be the unit line, i.e., the set of points  $[0, 1]$ . Let  $S$  be the unit square, i.e., the set of points  $L \times L$ . In these terms, Cantor proved that  $L \approx S$ . He wrote a note to Dedekind, essentially containing the following argument.

his:set:cantorplane:  
thm:cantorplane

**Theorem set.1.**  $L \approx S$

*Proof: first part..* Fix  $a, b \in L$ . Write them in binary notation, so that we have infinite sequences of 0s and 1s,  $a_1, a_2, \dots$ , and  $b_1, b_2, \dots$ , such that:

$$\begin{aligned} a &= 0.a_1a_2a_3a_4\dots \\ b &= 0.b_1b_2b_3b_4\dots \end{aligned}$$

Now consider the function  $f: S \rightarrow L$  given by

$$f(a, b) = 0.a_1b_1a_2b_2a_3b_3a_4b_4\dots$$

Now  $f$  is an **injection**, since if  $f(a, b) = f(c, d)$ , then  $a_n = c_n$  and  $b_n = d_n$  for all  $n \in \mathbb{N}$ , so that  $a = c$  and  $b = d$ .  $\square$

Unfortunately, as Dedekind pointed out to Cantor, this does not answer the original question. Consider  $0.\dot{1}0 = 0.10101010\dots$ . We need that  $f(a, b) = 0.\dot{1}0$ , where:

$$\begin{aligned} a &= 0.\dot{1}\dot{1} = 0.111111\dots \\ b &= 0 \end{aligned}$$

But  $a = 0.\dot{1}\dot{1} = 1$ . So, when we say “write  $a$  and  $b$  in binary notation”, we have to choose *which* notation to use; and, since  $f$  is to be a *function*, we can use only *one* of the two possible notations. But if, for example, we use the simple notation, and write  $a$  as “1.000...”, then we have no pair  $\langle a, b \rangle$  such that  $f(a, b) = 0.\dot{1}0$ .

To summarise: Dedekind pointed out that, given the possibility of certain recurring decimal expansions, Cantor’s function  $f$  is an **injection** but *not* a **surjection**. So Cantor has shown only that  $S \preceq L$  and *not* that  $S \approx L$ .

Cantor wrote back to Dedekind almost immediately, essentially suggesting that the proof could be completed as follows:

*Proof: completed..* So, we have shown that  $S \preceq L$ . But there is obviously an **injection** from  $L$  to  $S$ : just lay the line flat along one side of the square. So  $L \preceq S$  and  $S \preceq L$ . By Schröder–Bernstein (??),  $L \approx S$ .  $\square$

But of course, Cantor could not complete the last line in these terms, for the Schröder–Bernstein Theorem was not yet proved. Indeed, although Cantor would subsequently formulate this as a general conjecture, it was not satisfactorily proved until 1897. (And so, later in 1877, Cantor offered a different proof of **Theorem set.1**, which did not go via Schröder–Bernstein.)

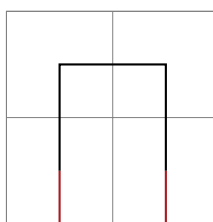
## set.6 Hilbert’s Space-filling Curves

his:set:hilbertcurve:  
sec

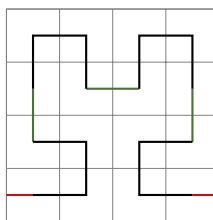
In chapter [section set.3](#), we mentioned that Cantor's proof that a line and a square have exactly the same number of points ([Theorem set.1](#)) prompted Peano to ask whether there might be a space-filling *curve*. He obtained a positive answer in 1890. In this section, we explain (in a hand-wavy way) how to construct Hilbert's space-filling curve (with a tiny tweak).<sup>2</sup>

We must define a function,  $h$ , as the limit of a sequence of functions  $h_1, h_2, h_3, \dots$ . We first describe the construction. Then we show it is space-filling. Then we show it is a curve.

We will take  $h$ 's range to be the unit square,  $S$ . Here is our first approximation to  $h$ , i.e.,  $h_1$ :



To keep track of things, we have imposed a  $2 \times 2$  grid on the square. We can think of the curve starting in the bottom left quarter, moving to the top left, then to the top right, then finally to the bottom right. Here is the second stage in the construction, i.e.,  $h_2$ :



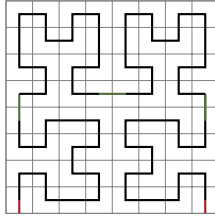
The different colours will help explain how  $h_2$  was constructed. We first place scaled-down copies of the non-red bit of  $h_1$  into the bottom left, top left, top right, and bottom right of our square (drawn in black). We then connect these four figures (with green lines). Finally, we connect our figure to the boundary of the square (with red lines).

Now to  $h_3$ . Just as  $h_2$  was made from four connected, scaled-down copies of the non-red bit of  $h_1$ , so  $h_3$  is made up of four scaled-down copies of the non-red bit of  $h_2$  (drawn in black), which are then joined together (with green lines) and finally connected to the boundary of the square (with red lines).

---

<sup>2</sup>For a more rigorous explanation, see [Rose \(2010\)](#). The tweak amounts to the inclusion of the red parts of the curves below. This makes it slightly easier to check that the curve is continuous.





And now we see the general pattern for defining  $h_{n+1}$  from  $h_n$ . At last we define the curve  $h$  itself by considering the point-by-point limit of these successive functions  $h_1, h_2, \dots$ . That is, for each  $x \in S$ :

$$h(x) = \lim_{n \rightarrow \infty} h_n(x)$$

We now show that this curve fills space. When we draw the curve  $h_n$ , we impose a  $2^n \times 2^n$  grid onto  $S$ . By Pythagoras's Theorem, the diagonal of each grid-location is of length:

$$\sqrt{(1/2^n)^2 + (1/2^n)^2} = 2^{(\frac{1}{2}-n)}$$

and evidently  $h_n$  passes through every grid-location. So each point in  $S$  is *at most*  $2^{(\frac{1}{2}-n)}$  distance away from some point on  $h_n$ . Now,  $h$  is defined as the limit of the functions  $h_1, h_2, h_3, \dots$ . So the maximum distance of any point from  $h$  is given by:

$$\lim_{n \rightarrow \infty} 2^{(\frac{1}{2}-n)} = 0.$$

That is: every point in  $S$  is 0 distance from  $h$ . In other words, every point of  $S$  lies *on* the curve. So  $h$  fills space!

It remains to show that  $h$  is, indeed, a *curve*. To show this, we must define the notion. The modern definition builds on one given by Jordan in 1887 (i.e., only a few years before the first space-filling curve was provided):

**Definition set.2.** A curve is a continuous map from  $L$  to  $\mathbb{R}^2$ .

This is fairly intuitive: a curve is, intuitively, a “smooth” map which takes a canonical line onto the plane  $\mathbb{R}^2$ . Our function,  $h$ , is indeed a map from  $L$  to  $\mathbb{R}^2$ . So, we just need to show that  $h$  is continuous. We defined continuity in [section set.2](#) using  $\varepsilon/\delta$  notation. In the vernacular, we want to establish the following: *If you specify a point  $p$  in  $S$ , together with any desired level of precision  $\varepsilon$ , we can find an open section of  $L$  such that, given any  $x$  in that open section,  $h(x)$  is within  $\varepsilon$  of  $p$ .*

So: assume that you have specified  $p$  and  $\varepsilon$ . This is, in effect, to draw a circle with centre  $p$  and radius  $\varepsilon$  on  $S$ . (The circle might spill off the edge of  $S$ , but that doesn't matter.) Now, recall that, when describing the function  $h_n$ , we drew a  $2^n \times 2^n$  grid upon  $S$ . It is obvious that, no matter how small  $\varepsilon$  is, there is some  $n$  such that some individual grid-location of the  $2^n \times 2^n$  grid on  $S$  lies wholly within the circle with centre  $p$  and radius  $\varepsilon$ .

So, take that  $n$ , and let  $I$  be the largest open part of  $L$  which  $h_n$  maps wholly into the relevant grid location. (It is clear that  $(a, b)$  exists, since we already noted that  $h_n$  passes through every grid-location in the  $2^n \times 2^n$  grid.) It now suffices to show to show that, whenever  $x \in I$  the point  $h(x)$  lies in that same grid-location. And to do *this*, it suffices to show that  $h_m(x)$  lies in that same grid location, for any  $m > n$ . But this is obvious. If we consider what happens with  $h_m$  for  $m > n$ , we see that exactly the “same part” of the unit interval is mapped into the same grid-location; we just map it into that region in an increasingly stretched-out, wiggly fashion.

## Photo Credits

# Bibliography

- Berkeley, George. 1734. *The Analyst; or, a Discourse Adressed to an Infidel Mathematician*.
- Giaquinto, Marcus. 2007. *Visual Thinking in Mathematics*. Oxford: Oxford University Press.
- Gouvêa, Fernando Q. 2011. Was Cantor surprised? *American Mathematical Monthly* 118(3): 198–209.
- Hilbert, David. 1891. Über die stetige Abbildung einer Linie auf ein Flächenstück. *Mathematische Annalen* 38(3): 459–460.
- Peano, Giuseppe. 1890. Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen* 36(1): 157–60.
- Rose, Nicholas J. 2010. Hilbert-type space-filling curves. URL <https://web.archive.org/web/20151010184939/http://www4.ncsu.edu/~njrose/pdfFiles/HilbertCurve.pdf>.