Chapter 1: The Reals

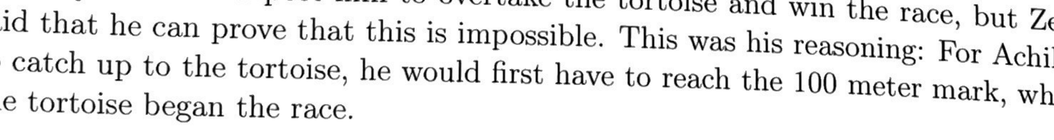
Real analysis was developed in the 17th century as a tool to answer some truly fundamental questions from physics, and is now applied throughout the sciences. We begin, however, long before this point, in Greece.

1.1 Zeno's Paradoxes

Nearly 2,500 years ago, Zeno of Elea wrote down a set of paradoxes to support the philosophy that *change* is an illusion. Specifically, one type of change is *motion*—the change of an object's position. Zeno set out to show that objects do not actually move, and our senses are lying to us when we perceive motion.[1]

Zeno's First Paradox

How did he try to prove that motion is an illusion? He created a thought experiment. He said to imagine a race between Achilles (the Greek warrior of legend) and a tortoise (just a regular tortoise). Suppose the race is 1,000 meters long and the tortoise is given a 100 meter head start.



The gun is fired and both begin. Achilles is much faster than the tortoise, so of course you would expect him to overtake the tortoise and win the race, but Zeno said that he can prove that this is impossible. This was his reasoning: For Achilles to catch up to the tortoise, he would first have to reach the 100 meter mark, where the tortoise began the race.

[1]. . . And to answer your first question, yes, Greek philosophers were believed to have experimented with psychedelic drugs.

happen. And it takes careful, detailed thought to avoid getting stuck in paradoxical situations. Indeed, this is essentially our goal for the book.

Textbook Goal. By studying the infinite, develop a ground-up understanding of the real numbers and functions on the reals. Also, improve one's mathematical maturity; that is, understand mathematical statements and arguments, construct proofs and find counterexamples, and appreciate the intrinsic beauty in the mathematics.

So, that's what we are going to do. We are going to think carefully about the infinite, and as we progress through this book we will see how that careful reasoning can illustrate some pretty amazing properties of real numbers and functions. And unlike calculus where you spent most of your time studying super nice functions and intuitive situations, and very little time on weird paradoxes like the above, in analysis we will instead spend a significant amount of time on the weird situations, where our intuition may deceive us. We will build analysis from the ground-up, retraining our intuition on solid — and increasingly higher — ground.

So get ready, because I think you will enjoy it.

1.2 Basic Set Theory Definitions

We begin about as basic as possible — with sets. Here is a quick review of basic set theory definitions that you learned in your intro-to-proofs class.

Definition.

Definition 1.1.

- A *set* is an unordered collection of distinct objects, which are called *elements.*[2]
- If x is an element of a set S, we write $x \in S$. This is read as "x in S."
- *Set-builder notation* looks like this:

$$S = \{\text{elements} \ : \ \text{conditions used to generate the elements}\}.$$

For example,

$$\{1, 4, 9, 16, 25, \dots\} = \{n^2 : n \in \mathbb{N}\}.$$

As a second example, the rationals can be built like so:

$$\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0 \right\}.$$

[2] Alternative definition: Everything. Everything is a set. Almost no definition in the world is as general as that of a set.

1.5 The Completeness Axiom

What's in a set? That which we call the reals,
By any other name would be as complete.

– Shakespeare in Dimension C-314[21]

We need just one final axiom to obtain the reals, and it is called *the completeness axiom*. Formally, it is an axiom about bounded sets, which requires a little motivation to understand. We begin there.

One of the first problems with the rationals that we mentioned was that if q is rational it need not be the case that $\sqrt{q}$ is rational—we worked out $q = 2$ in detail in Example 1.4. Said differently, if q is rational, there may not be a rational number x such that $x^2 = q$.

We want to include $\sqrt{2}$ in what will be the real numbers, but we also want all of the other irrational numbers. We could say "include all the numbers $\sqrt{q}$ where q is rational," but most irrational numbers are not of this form... We want an approach that gets all of them, identifies their ordering, etc. How do we use the rationals to identify *all* of the irrationals? There is a slick way to do this. Turning back to $\sqrt{2}$, consider the set

$$A := \{x \in \mathbb{Q} : x^2 < 2\}.$$

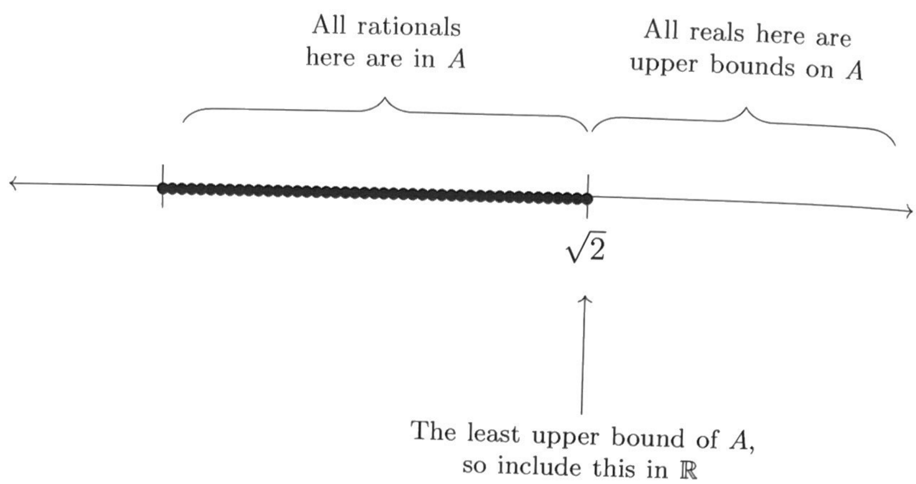
This set looks something like this:

Here are the important ideas:

- This set is *bounded above*. For example, 2 and 6 and 1.5 are all bigger than everything in A, so these are all *upper bounds* on A.
- Among all these upper bounds, there is one that is special—the smallest of them all. This is the *least upper bound*, and as you might imagine, this bound will in fact be exactly $\sqrt{2}$ (which by Example 1.4 is not rational).
- In order to get all of $\mathbb{R}$, we will start with $\mathbb{Q}$ and then we will add in all the least upper bounds from all sets that are bounded above, like A. E.g., due to

[21] The Shakespeare in the parallel universe where all his plays were about math. And Jerry's happy.

the above set A, we include $\sqrt{2}$ into our set $\mathbb{R}$. Doing this for all such A makes $\mathbb{R}$ complete.[22]



That is the blueprint for what we are going to do. Now here it is, formally.

Definition.

Definition 1.17. Let S be an ordered field (like $\mathbb{R}$) and $A \subseteq S$ be nonempty.

(i) The set A is *bounded above* if there exists some $b \in S$ such that $x \leq b$ for all $x \in A$; in this case, b is called an *upper bound* of A.

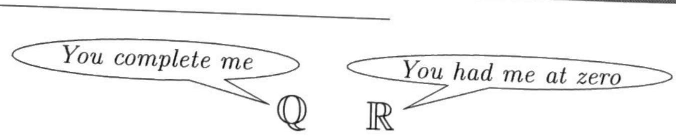
(ii) The *least upper bound* of A—if it exists—is some $b_0 \in S$ such that (1) b_0 is an upper bound of A, and (2) if b is any other upper bound of A, then $b_0 \leq b$. Such a b_0 is also called the *supremum* of A and is denoted $\sup(A)$.

(iii) Likewise, the set A is *bounded below* if there exists some $b \in S$ such that $x \geq b$ for all $x \in A$; in this case, b is called a *lower bound* of A.

(iv) Again, like above, the *greatest lower bound* of A—if it exists—is some $b_0 \in S$ such that (1) b_0 is a lower bound of A, and (2) if b is any other lower bound of A, then $b_0 \geq b$. Such a b_0 is also called the *infimum* of A and is denoted $\inf(A)$.

(v) If a set is both bounded above and bounded below, then it is simply called *bounded.*

---

22 

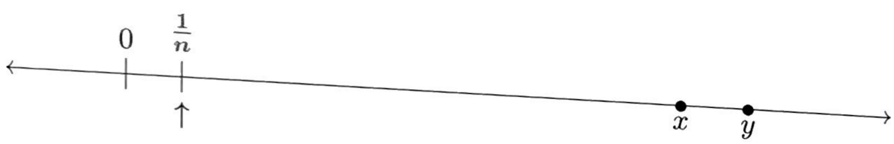
The next theorem says that between any two real numbers is a rational number. The rationals — which are super nice to describe and work with — are everywhere.[30]

Theorem.

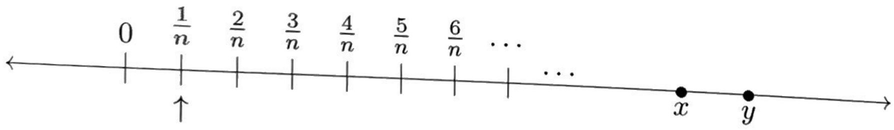
Theorem 1.31 ($\mathbb{Q}$ *is dense in* $\mathbb{R}$). The rational numbers are dense in the real numbers.

Proof Idea. The result follows from Lemma 1.30, and does so easily enough that it would be easy to miss what is what really going on. Assume that $x, y \in \mathbb{R}$ and assume both are positive with $x < y$. We aim to find a rational number $\frac{m}{n}$ between x and y. That is, with $x < \frac{m}{n} < y$.

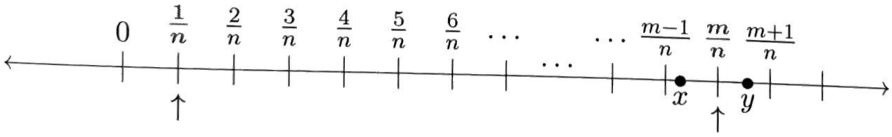
Now, $y - x$ is some positive number, and so by Archimedean principle there is some $n \in \mathbb{N}$ with $\frac{1}{n} < y - x$.



Now think about all integer multiples of this $\frac{1}{n}$.



Note that since each one is $\frac{1}{n}$ away from the next one, but $y - x > \frac{1}{n}$, it is impossible for these dashes to completely hop over the interval between x and y. That is, at least one of these must fall between x and y.



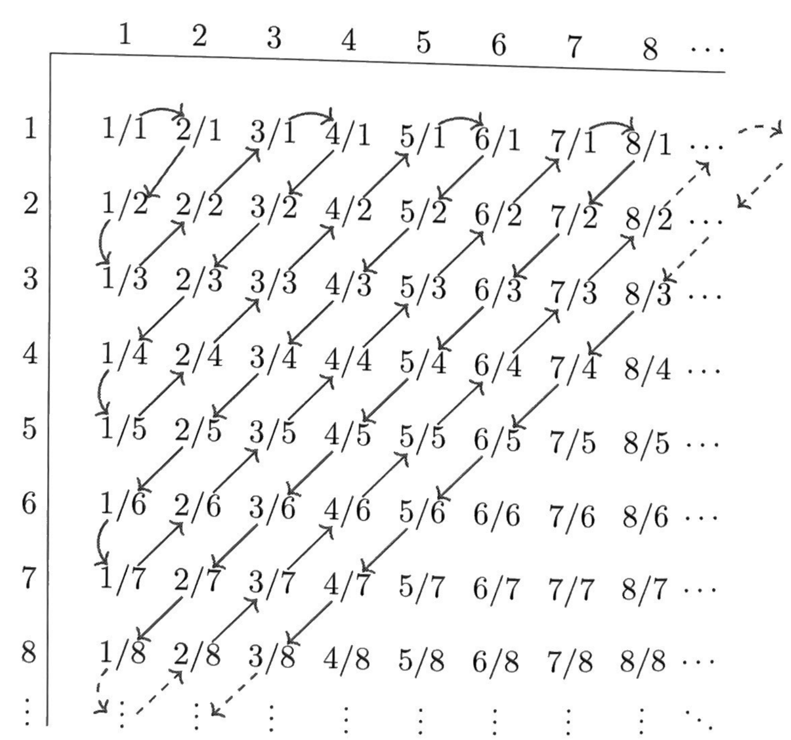
This is the idea, although by using Lemma 1.30 the proof hides some of these details. Now here's the proof of Theorem 1.31.

[30] In Chapter 2 when we study the cardinalities of these sets, this result will be even more impressive and important.

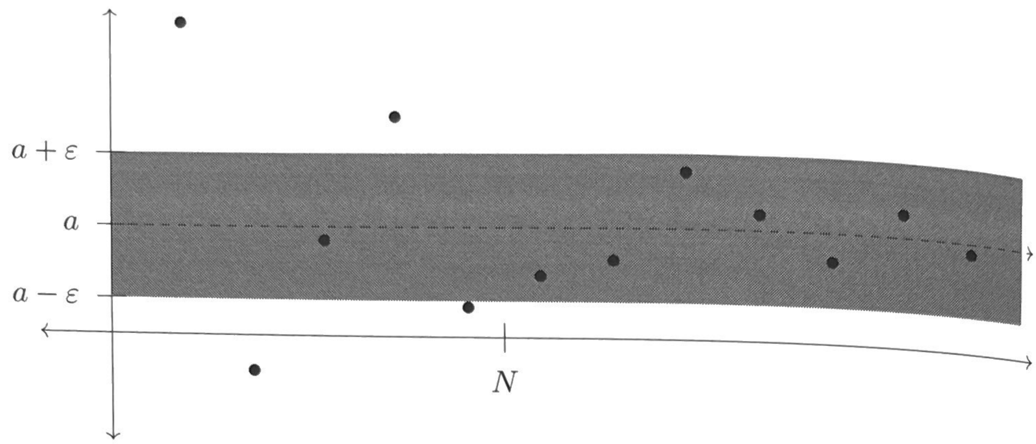
Consider the rational numbers (with some duplication) written in this way:

	1	2	3	4	5	6	7	8	⋯
1	1/1	2/1	3/1	4/1	5/1	6/1	7/1	8/1	⋯
2	1/2	2/2	3/2	4/2	5/2	6/2	7/2	8/2	⋯
3	1/3	2/3	3/3	4/3	5/3	6/3	7/3	8/3	⋯
4	1/4	2/4	3/4	4/4	5/4	6/4	7/4	8/4	⋯
5	1/5	2/5	3/5	4/5	5/5	6/5	7/5	8/5	⋯
6	1/6	2/6	3/6	4/6	5/6	6/6	7/6	8/6	⋯
7	1/7	2/7	3/7	4/7	5/7	6/7	7/7	8/7	⋯
8	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

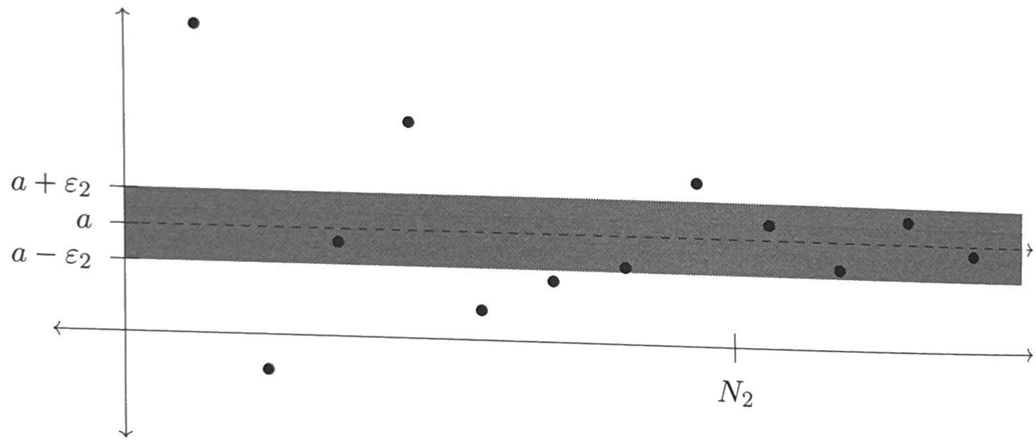
Our bijection, which we call *the winding bijection*, can be pictured like this:



Comment 2. Below is the important picture to keep in your head. Note that after the N, all the points are inside the shaded region between $a - \varepsilon$ and $a + \varepsilon$.



Now, if you picked a *smaller* value of ε, which we will call ε_2, you will need a *bigger* value of N (called N_2) to guarantee that all the points after N_2 are inside the shaded region.



So you see, smaller values of ε typically require larger values of N.

Comment 3. You can think about the converging sequence as the tragic story of a rebellious youth. When young, he may be all over the place on any given day, living his life and getting into mischief. But eventually society settles him down. When $\varepsilon = 100$ (miles) he has a car, and with his gang of misfits he can roam wherever he pleases, causing mayhem. But when $\varepsilon = 10$ (miles), the value of N_1 represents the age at which he is placed on at-risk watch and is not allowed to leave the city (maybe $N_1 = 17$ years old). Then $\varepsilon = 0.1$ (miles) represents the age N_2 at which he really gets into trouble and is placed under house arrest. And $\varepsilon = 0.01$ represents the age N_3 at which he is sentenced to life in prison and will forever more be contained within a small wing of prison.[5]

[5] When I started this paragraph I did not anticipate it getting so dark. I even dropped the solitary confinement stage (or beyond) for that reason. Poor guy. Stay in school, kids.

This next one looks a bit trickier, but the same procedure works.

Example 3.12. Let $a_n = \dfrac{3n+1}{n+2}$. Prove that $\lim_{n\to\infty} a_n = 3$.

Scratch Work. Again, we first play around. We start with where we want to get to (that $|a_n - a| < \varepsilon$), and then do some algebra to figure out which values of n would give this.

We want the following:

$$\begin{aligned}
|a_n - a| &< \varepsilon \\
\left|\frac{3n+1}{n+2} - 3\right| &< \varepsilon \\
\left|\frac{3n+1}{n+2} - \frac{3(n+2)}{n+2}\right| &< \varepsilon \\
\left|\frac{3n+1-3n-6}{n+2}\right| &< \varepsilon \\
\left|\frac{-5}{n+2}\right| &< \varepsilon \\
\frac{5}{n+2} &< \varepsilon \\
\frac{5}{\varepsilon} &< n+2 \\
\frac{5}{\varepsilon} - 2 &< n
\end{aligned}$$

So as long as we choose $N = \dfrac{5}{\varepsilon} - 2$, then for any $n > N$ we will have $n > \dfrac{5}{\varepsilon} - 2$, which by the above will imply that $\frac{3n-1}{n+2} \to 3$, as desired.

Solution. Fix any $\varepsilon > 0$. Set $N = \frac{5}{\varepsilon} - 2$. Then for any $n > N$,

$$\begin{aligned}
|a_n - a| &= \left|\frac{3n+1}{n+2} - 3\right| = \left|\frac{3n+1}{n+2} - \frac{3n+6}{n+2}\right| \\
&= \frac{5}{n+2} < \frac{5}{N+2} = \frac{5}{(\frac{5}{\varepsilon} - 2) + 2} \\
&= \frac{5}{5/\varepsilon} = \varepsilon.
\end{aligned}$$

That is, $|a_n - a| < \varepsilon$. So by Definition 3.7 we have shown that $\dfrac{3n+1}{n+2} \to 3$. $\square$

once we have proven 3, then 4 is similar. Those two are Exercise 3.13. So we will prove 3 next and leave the rest as exercises.

Scratch Work. Here's the scratch work for Law 3. We want to find an N such that $n > N$ implies $|a_n b_n - ab| < \varepsilon$. Our two big assumptions are that $a_n \to a$ and $b_n \to b$. Here's what we get from that:

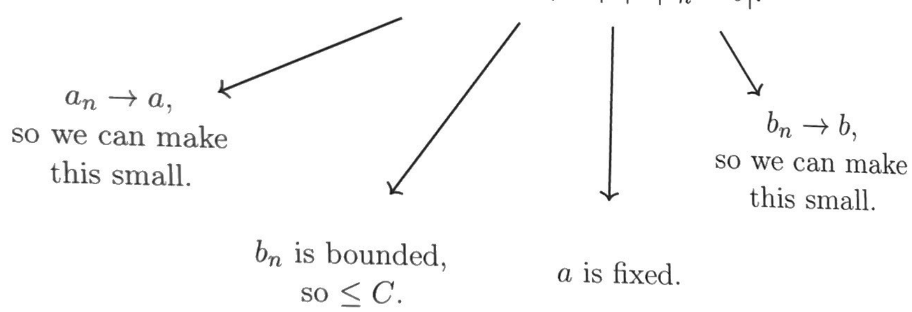
- Since $a_n \to a$, for any $\varepsilon_1 > 0$ there exists some N_1 such that $n > N_1$ implies that $|a_n - a| < \varepsilon_1$.
- Since $b_n \to b$, for any $\varepsilon_2 > 0$ there exists some N_2 such that $n > N_2$ implies that $|b_n - b| < \varepsilon_2$.

Now, going back to what we want to show, there is a clever trick that makes it all work. The idea is to first rewrite

$$|a_n b_n - ab| \quad \text{as} \quad |a_n b_n - ab_n + ab_n - ab|.$$

The middle terms cancel out, so those are the same. This is clever, though, because if you now use the triangle inequality, you get something we have control over:

$$\begin{aligned} |a_n b_n - ab_n + ab_n - ab| &\le |a_n b_n - ab_n| + |ab_n - ab| \\ &= |a_n - a| \cdot |b_n| + |a| \cdot |b_n - b|. \end{aligned}$$

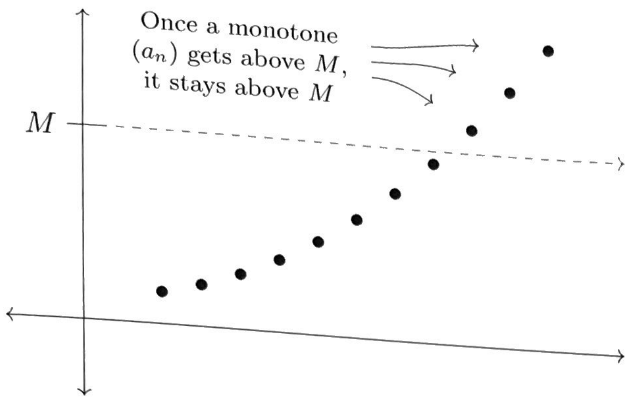


To get this all less than ε, we will make both of the products less than $\frac{\varepsilon}{2}$. To get $|a_n - a| \cdot |b_n|$ less than $\frac{\varepsilon}{2}$ we will have to ensure that $|a_n - a|$ is less than $\frac{\varepsilon}{2C}$. To get $|a| \cdot |b_n - b|$ less than $\frac{\varepsilon}{2}$ we will have to ensure $|b_n - b|$ is less than $\frac{\varepsilon}{2a}$. (And then be careful so we aren't dividing by 0.)

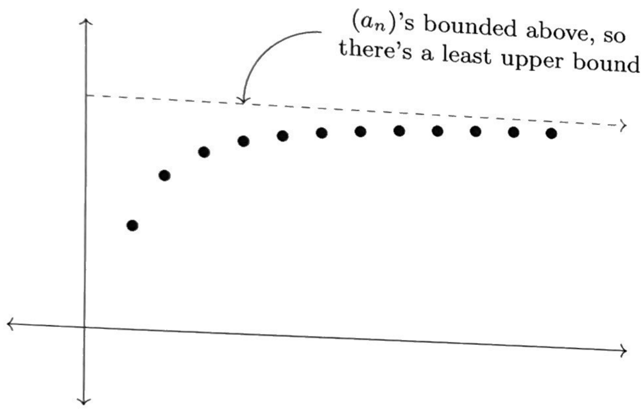
Proof. Let $\varepsilon > 0$. Since (b_n) converges, by Proposition 3.20 we know that (b_n) is bounded; that is, there exists some $C \in \mathbb{R}$ such that $|b_n| \le C$ for all n. Let $\varepsilon_1 = \frac{\varepsilon}{2C+1}$ and $\varepsilon_2 = \frac{\varepsilon}{2|a|+1}$. (Note: we need each "+1" to ensure we are not dividing by 0.)

Since $\varepsilon_1 > 0$ there exists some N_1 such that $|a_n - a| < \varepsilon_1$ for all $n > N_1$. And since $\varepsilon_2 > 0$ there exists some N_2 such that $|b_n - b| < \varepsilon_2$ for all $n > N_2$. Let

Proof Idea. Suppose (a_n) is monotonically increasing. If it's not bounded, then given any $M > 0$, this M is not an upper bound and so eventually (a_n) will get above it. And since it's monotonically increasing, once it gets above a number, it *stays* above that number.



If, on the other hand, (a_n) is bounded, then the sequence must be leveling out (it's monotone, so it can't go up and down).



Proof. Assume that (a_n) is monotonically increasing, and let's suppose first that a_n is not bounded. Then for any $M > 0$ there exists some N such that $a_N > M$. But since (a_n) is monotonically increasing, for $n > N$ we have $a_n \geq a_N > M$. And so, by Definition 3.15, (a_n) diverges to ∞.

Next suppose that (a_n) is bounded. Then we have that $\{a_n : n \in \mathbb{N}\}$ is a subset of $\mathbb{R}$ which is bounded above, which by the completeness of $\mathbb{R}$ implies that $\sup(\{a_n : n \in \mathbb{N}\})$ exists.[17] Call this supremum α. We want to show that $\lim_{n\to\infty} a_n = \alpha$; that is, we want to show that for any $\varepsilon > 0$ there exists some N such that $n > N$ implies $|a_n - \alpha| < \varepsilon$.

[17] Remember that so far only sets have been defined to have suprema. Sequences don't. So we have to turn the sequence $(a_1, a_2, a_3, \ldots)$ into the set $\{a_n : n \in \mathbb{N}\}$ in order to talk about their supremum.

that $n_j \geq j$.[28] In particular, $n_{J+1} > J$. And so, for any $j > J$,

$$\begin{aligned} |a_j - a| &= |a_j - a_{n_{J+1}} + a_{n_{J+1}} - a| \\ &\leq |a_j - a_{n_{J+1}}| + |a_{n_{J+1}} - a| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

□

Back to Axioms

In this text we assumed as an axiom that the reals were complete (AoC).[29] Using AoC we proved, in this order, the monotone convergence theorem (MCT) in Theorem 3.27, the Bolzano-Weierstrass theorem (B-W) in Theorem 3.37, and then the Cauchy criterion (CC) in Theorem 3.42. That is,

$$\text{AoC} \Rightarrow \text{MCT} \Rightarrow \text{B-W} \Rightarrow \text{CC}.$$

- In fact, AoC and MCT are equivalent! If we had started by supposing as an axiom that the reals were an ordered field containing $\mathbb{Q}$ which satisfies MCT, then we could have proved AoC and everything else and ended up right where we are now.
- Also, AoC and B-W are equivalent! So by supposing B-W instead of AoC, we again could have developed our entire theory.
- CC is almost enough, but not quite. Surprisingly, with CC you can not prove the Archimedean principle (AP). But if you assume both CC and AP, then you can prove everything.[30]

That is,

$$\text{AoC} \Leftrightarrow \text{MCT} \Leftrightarrow \text{B-W} \Leftrightarrow (\text{CC} + \text{AP}).$$

Neat stuff.

---

[28] For example, one subsequence is $(a_{n_j}) = (a_1, a_3, a_5, a_7, \dots)$, giving $n_1 = 1$, $n_2 = 3$, $n_3 = 5$, $n_4 = 7$, and so on. And so, $n_1 \geq 1$, and $n_2 \geq 2$, and $n_3 \geq 3$, and $n_4 \geq 4$, and so on.

[29] As someone whose mathematical interests are nearly matched by my interest in politics, the post-2018 dueling use of "AOC" has taken a toll on my casual readings.

[30] Suppose you tried to prove that an ordered field with CC (but not AP) has AoC. To prove AoC, you must prove that given a bounded set A, its least upper bound — $\sup(A)$ — exists. A reasonable approach that uses CC would be to try to approach it from above. Start with any upper bound b of A and then pick any $a \in A$. Construct a sequence which approaches where we know $\sup(A)$ to be by setting $x_1 = b$, and for any $n > 1$, letting $x_{n+1} = \begin{cases} x_n - \frac{b-a}{2^n} & \text{if this is an upper bound of } A; \\ x_n & \text{otherwise.} \end{cases}$
Try to convince yourself that this sequence should indeed converge to $\sup(A)$. But how would you prove it? Let $\varepsilon > 0$. We must show there exists a point in this sequence that gets within ε of $\sup(A)$. And showing that we can always get closer comes down to showing that there exists an n such that $\frac{1}{2^n} < \frac{\varepsilon}{b-a}$. And there it is — at this step we require the Archimedean principle.

4.4 Rearrangements

Definition.

Definition 4.21. A *rearrangement* of a series $\sum_{k=1}^{\infty} a_k$ is a series $\sum_{k=1}^{\infty} b_k$ for which there is a bijection $f : \mathbb{N} \to \mathbb{N}$ such that $b_{f(k)} = a_k$.

It is precise to define a rearrangement in terms of bijections (especially if the series has repeated terms) but, intuitively, a rearrangement of a series is simply the same series where you add up the terms in a different order.

Example 4.22. The alternating harmonic series is

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \dots.$$

A rearrangement of this series is

$$\frac{1}{5} - \frac{1}{2} - \frac{1}{152} + \frac{1}{57} - \frac{1}{1847604} - \frac{1}{8} + \dots.$$

A different rearrangement is

$$1 + \frac{1}{3} + \frac{1}{5} - \frac{1}{2} + \frac{1}{7} + \frac{1}{9} + \frac{1}{11} - \frac{1}{4} + \dots.$$

If you have *finitely* many numbers that you are adding/subtracting, then of course the order in which you add/subtract them does not matter — you get the same answer either way. Amazingly, though, if you have *infinitely* many numbers, then it *can* make a difference. It is possible to add up infinitely many numbers and have it equal one thing, but then by simply changing the order that you are adding up the numbers it is possible for the sum to equal something different! The order that you add the numbers can change what it equals! But if that's not freaky enough, it gets worse/better: In some cases, the sum can equal *anything at all.* This is one of my favorite theorems of this entire book, so pay attention.

Super Cool Theorem.

Theorem 4.23 (*Rearrangement theorem*). If a series $\sum_{k=1}^{\infty} a_k$ converges conditionally, then for any L ($L \in \mathbb{R}$ or $L = \pm\infty$) there exists some rearrangement of $\sum_{k=1}^{\infty} a_k$ which converges to L.

— Notable Exercises —

The exercises contain two more convergence tests that you probably learned in calculus — the ratio and the root tests. These tests are not as important as the others to this text's development of real analysis, but are nevertheless important results with particular utility in determining the convergence of specific series.

- Exercise 4.11 (c) is an important result from number theory. In fact, the converse is essentially true too: If a number is rational, then either it has a terminating decimal expansion or a repeating decimal expansion.
- In Exercise 4.17 we prove the *ratio test*. That is, given a series $\sum_{k=1}^{\infty} a_k$ with $a_k \neq 0$, we will show that if $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right|$ is less than 1 then the series converges.
- In Exercise 4.18 we prove the *root test*. That is, if $\sum_{k=1}^{\infty} a_k$ is a series where each $a_k \geq 0$ and the limit $\lim_{k\to\infty} (a_k)^{1/k}$ exists, we call this limit ρ. Then this series converges if $\rho < 1$ and diverges if $\rho > 1$. (The test is inconclusive if $\rho = 1$.)

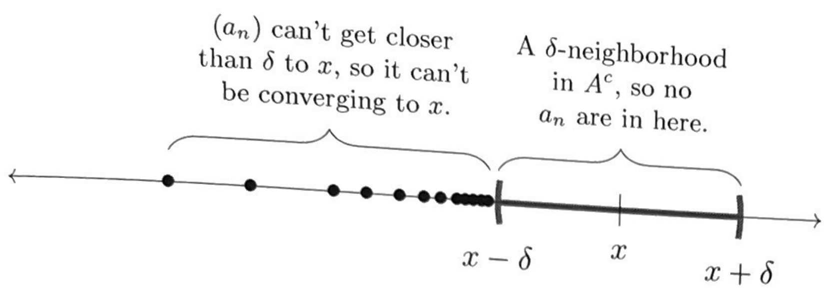
The rearrangement theorem (Theorem 4.23) stated that any conditionally convergent series can be rearranged to converge to anything at all — to any $L \in \mathbb{R}$ or to either ∞ or $-\infty$. After stating the theorem we sketched a proof for the case when $L \in \mathbb{R}$ and $L > 0$.

- In Exercise 4.23 we prove that there exists a rearrangement of this sum which diverges to ∞.

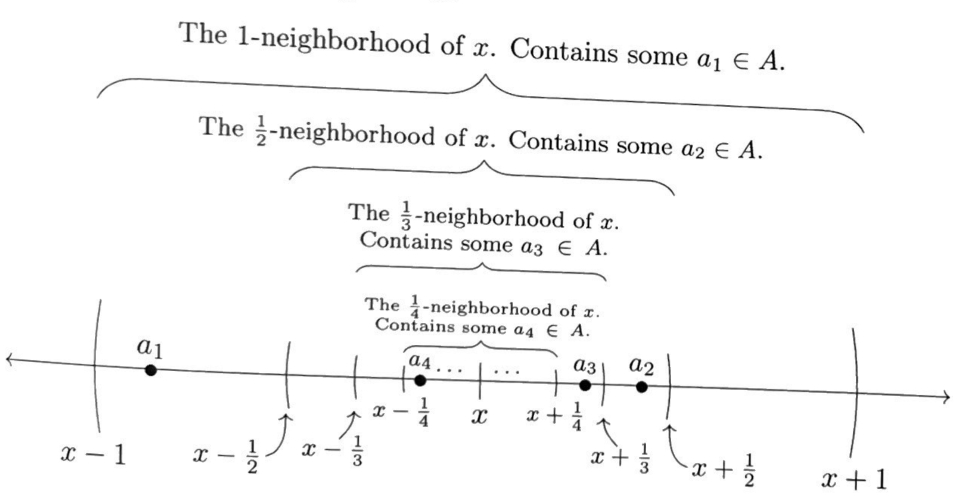
This roughly completes the proof of Theorem 4.23, and in the below we extend this result.

- In Exercise 4.24 we show that there exists a rearrangement of this sum whose limit does not exist.
- In Exercise 4.25 we prove a generalization of Theorem 4.23 called *Riemann's series theorem* in which in a rearrangement can be found whose *limit supremum* and *limit infimum* can be arbitrarily chosen.
- Notice that Theorem 4.23 says nothing about what happens for series which are not conditionally convergent. In Exercise 4.26 we prove that if a series converges absolutely, then every rearrangement converges to the same thing. And in Exercise 4.27 we prove that if each $a_k \geq 0$ and $\sum_{k=1}^{\infty} a_k = \infty$, then any rearrangement of this sum also diverges to ∞.

For the forward direction, when we assume for a contradiction that A does not contain all its limit points, this gives us an $x \notin A$ for which there is a sequence (a_n) converging to x, where each $a_n \in A$. Where's the contradiction? Well, converging to x means these points a_n from A are getting closer and closer to x. But since $x \notin A$, that means $x \in A^c$; and since A is closed, A^c is open. And being open means there is a δ-neighborhood of x which is entirely contained inside A^c; i.e., is disjoint from A. That's the contradiction: It's impossible for a sequence (a_n) be getting closer and closer to x, if all these points have to be from A, and hence have to stay outside this δ-neighborhood of x.



As for the reverse direction, by using the contrapositive we may assume that A is not closed, which is to say that A^c is not open. And being "not open" means that for some $x \in A^c$, every δ-neighborhood of x contains a member of A. So if we let $\delta = 1$, then this 1-neighborhood of x contains a member of A, call it a_1. Likewise, if we let δ be $1/2$, then $1/3$, then $1/4$, and so on, we get an a_2, a_3, a_4, and so on, each from A and contained inside these respective neighborhoods. These elements comprise a sequence from A, which converges to x.



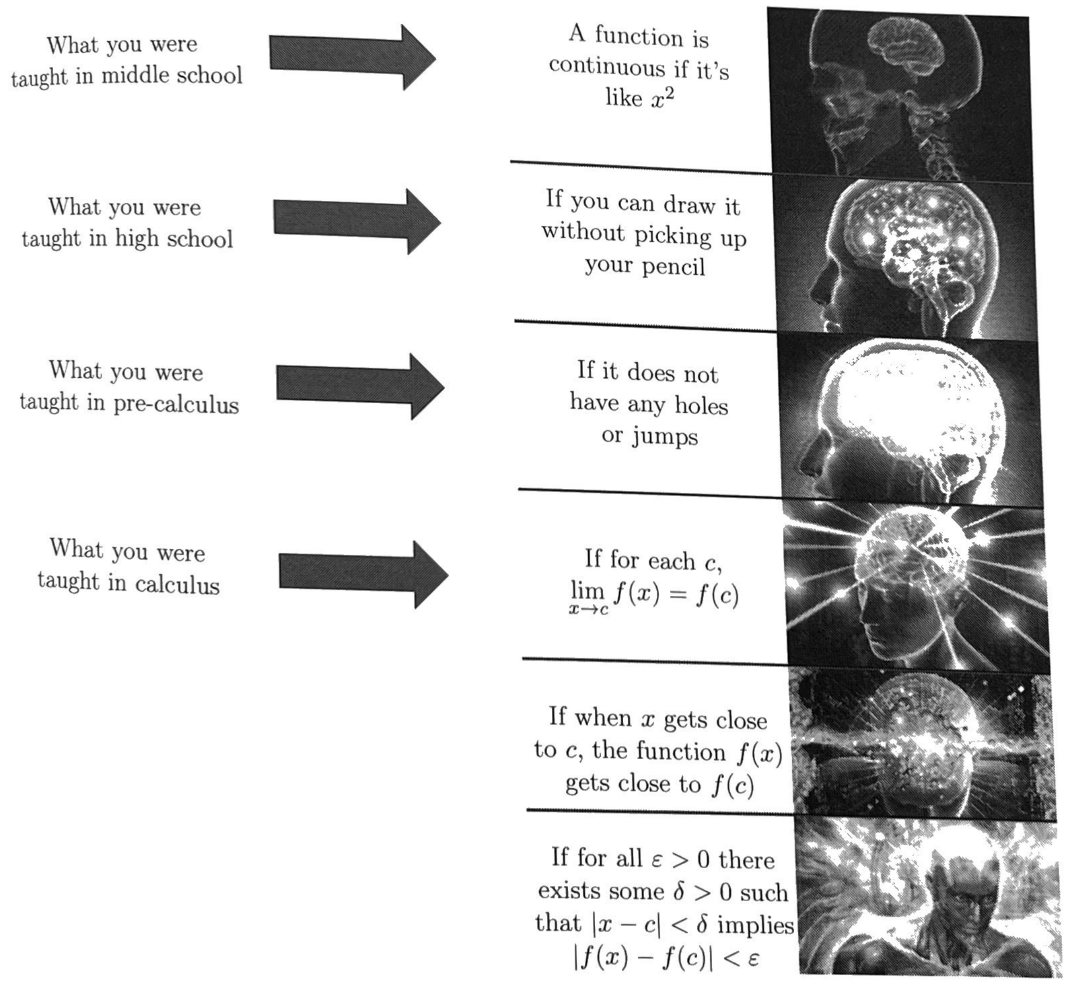
Proof. For the forward direction, suppose that A is closed, and assume for a contradiction that A does not contain all of its limit points. That is, suppose that there is a sequence $a_1, a_2, a_3, \dots$ from A where $a_n \to x$, but $x \notin A$. Observe that this implies that A^c is open and $x \in A^c$. And so, by openness, there is some δ-neighborhood $(x-\delta, x+\delta) \subseteq A^c$. We now show that this is a contradiction to the claim that $a_n \to x$.

Since $a_n \to x$ where each $a_n \in A$, for all $\varepsilon > 0$ there exists some N such that $|a_n - x| < \varepsilon$ for all $n > N$. In particular, if we let $\varepsilon = \delta$, then such an N exists, which

Chapter 6: Continuity

6.1 Approaching Continuity

I made a meme describing your discrete approach towards understanding continuity.[1]



[1]Here's to hoping that memes have a longer half-life than I fear!

Example 6.7. Define *Thomae's function*[4] to be

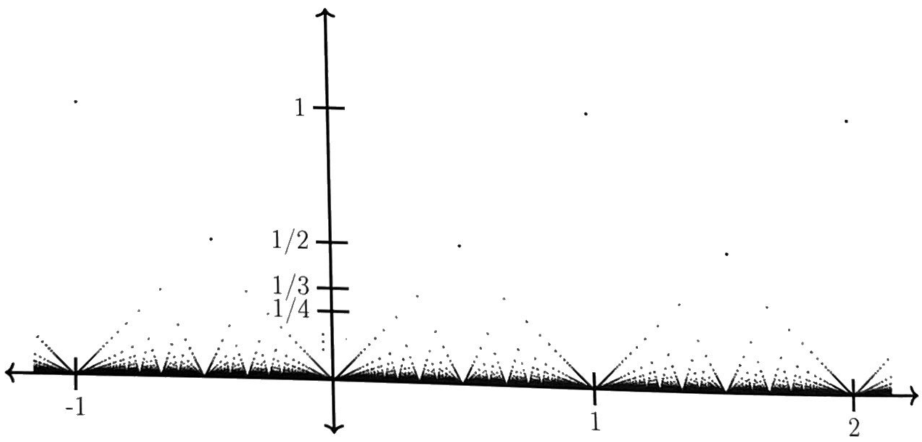
$$h(x) = \begin{cases} 1 & \text{if } x = 0 \\ 1/n & \text{if } x \in \mathbb{Q} \text{ and } x = m/n \text{ in lowest terms with } n > 0 \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Notice that at any rational point c, we have $h(c) > 0$; meanwhile, there are irrational numbers x arbitrarily close to c which have $h(x) = 0$. So h will indeed be discontinuous at every rational number.

At any irrational number c, though, we can show that h is actually continuous. We will prove this in detail later, but the basic idea is this: Let $\varepsilon > 0$, and find an N for which $\frac{1}{N} < \varepsilon$. We can show that, in some region around c, all x-values have $h(x) < \frac{1}{N} < \varepsilon$. To see this, note that there are only finitely many rational numbers which have a denominator of N or smaller and are also within, say, distance 1 of c. Find the one that is closest to c, and suppose it has distance $\delta > 0$ from c. Then, all rational numbers in $(c - \delta, c + \delta)$ have denominators that are larger than N. So within this range, all function values $h(x)$ are smaller than $1/N$. Awesome!

And, as you can imagine, as N gets large this forces all the function values to be converging to 0 in these shrinking regions around c, kind of like what happened at $c = 0$ in the modified dirichlet function. So yes, this will be an example of a function which is discontinuous on $\mathbb{Q}$, and continuous on $\mathbb{R} \setminus \mathbb{Q}$![5] □

This function's graph looks like this:[6]



So yes, with these functions in mind, we certainly are in need of a careful definition of continuity. But first, as we have discussed, we need a careful definition of a functional limit.

[4]AKA the *popcorn function*, AKA the *raindrop function*, AKA the *countable cloud function*, AKA the *ruler function*, AKA the *Riemann function*, AKA the *Stars over Babylon*.

[5]Make sure you take a moment to appreciate how remarkably, wonderfully weird this is.

[6]So cool that it made the cover.

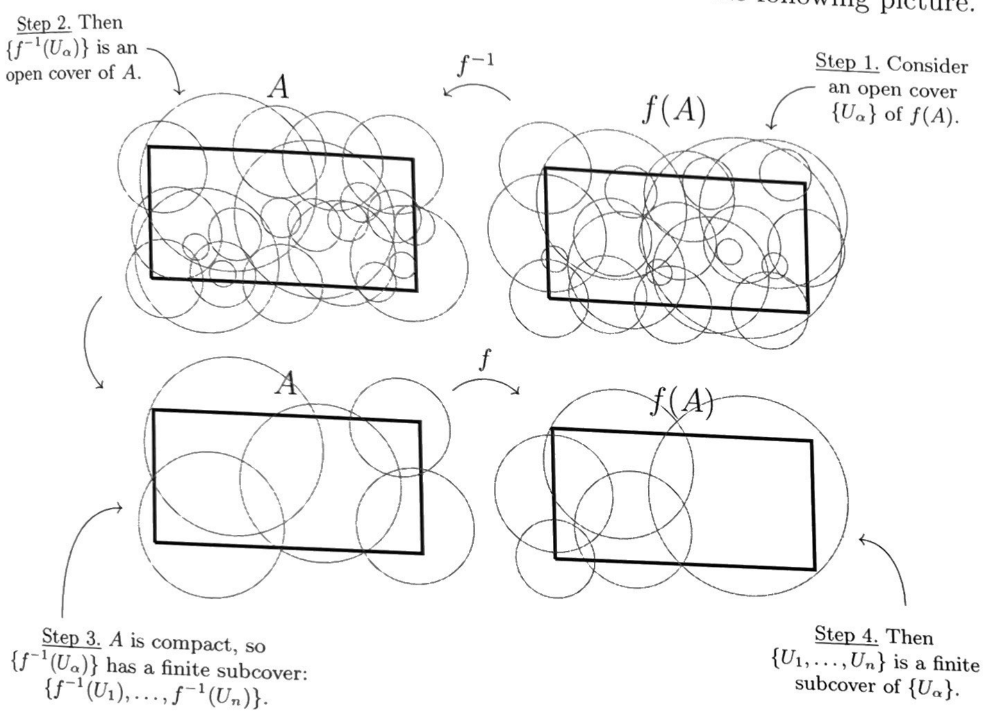
6.7 The Extreme Value Theorem

This next section begins with one more connection between continuity and one of the fundamental objects of topology: compact sets.

Theorem.

Theorem 6.30 (*The continuous image of a compact set is compact*)**.** Suppose $f: X \to \mathbb{R}$ is continuous. If $A \subseteq X$ is compact, then $f(A)$ is compact.

Proof Sketch. The idea behind the proof is contained in the following picture.



The only additional justification that is needed is in Step 2. Theorem 6.29 does not technically guarantee that $f^{-1}(U_\alpha)$ is open, but that $f^{-1}(U_\alpha) = V_\alpha \cap X$ for some open set V_α. So technically our open cover of A will be $\{V_\alpha\}$—but the above sketch will still essentially work.

Proof. Suppose A is compact. We aim to show that $f(A)$ is also compact, which we will show via the definition. To that end, let $\{U_\alpha\}$ be an open cover of $f(A)$. We will find a finite subcover of this cover, which will complete the proof.

By Theorem 6.29, each $f^{-1}(U_\alpha) = V_\alpha \cap X$ for some open set V_α. In particular, note that $f^{-1}(U_\alpha) \subseteq V_\alpha$.

We will show that $\{V_\alpha\}$ is an open cover of A. By the above, each V_α is open. To see that $\{V_\alpha\}$ is a cover of A, we will instead show that $\{f^{-1}(U_\alpha)\}$ is a cover. This is sufficient since we already noted that each $f^{-1}(U_\alpha) \subseteq V_\alpha$. Consider any $x_0 \in A$;

derivative, moving a couple things around, and applying a limit law.

$$\lim_{x\to c}\frac{(kf)(x)-(kf)(c)}{x-c} = \lim_{x\to c}\frac{k\cdot f(x)-k\cdot f(c)}{x-c}$$
$$= k\cdot\lim_{x\to c}\frac{f(x)-f(c)}{x-c}$$
$$= k\cdot f'(c).$$

□

Example 7.10. We have shown that $\frac{d}{dx}x^{12}=12x^{11}$ and $\frac{d}{dx}x^5=5x^4$. And so, by Proposition 7.9 part (i), we now know that

- $\frac{d}{dx}(x^{12}+x^5)=12x^{11}+5x^4$, and
- $\frac{d}{dx}(7\cdot x^5)=7\cdot\frac{d}{dx}x^5=7\cdot 5x^4=35x^4$.

Theorem.

Theorem 7.11 (*The product rule*). Let I be an interval and let $f,g:I\to\mathbb{R}$ be differentiable at $c\in I$. Then

$$(fg)'(c)=f'(c)g(c)+f(c)g'(c).$$

Proof Idea. At the moment we haven't proved much about derivatives that would be helpful here, so any proof of this will likely be via the definition of the derivative. When solving problems, it is always a good idea to keep in mind where you are trying to reach, and how you might reach that point. Indeed, let's write out where we will begin and start working down, and let's write where we will end and start working up, and let's see how far we can move inward before needing to be smart.

$$\downarrow\qquad (fg)'(c)=\lim_{x\to c}\frac{(fg)(x)-(fg)(c)}{x-c}$$
$$=\lim_{x\to c}\frac{f(x)g(x)-f(c)g(c)}{x-c}$$

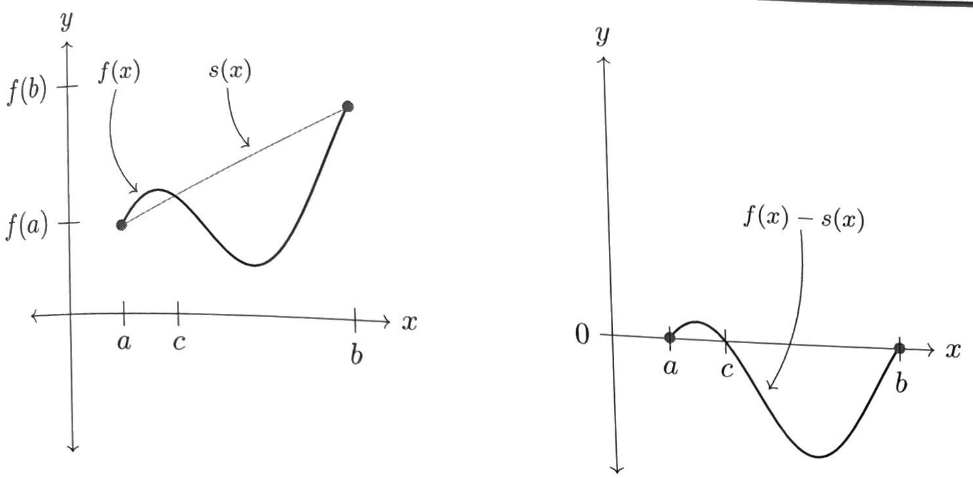
...then a miracle occurs...

$$\uparrow\qquad =\lim_{x\to c}\frac{f(x)\cdot g(c)-f(c)\cdot g(c)}{x-c}+\lim_{x\to c}\frac{f(c)\cdot g(x)-f(c)\cdot g(c)}{x-c}$$
$$=\lim_{x\to c}\frac{f(x)-f(c)}{x-c}\cdot g(c)+f(c)\cdot\lim_{x\to c}\frac{g(x)-g(c)}{x-c}$$
$$=f'(c)g(c)+f(c)g'(c).$$

Theorem.

Theorem 7.22 (*The (derivative) mean value theorem*). Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists some $c \in (a, b)$ where

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$



Our plan is to reduce this theorem to Rolle's theorem. If you start with $f(x)$, and then subtract off the secant line from $(a, f(a))$ to $(b, f(b))$ (call this secant line $s(x)$, pictured above), what you get is a function with $f(a) = f(b) = 0$, which the theorem can handle. The theorem will then give us what we seek. Let's Rolle.

Proof. Note that the equation of the secant line from $f(a)$ to $f(b)$ is

$$s(x) = \left(\frac{f(b) - f(a)}{b - a}\right)(x - a) + f(a).$$

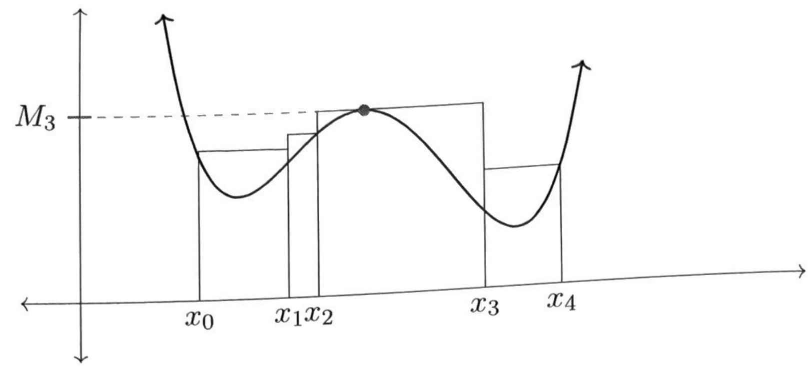
Let $g(x) = f(x) - s(x)$. That is, $g(x)$ is what you get when you subtract off this line from $f(x)$:

$$g(x) = f(x) - \left[\left(\frac{f(b) - f(a)}{b - a}\right)(x - a) + f(a)\right].$$

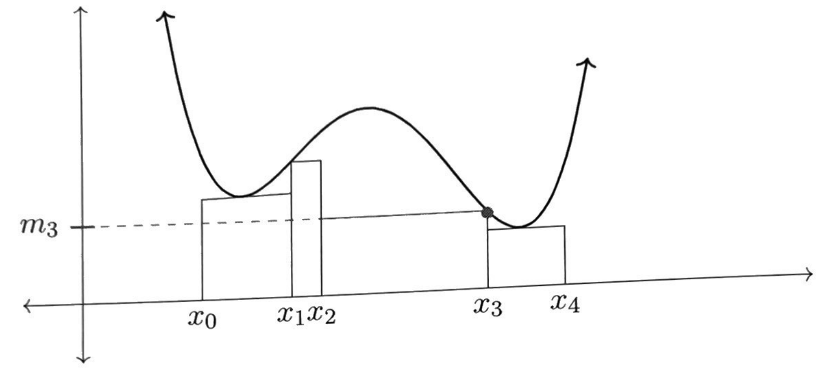
Note that g is differentiable on $[a, b]$ and continuous on (a, b), since f and s are. Also note that $g(a) = f(a) - s(a) = 0$, and $g(b) = f(b) - s(b) = 0$. So we may apply Rolle's theorem to g to say that there exists some $c \in (a, b)$ where $g'(c) = 0$. I.e., $f'(c) - s'(c) = 0$, or equivalently that $f'(c) = s'(c)$. And since

$$s'(x) = \frac{f(b) - f(a)}{b - a},$$

we have indeed found a $c \in (a, b)$ such that $f'(c) = \dfrac{f(b) - f(a)}{b - a}$. □



And here is m_3, the height of the third (lower) rectangle:



Now that we have a way to talk about the widths and heights of both the big rectangles and the small ones, we can write down what we are really interested in: a formula for the upper bound area and the lower bound area. These are called "upper sums" and "lower sums," respectively.

Definition.

Definition 8.4. Consider a function $f : [a, b] \to \mathbb{R}$, and consider a partition $P = \{x_0, x_1, \dots, x_n\}$ of $[a, b]$. Define

- the *upper sum* as

$$U(f, P) := \sum_{i=1}^{n} M_i \cdot (x_i - x_{i-1}), \text{ and}$$

- the *lower sum* as

$$L(f, P) := \sum_{i=1}^{n} m_i \cdot (x_i - x_{i-1}),$$

where M_i and m_i were defined in Notation 8.3.

string of inequalities that looks a little like the one we are trying to prove:

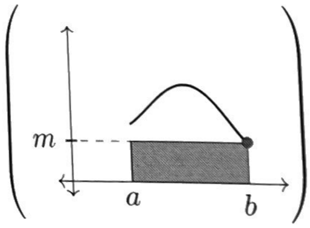
$$L(f, P_0) \leq L(f) \leq U(f) \leq U(f, P_0).$$

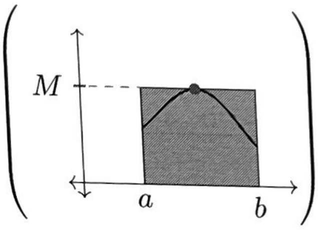
What's left? We must find a P_0 that gives the correct bounds.

What is the simplest partition P_0 of $[a, b]$ that we could apply? Remember that a partition of $[a, b]$ is a set of the form $\{x_0, x_1, \ldots, x_n\}$ where $x_0 = a$ and $x_n = b$. Since a and b are the only numbers that *must* be in the partition, the simplest partition would be $P_0 = \{a, b\}$! This is the partition that corresponds to a single rectangle, and since upper and lower sums correspond to area,

$$L(f, P_0) \leq L(f) \leq U(f) \leq U(f, P_0).$$

turns into

Area 

$\leq L(f) \leq U(f) \leq$ Area 

which gives

$$m(b - a) \leq L(f) \leq U(f) \leq M(b - a),$$

Proof. The inequality $L(f) \leq U(f)$ follows immediately from Proposition 8.7 part (b), and Exercise 8.3. Now we show the bounds on $L(f)$ and $U(f)$, which are a sup/inf over all partitions. The simplest partition possible is the 2-point partition $P_0 = \{a, b\}$ containing only the endpoints (that is, $n = 1$, $x_0 = a$, and $x_1 = b$); this in turn gives a single rectangle in the upper or lower Darboux sum. Thus,

$$\begin{aligned} L(f) &= \sup\{L(f, P) : P \in \mathcal{P}\} \\ &\geq L(f, P_0) \\ &= \sum_{i=1}^{n} m_i \cdot (x_i - x_{i-1}) \\ &= \sum_{i=1}^{1} m_i \cdot (b - a) \\ &\geq m \cdot (b - a). \end{aligned}$$

$$\begin{aligned} U(f) &= \inf\{U(f, P) : P \in \mathcal{P}\} \\ &\leq U(f, P_0) \\ &= \sum_{i=1}^{n} M_i \cdot (x_i - x_{i-1}) \\ &= \sum_{i=1}^{1} M_i \cdot (b - a) \\ &\leq M \cdot (b - a). \end{aligned}$$

Collectively,

$$m(b - a) \leq L(f) \leq U(f) \leq M(b - a).$$

□

What was the one error in the last paragraph? By Definition 8.2, a partition can not have infinitely many points in it; it must be a *finite* set, and so we can't put each point in its own rectangle. Indeed, some rectangle will include infinitely many points! So we must be more careful. Consider, for instance, the following example.

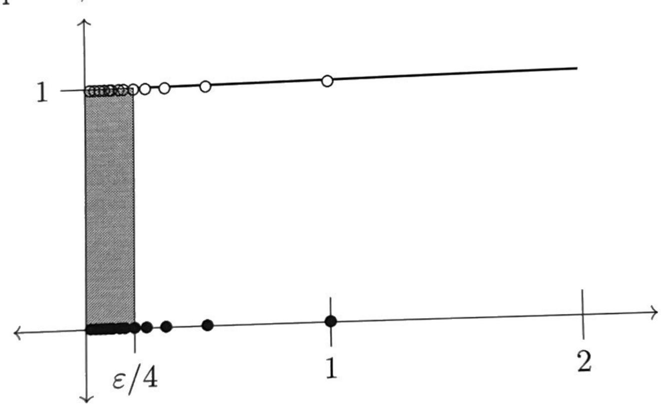
Example 8.20. The function $f_\infty : [0, 2] \to \mathbb{R}$ where

$$f_\infty(x) = \begin{cases} 1 & \text{if } x \neq \frac{1}{n} \text{ for any } n \in \mathbb{N} \\ 0 & \text{if } x = \frac{1}{n} \text{ for some } n \in \mathbb{N} \end{cases}$$

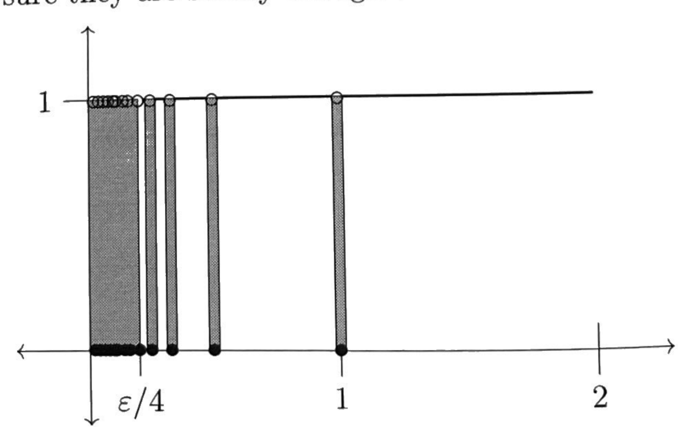
is integrable.

Proof Idea. Here is the idea behind the following proof. First, as we just mentioned, a partition must be a *finite* set. So we have to find a way to have a finite number of skinny rectangles which contain all the points of discontinuity and their areas collectively add up to less than ε (once again, they will add up to $\varepsilon/2$). And then we will use the integrals analytically theorem (Theorem 8.14).

So how do we get the "missing" areas to add up to $\varepsilon/2$? To do this, we will utilize the fact that the points of discontinuity are converging to 0. So for any $\varepsilon > 0$, only finitely many of them are outside of $[0, \varepsilon/4]$. So if we pick a partition that includes $\varepsilon/4$ (and, as is required, includes 0), then the rectangle



will capture all but finitely many of the points, and do so in a rectangle of area $(\varepsilon/4)\cdot 1 = \varepsilon/4$. Then we just add skinny rectangles around the finitely many remaining points, and make sure they are skinny enough that the sum of their areas equals $\varepsilon/4$.



9.3 Other Properties with Functional Convergence

We want to investigate which properties are preserved by pointwise and uniform convergence. We aim to complete the following chart.

Assume that each f_k has the below property	If $f_k \to f$ pointwise, must f satisfy property?	If $f_k \to f$ uniformly, must f satisfy property?
Continuous	No, by Example 9.6	Yes, by Proposition 9.8
Bounded	???	???
Unbounded	???	???
Uniformly Continuous	???	???
Differentiable	???	???
Integrable	???	???

A few of these will be left to the exercises, but ticking off the rest will help us better understand the differences between pointwise and uniform convergence. It's also pretty fun to think about finding a proof or counterexample of each of these. Let's start at the top — boundedness.

Boundedness

For each of these I strongly encourage you to think on your own about what you think the answer is. And if you think the answer is no, try to come up with an example demonstrating it.

Assuming you have spent time thinking about whether each f_k being bounded implies f is, you can read on. We will show that for pointwise convergence boundedness might not carry to the limit, but with uniform convergence it does. We begin with an example showing the pointwise claim.

Example 9.9. Assume that each f_k is bounded and $f_k \to f$ pointwise. It need not be the case that f is also bounded. Here is an example of that:

Define $f_k : (0,1] \to \mathbb{R}$ by

$$f_k(x) = \begin{cases} \frac{1}{x} & \text{if } x \in [1/k, 1] \\ 0 & \text{if } x \in (0, 1/k). \end{cases}$$

Clearly $0 \leq f_k(x) \leq k$ for each x, so each f_k is bounded.[2] However, for any $x \in (0,1]$, by the Archimedean principle there is some $N \in \mathbb{N}$ for which $\frac{1}{N} < x$. And thus for all $k \geq N$ we have $f_k(x) = 1/x$. So clearly for each x we have

$$\lim_{k\to\infty} f_k(x) = \frac{1}{x}.$$

[2]Recall again that for each particular f_k, the number k is fixed. So, e.g., $0 \leq f_{100}(x) \leq 100$. If we knew that $g : \mathbb{R} \to \mathbb{R}$ had the property $g(x) \leq x$, then this would not mean that g is bounded, because x is variable. But saying $f_k(x) \leq k$ does mean that f_k is bounded above, since k is fixed.

hope to do it. Also, this is a topic where seeing animations of the ideas makes a huge difference, and his animations are top notch. You won't be disappointed.

Ok good. Now that you have watched his excellent motivation and developed some intuition, we will get into the formal development of this theory. First up is a definition that is truly Talyor-made.

Definition.

Definition 9.36. Assume $f^{(k)}(c)$ exists for all $k \in \mathbb{N}$. The *Taylor series* of f about $x = c$ is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!}(x-c)^k.$$

If $c = 0$, then the series is called a *Maclaurin series*.[5]

Here is an example of finding a function's Taylor/Maclaurin series.

Example 9.37. Here we find the Maclaurin series of $f(x) = \cos(x)$.

$$\begin{array}{lll}
f(x) = \cos(x) & f(0) = 1 & a_0 = 1 \\
f'(x) = -\sin(x) & f'(0) = 0 & a_1 = 0 \\
f''(x) = -\cos(x) & f''(0) = -1 & a_2 = -\frac{1}{2!} \\
f'''(x) = \sin(x) & f'''(0) = 0 & a_3 = 0 \\
f^{(4)}(x) = \cos(x) & f^{(4)}(0) = 1 & a_4 = \frac{1}{4!} \\
f^{(5)}(x) = -\sin(x) & f^{(5)}(0) = 0 & a_5 = 0
\end{array}$$

It is evident that

$$f^{(k)}(0) = \begin{cases} 0, & \text{if } k \text{ is odd,} \\ (-1)^{k/2}, & \text{if } k \text{ is even.} \end{cases}$$

So the Maclaurin series for $f(x) = \cos(x)$ is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}(x)^k = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots.$$

5 Named after Scottish mathematician Colin Maclaurin, who was a pioneer in the field of analysis and, presumably, a pioneer in the art of rockin a dope afro:

(He was also considered the youngest professor in history, until 2008 when American materials scientist Alia Sabur claimed the crown 3 days before her 19th birthday.)

Appendix A: Construction of $\mathbb{R}$

"Far out in the uncharted backwaters of the unfashionable end of the western spiral arm of the Galaxy lies a small unregarded yellow sun. Orbiting this at a distance of roughly ninety-two million miles is an utterly insignificant little blue green planet whose ape-descended life forms are so amazingly primitive that they still think digital watches are a pretty neat idea." This is how Douglas Adams began his classic book *The Hitchhiker's Guide to the Galaxy.* He began its sequel, "The story so far: In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move." Douglas Adams can certainly set a tone.

"It was a pleasure to burn" is the first sentence of Ray Bradbury's *Fahrenheit 451.* "I'm pretty much fucked" is the start to Andy Weir's *The Martian.* And, "All happy families are alike; each unhappy family is unhappy in its own way" is how Leo Tolstoy began *Anna Karenina.*

If the first chapter of the *Harry Potter* series were a dry introduction to wand movements and cauldron thickness, few would read on. Yes, the spells and potions rely on those details, but it's not where the magic really is. Therefore J.K. Rowling's first chapter contained exhibitions of magic, mention of a dark wizard so terrifying that most fear to speak his name, allusions to this wizard's sadistic reign which mysteriously and abruptly just ended, and a significant baby boy, dropped on the doorstep of the least magical people possible. Lots of questions, few answers.

Where am I going with this? The material in Appendix A is the 'wand movements' and 'cauldron thickness' part of real analysis. The spells and potions we developed in this text rely on the foundational aspects we are about to discuss, but if I started the book with them you might have fallen asleep before Chapter 2. And just like how, after reading (and rereading[1]) the books, hardcore Harry Potter fans crave to know the small — but fundamental — details about the world they entered, perhaps now you will be motivated to learn the underpinnings of the world of real analysis.

Everything we proved relied on the existence of the real numbers. They needed to be there and to have the properties we discussed, and the rigor of mathematics demands we don't simply assume their existence. Which axioms are necessary to give us the world we want, and why? Doing this thoroughly (in long-form style) might take 50 pages, and even hardcore analysis fans would likely fall asleep. This appendix is just a 7 page sketch of it. Its goal is simply to illustrate for you the meticulous process which needed to be done to formally begin the theory of real analysis.[2]

[1] And rerereading

[2] And to make you appreciate the brave soul who took one for the team, and painstakingly did it.

Appendix B: Peculiar and Pathological Examples

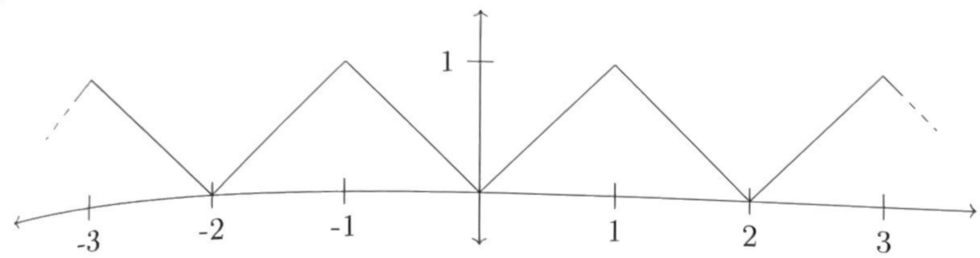
Jerry Seinfeld: Oh you're crazy.
Cosmo Kramer: Am I? Or am I so sane that you just blew your mind?
Jerry Seinfeld: It's impossible.
Cosmo Kramer: Is it? Or is it so possible that your head is spinning like a top?
Jerry Seinfeld: It can't be.
Cosmo Kramer: Can't it? Or is your entire world just crashing down all around you?
Jerry Seinfeld: Alright that's enough.
Cosmo Kramer: Yeaaaaaaaah!

The world of mathematics can be a strange place, and when your mind is exploring this world and finds itself in some dark, foreign region, you risk falling into pits, being eaten by monsters, and missing a beautiful forest for the boring trees.

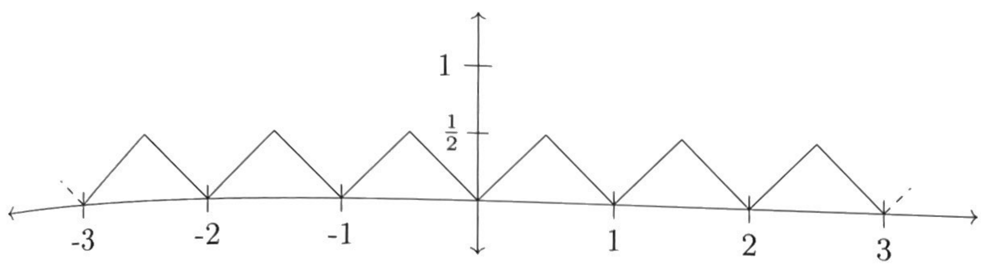
It has been said that one of the most important goals of learning real analysis is to collect as many bizarre examples as you can, and to keep them in your back pocket. From a practical standpoint they will inform your conjectures and guide your proofs, but they will also help to demonstrate why real analysis is such a great subject.

When little kids learn about the world, many are fascinated by dinosaurs because they show how strange and large and monstrous life on Earth can be. Flowers are also nice. I like flowers. Your course abstract algebra, for instance, is filled with theorems which are small and pretty and I am happy to see them and learn about them. But I prefer the monsters.[1] We have seen many awesome monsters already in this book, but if your back pocket has some room left, this appendix has a few more.

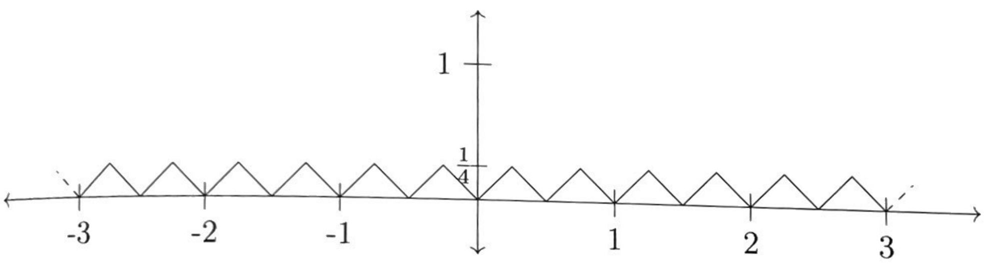
[1]Don't @ me, researchers of the Monster group.



But how non-differentiable can a continuous function be? We could of course make the non-differentiable points occur more often. Consider the function which we'll call A_1:



Note that A_1 is discontinuous at every point of the form $\frac{m}{2}$, where m is an integer. Better yet, here's A_2:



Note that A_2 is discontinuous at every point of the form $\dfrac{m}{2^2}$, where m is an integer. Indeed, you could keep doing this and A_k would have the property that it is discontinuous at every point of the form $\dfrac{m}{2^k}$.

In some sense this is making things weirder, but in another it's not doing much: Is it so different to have every 8^{th} term be discontinuous compared to every 64^{th} term? It's still a discrete set where you have fixed gap-sizes between the points.

The holy grail example would be to find a continuous function that is differentiable *nowhere*. Now take a minute to appreciate how peculiar such a function would be. In all our examples thus far, being non-differentiable at a point was obtained by having two lines (which are not only continuous, but also differentiable) meet at a single point, forming a peak. The other "elementary" technique to get a non-differentiable point is to have a jump discontinuity there, but this can't be used since we want our function to be continuous. In Example 7.15 we saw another approach to maintain continuity while losing differentiability, but this approach seems to have no advantages over a simple peak, while at the same time being much harder to grasp.

Peaks seem the best approach, and it is reasonable to think that in order to have a peak, you need to have a region around that peak where you have two nice,

B.18 Tarski's Terrific Talents Times Two

The great Ian Stewart imagined a *hyperdictionary*, which contains not only every word, but also contains every *possible* word. Sure, it contains "math" and "is" and "fun", but it also contains "ydac" and "faqir" and "galbhepvnx".[22] These are called *hyperwords*. The hyperdictionary contains infinitely many words, since any combination of finitely many letters counts as a hyperword. The hyperdictionary begins:

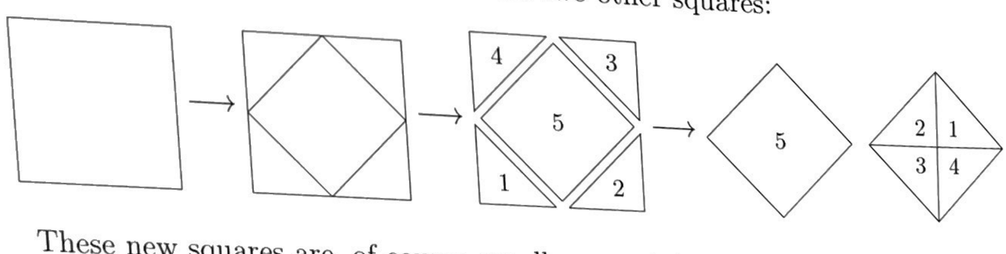
> a, aa, aaa, aaaa, aaaaa, ... <INFINITELY MANY HYPERWORDS> ..., aaab, aaaba, aaabaa, ... <INFINITELY MANY HYPERWORDS> ..., aab, aaba, ... <INFINITELY MANY HYPERWORDS> ..., ab, aba, abaa, ... <INFINITELY MANY WORDS> ..., b, ba, baa, baaa, ...

The words form what is called a total ordering — meaning that any two of them can be compared to determine which comes first in the hyperdictionary — but things are still a little weird in that most pairs have infinitely many words between them.

The hyperdictionary has another amazing property — in some sense it contains itself. This is what I mean: Suppose the hyperdictionary were divided into 26 volumes, based on each word's first letter. If you want a copy of the hyperdictionary but don't have enough money to buy all 26 volumes, do this: Buy Volume A and a lot of white-out. Every word in Volume A starts with an 'a' — now go through this volume and white out that first 'a' from each word. After doing so what you're left with is the entire 26 volume hyperdictionary! For example, we don't need Volume K to get the word "kdalghsdh", because the word "akdalghsdh" was in Volume A, and after removing the first 'a' with white-out, you're left with "kdalghsdh"! This is an interesting phenomenon, and it's only possible with an infinite dictionary. It might remind you of Chapter 2 where, for instance, we showed that a set (like $\mathbb{N}$) can be in bijection to a proper subset of it (like $2\mathbb{N}$).

Now we move on to an amazing theorem, named after Stefan Banach and Alfred Tarski who in 1924 gave a construction that shocked the mathematical word.[23]

We begin by noting that a square can be divided up into five pieces, and those five pieces can then be rearranged to form two other squares:



These new squares are, of course, smaller — each has half the area of the original square. However, of course, you should always be careful when saying "of course" in

[22] Actually, one of those is a real word!

[23] If I could have either a construction, lemma or theorem named after me, I think my preferences would be in that order. To me, proving something is cool, but proving something that is fundamental enough and used often enough to be called a lemma is even cooler, but the best would be to construct a concrete mathematical object that shocks and awes.