

IAP Proofs Workshop 2021 Lectures 6-7

Andrew Lin

January 2021

6 Countable and Uncountable Sets

In the last two days of this class, we'll continue to explore proofs through different kinds of examples, and the focus here will be on **analysis**: understanding concepts like infinities, limits, and continuity. We're introducing more abstraction now, and even though something like "infinity" doesn't technically exist in the real world, they're still necessary to make techniques like calculus fully justified!

The reason that mathematics starts becoming trickier when we introduce infinitely large "numbers" or infinitely large sets is that some of the usual tools and calculations that we use don't work anymore:

Example 6.1

- Suppose we're trying to evaluate the sum $1 - 1 + 1 - 1 + 1 - \dots$. We can pair up adjacent -1 s and 1 s and cancel them out, and thus the sum can seemingly be grouped to add to either 1 or 0 . In fact, this sum is not actually well-defined as stated, even though we're just adding a bunch of integers together.
- Similarly, we can do a silly calculation as follows: suppose that $S = 1 + 2 + 3 + \dots$. Then we can do a bit of manipulation to "find" that

$$S = 1 + (2 + 3 + 4) + (5 + 6 + 7) + (8 + 9 + 10) + \dots = 1 + 9 + 18 + 27 + \dots,$$

$$\text{so } S = 1 + 9(1 + 2 + 3 + \dots) = 1 + 9S, \text{ meaning } S = -\frac{1}{8} \text{ (which is absurd)}$$

So we'll need to be more precise with our language to avoid those issues:

Definition 6.2

(Positive) **infinity** is a formal symbol ∞ which is larger than every (real) number.

We'll come back to this a bit more when we discuss limits in the next class, but the important thing is that if we write something like $1+1+1+\dots = \infty$, it doesn't mean that that infinite sum is actually a number, so we can't directly work with it and do things like addition and subtraction without being careful. And a big focus of today's class is going to be talking about why infinity doesn't exactly behave like a number.

Example 6.3

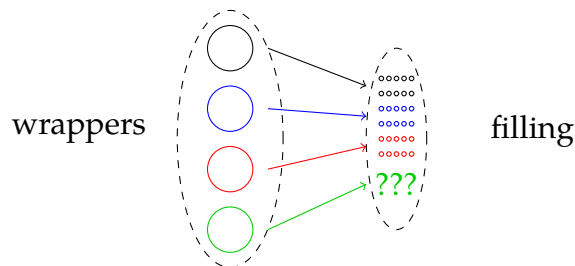
When we define the **extended real numbers**, which means we add infinity and negative infinity in with the "normal numbers," we define $c + \infty = \infty$ for any number c , and it's also okay to say that $c \cdot \infty$ is ∞ for positive c and $-\infty$ for negative c . But we cannot say that $0 \cdot \infty$ evaluates to anything, and $\infty - \infty$ isn't defined either.

(We'd start running into a lot of $1 = 2$ funny business if we allowed ourselves to do those kinds of forbidden calculations.) And the main attraction of today's class is actually the idea that not only can we not know exactly how large infinity is, but furthermore there are different levels of "hugeness" that are hiding behind all of the ordinary mathematics that we do. (A motivating question here might be something like "is $\infty \cdot \infty$ larger than ∞ in any rigorous sense?")

Example 6.4

If I have a bunch of dumpling wrappers and a bunch of dumpling filling, and my friend asks me whether there's too many wrappers or too much filling, the easiest way to resolve the question is to make dumplings until we run out.

This might seem like a silly idea, but the fundamental concept here is that we can **compare how big different sets are by pairing things up with each other**, and we can do this without calculating the exact number of dumpling wrappers or exact quantity of filling.



And this pairing is very similar to how we think about functions between sets in mathematics in general, so let's study that a bit!

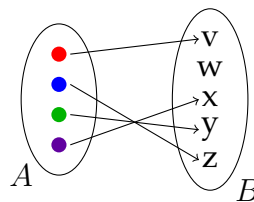
Definition 6.5

Suppose we have two sets A and B . A **function** $f : A \rightarrow B$ assigns, for every element $a \in A$, an element $f(a) \in B$. (A is called the **domain**, and B is called the **codomain**.) The function f is

- **one-to-one** (also **injective**) if no two elements in A map to the same element of B ,
- **onto** (also **surjective**) if every element in B is mapped to by some element of A ,
- **bijective** if both of the two conditions above are satisfied.

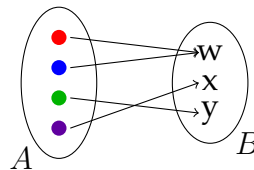
These words come up very often in mathematics, so let's try looking at them more carefully to understand when each one is applicable.

- **Injective:** if we think about our functions as drawing arrows between the sets A and B , then there is exactly one arrow outgoing from each element of A . And being injective means that no two of these arrows can collide and end up at the same endpoint in B :



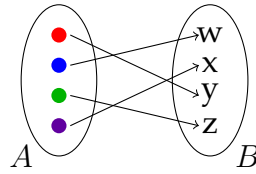
In mathematical language, we can equivalently say that if $f(a) = f(b)$ for our function f , then $a = b$ (we must have started with the same element at the beginning).

- **Surjective:** again thinking about our arrows, mapping to every element in B means that there is at least one arrow that points to any element of B , and we get a picture like this:



In mathematical language, we can say that for all elements $b \in B$, there is an element $a \in A$ such that $f(a) = b$.

- **Bijective:** Combining the previous two ideas together, a function which is both injective and surjective has both at most one arrow and at least one arrow to any element of B , so exactly one arrow points to each element of B .



In particular, this is a way of “pairing up” elements in A and B with each other – any element of A corresponds exactly with an element of B . And here’s another way to think about this: a function is bijective whenever we can construct an **inverse function** (that goes backwards along the arrows and always makes us go back to where we started).

If you prefer to think about functions $f : \mathbb{R} \rightarrow \mathbb{R}$ (which we tend to represent with graphs on the xy -plane), here are some additional examples:

Example 6.6

Real-valued functions are injective if they pass the horizontal line test (in other words, no horizontal line intersects the graph of $y = f(x)$ twice), and they’re surjective if there is a value of x such that $f(x) = c$ for any x . So if we’re considering functions $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = e^x$ is injective but not surjective, and $f(x) = \tan x$ is surjective but not injective. (But restricting the domain and codomain can make either function both injective and surjective, and thus bijective.)

Going back to the idea of sizes of sets, suppose that we have a set A with six elements, and a set B with seven elements. There’s no way to use A to “cover B ” if each element of A can only hit one element of B , because “seven is larger than six.” So the idea here is that **because there’s no bijective map between A and B , A and B must have different sizes.**

And defining largeness through functions is exactly what we’re going to do now. We know that there’s one set that is supposed to be used for counting: the counting numbers $\mathbb{N} = \{0, 1, 2, \dots\}$. And in some sense, this is the “smallest” infinity, because it’s (informally) counting “barely past” any number.

Definition 6.7

Let S be a set. If there are finitely many elements in S , we call the set **finite**; otherwise, it is **infinite**. For infinite sets, if there is a bijective function $f : S \rightarrow \mathbb{N}$, then S is **countable**; otherwise it is (bigger and called) **uncountable**.

Problem 6.8

Some people like to use the positive integers $S = \{1, 2, 3, \dots\}$ instead of the natural numbers $\mathbb{N} = \{0, 1, 2, \dots\}$; let’s prove that these are the “same size” by showing that S is countable.

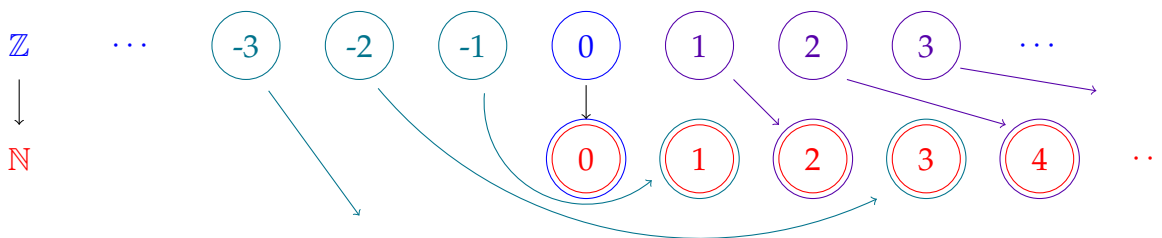
Proof. The function $f : S \rightarrow \mathbb{N}$ given by $f(x) = x - 1$ is a bijection. This is because associating x with $x - 1$ is the same as associating y with $y + 1$, so there is an inverse function $f^{-1} : \mathbb{N} \rightarrow S$, specifically $f^{-1}(x) = x + 1$. So f is a bijection, and S is countable. \square

So the takeaway here is that even though one of the sets is a (strict) subset of the other, it's still possible for them to be the same size – we just need to find **one** bijection between them! (It makes sense for this to feel a bit weird, but make sure you can follow the proof.)

Problem 6.9

Next, let's show that the set of **all** integers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ is countable. (Notice now that unlike $\mathbb{N} = \{0, 1, 2, \dots\}$, this set is "infinite in two directions.")

Proof. In order to show that a set is countable, we need to explicitly construct a bijection. Here's a picture to generally illustrate how we'll do that:



In other words, send 0 to 0, send a positive integer x to $2x$, and send a negative integer $-x$ to $2x - 1$. This is a bijection, because given any **red natural number**, I can tell you exactly which **blue integer** it corresponds to. So \mathbb{Z} is indeed countable. \square

Another way to think about this proof is that we showed \mathbb{Z} is countable by "counting off" integers, saying them in a list, and proving that we always hit every single one. (In this case, we're counting like 0, 1, -1, 2, -2, ...) So now with this "counting off" framework, we can show that the set of primes is countable (by assigning the primes 2, 3, 5, 7, ... to the integers 0, 1, 2, 3, ... respectively, and so on). But let's move up in complexity a step:

Problem 6.10

Think about how you might prove that \mathbb{Q} , the set of rational numbers, is also countable! As a hint, one way to think about this is to find a way to "order" the fractions from "small" to "large" (for some definition of "small" and "large" different from usual), and making the "smallest" fraction map to 0, the next smallest map to 1, and so on.

But the real star of today's class is the most important proof of the day:

Theorem 6.11

The set of real numbers is **uncountable**, even though the rationals are countable and seem to take up every spot on the number line.

Cantor's diagonalization argument. Suppose **for the sake of contradiction** that we had a bijective function $f : \mathbb{R} \rightarrow \mathbb{N}$, meaning that we associate exactly one real number with each of 0, 1, 2, and so on. Write out their decimal expansions, for instance like this:

0 : 0020.12398713928
1 : 4014.238400000000
2 : 0244.000000000000
3 : 1022.24937842983
4 : 0005.14978210000
⋮

(The point is to line up the decimal points – don't pay too much attention to the trailing or leading zeros.) We're trying to show that this bijective function f cannot actually exist, and the way we do this is by showing that there is always a real number that is not in the infinite list. To do that, for the number corresponding to the natural number k , look at the k th decimal place.

0 : 002**0**.12398713928
1 : 4014.**2**384000000000
2 : 0244.**0**000000000000
3 : 1022.24**9**37842983
4 : 0005.149**7**8210000
⋮

And now we make a new number which has **all decimal digits different from the ones listed in blue**: for example, we can start the number with 2.3186, which doesn't have a 0 in the ones digit, a 2 in the tenths digit, and so on. Since this number is different from **every single number in the list**, it must have been left out, and thus our original bijective function f is invalid. Thus no bijective function exists, and \mathbb{R} cannot be countable.

There's a bit of subtlety in the way we do this, though: some real numbers may seem to have two different decimal representations, like $0.9999 \dots$ and 1. Avoiding these issues is worth thinking about, so we'll leave it as an exercise! (A general hint: just avoid having infinite trailing 9s, and things should generally work out.) \square

Studying more set theory will reveal that this distinction between countable and uncountable sets can be expanded to the study of ordinal numbers and much more. But because we're focusing on proofs, we'll steer away from too much set theory, and hopefully in the last recitation there will be more practice with this idea of bijections and functions!

7 Basics of Real Analysis

This final lecture is intended to give a teaser into the material of 18.100 (real analysis), which discusses the ideas of calculus in a more rigorous (and abstract) way. Throughout this, try to focus on understanding all of the definitions and how we might work with them to write proofs.

Calculus is (basically) the study of derivatives and integrals, and we start defining all of these using limits. But historically, there was quite a bit of discussion that happened before all of these concepts could be put on rigorous footing.

Example 7.1 (Zeno's paradox)

Suppose we wanted to walk down the Infinite Corridor. To do this, we have to first walk down the first half of the path, but before that, we need to first walk the first quarter of the path, and before that, the first eighth, and so on. There's always an "earlier" action to take, and Zeno claims that "it's impossible to do infinitely many actions."

The infinite sum $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$ is the well-known geometric series, but the way we actually arrive at this result requires us to add infinitely many numbers together, and that's what the Greeks weren't super happy about. So we need an idea of what happens when we keep trying to do something infinitely many times, or get infinitely (we'll also use the word "arbitrarily") close to some final answer.

But (part of) what really started analysis was that in the late 19th century, people started taking a closer look at the foundations of calculus. (Take a look at https://en.wikipedia.org/wiki/Arithmetization_of_analysis, for example.) Basically, it was odd that the techniques used in calculus (drawing rectangles and tangent lines) were so geometric compared to the more algebraic ideas in other areas of math (like number theory or discrete math, as we've discussed previously). And in fact, it wasn't even clear how to define the real numbers rigorously to have all of the properties that we normally want! So we'll go through some of those steps in the rest of this lecture.

Example 7.2

Here's an analogy similarly silly to the dumpling one from last lecture: suppose that I needed to make a meter stick (which is 1 meter long). We can make formal definitions of what a meter is, but in practice there's always some **desired precision** or **margin of error** that the meter sticks need to fall within. (We could imagine that there's an evil construction worker who wants to test our meter sticks, and no matter how small the margin of error is, we need to make sure we meet the demands.)

The "principle" to keep in mind here is that analysis is often about **being close enough** to some final number or other quantity, so many of our definitions will have this "wiggle room" built in. But let's see a few examples to make this point more clear.

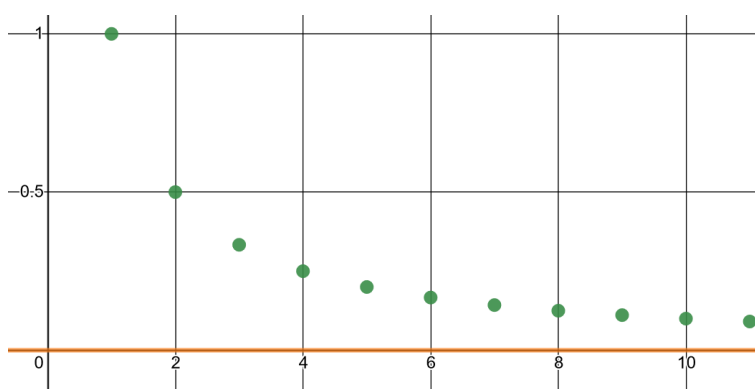
Example 7.3

We've often seen in calculus that we make statements like $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$. In other words, if we have a sequence of numbers like

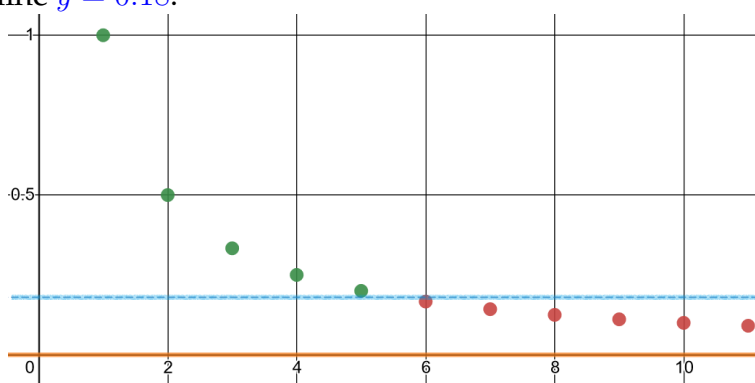
$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots,$$

this sequence has **limit 0**.

We could imagine graphing the sequence as shown below: one way to describe this formally is that on the x -axis, we plot the integer n , and on the y -axis, we plot the integer a_n (the n th term of the sequence).



Notice that none of the numbers are exactly zero, but as we go down the list, the numbers get arbitrarily close to zero. To make that precise, we can say (for example) that we want our threshold to be the line $y = 0.18$.

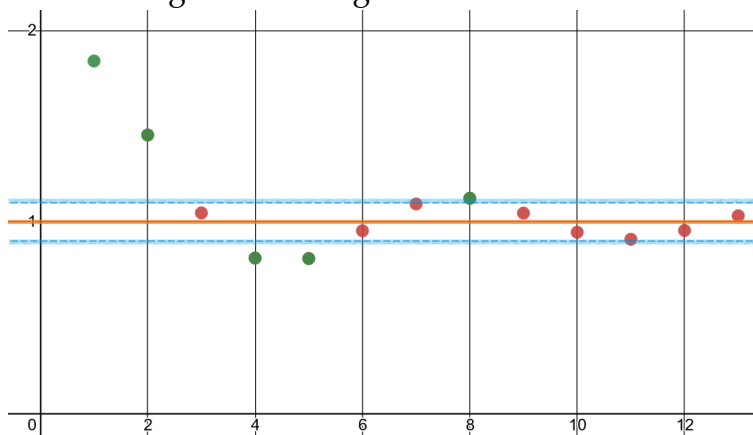


The first few terms of our sequence are above this line, but as long as we're on the **6th number or later**, the whole sequence is indeed **below the blue line**. And even if the blue line gets closer, our sequence will still **eventually be within the margin of error**— we just have to wait longer for the sequence to “settle down.”

Example 7.4

For our next example, I'll use the sequence $a_n = 1 + \frac{\sin n}{n}$, which we want to show converges to 1.

One silly way to describe this is that I'm trying to make meter sticks (where a_n describes the length of the n th attempt), and as we can see in the picture below, the meter sticks are generally getting closer and closer to being 1 meter long.

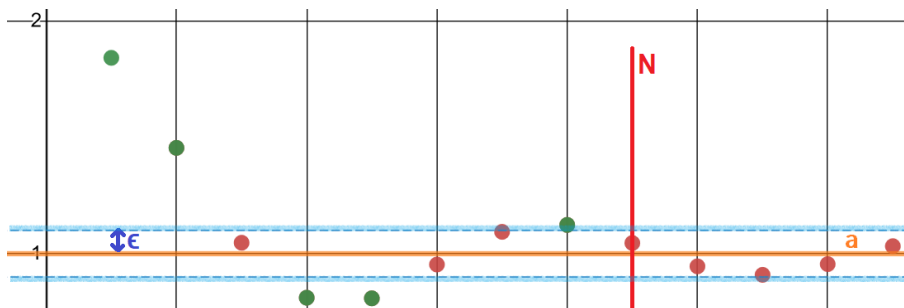


We've drawn in blue lines at $y = 0.9$ and $y = 1.1$ here, and one way to interpret this is that "the evil construction worker wants our meter sticks to be within 0.1 meters of the correct length." And at first, some of my meter sticks aren't good enough – the first, second, fourth, fifth, and eighth sticks are outside the blue lines – but once I get to the **9th stick**, all of my attempts are **within the margin of 0.1**.

To make this into a formal definition, the idea is that no matter how tight my **margin of error** needs to be, I can show that **starting from some number in the sequence**, we're within that margin. Here's a formal definition:

Definition 7.5 (Limits, part 1)

Suppose we have an infinite sequence of numbers $\{a_1, a_2, a_3, \dots\}$. This sequence **converges** to a number a (or **has limit** a) if **for every number** $\varepsilon > 0$, the sequence is eventually within ε of a . In other words, **we can find some integer** N so that $|a_n - a| < \varepsilon$ for all $n \geq N$.



Here, ε can be thought of as the demand from the evil construction worker, and N can be chosen in terms of this demand. So now let's return to our original example and write out something more rigorous:

Problem 7.6

Prove that the sequence $\{1, \frac{1}{2}, \frac{1}{3}, \dots\}$, which we can write as $a_n = \frac{1}{n}$, converges to 0 (notably, even though none of the terms in the sequence are 0).

Let's first see the structure of the proof with some concrete numbers:

The structure of the proof. Suppose we are given (for example) $\varepsilon = 0.01$. We need to find a way to make sure all of the terms a_n are eventually less than 0.01 away from 0: this happens whenever $\frac{1}{n} < 0.01 \iff n > 100$.

So let's pick $N = 101$ (larger cutoffs work too). Now we can show that $\frac{1}{n}$ is eventually close enough to 0. Indeed,

$$\frac{1}{n} \leq \frac{1}{101} < 0.01$$

whenever $n \geq 101$, which implies that our sequence is eventually always within 0.01 of 0. \square

And now we just replace numbers with variables, so that things work for **any** value of ε :

Proof. Suppose we are given some **arbitrary** ε . We need to find a way to make sure all of the terms a_n are eventually less than ε away from 0: this happens whenever $\frac{1}{n} < \varepsilon \iff n > \frac{1}{\varepsilon}$.

So let's pick $N = 1 + \lceil \frac{1}{\varepsilon} \rceil$ (larger cutoffs work too, and the $\lceil \cdot \rceil$ "ceiling function" is rounding the number up just to make sure it's an integer). Now we can show that $\frac{1}{n}$ is eventually close enough to 0. Indeed,

$$\frac{1}{n} \leq \frac{1}{N} < \varepsilon$$

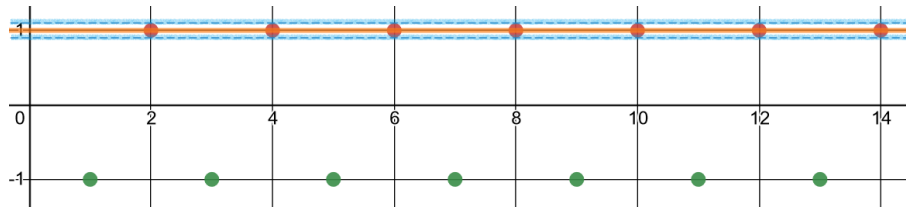
whenever $n \geq N$, which implies that our sequence is eventually always within ε of the point 0 (formally, $|\frac{1}{n} - 0| < \varepsilon$). Since ε was arbitrary, this completes the proof. \square

This kind of proof often takes many passes through to internalize, so it's highly encouraged to try to write it up yourself!

Problem 7.7

Next, let's try to prove that the sequence $\{-1, 1, -1, 1, \dots\}$, which we can write as $a_n = (-1)^n$, does **not** converge to 1.

Much like how we show a number is irrational using contradiction (because it's easier to work with rational numbers than irrational ones), we'll prove that the sequence does not converge by contradicting the definition of convergence. In other words, "we get to be the evil construction worker" here!



Proof. Suppose for the sake of contradiction that a_n converges to 1. **Now we are the evil construction worker:** we show that this fails by trying $\varepsilon = 0.1$. (Other choices also work here.)

But it's impossible to eventually always be within 0.1 of 1 in this sequence (because of the -1 s). In other words, it is impossible to choose N so that after the N th term, we're always between the blue lines. So convergence fails, as desired. \square

We can also refine this line of thinking by discussing what it means to converge to positive or negative infinity, and the definition here is a little different because positive and negative infinity aren't real numbers. One way to think about this next part is that an "infinitely long meter stick" just needs to be longer than any finite measurement:

Definition 7.8

A sequence $\{a_n\}$ **diverges to** $+\infty$ if for every real number x , the sequence is eventually larger than x : we can pick an N (in terms of x) so that $a_n > x$ for all $n \geq N$. (The definition for divergence to $-\infty$ is very similar; try it out yourself.)

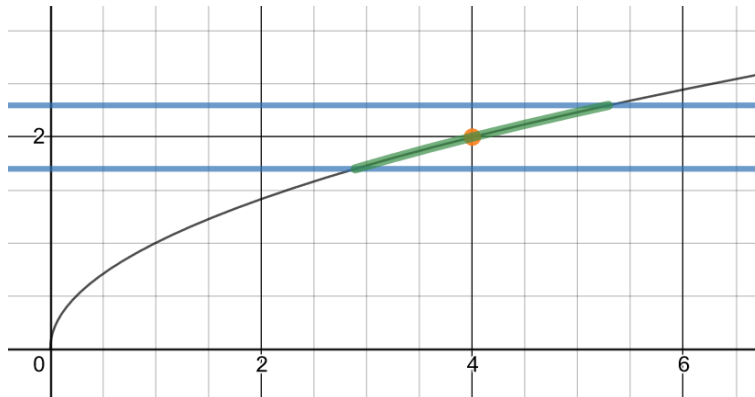
But the definition we've been working with here is for sequences, not real-valued functions – we're evaluating our sequences a_n by plugging in an integer n , while normally with a function $f : \mathbb{R} \rightarrow \mathbb{R}$, we'd want to plug in a real number x . And before we jump into the definition for real-valued functions, it's worth discussing what real numbers really are first:

Example 7.9

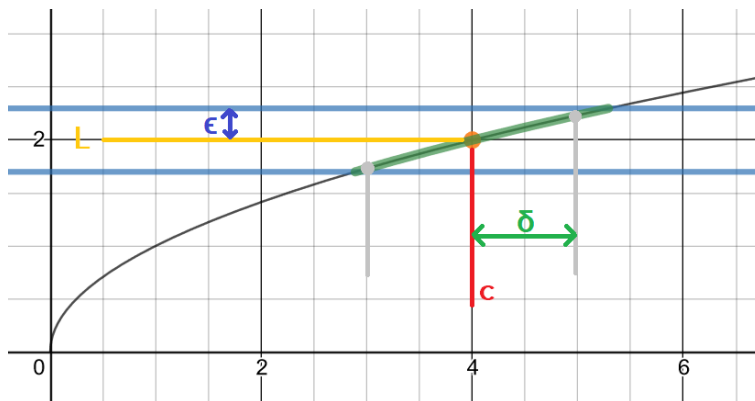
The issue with rational numbers is that a sequence of rational numbers can converge to something that's not rational. For example, the sequence $\frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \dots$ (ratios of consecutive Fibonacci numbers) converges to the golden ratio $\phi = \frac{1+\sqrt{5}}{2}$, which is irrational.

Generalizing this a bit, it makes sense to say that real numbers are basically **all of the possible convergence points for rational sequences**, and in fact, this is one way to formally define the real numbers! (But that's something to see in 18.100.)

With that, we'll get into limits and continuity for real-valued functions. An informal definition for a function being **continuous** is that we can draw the graph without picking up our pencil, but a more mathematical way of saying this is that **values of outputs $f(x)$ are close to each other if the inputs x are close.**



For example, the graph of $f(x) = \sqrt{x}$ shown above has a point $(4, 2)$. And if we change the x -value by a little, the y -value doesn't change very much: in particular, we can stay between the blue horizontal lines (which should look familiar) as long as we're close enough to $x = 4$. So this time, instead of eventually being **close enough** because **we go far enough in the sequence**, we make sure that we're **close enough** once **our inputs are close enough**. And that means we're ready for the full definition now:

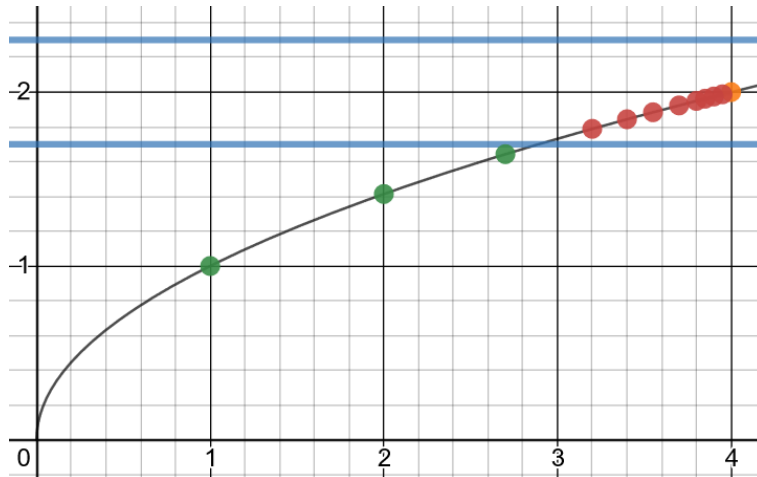


Definition 7.10 (The “epsilon-delta”)

Suppose we have a function $f : \mathbb{R} \rightarrow \mathbb{R}$. This function has **limit** L at a point $x = c$ (denoted $\lim_{x \rightarrow c} f(x) = L$) if for every threshold $\epsilon > 0$, we can pick a $\delta > 0$ (in terms of ϵ) so that whenever $0 < |x - c| < \delta$, we have $|f(x) - L| < \epsilon$. And f is **continuous** at $x = c$ if $\lim_{x \rightarrow c} f(x) = f(c)$.

Just like with the previous definition, it can often take a few tries before this definition really sinks in! But this formal version of continuity is necessary if we want to use continuity to prove facts about functions or other mathematical objects.

Furthermore, we can actually connect these two definitions of convergence with each other. Saying that we get closer to $x = 4$ on the graph above can be visualized by taking a sequence of numbers that converge to 4, as shown below:



It turns out that a function is continuous at some $x = c$ **if and only if** for any sequence $\{x_1, x_2, \dots\}$ of x -coordinates that converges to c , the values $\{f(x_1), f(x_2), \dots\}$ converge to $f(c)$ (follow the dots in the picture above, and see if their heights converge to the correct height). This is called **sequential continuity**, and the reason that it connects our two previous definitions is that in both cases (thinking about this as a real-valued sequence or as a function), we care about being within some ε of the final goal.

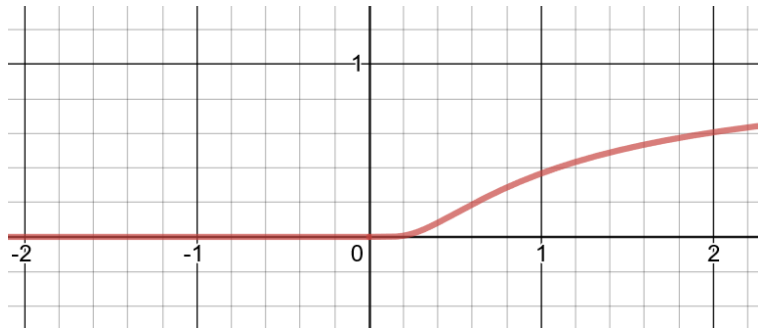
To bring everything back into perspective, we can think about a statement we'd make in 18.01 like

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Implicitly, when we say that $h \rightarrow 0$ here, what we're claiming is that the limit exists **no matter how we approach 0** (in other words, no matter which sequence $\{h_1, h_2, \dots\}$ converging to 0 we use). And the fact that this derivative exists makes many theorems (like the power rule, chain rule, u -substitution, and the fundamental theorem of calculus) provable! So in a more detailed study of analysis, this is the point from which we build calculus up, and it's what brings it into the same standard of rigor as the other mathematical topics we've discussed in this class.

Finally, to finish the class, we'll mention a few things that come up as we get further into analysis. Basically, there are two ways to keep going after we re-establish the rules of calculus: either we can increase abstraction (thinking about n -dimensional space, or talking about how these properties apply to general **metric spaces**), or we can study more properties of our functions, classifying how they behave and understanding how to work with them. Here are three "sneak peeks" for that:

- We often like to work with Taylor series, which are "infinite polynomial series" that represent a function. But this representation isn't always accurate: consider the function below, which is $f(x) = e^{-1/x}$ for $x > 0$ and 0 otherwise.



It turns out that this function is infinitely differentiable, but all of its derivatives at $x = 0$ are just 0, so the Taylor series is zero! So it's not enough just to know whether a function is differentiable or continuous – there are many more subtleties to consider.

- A natural next step after we've defined convergence of numbers is to think about convergence of functions (what happens if we have a sequence of functions and try to take their "limit?"). For that, consider the following diagram:



Here, we've graphed $f(x) = x, x^2, x^5$ for $x \in [0, 1]$. As the exponent increases, the function "droops" more and more, and if we take the limit of the function values at every x -coordinate, we find that the "pointwise limit" evaluates to 0 for all $x \in [0, 1)$, but stays at 1 for $x = 1$. In other words, even though we have a bunch of continuous functions to start with, the final limiting shape is not even continuous! So there are some subtleties here as well (and this is connected to the concept of **uniform continuity** if you're curious).

- Finally, going back to the "arithmetization of analysis" idea from mentioned earlier, there are still other questions of a similar flavor that can be asked. For example, how do we measure area and volume formally, or explain why most real numbers are irrational (measure theory)? Is there more than can be done with this meta-study of functions (functional analysis)? And can we derive everything in math from the axioms, and is it possible to prove all of these statements (mathematical logic)? The point here is that **new mathematical ideas are invented by coming up with new perspectives**, and we hope that you'll continue to explore some of them after this class!