FA25 Introduction to Single Variable Calculus (Honors)

Anish Chedalavada

November 6, 2025

Contents

1 Introduction

Course methodology

Officially, this course services both a year-long calculus requirement as well as an introduction to proofs (IP) prerequisite. I'm assuming that you enrolling in this course means that you're interested in its contents: before deciding to be here, you ought to have read what's on the cereal box. As stated in the syllabus, by the end of this course you should be able to:

1. Understand, formulate, and justify mathematical assertions.

2. Apply the above skills broadly towards setting up and rigorously studying the differential and integral calculus of functions in a single variable.

This doesn't really address the question of *why* we're here though. Take for example the phrase "justify mathematical assertions". I have not unpacked the words *justify*, *assertion*, or even *mathematical*. Maybe each of you already has a comprehensive idea of what these are, but I think it's safe to assume that this is work in progress. Yet somehow in advance of this, we've committed to assemble here on Mondays and Wednesdays this semester at 4:30-5:45 in Krieger 309 or in my office hours at TBA. I think it's best to update our learning objectives:

1. Define the following words: "Understand", "formulate", "justify", "mathematical" "assertions", and what the concatenations thereof mean. Develop a framework their execution.

2. We then formulate an objective, which will generally involve applying the above skills towards the differential and integral calculus of functions in a single variable.

Once again, I'm conflicted. What are "necessary preliminaries"? The differential and integral calculus of functions in a single variable has a bunch of foundational approaches and differing scopes, even if you fix what it means to "use proof-based techniques". Where do I start to set up the theory, and where even do the boundaries of "the theory" lie? At the future point where you develop your own mathematical practice, any approach I pick will seem ridiculous to at least some of you. Maybe we update our objectives again.

1. Define the following words: "Understand", "formulate", "justify", "mathematical" "assertions", and what concatenations thereof mean. Develop a framework for their execution.

2. Figure out what each of us might want to call "differential and integral calculus" for the purposes of this course, then figure out a way to formulate it in a fashion best suited for the application of the techniques of part 1 towards our objectives (which we will determine).

Remark 1.0.1. At this point I began to question why I even endorsed formatting this as a university course; surely one should be discovering one's own calculus out in the world, not confining oneself to the narrow domestic walls of institutional practice. I quickly found myself working out the logistics of running a commune; for time reasons, I will not pursue this here.

Clearly, the above is much too ambitious in scope. The reality of our situation is that I'm going to have to intervene at some point and guide our investigations in this course. For the moment then, let's skip answering *why* you're in this course, and instead start by focusing on *how one might get here*.

1.1 Lecture 1: Issues of convergence [08/24/25]

Example 1.1.1 (Zeno's "Dichotomy" Paradox)**.** Let's say a runner (for this example, the mythical Atalanta the huntress) needs to cross a road of length L meters. Zeno argues first that to cover this length L, Atalanta must first cover half the length $\frac{L}{2}$. At the instant Atalanta reaches this point, there is exactly $L - \frac{L}{2} = \frac{L}{2}$ remaining of the path: and in order to clear this amount, she must first cover half of the remaining amount $\frac{L}{2}/2 = \frac{L}{4}$. Applying the argument once again at the point of covering $\frac{L}{2} + \frac{L}{4}$ of the remaining distance, we see that she must again cover at least $\frac{L}{8}$ more. Iterating this process, we see Atalanta must hit an infinite number of points at distances $p_k = \frac{L}{2} + \ldots + \frac{L}{2^k} < 1$ before she crosses the entire path.

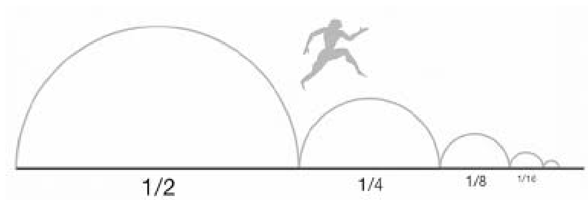

Figure 1: Zeno's Dichotomy Paradox: Atalanta running a length which is subdivided into half a length, then an additional quarter, then an additional eighth, and so on, following the text of the example.

Zeno now argues that this motion must have *at least* entailed performing an infinite number of tasks: and clearly, an infinite number of tasks must take an infinite time to perform. He concludes that the concept of "motion" itself is impossible.

Remark 1.1.2. This example has also been written in the chapter *Under Heaven* of the *Zhuangzi*, see `https://plato.stanford.edu/entries/school-names/paradoxes.html`.

Well, something about this is a little fishy. As I deliver the lesson, you will probably notice me pacing the length of the classroom and marvel at my ability to undertake such godly feats. Clearly, there is some worth in treating infinitary processes without outright dismissal. It looks a lot as though assuming she can perform infinite "supertasks"[1] nets us the ability to say the following:

Proposition 1.1.3. *The following two quantities are equal for any number L:*

$$L = \frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \ldots \tag{1.1.4}$$

Note exactly how strange Proposition 1.1.3 is: I am writing a mathematical equality between "quantities" when I'm not even sure that I can fully write down the second quantity! Let's try and **convince** ourselves of this fact arithmetically, making the assumptions below.

Assumption 1.1.5. Let's pretend that I can perform "infinite sums" (henceforth we'll call them *infinite series* or just *series*) which sum to fixed quantities.

[1] see here

However, at this point we're still doing finite processes: we haven't actually jumped into doing a supertask quite yet. How can we actually show something is L? Since we're dealing with a supertask, trying to "directly compute it" is a little out the window; we'll have to find L another way.

Observation 1.1.6. We can show something is L by showing that it satisfies an equation that only L satisfies.

1. If a given quantity x satisfies the equation $x - 1 = 0$ then $x = 1$, the same is true if it satisfies any of the equations $2x - 1 = 1$, or $3x - 2 = 1$, ...

2. Similarly, if a given equantity x satisfies the equation $x - L = 0$ then $x = L$, but similarly for $2x - L = x$, or in fact $2x - x = L$.

This last observation will be the operative one for the argument below.

Argument Idea 1 for Proposition 1.1.3. Let's try to do a little reduction with this sum. Set

$$S = \frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots$$

We have:

$$S = \frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots \implies 2 \cdot S = 2 \cdot \left(\frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots\right) \implies 2S = L + \frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots \quad (1.1.7)$$

Notice that we're acting as though we can distribute the multiplication by 2 across the "infinite addition" in the final step of (0.6). Substituting $S = \frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots$ again, we get:

$$2S = L + \frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots \implies 2S = L + S \implies 2S - S = L \quad (1.1.8)$$

and so $S = L$. □

At first blush, *Argument Idea 1* seems to us a completely valid use of standard arithmetic techniques. However, there are some problems with concluding the discussion above as we just did.

Discussion .

What are some obvious issues with doing this?

1. **The infinity issue:** Sometimes I can write down formal sums (eg $1 + 1 + 1\dots$) that look like they're growing without bound. In this case, I'll pretend I'm dealing with a formal symbol ∞ (which behaves how you might expect it to).

2. **Do the standard rules of arithmetic actually work?:**

Example 1.1.9 (Grandi's Series)**.** In this example we're going to examine the *associativity of infinite addition and subtraction*, namely if we can always add and subtract objects in any order in an infinite process.

Consider the infinite series[2] given by

$$1 - 1 + 1 - 1 + 1 + \dots \tag{1.1.10}$$

Discussion .

If you had to pick a value of this series, what would it be? Argue it for me. Here are some options:

1. 0
2. 1
3. $\frac{1}{2}$
4. Impossible to say

We will show below that every single one of these answers has a reasonable justification.

First, pretending that "infinite arithmetic" works as though ordinary arithmetic does, we may use the "associativity of infinite addition" to group Equation 1.1.10 in either of the following ways

$$(1-1) + (1-1) + (1-1) + \dots \qquad\qquad 1 + (-1+1) + (-1+1) + \dots$$

Notice that the first series looks like $0 + 0 + 0 + \dots$ which should clearly be 0, while the second looks like $1 + 0 + 0 + \dots$, which should clearly be 1. Maybe this eliminates the ability to deal with the "associativity of infinite addition", but the news gets a little worse: if we try to argue as in *Argument Idea 1*, by studying the equations this series satisfies, we run into the following conundrum

$$S = 1 - 1 + 1 - \dots \implies 1 - S = 1 - (1 - 1 + 1 - \dots) \implies 1 - S = 1 - 1 + 1 - 1 \dots = S$$

where in the last step we once again assumed multiplication "distributes over infinite sums". However, $1 - S = S \implies 2S = 1$ or that $S = \frac{1}{2}$.

You might here be tempted to answer 4; if this is you, you might also completely abandon hope for Assumption 1.1.5. A natural next step would be to denounce the blasphemies of Zeno's Paradox and other mathematical brain-teasers like it. On your homework this week we're going to learn why it's not wise to immediately discount it.

Observation 1.1.11. Let's reexamine the mathematical observation that Figure 1 was trying to get at. Let's start adding up the first few terms of (1.1.4) and try to reexpress them in

[2] Often referred to as Grandi's Series after Italian mathematician Guido Grandi, see this Wikipedia article for some history and approaches to its summation

terms of L.

$$\begin{aligned}
&\frac{L}{2} = \frac{L}{2} \\
&\frac{L}{2} + \frac{L}{4} = \frac{3L}{4} \\
&\frac{L}{2} + \frac{L}{4} + \frac{L}{8} = \frac{7L}{8} \\
&\frac{L}{2} + \frac{L}{4} + \frac{L}{8} + \dots
\end{aligned} \tag{1.1.12}$$

Clearly our big sum grows closer and closer to L as we sum more terms, which matches with our visual intuition; again, you can envision Atalanta as running along a big number line. Of course, these are still finite processes: we haven't jumped into the realm of trying to do a supertask quite yet. During our last attempt, we rectified this issue by just *assuming* that a quantity called an "infinite sum" existed and that some entity had done the supertask for us. To proceed, let's instead try to isolate the *specific* observation that actually seems true: namely, direct observation seems to tell us that every time we add another term in the infinite series above, we close the gap to L.

Definition 1.1.13. An infinite list of numbers $\{a_0, a_1, a_2, \dots\}$ is called a *sequence*; we will write this with the notation $\{a_i\}_{i\in\mathbb{N}}$[3], with $\{a_i\}_i$ as shorthand. The *nth partial sum* of the sequence $\{a_i\}_i$ is the following quantity

$$s_n := a_0 + \dots + a_n$$

given by summing up the terms of the sequence $\{a_i\}_i$ from 0 to n. Using summation notation, one has the equality

$$s_n := \sum_{i=0}^{n} a_i$$

Let's try to write down what our new, restricted goal should be, using the language above.

Question 1.1.13.

What is the strongest thing we can convincingly assert about the relation between L and the collection of finite processes s_n for each n?

Proposition 1.1.14. *For any small quantity $\epsilon > 0$, there is a big enough number N so that if $n > N$ then $L - s_n < \epsilon$. Equivalently, for any number $a < L$, there is an $N > 0$ so that $a < s_n < L$ for every $n > N$.*

Why? Essentially because it looks like $s_n = L - \frac{L}{2^{n+1}}$ for every n, and $\frac{L}{2^{n+1}}$ becomes very small as n becomes very big. □

Note that to make this argument, we still need to find a convincing principle to argue about the behavior of infinitely many finite processes. Here, we come up against the *principle of induction* as one of our main tools for doing so.

[3] we'll explain this notation later

1.2 Lecture 2: Sets, natural numbers, induction [08/26/25]

Discussion (Inference, proof, and "first principles").

1. What is a mathematical observation?
2. What forms a convincing argument?
3. What is the role of a proof?
4. Why do we start from "first principles"?

Recommended Reading

1. Introduction to logical statements and connectives: https://www.chilimath.com/lessons/introduction-to-number-theory/intro-to-truth-tables-statements-and-connectives/
2. Inversion and contrapositive: https://www.chilimath.com/lessons/introduction-to-number-theory/converse-inverse-and-contrapositive-of-conditional-stat

Let's start by defining our basic objects, the natural numbers. We will not try to show that they exist or construct a model of them, but we will take the approach that a good theory of the natural numbers will satisfy the following *axioms*.

Key Definition 1.2.0 (The Peano Axioms).

The set $\mathbb{N}$ of natural numbers is uniquely characterized by the following axioms:

1. There is an element $0 \in \mathbb{N}$.
2. For any element $n \in \mathbb{N}$, there is an element $S(n) \in \mathbb{N}$, called the *successor* of $\mathbb{N}$.
3. 0 is not the successor of any element in $\mathbb{N}$.
4. If $n, m \in \mathbb{N}$ satisfy $S(n) = S(m)$, then $n = m$.
5. The Principle of Induction, deferred until later.

We'll just assume there's some set $\mathbb{N}$ of objects which satisfies the axioms above, and not try to delve into actually building one from an axiomatic framework of what *sets* are. We will write

$$1 := S(0),\ 2 := S(S(0)),\ 3 := S(S(S(0))),\ 4 := S(S(S(S(0)))),\ \ldots$$

Question 1.2.0.

Is $3 = 0$? How might I say that it isn't?

Definition 1.2.1 (Informal)**.** A *recursive* definition is one which defines elements in a set in terms of other elements in a set. In particular, this is a definition that does not happen "all at once".

Let's look at some examples.

Definition 1.2.2 (Addition)**.** We define the operation of *addition*, denoted $+$, recursively as follows. For any $m \in \mathbb{N}$, we set

$$0 + m = m$$

Suppose that $n + m$ has been defined for $n, m \in \mathbb{N}$. Then we set

$$S(n + m) = S(n) + m.$$

Addition is a *binary operation*.

Recursive definitions that work in the way above, i.e., defining elements by starting at 0 and then defining them for all n using the successor function, are usually best reasoned about using induction.

Key Definition 1.2.2 (Principle of Induction)**.**

Let P be a property of natural numbers, such that

1. 0 satisfies P.
2. If n satisfies P, then $S(n)$ satisfies P.

Then all natural numbers satisfy P.

Notation 1.2.3. We write $P(n)$ to indicate the property P evaluated at a particular number n.

Remark 1.2.4. **General principle:** Recursion provides a method to perform computations about elements in a set by assuming the computations about certain subcollections of elements in a set; for example, it gives us a recipe to compute information (do a finite "task") about any natural number if we assume that we know the computations for the smaller numbers which build it (here, out of the successor function). Induction is the ingredient that lets us prove assertions about the infinite collection of finite tasks.

Let's try to prove a basic fact about addition, the fact that 0 is an *additive identity*, which is the following fact.

Proposition 1.2.5. *For every* $m \in \mathbb{N}$*,* $0 + m = m + 0$ *and both are equal to* m*.*

Proof. Let us prove this by induction. Let $P(n)$ refer to the property that $n + 0 = 0 + n$. In symbols, we write:

$$P(n) : 0 + n = n + 0.$$

We start with the case $n = 0$, known as the *base case*.

- **Case $n = 0$:** $P(0) : 0 + 0 = 0$, is true from the definition of addition.

We now show that if $P(n)$ is true, then $P(n + 1)$ is true. This is known as the *inductive case*.

- **Inductive case:** Suppose $P(n)\colon n+0=0+n$ is true. To show $P(S(n))$, note first that $S(n)+0=S(n+0)$ by the definition of addition. By the inductive case, $n+0=0+n=n$. It follows that $S(n)+0=S(n)$ and the latter is equal to $0+S(n)$ by the definition of addition. Thus, $P(S(n))$ is true, and we finish the claim.

□

The following is another basic property of addition that we will use.

Proposition 1.2.6 (Associativity of addition)**.** *For any elements $l, m, n \in \mathbb{N}$, one has*

$$l+(m+n)=(l+m)+n$$

Proof. Exercise 1.4.3. □

Using the above properties, let us try to prove another fundamental feature, the *commutativity* of addition.

Proposition 1.2.7 (Commutativity of addition)**.** *For any $n, m \in \mathbb{N}$, we have that $m+n=n+m$.*

In the proposition above, there are seemingly *two* variables to induct on; if you like, this is because the statement $m+n=n+m$ is a statement about infinitely many objects m and infinitely many objects n. That is, given a *fixed* m, there is the statement that every number $n \in \mathbb{N}$, satisfies $m+n=n+m$; this is a statement about infinitely many objects n. However, there is the statement that this commutation works for every number m that one could have picked.

This suggest that we need two inductions, one happening on m and then a "nested" one happening on n for the fixed m. Schematically, this looks like:

Figure 2: A schematic illustration of the two-variable induction described above

Proof of Proposition 1.2.7. We prove this using induction on m. For each fixed m, we will prove by induction on n that $m + n = n + m$.

Let $Q(m)$ be the property that for every $n \in \mathbb{N}$, one has $m + n = n + m$.

- **Base case $m = 0$:** We need to show $Q(0)$, i.e., for every $n \in \mathbb{N}$, $0 + n = n + 0$.

  We prove this by induction on n. Let $P(n)$ be the property that $0 + n = n + 0$.

  - **Base case $n = 0$:** $0 + 0 = 0 = 0 + 0$
  - **Inductive step:** Assume $P(n)$, i.e., $0 + n = n + 0$. We show $P(S(n))$:

    $$\begin{aligned} 0 + S(n) &= S(n) && \text{(by definition of addition)} \\ S(n) + 0 &= S(n) && \text{(by Proposition 1.2.5)} \end{aligned}$$

    Therefore $0 + S(n) = S(n) + 0$, so $P(S(n))$ holds.

  Thus $Q(0)$ is proven.

- **Inductive step:** Assume $Q(m)$, i.e., for every $n \in \mathbb{N}$, $m + n = n + m$. We show $Q(S(m))$, i.e., for every $n \in \mathbb{N}$, $S(m) + n = n + S(m)$.

  We prove this by induction on n. Let $R(n)$ be the property that $S(m) + n = n + S(m)$.

  - **Base case $n = 0$:**

    $$\begin{aligned} S(m) + 0 &= S(m) && \text{(by Proposition 1.2.5)} \\ 0 + S(m) &= S(m) && \text{(by definition of addition)} \end{aligned}$$

    Therefore $S(m) + 0 = 0 + S(m)$, so $R(0)$ holds.

- **Inductive step:** Assume $R(n)$, i.e., $S(m) + n = n + S(m)$. We show $R(S(n))$:

$$\begin{aligned} S(m) + S(n) &= S(S(m) + n) && \text{(by definition of addition)} \\ &= S(n + S(m)) && \text{(by inductive hypothesis } R(n)\text{)} \\ &= S(S(n + m)) && \text{(by definition of addition)} \\ &= S(S(m + n)) && \text{(by outer inductive hypothesis } Q(m)\text{)} \\ &= S(m + S(n)) && \text{(by definition of addition)} \\ &= S(n) + S(m) && \text{(by definition of addition)} \end{aligned}$$

  Therefore $S(m) + S(n) = S(n) + S(m)$, so $R(S(n))$ holds.

Thus $Q(S(m))$ is proven.

By induction on m, we have shown $Q(m)$ for all $m \in \mathbb{N}$, completing the proof. □

Let us make another recursive definition.

Definition 1.2.8 (Multiplication)**.** We define the operation of *multiplication*, denoted $\cdot$, recursively as follows. For any $m \in \mathbb{N}$, we set

$$0 \cdot m = 0.$$

Suppose $n \cdot m$ has been defined for some $n, m \in \mathbb{N}$. Then we set

$$S(n) \cdot m = (n \cdot m) + m.$$

Multiplication is a *binary operation.*

Proposition 1.2.9 (1 is a multiplicative identity)**.** *For every $m \in \mathbb{N}$, $m \cdot 1 = 1 \cdot m$ and both are equal to m.*

Definition 1.2.10 (Exponentiation)**.** We define the operation of *exponentiation*, recursively as follows. For any $m \in \mathbb{N}$, we set

$$m^0 = 1.$$

Suppose m^n has been defined for some $m, n \in \mathbb{N}$. Then we set

$$m^{S(n)} = m^n \cdot m$$

Exponentiation is, you guessed it, a binary operation.

You will prove properties of the above in this week's homework.

1.3 Lecture 3: The Rational Numbers and Convergence [09/03/25]

Learning Objectives

1. Define the integers and the rational numbers *axiomatically*, without *constructing* them.
2. Define what it means for a sequence to converge, and finally provide a resolution to Zeno's Dichotomy Paradox.
3. (Next time) Explain why the real numbers appear, and how to construct them rigorously.

Last time we defined the set $\mathbb{N}$ of the natural numbers following the Peano Axioms. Let's define some other basic objects. We will define the following *axiomatically*; this means that we will assert their basic properties without explicitly constructing a set which satisfies those properties.

Definition 1.3.1. The *integers* are a set, denoted $\mathbb{Z}$, satisfying the following properties.

1. The set $\mathbb{N}$ is a subset of $\mathbb{Z}$, i.e., $\mathbb{N} \subset \mathbb{Z}$. Moreover, addition and multiplication extend to binary operations on all of $\mathbb{Z}$.
2. For every $b \in \mathbb{Z}$, there exists an object $-b \in \mathbb{Z}$ satisfying $b + (-b) = 0$.
3. For every $b \in \mathbb{Z}$, either $b \in \mathbb{N}$ or $-b \in \mathbb{N}$.

The above properties force $\mathbb{Z}$ to look like $\mathbb{N} \cup \{-1, -2, -3, \ldots\}$ as a set. This usually takes the familiar picture of a "number line", where we can write elements of $\mathbb{Z}$ from left to right going from "less than" to "greater than". Let's try to make this latter picture mathematically precise:

Definition 1.3.2. An order relation on a set S is a relation $<$ which satisfies:

1. If $a, b \in S$, then either $a < b$, $b < a$, or $a = b$. However, only one of these may be true.
2. If $a < b$ and $b < c$ in S, then $a < c$.

In the homework, I introduced the order relation on $\mathbb{N}$. This was given by saying $m, n \in \mathbb{N}$ satisfy $m < n$ if $\exists r \in \mathbb{N}$ so that $m + r = n$. This isn't going to work to extend this order relation to the integers, however, as for any two numbers $a, b \in \mathbb{Z}$, there is always some r so that $a + r = b$; namely, we can take r to be $b - a$. We'll have to work a bit harder to introduce the order relation on $\mathbb{Z}$.

Example 1.3.3. $\mathbb{Z}$ carries an order relation $<$, given by $m < n$ if $n - m \in \mathbb{N}$. Checking that this is indeed an order relation is an exercise.

With the above in hand, we have the following pictures of the integers.

Proposition 1.3.4. *The order relation on $\mathbb{Z}$ satisfies the following properties.*

1. (Translation Invariance) *Given $m, n \in \mathbb{Z}$, $m < n$ if and only if for every $r \in \mathbb{Z}$ one has that $m + r < n + r$.*

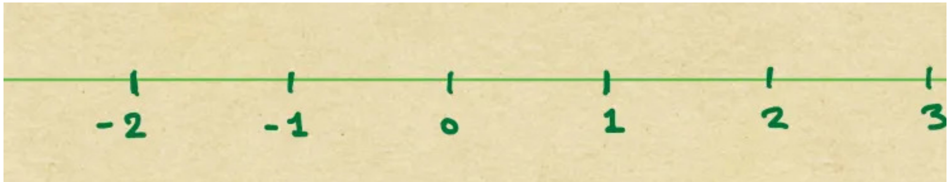

Figure 3: A depiction of the integers written left to right as {..., -2, -1, 0, 1, 2, ...} where terms on the left are less than terms on the right

2. ($\mathbb{N}$-Scaling Invariance) *Given* $m, n \in \mathbb{Z}$, $m < n$ *if and only if for every* $r \in \mathbb{N}$ *one has that* $m \cdot r < n \cdot r$.

Proof. Exercise. □

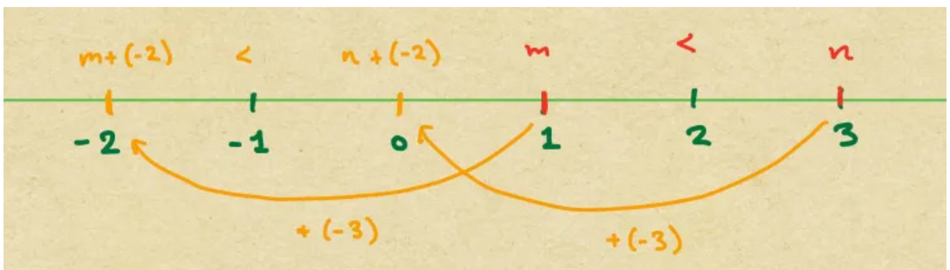

Figure 4: A pictorial depiction of the translation invariance of the integers

Let us now construct the rational numbers, once again doing so *axiomatically.*

Definition 1.3.5. The *rational numbers* are a set, denoted $\mathbb{Q}$, satisfying the following properties.

1. $\mathbb{Z}$ is a subset of $\mathbb{Q}$, in symbols $\mathbb{Z} \subseteq \mathbb{Q}$, and the addition and multiplication on $\mathbb{Z}$ extend to binary operations on $\mathbb{Q}$.

2. For any $m \in \mathbb{Q}$, if $m \neq 0$ then $\exists$ an object $\frac{1}{m} \in \mathbb{Q}$ such that $m \cdot \frac{1}{m} = 1$.

3. For any $m \in \mathbb{Q}$, if $m \neq 0$ then $\exists$ integers $a, b \in \mathbb{Z}$ such that $m = \frac{a}{b}$.

The rational numbers are also often depicted pictorially as a number line, interspersed as a "fine dust" between the integers. In order to do this, we'll need to figure out a way to extend the order on $\mathbb{Z}$ to an order on $\mathbb{Q}$; note that this is also not immediate, because we need to figure out how we want inverses (objects of the form $\frac{1}{a}$ for $a \in \mathbb{Z}$) to behave with respect to the ordering. We'll do this in two steps.

Definition 1.3.6. Let $a, b \in \mathbb{Z}$.

1. We say that $\frac{a}{b} > 0$ if $a \cdot b > 0$.

2. For $m, n \in \mathbb{Q}$, we say that $m < n$ if $n - m > 0$ using the definition of being > 0 given in the previous item.

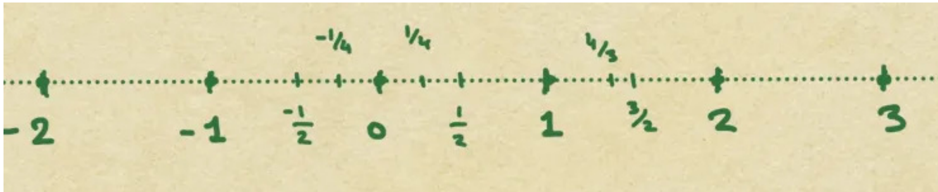

Figure 5: A depiction of the integers written left to right as {..., -2, -1, 0, 1, 2, ...} where terms on the left are less than terms on the right, along with rational numbers interspersed as a dotted line between the integer points. Some rational numbers are labelled, for example $0 < \frac{1}{2} < 1$.

With the definition of the order relation above, we obtain the familiar picture of $\mathbb{Q}$ as a "number line":

Proposition 1.3.7. *The order relation on $\mathbb{Q}$ satisfies the following properties.*

1. (Translation Invariance) *Given $m, n \in \mathbb{Q}$, $m < n$ if and only if for every $r \in \mathbb{Q}$ one has that $m + r < n + r$.*

2. (Positive Scaling Invariance) *Given $m, n \in \mathbb{Q}$, $m < n$ if and only if $r \in \mathbb{Q}$, if $r > 0$ then $m \cdot r < n \cdot r$.*

Proof. Exercise: will be on your next homework. □

The integers being a "fine dust" has a very particular interpretation, one which is exploited greatly in proving convergence of series/approximating objects by series. This particular technique was pioneered by Archimedes, and the following property carries his name:

Theorem 1.3.8 (Archimedean Property of the Rationals)**.** *For any $m, n \in \mathbb{Q}$, if $m < n$ then $\exists c \in \mathbb{Q}$ such that $m < c < n$.*

We will prove the above in the next lecture; however, it gives us the familiar picture of being able to find a rational number in between any two other rational numbers in the number line:

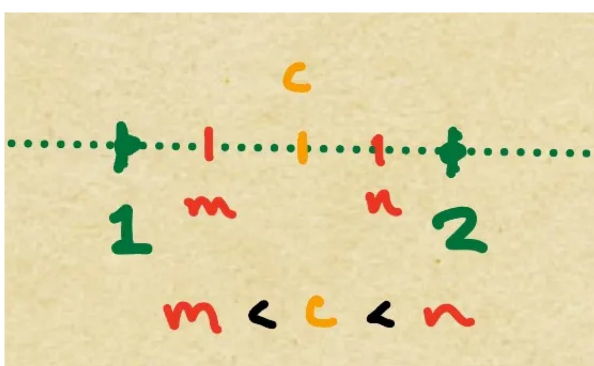

Figure 6: A pictorial depiction of the Archimedean property on the rational number line

1.3.1 Back to our scheduled program on the Dichotomy Paradox

Recall that our resolution to trying to compute the "supertask" involved in Zeno's Dichotomy Paradox (1.1.4) was to instead show that the following infinite collection of *finite* subtasks gets arbitrarily close to L as we compute larger and larger portions:

$$s_0 = \frac{1}{2}$$
$$s_1 = \frac{1}{2} + \frac{1}{4}$$
$$s_2 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$$
$$\ldots$$
$$s_n = \sum_{k=0}^{n} \frac{1}{2^{k+1}}$$

The concrete criterion that we settled on as a class was the following, which appeared as a Proposition 1.1.14:

Proposition 1.3.9. *For any small quantity $\epsilon > 0$, there is a big enough number N so that if $n > N$ then $L - s_n < \epsilon$. Equivalently, for any number $a < L$, there is an $N > 0$ so that $a < s_n < L$ for every $n > N$.*

More generally, the above is an example of the general phenomenon of *convergence* of sequences. In order to define this, first recall the *absolute value* function:

$$|-| : \mathbb{Q} \longrightarrow \{x \in \mathbb{Q} : x > 0\}$$

given by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \text{ (read as } \textit{greater than or equal to}\text{)} \\ -x & \text{if } x < 0. \end{cases}$$

The absolute value function encodes the "distance" on the number line between the point x and the point 0. Similarly, given two elements $m, n \in \mathbb{Q}$, the absolute value $|m - n|$ is a measure of the distance between m and n. With this notion in mind, we are ready to define the following.

Definition 1.3.10. We say a sequence of rational numbers $\{a_n\}$ *converges* to a if for every rational number $\epsilon > 0$, there exists some $N \in \mathbb{N}$ so that if $m > N$ then $|a - a_m| < \epsilon$.

In particular, for the sequence s_n of finite subtasks implicated in Zeno's Dichotomy Paradox, we are trying to show concretely that s_n *converges to* the length L.

Argument that the series s_n converges to L. Let's first start with an intermediate claim:

Claim: For every $n \in \mathbb{N}$, $L - s_n = \frac{L}{2^{n+1}}$.
Proof of claim. Note that we're making an argument about infinitely many tasks indexed over the naturals, and we have exactly one tool to use here: the principle of induction.

- **Base case:** $L - s_0 = L - \frac{L}{2} = \frac{L}{2}$ and so the base case holds.
- **Inductive case:** Suppose we know already that $L - s_n = \frac{L}{2^{n+1}}$. We need to show that given this fact, $L - s_{n+1} = \frac{L}{2^{n+1}}$. Let's make the following simplifications.

$$\begin{aligned} L - s_{n+1} &= L - (s_n + \frac{L}{2^{n+2}}) \quad \text{(Expanding out the summation notation)} \\ L - (s_n + \frac{L}{2^{n+2}}) &= L - s_n - \frac{L}{2^{n+2}} \quad \text{(Fact about taking negatives)} \\ L - s_n - \frac{L}{2^{n+2}} &= \frac{L}{2^{n+1}} - \frac{L}{2^{n+2}} \quad \text{(By the inductive hypothesis)} \\ \frac{L}{2^{n+1}} - \frac{L}{2^{n+2}} &= \frac{2L}{2^{n+2}} - \frac{L}{2^{n+2}} = \frac{1}{2^{n+2}}(2L - L) \\ &= \frac{L}{2^{n+2}} \end{aligned}$$

  which concludes the inductive case, showing that the case for n implies the case for $n+1$.

This concludes the proof of the intermediate claim above.

Let's now try to show the convergence directly. Namely, we need to show that given any $\epsilon > 0$, we need to show that there is an $N \in \mathbb{N}$ so that $m > N$ implies that $|L - s_m| < \epsilon$. Henceforth, let $\epsilon > 0$ be any choice. Using the claim above, we know that $|L - s_m| = \frac{L}{2^{m+1}}$. It follows that we just need to show that there is an N so that $\frac{L}{2^{m+1}} < \epsilon$ for every $m > N$.

Note that since ϵ is a rational number, we have that $\epsilon = \frac{a}{b}$ for some integers $a, b \in \mathbb{Z}$. Moreover, since $\epsilon > 0$, we have that $a \cdot b > 0$. This breaks us into two possible cases.

- **Case 1:** $a, b > 0$. In this case, we need to find an N so that $\frac{L}{2^{m+1}} < \frac{a}{b}$ for any $m > N$. We'll use the following steps.

  - For any $n \in \mathbb{N}$, using the scaling rule for the order relation, we know that $\frac{L}{2^{n+1}} < \frac{a}{b}$ if and only if $b \cdot \frac{L}{2^{n+1}} < b \cdot \frac{a}{b} = a$.
  - Similarly, we know that $b \cdot \frac{L}{2^{n+1}} < a$ if and only if $b \cdot L < 2^{n+1} \cdot a$.
  - Given a, b bigger than 0, there is always some $N \in \mathbb{N}$ so that $b \cdot L < 2^{N+1} \cdot a$. Furthermore, for any $m > N$, it must be the case that $b \cdot L << 2^{N+1} \cdot a < 2^{m+1} \cdot a$.

  It follows that in this case, there exists some N for which any $m > N$ satisfies $|L - s_m| = \frac{L}{s_m} < \frac{a}{b} = \epsilon$.

- **Case 2:** $a, b < 0$. Exercise!

As we have demonstrated the above for any particular choice of $\epsilon > 0$, we may conclude. □

1.4 Homework 1 (Due Wednesday, September 10th)

3 out of the following 5 problems will be randomly graded for correctness, the remainder will be graded for completeness. This homework is out of a possible **24 points**, with the graded problems worth 6 points and the ungraded problems worth 3 points each.

For the problems below, keep in mind the distinction between inference for yourself, and inference for others. Start by trying to come up with visualizations or explanations why such a thing might be true for yourself. Then try to convince the reader using the logical framework that we have agreed upon in class.

Finally, don't expect to be able to do everything in this homework immediately! I expect you to return to this once or twice with a group as we progress through the course over the next two weeks.

Exercise 1.4.1. *Prove that $n \neq S(S(S(n)))$[4] for every $n \in \mathbb{N}$ using the Peano axioms. (Notice that you are saying a thing about infinitely many objects, and you need to justify it for all of them).*

Exercise 1.4.2. *Prove DeMorgan's Laws for propositions: Let A and B be two mathematical propositions. Then*

1. $\neg(A \wedge B) \iff \neg A \vee \neg B$.
2. $\neg(A \vee B) \iff \neg A \wedge \neg B$.

For the above, I will accept a proof using truth tables. It may also help to unpack, in words, what the expressions above are trying to say. Now show it for sets: Let X be a set, and let A, B be subsets of X. Then

3. $(A \cup B)^c = A^c \cap B^c$.
4. $(A \cap B)^c = A^c \cup B^c$.

For 3. and 4., it might help to draw a picture to orient yourself. Then start your argument as follows: "Suppose an element $x \in (A \cup B)^c$. Then by definition, x satisfies..."

Exercise 1.4.3. *Prove the following properties of addition and multiplication, as defined recursively in the notes. You will only be allowed to use properties proved in the notes and the Peano axioms.*

1. ***(Associativity of addition)*** *For every choice of elements $n, m, l \in \mathbb{N}$ the following quantities are equal*
$$n + (m + l) = (n + m) + l.$$
Namely, the order in which I evaluate does not matter.

2. ***(Commutativity of multiplication)*** *For every choice of elements $n, m \in \mathbb{N}$ the following quantities are equal*
$$n \cdot m = m \cdot n.$$

[4]read: n is not equal to $S(S(S(n))))$

3. ***(Associativity of multiplication)*** *For every choice of elements* $n, m, l \in \mathbb{N}$ *the following quantities are equal*
$$n \cdot (m \cdot l) = (n \cdot m) \cdot l.$$
Namely, the order in which I evaluate does not matter.

4. ***(Additive law for exponentiation)*** *For every choice of elements* $n, m, l \in \mathbb{N}$*, the following quantities are equal*
$$n^{(m+l)} = n^m \cdot n^l.$$

Exercise 1.4.4. *Use mathematical induction to show the following expressions are true. For this problem you are allowed to use all standard facts about division and fractions, even though we have not defined division or proved those properties yet.*

1. $\sum_{k=0}^{n} k = 0 + 1 + 2 + \cdots + n = \frac{n\cdot(n+1)}{2}$.
2. $\sum_{k=0}^{n} k^2 = 0 + 1^2 + 2^2 + \cdots n^2 = \frac{n\cdot(n+1)\cdot(n+2)}{6}$.
3. $\sum_{k=0}^{n} 2k + 1 = (n+1)^2$.

Definition 1.4.5. Let us define an *order relation* on $\mathbb{N}$ as follows: We say $m \leq n$ in $\mathbb{N}$ if there exists $a \in \mathbb{N}$ so that
$$m + a = n.$$

Exercise 1.4.6 (Well-Ordering Principle)**.** *Show that any subset* $S \subseteq \mathbb{N}$ *contains a minimal element for the order relation* $\leq$*, namely there exists an element* $x \in S$ *so that* $x \leq s$ *for every other element* $s \in S$*.*

2 Convergent Sequences and the Real Numbers

2.1 Lecture 4: The Uniqueness of Limits [09/08/25]

Learning Objectives

1. Prove the Archimedean Property of the rational numbers.
2. Learn what a proof by contradiction is.
3. Practice arguing about convergent sequences, and use those techniques to prove arithmetic properties of convergent sequences.

Let's do a quick recap of the narrative arc of the course so far.

1. In Lecture 1, we investigated supertasks which involve addition and subtraction, and discovered that just "assuming someone could do the infinite addition" led to problems in deciding which supertasks had answers.
2. In Lecture 2, we decided to be very careful about defining our objects and our reasoning tools; we discussed how to define the natural numbers and its basic operations, in addition to how to reason about them using the principle of induction.
3. In Lecture 3, we defined the integers and the rationals, and finally came up with a framework to decide which supertasks were "doable"; these correspond to infinite sequences of finite subtasks which *converge* to the right answer.

This last point is our answer to Zeno; while not every big infinite supertask can be done, sometimes these consist of infinite sequences of finite subtasks which eventually converge to a well-defined answer. Let's recall these definitions below.

Definition 2.1.1. We say a sequence of rational numbers $\{a_n\}$ *converges* to a if for every rational number $\epsilon > 0$, there exists some $N \in \mathbb{N}$ so that if $m > N$ then $|a - a_m| < \epsilon$. In this case, we say that a is a *limit* of $\{a_n\}$.

Discussion

What is the negation of the statement above?

Definition 2.1.2. We say a sequence of rational numbers $\{a_n\}$ *diverges* if it does not converge to any rational number $a \in \mathbb{Q}$.

Exercise 2.1.3. *Convince yourself that the statement of divergence may be logically represented as*

$$\forall a \in \mathbb{Q}, \exists \epsilon > 0 \text{ such that } \forall N \in \mathbb{N}, \exists m > N \text{ with } |a - a_m| > \epsilon.$$

Now try to directly show the statement above for Grandi's Series $s_n = \sum_{k=0}^{n} (-1)^n$. This should square with our discussion of this being a "heretical" supertask.

Our first goal is to show that "doable" supertasks have well-defined answers. Following our idea above, this boils down to the following statement.

Theorem 2.1.4 (Uniqueness of Limits)**.** *Given a convergent sequence of rational numbers $\{a_n\}$ with limits $a, b \in \mathbb{Q}$, it must be the case that $a = b$. That is, if a sequence of rational numbers converges, it converges to only one limit.*

We're going to need to develop some tools to show this statement. Our first tool comes from a statement that we punted from the last lecture:

Theorem 2.1.5 (Archimedean Property of the Rationals)**.** *For any $m, n \in \mathbb{Q}$, if $m < n$ then $\exists c \in \mathbb{Q}$ such that $m < c < n$.*

A pictorial depiction of the above is in Figure 6.

Proof. We can just use Zeno's trick: given some choice of $m < n$, then we can try to find a point exactly halfway between them. Recall that the distance between m and n is given by the function $|m - n|$. It should follow that the midpoint between the two is given by $m + \frac{|m-n|}{2}$. Let's try to show this actually satisfies the property we care about:

1. **Claim 1:** $m < m + \frac{|m-n|}{2}$: Note that $0 < |m - n|$ by construction, and so $0 < \frac{|m-n|}{2}$ by multiplying $\frac{1}{2}$ to both sides. The claim follows by adding m to both sides.

2. **Claim 2:** $m + \frac{|m-n|}{2} < n$: First, note that $n = m + |m - n|$ (You should check this!). Thus, this claim is equivalent to showing that $m + \frac{|m-n|}{2} < m + |m - n|$. Subtracting m from both sides, we are left to show that $\frac{|m-n|}{2} < |m - n|$, which you can show from the definition by multiplying both sides by 2 and directly using the definition of the order relation on the natural numbers.

Thus, we have that $c = \frac{|m-n|}{2}$ satisfies the property. □

In a strong sense, the Archimedean property says that there is no such thing as an "infinitely small number" in the rationals: between 0 and any positive rational m, you can always find a smaller rational number between the two. This realization gives us our first important tool for this section:

Corollary 2.1.6. *For every $m, n \in \mathbb{Q}$, if for every $\epsilon > 0$ one has that $|m - n| < \epsilon$, then $m = n$.*

To show this, we'll have to introduce a new tool, known as a *proof by contradiction.*

Key Definition 2.1.7 (Proof by Contradiction)**.**

Given mathematical statements A and B, the following statement is always true:

$$((A \implies B) \land (A \implies \neg B)) \implies A$$

Said in words, "if assuming A showed that B was true and and it showed that B was not true, then one could not have assumed A was true. Thus, A must have been false.

Proofs by contradiction will be a vital tool in this course, and the argument below will show how they work.

Proof of Corollary 2.1.6. Assume, towards a contradiction, that there is some pair $m, n \in \mathbb{Q}$ satisfying $\forall \epsilon > 0, |m - n| < \epsilon$ and that $m \neq n$. We will demonstrate that this assumption leads to a contradiction.

Claim: There is some c for which $0 < c < |m - n|$.

Proof of claim: Because $m \neq n$, it must be the case that $m > n$ or $m < n$, using the fact that $<$ is an order relation on $\mathbb{Q}$. Thus, either $m - n > 0$ or $n - m > 0$, and thus $|m - n| > 0$. However, by the Archimedean Property of the rationals, there must be some $c \in \mathbb{Q}$ satisfying $0 < c < |m - n|$, yielding the claim.

However, we had already assumed that $\forall \epsilon > 0, |m - n| < \epsilon$, and the claim above proves the negation of this sentence. By contradiction, it could not be the case that $\forall \epsilon > 0, |m - n| < \epsilon$ and $m \neq n$. This concludes the proof. $\square$

Sketch

This last step might seem a little suspect to the discerning reader. Let's try to elucidate exactly what we did, in logical notation. Let

$$P : \text{For every } \epsilon > 0,\ |m - n| < \epsilon$$
$$Q : m = n$$

We are trying to show $P \implies Q$, which is always the same as $\neg(P \wedge \neg Q)$: you can check this with a truth table, and sound it out in words to make sense of this. To proceed with proof by contradiction to show $\neg(P \wedge \neg Q)$, let's assume, towards a contradiction, that $(P \wedge \neg Q)$ is true.

1. First, it's always the case that $(P \wedge \neg Q) \implies P$; namely, if P and *anything* is true, then P is true, so this is essentially by definition (the mathematical term for this is *tautological*).

2. Second, over the course of the proof, we showed the claim that $Q \implies \neg P$, and so $(P \wedge Q) \implies \neg P$.

Thus, we've shown that $((P \wedge \neg Q) \implies P) \wedge ((P \wedge \neg Q) \implies \neg P)$. Applying the proof by contradiction, it must be the case that $\neg(P \wedge \neg Q)$, which is what we wanted to show.

Exercise 2.1.8. *Go through the proof above and annotate it by where we've done each step.*

Now we introduce our second main tool, which is going to be a major staple of the course.

Key Theorem 2.1.9 (Triangle Inequality (Or the Importance of Good Public Transit))**.**

For any $a, b, c \in \mathbb{Q}$, one has the relation

$$|a - c| \leq |a - b| + |b - c|$$

Proof. Our first claim is that for any $m, l \in \mathbb{Q}$, $|l| + |m| \geq |l + m|$. The proof is left as an exercise to the reader, which is short for "I don't want to do it". However, you can now set $m = a - b$ and $l = b - c$, and you're done! $\square$

With these in mind, we may now proceed to the proof of Theorem 2.1.4.

Proof of Theorem 2.1.4. Let $\epsilon > 0$ be any choice. The following two statements we have by assumption:

- $\exists N_1 \in \mathbb{N}$ so that $\forall m \in \mathbb{N}$, if $m > N_1$ then $|a - a_m| < \frac{\epsilon}{2}$.
- $\exists N_2 \in \forall \epsilon >, \mathbb{N}$ so that $\forall m \in \mathbb{N}$, if $m > N$ then $|b - a_m| < \frac{\epsilon}{2}$.

Set $N = \max(N_1, N_2)$, where the notation "max$(-,-)$"" means the largest of the inputs. Then the above two statements collectively imply:

- $\forall m \in \mathbb{N}$, if $m > N$ then $|a - a_m| < \frac{\epsilon}{2}$ **and** $|b - a_m| < \frac{\epsilon}{2}$.

Now, let's use the triangle inequality:

$$|a - b| \leq |a - a_m| + |a_m - b| = |a - a_m| + |b - a_m| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

In particular, reading off the start and the end, we've shown that $|a - b| < \epsilon$. However, $\epsilon > 0$ was just some random choice, and the above argument should work for *any* choice of $\epsilon > 0$. Really, what we've shown is:

- $\forall \epsilon > 0$, $|a - b| < \epsilon$.

And hence $a = b$ by our argument above. $\square$

Sketch

What's our strategy upstairs? We're trying to show that if:

- a_m gets arbitrarily close to a
- a_m also gets arbitrarily close to b

then by triangle inequality we can show that a and b must have been arbitrarily close together. So close, in fact, that they must have been equal.

Exercise 2.1.10. *Try to annotate the proof of Theorem 2.1.4 in logical notation, similar to how we annotated Corollary 2.1.6.*

In this, we see the basic flow of a convergence argument: we're trying to show some statement $\forall \epsilon > 0$, so we start our proofs by letting $\epsilon > 0$ be some arbitrary choice. The rest of our argument is now not allowed to make any references to the specific properties of ϵ used, outside of any reductions to specific choices of ϵ. In this way, we show that the argument works for every possible choice. We'll see this technique a lot in the next lecture.

2.2 Lecture 5: The Algebraic Limit Theorem [09/10/25]

Learning Objectives

Prove the Algebraic Limit Theorem.

The aim of this class is to prove the following.

Key Theorem 2.2.1 (Algebraic Limit Theorem)**.**

Given convergent sequences a_n and b_n with limits a and b, the following are true:

1. Given any c, $\lim_{n\mapsto\infty}(c \cdot a_n) = c \cdot a$
2. $\lim_{n\mapsto\infty}(a_n + b_n) = a + b$
3. $\lim_{n\mapsto\infty}(a_n \cdot b_n) = a \cdot b$
4. $\lim_{n\mapsto\infty}\left(\frac{a_n}{b_n}\right) = \frac{a}{b}$.

Proof. [Abb15, §2.3 The Algebraic and Order Limit Theorems]. □

2.3 Lecture 6: Cauchy Sequences, Completeness, and the Real Numbers [09/15/25]

Learning Objectives

- Explain the need for the real numbers
- Define what the real numbers are, using a modification of the Axiom of Completeness
- Prove that least upper bounds exist in the real numbers.
- Prove that bounded monotone sequences converge in the real numbers.

This week we'll finally expand the scope of what we're discussing from sequences of rational numbers to sequences of *real numbers*. Before we do that, let's take a quick second to explain why one might want to think about them, using a very classical argument.

Proposition 2.3.1 (There is no rational square root of 2)**.** *There does not exist an $a \in \mathbb{Q}$ satisfying $a^2 = 2$. Equivalently, for every $a \in \mathbb{Q}$, $a^2 \neq 2$.*

Proof. Assume, towards a contradiction, that such an a exists. Then $a \in \mathbb{Q}$, so there are integers $m, n \in \mathbb{Z}$ so that $a = \frac{m}{n}$, and we may furthermore assume that m and n have no common factors by reducing to lowest terms.

By assumption, $(\frac{m}{n})^2 = \frac{m^2}{n^2} = 2$, so $m^2 = 2n^2$. Since m and n have no common factors, we have that 2 divides m^2 and does not divide m. However, since 2 is prime, it must divide m itself, and so $2^2 = 4$ must divide m^2. However: $m^2 = 2n^2$, so we have shown that 4 must divide

$2n^2$ and hence 2 must divide n. This is a contradiction: m and n were selected to have no common factors!

Thus, $a \in \mathbb{Q}$ satisfying $a^2 = 2$ could not have existed. In other words, there is no rational square root of 2. □

However, there's reason for us to want to write down something like a $\sqrt{2}$. For one, there's an easy visual representation of it using the Pythagorean theorem:

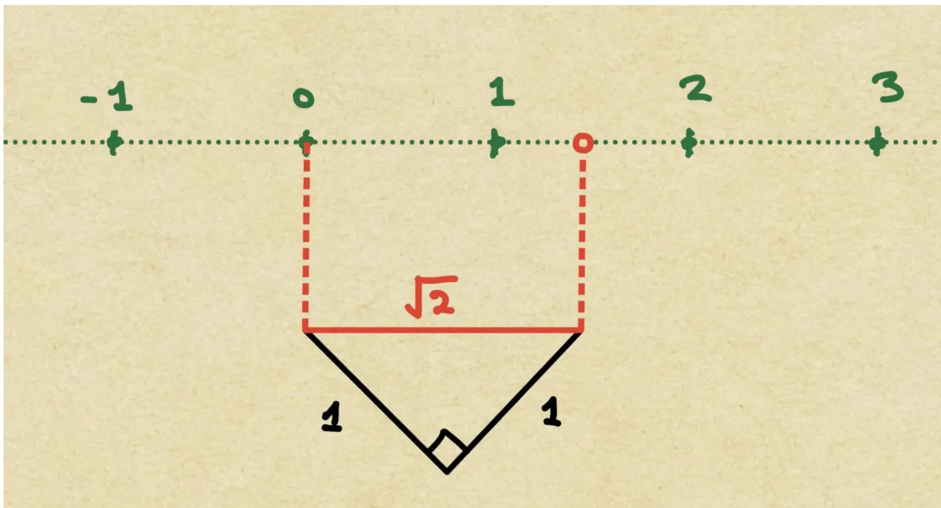

Figure 7: A depiction of the square root of 2 appearing as the hypotenuse of a right triangle with sides of length 1, along with a depiction of the square root of 2 as a "missing" point in the rational number line

For another, the picture above suggests that it's possible to uniformly approximate $\sqrt{2}$ by a sequence of rational numbers. You'll see an example of this in Homework 2 this week.

Proposition 2.3.2. *There is a sequence of positive rational numbers* $\{a_n\}$ *which does not converge to a rational number, but which satisfies* $\lim_{n\to\infty}\{a_n^2\} = 2$.

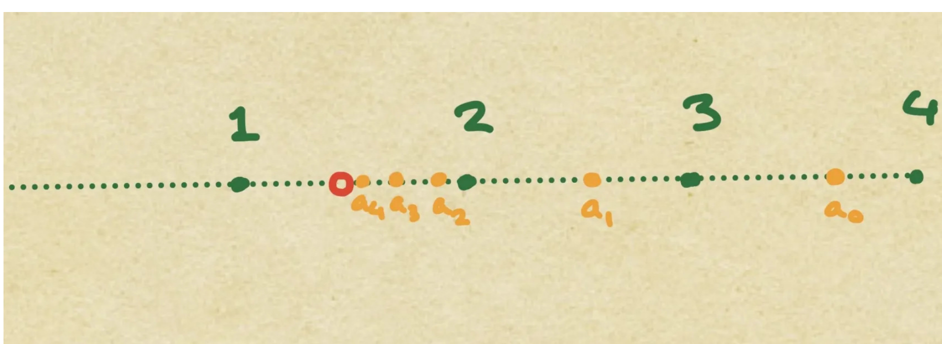

Figure 8: A depiction of a sequence which appears to be converging to the square root of 2.

Such a sequence looks as though it's converging to *something*, but clearly whatever it converges to cannot be a rational number, since the Algebraic Limit Theorem would imply that whatever it converged to would be a square root of 2. Instead, what it converges to is a *real number*. But in order to talk about real numbers, we need to characterize the kind of sequence that this is.

Definition 2.3.3. We say that a sequence $\{a_n\}$ is a *Cauchy sequence* if it satisfies the following: $\forall \epsilon > 0$, there exists an $N \in \mathbb{N}$ so that for every $m, n > N$, we have that $|a_m - a_n| < \epsilon$.

Take a second to internalize this definition: it says that the elements of a sequence eventually all get very close to *each other*, instead of making reference to another element, a limit, that it gets closer to. The sequence we just constructed turns out to be an example of a Cauchy sequence: this is instructive to check.

Exercise 2.3.4. *Prove that any convergent sequence is a Cauchy sequence.*

We are now ready to discuss the real numbers.

Key Definition 2.3.5 The Reals.

The *real numbers* are a set, denoted $\mathbb{R}$, satisfying the following properties.

1. The rational numbers are contained in $\mathbb{R}$, i.e. $\mathbb{Q} \subseteq \mathbb{R}$.
2. Addition and multiplication extend to binary operations on $\mathbb{R}$, while the order relation also extends to $\mathbb{R}$. In particular, the order relation on $\mathbb{Q}$ extends to $\mathbb{R}$, and so does the absolute value function.
3. Every real number is the limit of a sequence of rational numbers.
4. (Axiom of Completeness) Every Cauchy sequence converges in $\mathbb{R}$.

The rational numbers satisfy every one of the conditions above up until the Axiom of Completeness: it is this last feature which makes the real numbers useful. Let's first note that convergent sequences have the same nice behaviour in the real numbers that we're used to in the rationals:

Theorem 2.3.6. *The following results hold for the real numbers.*

1. *If $a, b \in \mathbb{R}$ satisfy $|a - b| < \epsilon$ for every $\epsilon > 0$, then $a = b$.*
2. *The Triangle Inequality.*
3. *Convergent sequences have unique limits.*
4. *The Algebraic Limit Theorem for convergent sequences.*

Proof. Part 1. of the above follows from the density of the rationals in $\mathbb{R}$, see Problem 2 of Homework 2. The remaining parts have the exact same proofs as their analogues in $\mathbb{Q}$. In particular, convergent sequences have the exact same nice behaviour as usual. $\square$

Definition 2.3.7. Let's say a sequence $\{a_n\}$ is monotone increasing if $\forall n \in \mathbb{N}$, $a_n < a_{n+1}$, and monotone decreasing if $\forall n \in \mathbb{N}$, $a_{n+1} < a_n$. We say a sequence is simply monotone if it is either monotone increasing or descreasing.

Definition 2.3.8. Recall that a sequence $\{a_n\}$ is *bounded* if there is a positive real number $M \in \mathbb{R}$ such that $0 \leq |a_n| \leq M$ for every $M \in \mathbb{R}$.

We now note the following new behaviour of convergent sequences:

Theorem 2.3.9 (Monotone Convergence Theorem)**.** *If a sequence of real numbers $\{a_n\}$ is bounded and monotone, then it must converge.*

Proof. By the Axiom of Completeness and Exercise 2.3.4, it suffices to show that the bounded monotone sequence $\{a_n\}$ is a Cauchy sequence. We'll first handle the case when $\{a_n\}$ is monotone *increasing*. In the other case, we can just multiply by -1 to get something which is increasing, find a limit of the resultant sequence, and then use the Algebraic Limit Theorem to multiply by -1 again.

Now assume towards a contradiction that $\{a_n\}$ isn't a Cauchy sequence. Then there exists some $\epsilon > 0$ such that for any choice of $N \in \mathbb{N}$, there are $m > n > N$ satisfying $|a_m - a_n| > \epsilon$. Since we assumed the sequence was monotone increasing, we learn that $a_n + \epsilon < a_m$.

Thus, for a_0 there is an a_{n_1} with $a_0 + \epsilon < a_{n_1}$, and similarly for every a_{n_i} there is an $a_{n_{i+1}}$ satisfying $a_{n_i} + \epsilon < a_{n_{i+1}}$. In this way, we can construct an ascending sequence

$$a_0 < a_{n_1} < a_{n_2} < \ldots < a_{n_k} < \ldots$$

where $a_0 + k \cdot \epsilon < a_{n_k}$. Since $\epsilon > 0$, we may eventually choose k to be so large that $a_0 + k \cdot \epsilon$ is positive and also $a_0 + k \cdot \epsilon > M$; however, this is a contradiction, since we constructed a_{n_k} in the sequence satisfying $a_{n_k} > a_0 + k \cdot \epsilon$, but by the assumption of boundedness, every element of the sequence had absolute value less than M.

It follows that the sequence must be Cauchy, and hence that it converges. □

Exercise 2.3.10. *Annotate the logic of the proof by contradiction above. Then try to find a least possible value for k in the last part of the argument.*

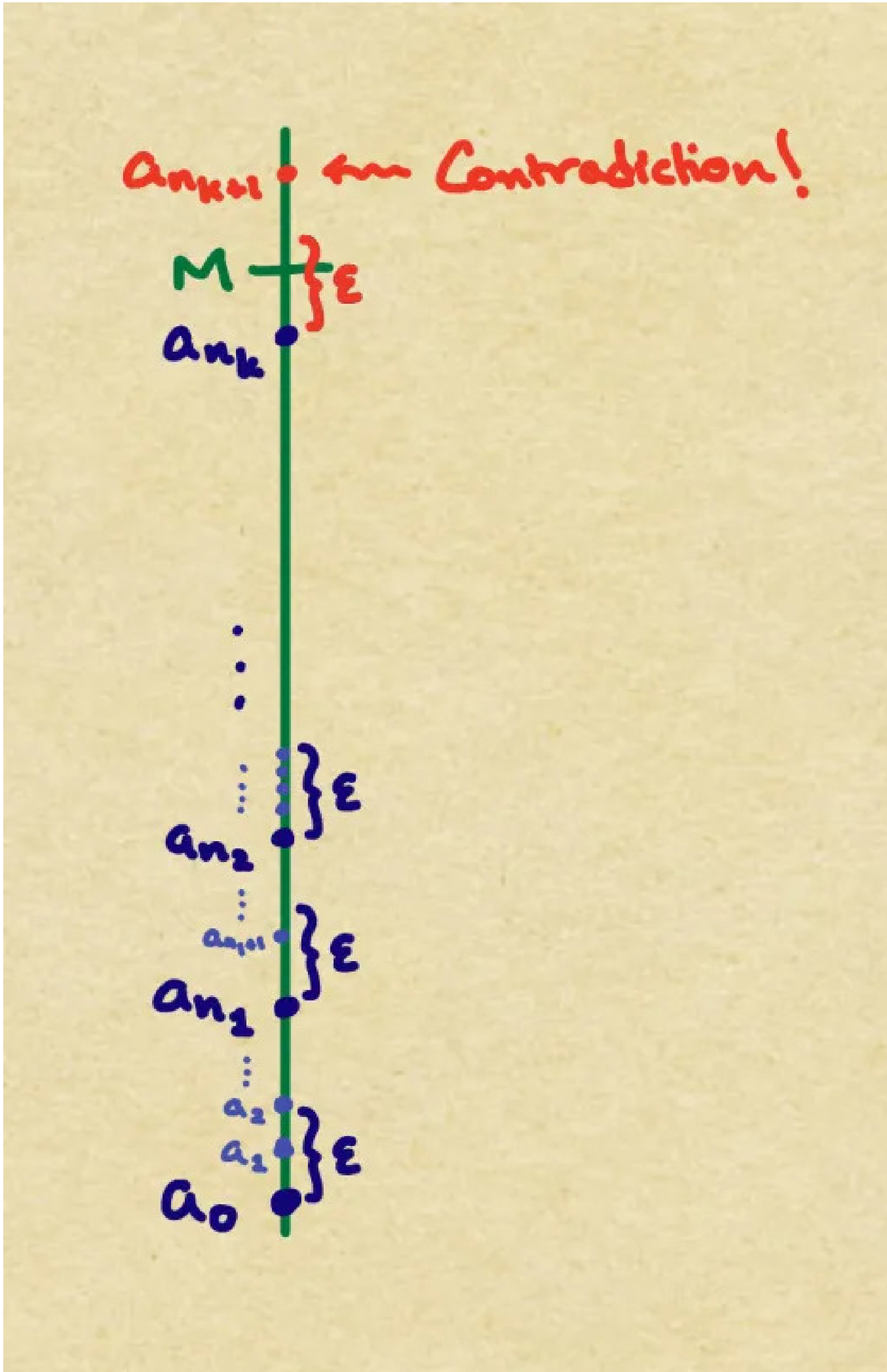

Figure 9: A graphical depiction of the Monotone Convergence Theorem and its disproof.

2.4 Lecture 7: Bolzano-Weierstrass and Least Upper Bounds [09/17/25]

Let's examine some consequences of the monotone convergence theorem.

Proposition 2.4.1 (Direct Comparison Test)**.** *Let a_n, b_n be sequences of* positive *real numbers. Assume moreover that $\forall v \in \mathbb{N}, a_n \leq b_n$. Then we have the following.*

1. *If the series $\sum_{k=0}^{\infty} b_k$ converges (i.e., if the sequence of partial sums $s_n = \sum_{k=0}^{n} b_k$ converges), then the series $\sum_{k=0}^{\infty} a_k$ converges.*
2. *If the series $\sum_{k=0}^{\infty} a_k$ diverges (i.e., if the sequence of partial sums $t_n = \sum_{k=0}^{n} a_k$ diverges), then the series $\sum_{k=0}^{\infty} b_k$ converges.*

Proof. For part 1, note that since every element of the sequence $\{a_k\}$ is positive, the sequence $t_n = \sum_{k=0}^{n} a_k$ is monotone increasing, and similarly the sequence $s_n = \sum_{k=0}^{n} b_k$ is monotone increasing. We realise the following two facts.

1. Set $s := \lim_{n\to\infty} s_n = \lim_{n\to\infty}(\sum_{k=0}^{n} b_k)$. It is not hard to see that $\forall k, s_k \leq s$; the sequence s_k is monotone increasing, and thus the limit is always greater than any individual finite sum (Check this formally!).
2. Since we assumed that $\forall k, a_k \leq b_k$, we have that $\forall n, t_n \leq s_n$ and thus t_n is less than s.

Since t_n is bounded above by s and monotone, we may apply Theorem 2.3.9 and conclude.

Part 2 is exactly the contrapositive of part 1, and hence we've already shown it. (Verify this for yourself!). □

Example 2.4.2 (Harmonic Series)**.** Let's use the above to show that the following series diverges:

$$\{a_k\} = \frac{1}{k}\ ,\ s_n = \sum_{k=1}^{n} a_k = \sum_{k=0}^{n} \frac{1}{k}$$

which is known as the *harmonic series*. We'll do this by direct comparison: for any positive number n, the number $2^{\lceil \log_2 n \rceil}$ represents the smallest natural number power of 2 that is still larger than n.[5] In particular, we have the relation $n \leq 2^{\lceil \log_2 n \rceil}$. Let's define the sequence

$$\{b_n\} := \frac{1}{2^{\lceil \log_2 n \rceil}}$$

which, when written out explicitly starting from $n = 1$, is the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \ldots\}$.

By the discussion above, it's clear that for every n, $b_n \leq a_n$ because $\frac{1}{2^{\lceil \log_2 n \rceil}} \leq \frac{1}{n}$. In order to show that $\sum_{k=0}^{n} \frac{1}{n}$ diverges, it will be sufficient by the Direct Comparison Test to show that $t_n := \sum_{k=0}^{n} \frac{1}{2^{\lceil \log_2 n \rceil}}$ diverges.

[5] The notation $\rceil x \lceil$ is the *ceiling* function, and is the function sending a real number x to the smallest integer that is greater than x; the identification in the sentence above should follow.

Why is this latter fact true? Well, notice the following pattern:

$$\begin{aligned}
t_1 &= 1 + \frac{1}{2} \\
t_4 &= 1 + \frac{1}{2} + \frac{1}{2} = 1 + 1 = 2 \\
t_8 &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1 + 1 + 1 = 3 \\
&\ldots
\end{aligned} \tag{2.4.3}$$

It seems a lot like there's a pattern emerging. Here's a guess for what it might be, and I'm going to throw it below for you to do.

Exercise 2.4.4. *Prove, by induction, that* $t_{2^n} = n$. *Deduce that the sequence of partial sums* $t_n := \sum_{k=0}^{n} \frac{1}{2^{\lceil \log_2 n \rceil}}$ *diverges. Applying the direct comparison test, conclude that the Harmonic Series diverges.*

While the Monotone Convergence Theorem is a useful fact, it turns out that one can get away without monotonicity as long as one is willing to work with a weakened version of convergence instead. For this, we'll need to introduce the notion of a *subsequence*.

Definition 2.4.5. Let $\{a_n\}$ be a sequence of real numbers. A *subsequence* of $\{a_n\}$ is a sequence of the form $\{a_{n_1}, a_{n_2}, a_{n_3}, \ldots\}$ where $\{n_1 < n_2 < n_3 < \ldots.\}$ is an infinite subset of the naturals.

Discussion

Write an example of a sequence which:

1. Does not converge to anything, but has a subsequence that does.

2. Is not monotone increasing/decreasing, but has a subsequence that is.

Bonus points if you can do one that *isn't* Grandi's series.

Our last main result for today is the following fact.

Theorem 2.4.6 (Bolzano-Weierstrass)**.** *Any bounded sequence has a convergent subsequence.*

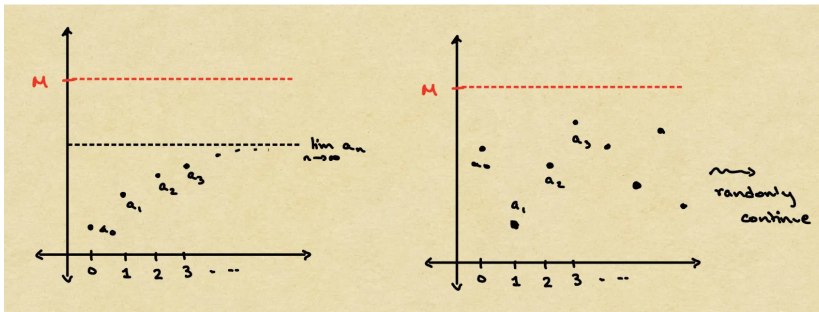

Figure 10: A depiction of a bounded monotone sequence converging on the left hand side, and on the right hand side a depiction of a bounded sequence that does not converge, but wanders around a bit.

Our strategy will be to show the following instead:

Proposition 2.4.7. *Any infinite sequence has a monotone subsequence.*

Remark 2.4.8. How might we prove a statement like the above? Given an arbitrary infinite sequence, it should be unlikely that we are able to manually construct a monotone subsequence: we are unable to assume anything that lets us construct such a thing! We'll instead go about by trying to prove this by contradiction.

Proof. Assume, towards a contradiction, that there is an infinite sequence a_n without any monotone subsequences. In particular, it has no monotone *decreasing* subsequences.

> **Claim:** There is a $k \in \mathbb{N}$ so that $\forall l > k$, $a_k \geq a_l$. In other words, a_k is a minimum. Let's call such an element a "valley" (or if you like, a minimum).
>
> *Proof of claim:* By a nested contradiction: suppose that for every a_l, there was an $m > l$ with $a_l \geq a_m$. Then let us construct a monotone decreasing subsequence as follows: set $a_{n_0} = a_0$, set a_{n_1} to be some element such that $n_1 > 0$ and $a_0 \geq a_{n_1}$, set a_{n_2} to be some element with $n_2 > n_1$ and $a_{n_1} \geq a_{n_2}$, and so on. Thus, we have produced a monotone decreasing subsequence
> $$a_0 = a_{n_0} \geq a_{n_1} \geq a_{n_2} \geq a_{n_3} \geq \ldots$$
> which contradicts our assumption that a_n has no monotone decreasing subsequences. Thus, the claim must follow.

Thus, we have found a valley, i.e., an element a_k for which $a_k \geq a_l$ for any $l > k$. I've included a picture below.

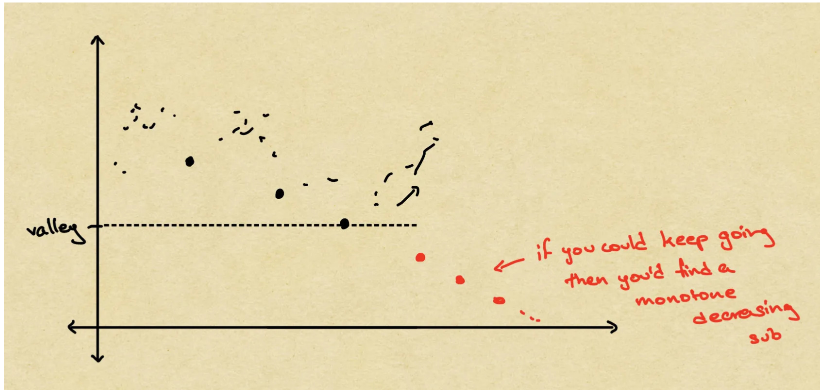

Figure 11: A depiction of a valley, along with the proof by contradiction that one must exist.

Claim: For any valley a_l, there is another valley a_m with $m > l$ and $a_l \leq a_m$.

Proof of claim: Since the sequence $\{a_n\}$ has no monotone decreasing subsequences, we have that the subsequence $\{a_{k+1}, a_{k+2}, a_{k+3}...\}$ which starts from $k+1$ must also have no monotone subsequences. The same proof as above must produce an valley a_m with $m > l$. Since a_l is a valley we have that $a_l \leq a_m$. This shows the following claim.

Set $a_{m_0} = a_l$ the first valley that we found. Set a_{m_1} to be any valley with $m_1 > m_0$. Similarly, set a_{m_k} to be any valley with $m_k > m_{k-1}$. It follows that we've produced an infinite subsequence

$$a_{m_0} \leq a_{m_1} \leq a_{m_2} \leq a_{m_3} \leq \ldots$$

which is monotone increasing by construction. Contradiction! $\{a_n\}$ was assumed to have no monotone subsequences at all; it follows that no such sequence $\{a_n\}$ can exist.

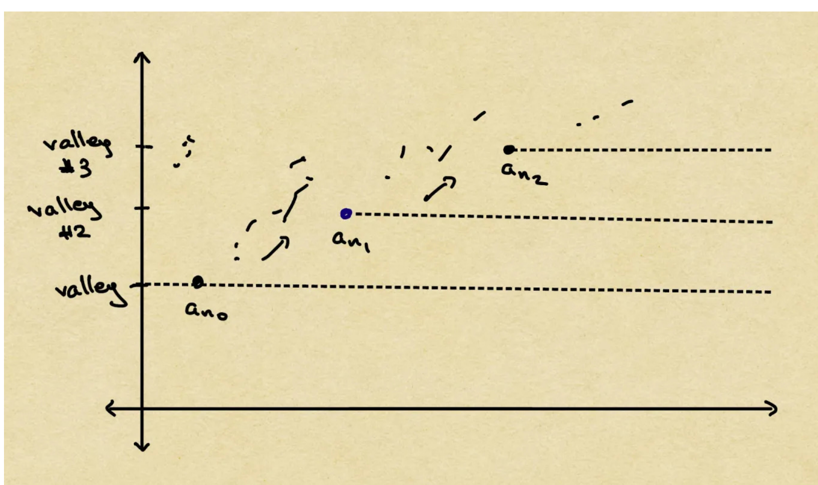

Figure 12: A depiction of a monotone increasing subsequence of valleys.

□

The Bolzano-Weierstrass Theorem now follows:

Proof of Theorem 2.4.6. By combining Proposition 2.4.7 and Theorem 2.3.9, we have that any bounded subsequence has a bounded monotone subsequence which is thus convergent. $\square$

2.5 Homework 2 (Due Friday, September 26th)

3 out of the following 6 exercises graded for correctness, the remainder will be graded for completeness. I have marked the graded ones with an asterisk $(*)$. This homework is out of a possible **24 points**, with the graded exercises worth 6 points and the ungraded exercises worth 2 points each.

Finally, don't expect to be able to do everything in this homework immediately! I expect you to return to this once or twice with a group as we progress through the course over the next two weeks.

The following exercise should be doable with what you know as of Lecture 5, 09/10/25:

Exercise 2.5.1. ***(Cesaro Means)*** *Show that if $\{x_n\}$ is a convergent sequence, then the sequence given by the averages*

$$y_n = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

is convergent, and converges to the same limit.

The following exercises will require us to talk about the real numbers, Cauchy Sequences, and the Monotone Convergence Theorem, which will be the subject of Lecture 6:

Exercise 2.5.2 $(*)$**.** [Density of the Rationals] *Using the axiomatic characterization of the real numbers, show that for any real numbers $a \in \mathbb{R}$ and $b \in \mathbb{R}$ that there is a rational number r satisfying $a < r < b$.*

Definition 2.5.3. Let $\{a_n\}$ be a sequence of real numbers. The sequence $s_n = \sum_{k=0}^{n} a_k$ is referred to as its *associated sequence of partial sums*. For example, the sequence $a_n = \frac{1}{n}$ has associated sequence of partial sums $s_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$.

As shorthand, we usually write $\sum_{k=0}^{\infty} a_n$ to refer to the limit of the sequence s_n, and refer to this as its *infinite series*: note that this is consistent with us viewing this as a supertask.

The following two items are referred to as the *divergence test* and the *absolute convergence test* for series respectively. These are easy ways of checking whether or not a given series is "summable", i.e., whether the infinite summation supertask has an answer, without actually computing what the answer is.

Exercise 2.5.4 $(*)$**.** *Let a_n be a sequence of real numbers, and $s_n = \sum_{k=0}^{n} a_k$ be its associated series. Then show the following.*

1. *If the sequence $\{a_n\}$ does not converge to 0, then its associated series $\{s_n\}$ does not converge to anything.*

2. *Let $a_n' = |a_n|$, and let its associated sequence of partial sums be denoted $s_n' = \sum_{k=0}^{n} |a_n|$. Show that if s_n' converges, then s_n also converges (although not necessarily to the same limit).*

Afterwards, give an example of a sequence that converges to 0 *whose associated series does not converge, and prove that it does not (hint: you can find the most well-known one in Abbott).*

Definition 2.5.5. Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say f is a *contraction mapping* if it satisfies the following: $\exists 0 < C < 1$ some constant so that for every pair $x, y \in \mathbb{R}$, $|f(x)-f(y)| < C\cdot|x-y|$.

Exercise 2.5.6 $(*)$**.** *Given a contraction mapping f, show that for every $x \in \mathbb{R}$ the sequence $\{x, f(x), f(f(x)), ..., f^{(\circ n)}(x), ...\}$ converges. Then show that the result $L \coloneqq \lim_{n\mapsto\infty} f^{(\circ n)}(x)$ satisfies $f(L) = L$. This is known as a* fixed point *of f.*

Remark 2.5.7. What's the point of the above? We want to find a fixed point of f by using the following general idea:

$$f \circ f^{(\circ\infty)}(x) = f^{(\circ\infty+1)}(x) = f^{(\circ\infty)}(x).$$

This is another example of an infinite supertask, this time not involving addition: we tried to "apply f infinitely many times": of course, this isn't a well-defined thing that you or I could do. Instead, we reasoned about the infinite sequence of finite tasks, which involved composing f finitely many times.

To caution you that supertasks usually have bad answers unless you assume something nice, let's do the following:

Exercise 2.5.8. *Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ and an input value $x \in \mathbb{R}$ for which the sequence $\{x, f(x), f(f(x)), ...\}$ does not converge. If you haven't already, give an example for which this corresponding sequence is* bounded*, but still does not converge to a limit.*

Exercise 2.5.9 (**Heron's Method**)**.** *Using Exercise 2.5.6, show that the following sequence converges and find its limit:*

$$x_0 \in \mathbb{R},\ \ x_{n+1} = \frac{1}{2}\cdot(x_n + \frac{2}{x_n}).$$

3 Continuity

3.1 Lecture 9: Defining Continuous Functions [9/29/25]

Learning Objectives

1. Define what a continuous function is, using our intuition about "not lifting our pencil".
2. Prove that a continuous function is *exactly* one which preserves limits of convergent sequences.

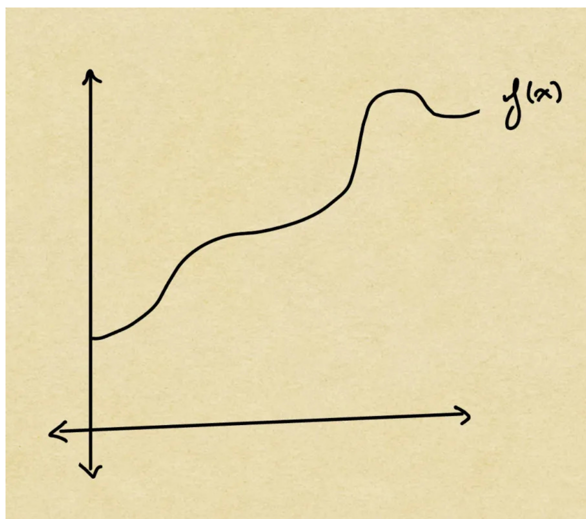

Figure 13: A depiction of the graph of a function which is "continuous" in the sense that drawing it did not require one to lift their pencil.

Generally, we regard the above to be the graph of a "continuous" function, because I did not have to lift my pencil.

Discussion (Not lifting my pencil)

Why did I not have to lift my pencil to draw this? What precise quality of the function is this getting at?

Key Definition 3.1.1 (Continuity of a Function)**.**

Let $A \subseteq \mathbb{R}$ be some domain, and $c \in A$ an element. A function $f : A \longrightarrow \mathbb{R}$ is *continuous at c* if it satisfies the following condition:

$$\forall \epsilon > 0,\ \exists \delta > 0 \text{ such that for any } x \in A \text{ if } x \text{ satisfies } |x - c| < \delta, \text{ then } |f(x) - f(c)| < \epsilon$$

The statement above is best understood through its negation:

"$\exists \epsilon > 0$ such that $\forall \delta > 0$, there exists some $x \in A$ satisfying $|x-c| < \delta$ but $|f(x)-f(c)| > \epsilon$."

What is this saying? This says that there's some distance ϵ around the value $f(c)$ for which any small neighborhood around c will always find some point x whose output value $f(x)$ is further than ϵ away from c.

Exercise 3.1.2. *Write down the correct negation for yourself, then check that the statement above is the correct negation.*

Let's try to understand this in the specific example below.

Example 3.1.3 (A Discontinuous Function)**.** Consider the piecewise function $f : (0, \infty) \to \mathbb{R}$ given by:

$$f(x) = \begin{cases} x^2 & \text{if } 0 < x < 2 \\ x & \text{if } x \geq 2 \end{cases}$$

Let's look at the graph of this function to convince ourselves it should look *discontinuous* at 2, i.e., it is not continuous at 2. Take a second to look at the figure below carefully.

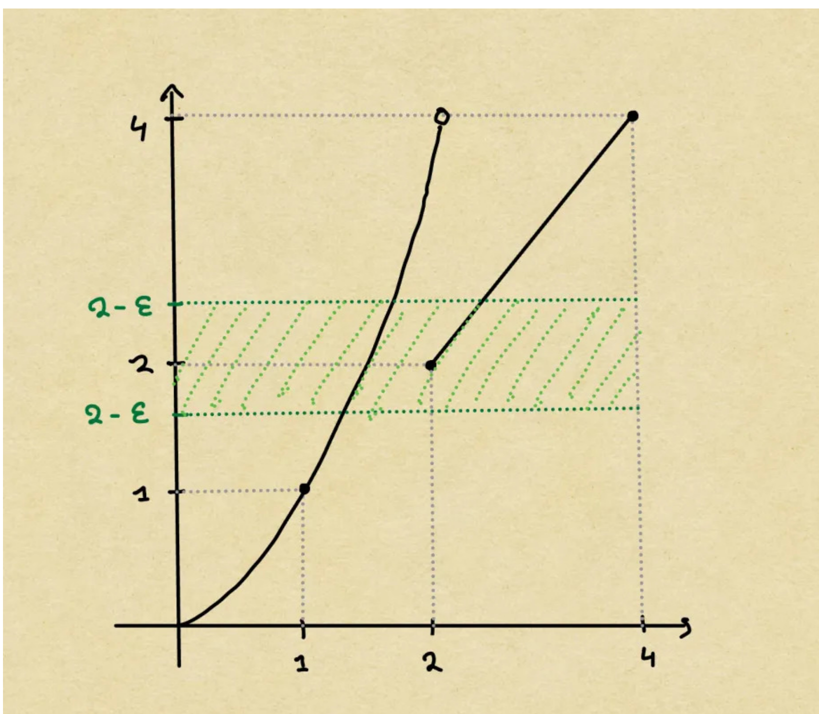

Figure 14: A depiction of the graph of the function described in Example 3.1.3, along with a shaded in region indicating a neighborhood of size ϵ around $f(2) = 2$ on the y-axis

Let's orient ourselves with what we're looking at:

- Here the point under consideration is $c = 2$, and we've fixed some given distance ϵ.
- In order to graph this function, I had to "jump" from the values at points on the left hand side of $x = 2$ to the actual value at $x = 2$.
- In doing so, I specifically had to jump across some vertical distance, and that given distance is *bigger* than the ϵ graphed above.

In our parlance above, using our notion of *discontinuity* I want to try to establish the following: For *every* possible choice of $\delta > 0$, there is a point x satisfying

- $|x - 2| < \delta$.
- $|f(x) - f(2)| > \epsilon$

This is implemented in the figure below, where I've picked some random choice of δ. Notice how the "jumping" behaviour shows up below. No matter how close I got to 2 on the x-axis, I always find a point whose output under f is further than ϵ away from $f(2)$. Here:

1. The statement "no matter how close you get to 2" is implemented by the statement $\forall \delta > 0$.
2. The fact that I can find a point whose output under f is further than ϵ away from $f(2)$ is implemented by the subsequent statement $\exists x$ satisfying $|x - 2| < \delta$ but $|f(x) - f(2)| > \epsilon$.

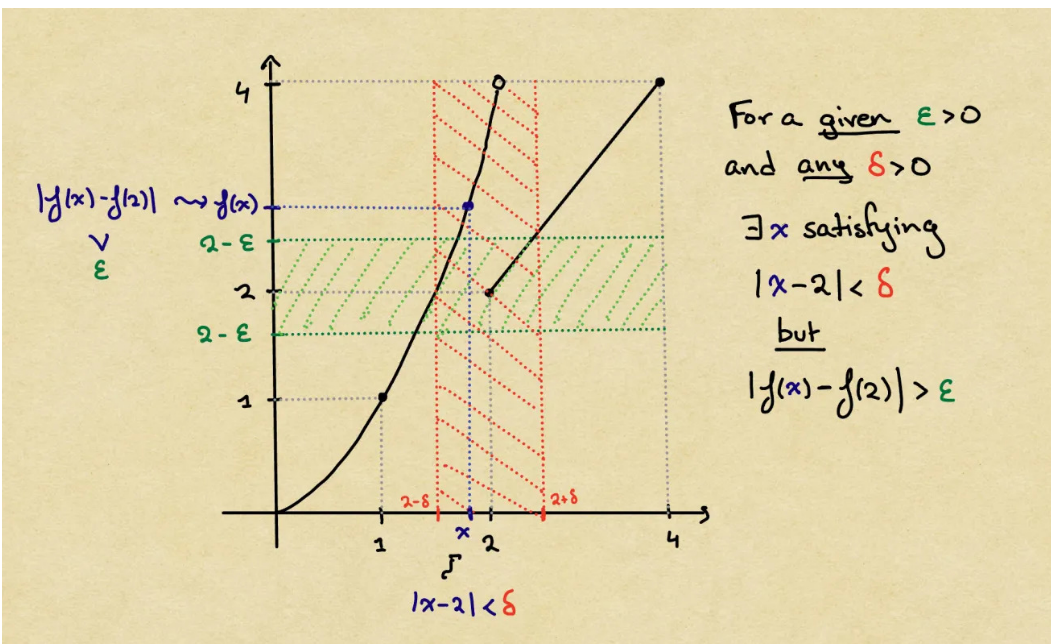

Figure 15: A depiction of the graph of the function described in Example 3.1.3, now with a shaded-in region indicating a neighborhood of size δ around 2 on the x-axis

Continuous functions are defined exactly to eliminate the behaviour above, and in this way they eliminate "jumping". In the next lecture we will prove the Intermediate Value Theorem, which will fully encode this connection for us. The final part of this lecture is the proof of the following theorem, which provides us with an alternate view on how to regard continuous functions.

Key Theorem 3.1.4 (Continuous Functions Are Exactly Those Which Respect Limits)**.**

Let $f : A \to \mathbb{R}$ be some function, and $c \in A$ some point. Then f is continuous at c *if and only if* $\forall$ sequences $\{a_n\} \subseteq A$ which converge to c, $\lim_{n\to\infty} f(a_n) = f(c)$.

In short, f is continuous if and only if f sends all sequences converging to c to sequences converging to $f(c)$. The proof will be instructive practice in using the definition.

Proof. Let's first show that if a function f is continuous at c and $\{a_n\}$ is a sequence converging to c, we have that $\lim_{n\to\infty} f(a_n) = c$. Concretely, we need to show the following.

> **Claim:** $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ so that if $n > N$ then $|f(a_n) - f(c)| < \epsilon$.
>
> *Proof of claim:* Fix some choice of $\epsilon > 0$. We know that by the assumption of continuity, there exists some particular $\delta > 0$ so that if $|x - c| < \delta$ then $|f(x) - f(c)| < \epsilon$. By the assumption of converging of $\{a_n\}$, there exists an $N \in \mathbb{N}$ so that if $n > N$ then $|a_n - c| < \delta$. By the assumption on δ, we learn that $|f(a_n) - f(c)| < \epsilon$. Thus, for this choice of $\epsilon > 0$ there exists $N \in \mathbb{N}$ so that for every $n > N$, we have that $|f(a_n) - f(c)| < \epsilon$, yielding the claim.

The above demonstrates the claim that

$$(f \text{ is continuous at } c) \implies (f \text{ sends all sequences converging to } c \text{ to sequences converging to } f(c)).$$

Since this is an *if and only if* statement, we now need to show the other direction, namely that:

$$(f \text{ sends all sequences converging to } c \text{ to sequences converging to } f(c)) \implies (f \text{ is continuous at } c)$$

We're going to prove this statement by *contrapositive.* The major reason one might want to do this is because negation turns $\forall$ statements into $\exists$ statements. Sometimes it's easier to work with $\exists$ statement, because these come down to *constructing* something specific. This brings us to trying to show the following claim.

> **Claim:** If f is not continuous at c, then there exists a sequence $\{a_n\}$ converging to c such that $\{f(a_n)\}$ does not converge to $f(c)$.

- *Proof of claim:* Let's try to build such a sequence by hand. From the definition if f is not continuous at c, then $\exists$ some *particular* $\epsilon > 0$ so that for every $\delta > 0$, there's an $x \in A$ satisfying $|x - c| < \delta$ but $|f(x) - f(c)| > \epsilon$.

  Thus, if we pick a sequence of distances $\delta_1 > \delta_2 > \delta_3 > \ldots$ all going to 0, we can find points x_n satisfying $|x_n - c| < \delta_n$ but $|f(x_n) - f(c)| > \epsilon$ stays a fixed distance away from $f(c)$. This is the key insight in building the desired sequence.

  Let's build the sequence as follows:

  - Let $x_0 \in A$ be some point satisfying $|x_0 - c| < 1$ and $|f(x_0) - f(c)| > \epsilon$.

– Let $x_1 \in A$ be some point satisfying $|x_1 - c| < \frac{1}{2}$ and $|f(x_1) - f(c)| > \epsilon$.
– Let $x_2 \in A$ be some point ...
– Let $x_n \in A$ be some point satisfying $|x_n - c| < \frac{1}{2^{n+1}}$ and $|f(x_1) - f(c)| > \epsilon$.

And so on and so forth. By construction, the sequence $\{x_n\}$ converges to c, because for every $N \in \mathbb{N}$ we have that $n > N$ implies that $|x_n - c| \leq \frac{1}{2^{n+1}} < \frac{1}{2^{N+1}}$ and for any $\epsilon_1 > 0$ we can find a $\frac{1}{2^m}$ smaller than ϵ_1. However, for every $n \in \mathbb{N}$ the value of $|f(x_n) - f(c)| > \epsilon$; this means that the sequence $\{f(x_n)\}$ *can't* converge to $f(c)$, because it stays at least a fixed distance away from $f(c)$. This completes the claim: we've built the desired sequence.

This completes the proof of the theorem. □

Warning 3.1.5.

Note that continuous functions only respect limits of sequences *which already converge*. Namely, continuous functions do NOT have to send divergent sequences to divergent sequences. Consider the example of the constant function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 0$. This function sends *any* sequence to a sequence converging to 0; in particular, it sends all divergent sequences to convergent sequences. However, it's clearly continuous.

3.2 Lecture 10: Least Upper Bounds and The Intermediate Value Theorem [10/1/25]

Learning Objectives

Objectives for today:

1. Recall what the least upper bound property is.
2. Explore multiple ways to characterize least upper bounds.
3. Use their existence to prove the Intermediate Value Theorem.

In this lecture, we will provide another reconciliation of our definition of continuous functions with the idea of "not requiring one to lift their pencil while graphing". This will be supplied by the following Very Important Theorem about continuous functions.

Key Theorem 3.2.1 (Intermediate Value Theorem)**.**

Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If $f(a) \neq f(b)$, then for any value d in between $f(a)$ and $f(b)$, there is an element $c \in [a, b]$ so that $f(c) = d$.

Before we show this, we'll need to recall some preliminaries on least upper bounds.

3.2.1 The Least Upper Bound Property of the Reals

Key Definition 3.2.2 (Least Upper Bounds)**.**

Let $A \subseteq \mathbb{R}$ be a nonempty subset of real numbers.

1. We say that A *admits an upper bound* if there exists some $M \in \mathbb{R}$ so that $\forall x \in A, x \leq M$. Any choice of such an M is called an *upper bound* of A.

2. We say that $M \in \mathbb{R}$ is a *least upper bound* for A if it is an upper bound of A, and for every upper bound N of A, $M \leq N$.

The following trick is a very useful reformulation of the property of being a least upper bound.

Proposition 3.2.3. *Let $A \subseteq \mathbb{R}$ be some nonempty subset of the real numbers. $M \in \mathbb{R}$ is a least upper bound for A if and only if:*

1. *M is an upper bound of A.*

2. *For every $\epsilon > 0$, there exists an $a \in A$ so that $M - \epsilon < a$.*

Proof. Let's show the forward direction first: namely, if M is a least upper bound of A, then it satisfies conditions (1) and (2) above.

M is a least upper bound of A $\implies$

M satisfies condition (1): This is part of the definition of being a least upper bound.

M satisfies condition (2): Fix $\epsilon > 0$. Since M is a least upper bound, $M - \epsilon$ cannot be an upper bound of A, since $M - \epsilon < M$. Thus, there must exist some $a \in A$ so that $M - \epsilon < a$.

It remains to show the backward direction: namely, that satisfying conditions (1) and (2) of the proposition is enough to imply being least upper bound. To this end, let $M \in \mathbb{R}$ be some element satisfying conditions (1) and (2).

M satisfies conditions (1) and (2) $\implies$ M is a least upper bound of A: By assumption, M is an upper bound of A. It remains to show that if $N \in \mathbb{R}$ is any other upper bound of A, then $M \leq N$. By contrapositive, this amounts to showing that if $N < M$, then N cannot be an upper bound of A. From this it is clear: if $N < M$, then $N = M - \epsilon$ for some $\epsilon > 0$, and condition (2) implies that there exists $x \in A$ such that $N = M - \epsilon < a$. Thus, N cannot be an upper bound unless it is greater than M.

□

From the above, we can immediately deduce the following rather useful corollary.

Corollary 3.2.4. *Let $A \subseteq \mathbb{R}$ be some nonempty subset of the real numbers, and $M \in \mathbb{R}$ be an upper bound of A. Then M is a least upper bound if and only if there exists some sequence $\{a_n\}$ of elements of A such that $lim_{n \to \infty} a_n$ converges to M.*

Proof. Let's first try to show the forward direction: namely, let's first assume M is a least upper bound, and try to build this sequence by hand.

> *M is a least upper bound $\implies$ there is a sequence in A converging to M:* From Proposition 3.2.3, we know that M is a least upper bound if and only if for every $\epsilon > 0$, there exists an $a \in A$ so that $M - \epsilon < a$. Using this, we can build a sequence by hand as follows:
>
> Let $a_1 \in A$ be selected so that $M - 1 < a_1 \leq M$.
>
> Let $a_2 \in A$ be selected so that $M - \frac{1}{2} < a_2 \leq M$.
>
> Let a_3 ...
>
> Let $a_n \in A$ be selected so that $M - \frac{1}{n} < a_n \leq M$.
>
> ...
>
> We claim that the constructed sequence a_n converges to M. Fix any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ so that $0 < \frac{1}{N} < \epsilon$. By construction, for every $n > N$, we have that $M - \frac{1}{n} < a_n \leq M$ and hence that $|M - a_n| < \epsilon$. Such a selection worked for any arbitrary choice of $\epsilon > 0$, and thus this sequenve must converge to M.

We now show the other direction, namely that if M is an upper bound and the limit of a sequence in A then it is a least upper bound of A.

> *M is an upper bound of A and M is the limit of a sequence in A $\implies$ M is the least upper bound of A:* We again invoke Proposition 3.2.3 here. Suppose that there exists a sequence $\{a_n\}$ of elements of A with limit M. Fix any $\epsilon > 0$. Convergence of a_n implies a weaker statement, namely that there exists some a_n satisfying $|M - a_n| < \epsilon$. However, since M is an upper bound of A, it must be the case that $a_n \leq M$. Altogether, we have that $M - \epsilon < a_n \leq M$. Since $\epsilon > 0$ was an arbitrary choice, we learn that for any $\epsilon > 0$ we may find $a \in A$ so that $M - \epsilon < a \leq M$. By Proposition 3.2.3, we may conclude that M is a least upper bound.

□

In your Discussion Worksheet 2 from last week, you in fact proved the following wonderful feature of the real numbers, which will be of critical importance in our next couple lectures.

Key Theorem 3.2.5 (The Least Upper Bound Property)**.**

If $A \subseteq \mathbb{R}$ is any nonempty subset admitting an upper bound, then it admits a least upper bound.

The argument is another "sequence-construction" style of proof. I highly recommend going back and re-attempting the proof for a better understanding of why this works with our current formulation of the reals.

Exercise 3.2.6. *Show that the least upper bound property implies that Cauchy sequences converge in the reals. In this way, the least upper bound property is an alternate formulation of "completeness", known as Dedekind-completeness.*

Notation 3.2.7. The least upper bound of a nonempty subset $A \subset \mathbb{R}$ is denoted $\sup A$, and referred to as its *supremum*.

Remark 3.2.8. There is a dual notion of "greatest lower bound" of a set A, where lower bounds and greatest lower bounds are exactly flipped. The "greatest lower bound" of A is always identified with $(-1) \cdot \sup((-1) \cdot A)$, since multiplying by (-1) reverses orders (check this!).

Notation 3.2.9. The greatest lower bound of a nonempty subset $A \subseteq \mathbb{R}$ is denoted $\inf A$, and referred to as its *infimum*.

3.2.2 A Useful Exercise

Exercise 3.2.10. *Show that if $\{a_n\}$ is a convergent sequence consisting of elements in some closed interval $[a, b]$, them $\lim_{n\to\infty} a_n \in [a, b]$. Concretely, show that if $\forall n \in \mathbb{N}$ $a \leq a_n \leq b$, then the same is true of the limit.*

3.2.3 The Proof of the Intermediate Value Theorem

Let's go back to trying to show the intermediate value theorem, whose statement we recall below.

Theorem 3.2.11 (Intermediate Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. If $f(a) \neq f(b)$, then for any value d in between $f(a)$ and $f(b)$, there is an element $c \in [a, b]$ so that $f(c) = y$.*

What's the idea? If I "draw the graph of the function without lifting my pencil", then to get from $f(a)$ to $f(b)$, I needed to intersect the line at $y = d$ somewhere. This heuristic is illustrated below.

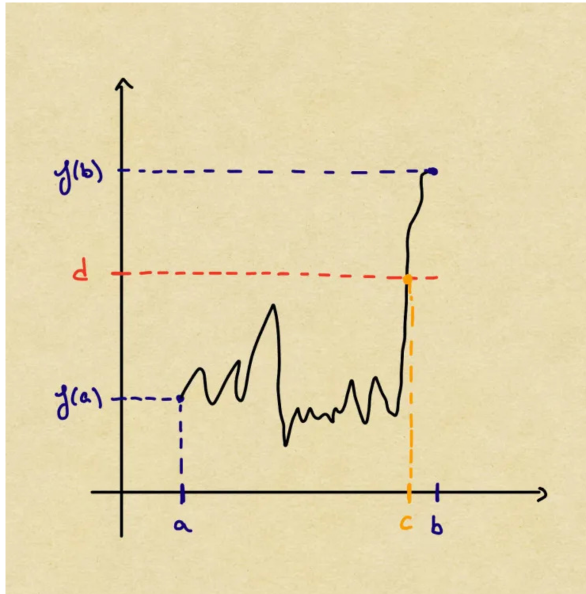

Figure 16: A depiction of the graph of a continuous function $f : [a, b] \to \mathbb{R}$ with a value $f(a) < y < f(b)$, and the graph intersecting the line at $y = d$

We're going to try to prove this by contradiction. If we assume that a function did not intersect $y = d$, then our intuition will be that it failed to do so for one of the following reasons:

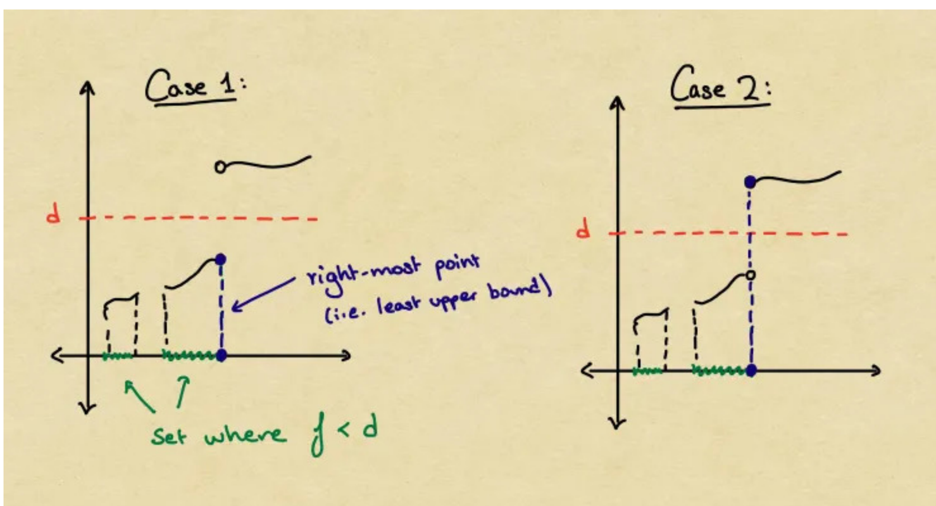

Figure 17: A depiction of the "right-most point" of the set where where f takes values less than d. There are two cases: either the value of f at this right-most point is strictly less than d, or it is strictly greater than d.

Turning this into mathematics, the "right-most point" of the set of points where f is still less than d is concretely given by $\sup\{x \in [a, b] \mid f(x) < d\}$; the ability to even refer to such a point is given by being able to refer to least upper bounds. This is an important way in which we'll use them, and we'll see it in action below.

Proof of Theorem 3.2.11. Assume, towards a contradiction, that there does not exist $c \in [a, b]$ so that $f(c) = d$. Then for every $x \in [a, b]$, we have that either $f(x) > d$ or $f(x) < d$. Consider the set $A := \{x \in [a, b] \mid f(x) < d\}$ of points where f takes value strictly less than d. Since $f(a) < d$, we know that A is nonempty. By the least upper bound property, A admits a least upper bound, denoted $\sup A$.

Now note that since $A \subseteq [a, b]$, the element b is an upper bound of A and thus $\sup A \leq b$. Furthermore, since $a \in A$, we have that $\sup A \geq a$. Thus, $\sup A \in [a, b]$, and thus the function f must be defined on it. We have two cases for the value of $\sup A$.

> **Case 1:** $f(\sup A) < d$ In this case, we have that $\sup A < b$ since $f(b) > d$. Fix any $0 < \epsilon < |b - \sup A|$. Using the exact same method of proof as in Corollary 3.2.4, we may construct a sequence of elements $\{b_n\}$ in $[a, b]$ such that $b_n > \sup A$ and $\lim_{n\to\infty} b_n = \sup A$.
>
> By construction, we have that $f(b_n) > d$ for every $n \in \mathbb{N}$, as $b_n > \sup A$ and thus $b_n \notin A$. Since f is assumed to be continuous, it must be the case that $\lim_{n\to\infty} f(b_n) = f(\lim_{n\to\infty} b_n) = f(\sup A)$ by the main result of Lecture 9.
>
> *Contradiction!* By Exercise 3.2.10 if $\{f(b_n)\}$ is a sequence of elements strictly greater than d, it must be the case that $\lim_{n\to\infty} f(b_n) = f(\lim_{n\to\infty} b_n) = f(\sup A) \geq d$. However, this case assumed that $f(\sup A) < d$.
>
> **Case 2:** $f(\sup A) > d$ This is the exact flip of Case 1, except we may construct a sequence which breaks continuity as a direct consequence of Corollary 3.2.4. The details are left as an exercise (but I must emphasize again that this is a direct flip of the first case).

As in both cases we may derive a contradiction, it must have been the case that the premise, i.e. $\nexists\, c \in [a, b]$ such that $f(c) = d$, was false. The theorem follows. □

3.3 Lecture 11: The Extreme Value Theorem and Uniform Continuity [10/6/25]

3.3.1 The Boundedness and Extreme Value Theorems

Learning Objectives

Prove the Extreme Value Theorem, using the Bolzano-Weierstrass Theorem.

The following is a useful fact, and follows from the Bolzano-Weierstrass Theorem coupled with Exercise 3.2.10.

Exercise 3.3.1. *Show that if $\{a_n\}$ is an arbitrary sequence of real numbers in $[a, b]$, then it has a convergent subsequence $\{a_{n_k}\}$ whose limit is also in $[a, b]$.*

Using the above, we will prove today's major result:

Key Theorem 3.3.2 (Extreme Value Theorem)**.**

Let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function. Then there exists a point $c \in [a, b]$ so that $\forall x \in [a, b]$, $f(x) \leq f(c)$. Such a point will be called a *maximum* of the function f on $[a, b]$.

The proof of the above will witness two applications of the same general trick, namely the use of Exercise 3.3.1. Let's remark a little about the difference in strategy to the previous lecture.

- In Lectures 8 and 9 we directly constructed *convergent sequences* which satisfied or flouted certain properties, see for example the proofs of Corollary 3.2.4 and Theorem 3.2.11.
- In this lecture, however, we will find that at best we can only build an *arbitrary* sequence of numbers in $[a, b]$.

Our key insight will be the use of Exercise 3.3.1 to guarantee a convergent *subsequence* of the original sequence, and often we have enough control to make this subsequence "do what we want". Let's see this first in the following example.

Theorem 3.3.3 (Boundedness Theorem)**.** *Let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function. Then f is bounded; that is to say, there exists an $M > 0$ so that $\forall x \in [a, b]$, $|f(x)| \leq M$.*

Proof. Assume, towards a contradiction, that f is not bounded. Then for every $n \in \mathbb{N} \subseteq \mathbb{R}$, there exists an element $x_n \in [a, b]$ satisfying $|f(x_n)| > n$. In this way, we may build a sequence $\{x_n\}$ of elements in $[a, b]$.

If this sequence itself converged, we'd be done: if x_n converged, then it had to converge to some $x \in [a, b]$, and by continuity we'd have $lim_{n\to\infty} f(x_n) = f(x)$. However, the sequence $f(x_n)$ is not bounded and thus has no limit, i.e., it cannot converge. Contradiction! It must be the case that no such convergent sequence existed.

However, we *do not know* whether or not $\{x_n\}$ converges. This is where the ingenuity of Exercise 3.3.1 comes in: we know that $\{x_n\}$ had to have a convergent *subsequence* $\{x_{n_k}\}$. Moreover, by our assumption on the original sequence, we know that $|f(x_{n_k})| > n_k \in \mathbb{N}$. In particular, since n_k is an increasing sequence of natural numbers indexed over k, the sequence $f(x_{n_k})$ must not have been bounded.

Contradiction! If $\{x_{n_k}\}$ converges to some $x \in [a, b]$, then $f(x) = \lim_{k\to\infty} f(x_{n_k})$. However, a limit of the sequence $\{f(x_{n_k})\}$ cannot exist, as by construction it is not a bounded sequence. $\square$

Warning 3.3.4. The conclusion of Theorem 3.3.3 *really really* needs it to be the case that we're working with the closed interval $[a, b]$. Namely, the boundedness theorem fails in the cases (a, b), $(a, b]$, $[a, b)$. Let's explore an example.

Example 3.3.5 (Failure of the Boundedness Theorem)**.** Consider the function $f(x) = \frac{1}{x}$ on the interval $(0, 1)$. This is continuous everywhere that it is defined; however, it is very much not bounded, as the following diagram indicates.

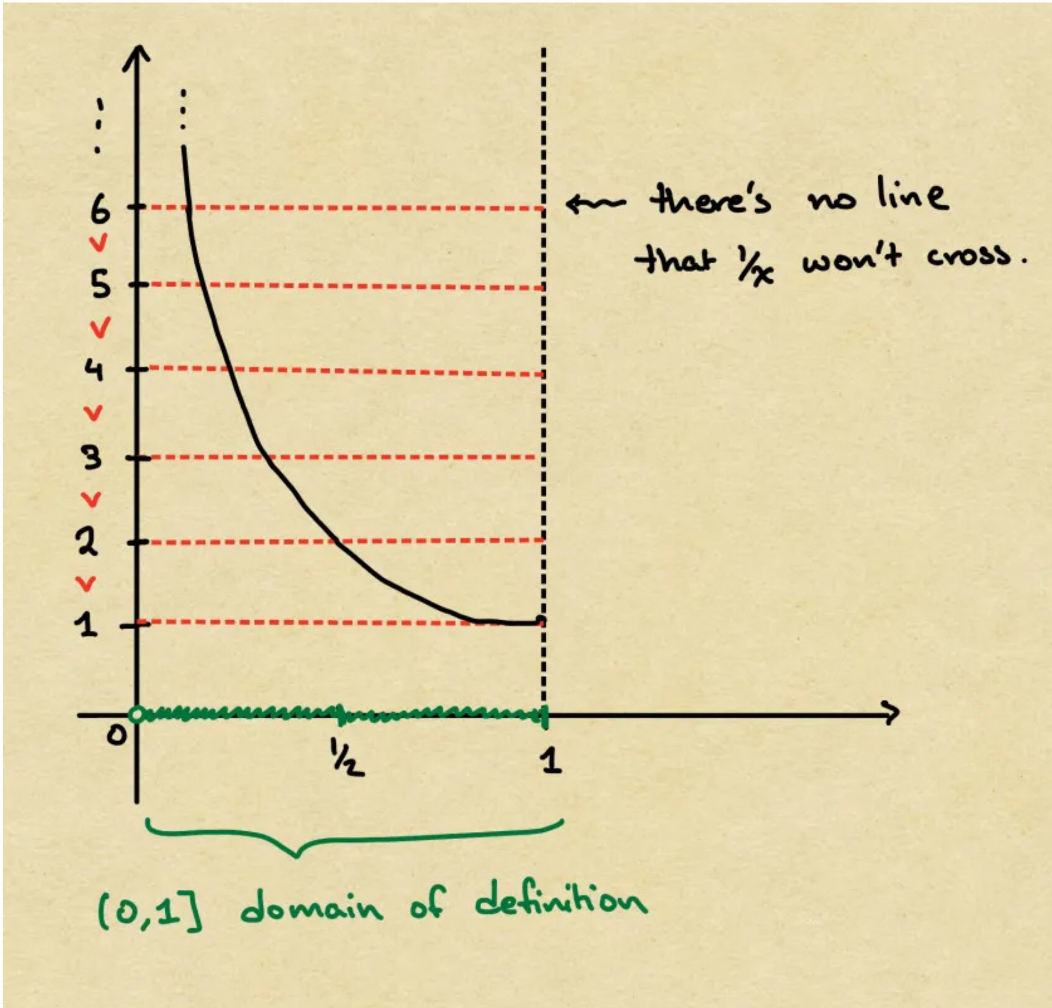

Figure 18: A depiction of the failure of $f(x) = \frac{1}{x}$ to be bounded on the open interval $(0, 1)$.

Why did the boundedness theorem fail in this example? Looking at the proof, given a sequence in $[a, b]$, we needed to be able to find a convergent sequence with a limit inside $[a, b]$; of course, this is just not true in $(0, 1)$. Consider the sequence $\{\frac{1}{n}\}$, which converges to 0, a point outside the range of definition of $f(x)$.

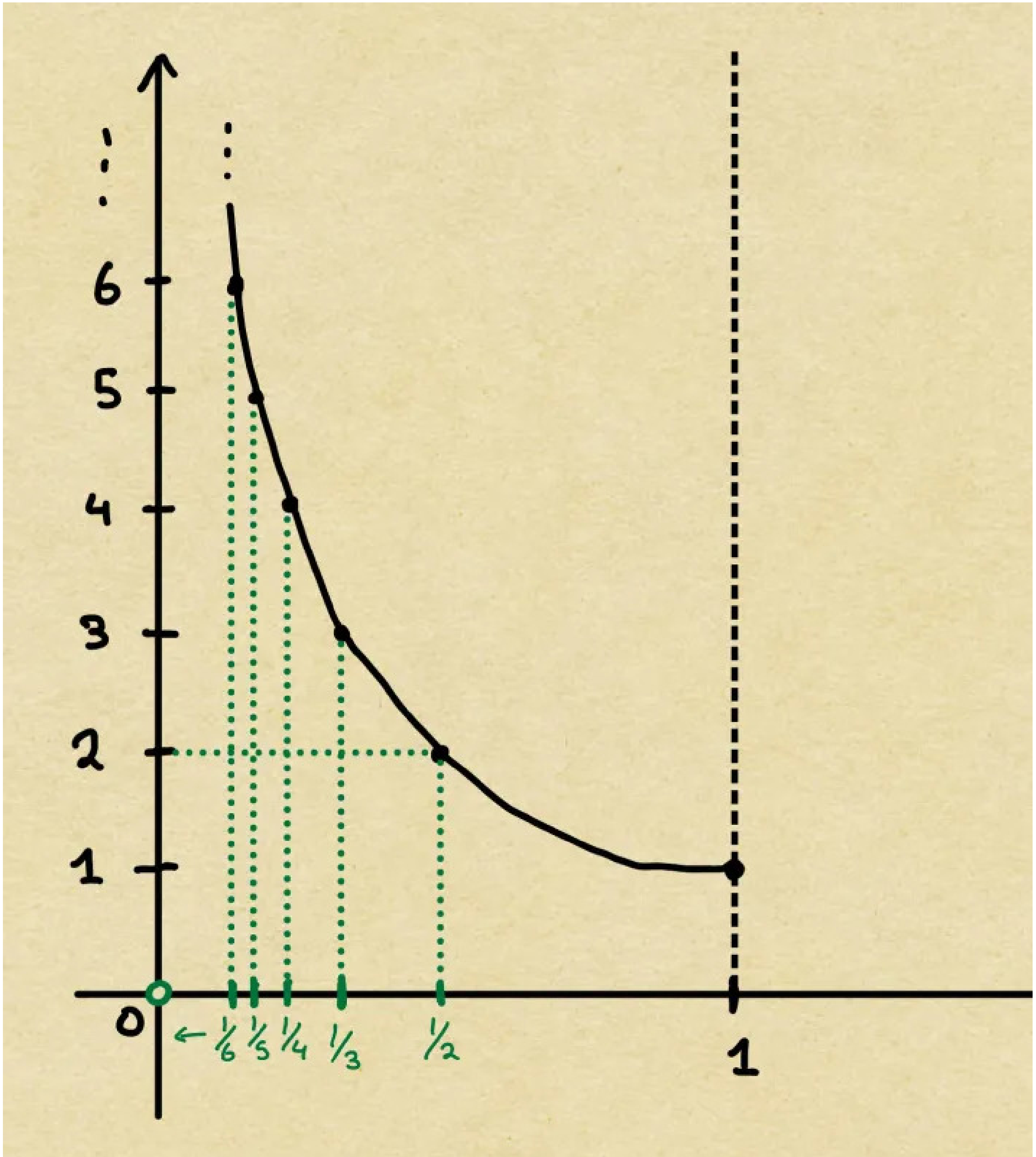

Figure 19: A depiction of the unbounded sequence $\{f(\frac{1}{n})\}$.

Thus, the boundedness theorem follows from:

1. Continuity being well-behaved with respect to limits
2. The domain of definition containing limits of convergent sequences.

However, it is also possible to say more in the case where the domain does *not* contain limits of convergent sequences, by asking for our continuous functions to be well-behaved with respect to *Cauchy* sequences. This will be the notion of *uniform continuity*, discussed below.

Our next order of business is to finish the proof of the Extremal Value Theorem, the statement of which we recall below.

Theorem 3.3.6 (Extreme Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then there exists a point $c \in [a, b]$ so that $\forall x \in [a, b]$, $f(x) \leq f(c)$. Such a point will be called a* maximum *of the function f on $[a, b]$.*

Proof. We're going to do the same sequence-construction argument. By the Boundedness Theorem Theorem 3.3.3, we know that the range of f, given by $\text{Range}(f) := \{y \in \mathbb{R} \mid \exists\, x \in [a, b] \text{ such that } f(x) = y\}$ admits an upper bound, and hence admits a least upper bound. Let us refer to the least upper bound of Range(f) with the notation $\sup f$, and refer to it as the *supremum of f*.

Our goal will be to show that $\exists c \in [a, b]$ so that $f(c) = \sup f$. To this end, note that Corollary 3.2.4 supplies a sequence of the form $\{y_n\}$ converging to $\sup f$ with each $\{y_n\} \in \text{Range}(f)$. By definition of the range, we have that there is some sequence $\{x_n\}$ in $[a, b]$ such that $f(x_n) = y_n$ for every n.

Now, using our result Exercise 3.3.1, we have that $\{x_n\}$ admits a convergent subsequence $\{x_{n_k}\}$ with limit $L \in [a, b]$. Furthermore, by continuity, we know that $f(L) = \lim_{k\to\infty} f(x_{n_k})$. Since $f(x_{n_k})$ is a subsequence of the convergent sequence $f(x_n)$, we know that $\lim_{k\to\infty} f(x_{n_k}) = \lim_{n\to\infty} f(x_n) = \sup f$. Altogether, we find that $f(L) = \sup f$, and we have constructed the desired point by setting $c = L$. □

Note that the conclusions of the Extreme Value Theorem also fail in the cases that the conclusions of the Boundedness Theorem fail, for example $f(x) = \frac{1}{x}$.

3.3.2 Uniform Continuity

Let us now introduce a new notion, meant to rectify what's going on with $f(x) = \frac{1}{x}$. This will also be our first introduction to bounding "rates of change", as a prelude of what's to come next lecture. Let's first understand what exactly went wrong with $f(x) = \frac{1}{x}$.

Example 3.3.7 (Failure of Boundedness Revisited)**.** Why did $f(x)$ fail to be bounded on $(0, 1)$? One answer is that its domain of definition didn't contain all of its limits; however, on the interval $(1, 2)$ it was bounded. What exactly happened here? One answer is provided by looking at "how fast the function $f(x) = \frac{1}{x}$ changed" near the point 0. Namely, even though it was continuous near 0, the Cauchy sequence $\{\frac{1}{n}\}$ was not sent to a Cauchy sequence in the reals, as is shown in Figure 3.3.2

So although continuity sent convergent sequences to convergent sequences *whenever the function was defined at its limit*, they do not have to send Cauchy sequences to Cauchy sequences in general. Let us introduce a finer notion than continuity to remedy this deficit.

Definition 3.3.8. A function $f : A \to \mathbb{R}$ is *uniformly continuous* on A if for every $\epsilon > 0$, there exists a $\delta > 0$ so that if $x, y \in A$ is *any pair* satisfying $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.

Note the contrast in definition to ordinary continuity:

- With ordinary continuity, we say that at a *particular point* $c \in A$, there is some choice of $\delta > 0$ so that $|x - c| < \delta$ implies $|f(x) - f(c)| < \epsilon$. In some sense, this is *local choice*: the choice of δ at some c may not work for other points in A.

- For uniform continuity, we demand to be able to find a choice of δ that works for *every possible pair* $x, y \in A$. This is a *global choice*

Let's look back at the example of $f(x) = \frac{1}{x}$.

Example 3.3.9 ($f(x) = \frac{1}{x}$ is not uniformly continuous)**.** This function fails uniform continuity on $(0, 1)$: for example, taking $\epsilon = \frac{1}{2}$, for any $\delta > 0$ it is always possible to find two points x, y satisfying $|x - y| < \delta$ but $|\frac{1}{x} - \frac{1}{y}| > \frac{1}{2}$. For example, taking $x = \frac{1}{n}$ and $y = \frac{1}{n+1}$, one has that

$$|x - y| = \left| \frac{1}{n} - \frac{1}{n+1} \right| = \left| \frac{1}{n(n+1)} \right|, \quad |f(x) - f(y)| = |n - (n+1)| = 1 > \epsilon$$

and for any $\delta > 0$ we can always find some big enough n so that $\frac{1}{n(n+1)} < \delta$, showing that no δ can work for all pairs. This is illustrated below.

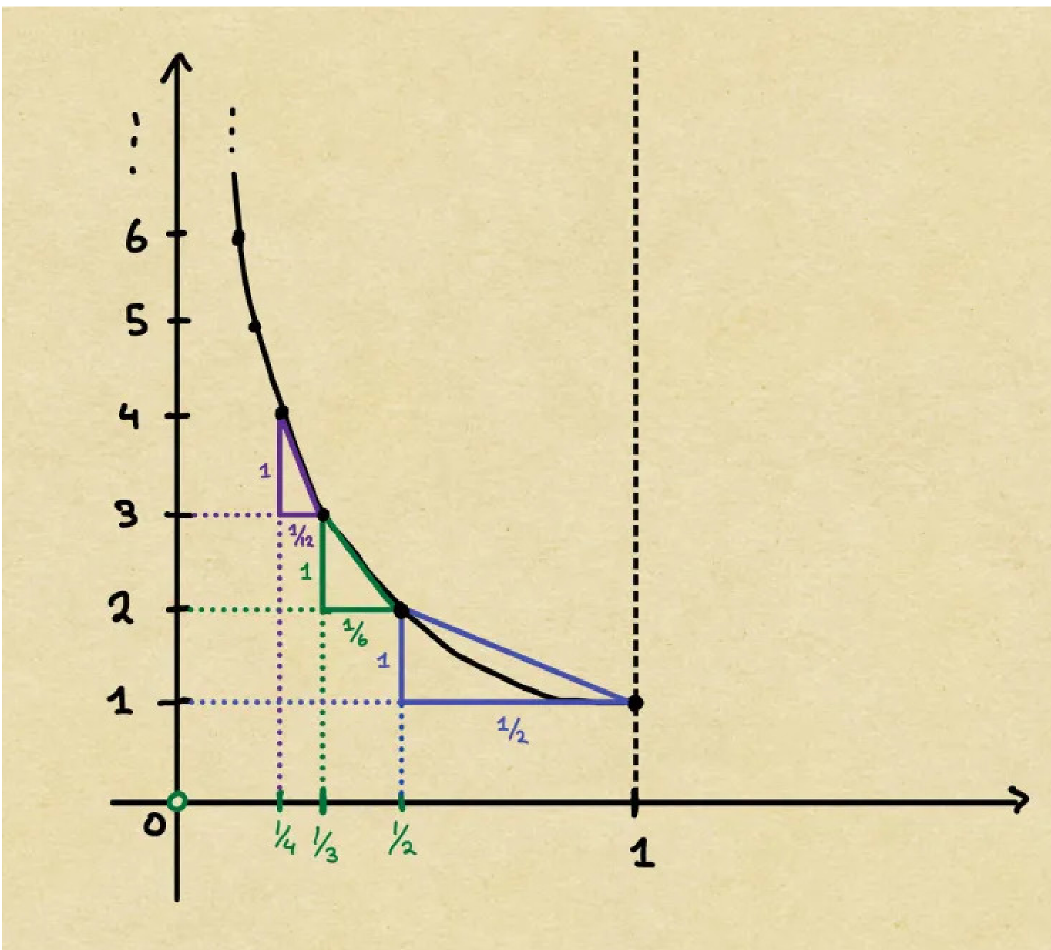

Figure 20: A depiction of the failure of $f(x)$ to be uniformly continuous: it is always possible to find two points x and y so that $|f(x) - f(y)| = 1 > \epsilon = \frac{1}{2}$, but x and y can be chosen to be arbitrarily close together.

Exercise 3.3.10. *Show that if* $f : A \to \mathbb{R}$ *is uniformly continuous and* $\{a_n\}$ *is a Cauchy sequence in* A*, then* $\{f(a_n)\}$ *is a Cauchy sequence in* $\mathbb{R}$*.*

We will explore more properties of uniformly continuous functions on this week's homework.

3.4 Homework 3 (Due Friday, October 10th)

3 out of the following 6 exercises graded for correctness, the remainder will be graded for completeness. I have marked the graded ones with an asterisk ($*$). This homework is out of a possible **24 points**, with the graded exercises worth 6 points and the ungraded exercises worth 2 points each.

Finally, don't expect to be able to do everything in this homework immediately! I expect you to return to this once or twice with a group as we progress through the course over the next two weeks.

Exercise 3.4.1. *Prove that if* $f : A \to \mathbb{R}$ *and* $g : B \to \mathbb{R}$ *are continuous functions, then* $f \circ g : f^{-1}(B) \to \mathbb{R}$ *is a continuous function.*

Here we write $f^{-1}(B)$ to refer to the collection of points of A which are sent to B by the function f. This is called the *preimage* of f, in case you want to Google it.

Exercise 3.4.2. ($*$) *Prove that the function* $f(x) = x^{\frac{m}{n}}$ *is continuous for any natural numbers* m, n*. You will probably benefit from using Exercise 3.4.1.*

Exercise 3.4.3. ($*$) *Solve* [Abb15, Exercise 4.3.2].

The following exercises will need to invoke the Intermediate Value Theorem, which will be proved in Lecture 10 on **[10/1/25]**.

Exercise 3.4.4. *Show that any polynomial function with odd degree has at least one root in* $\mathbb{R}$.

Exercise 3.4.5. *Show that any continuous one-to-one function* $f : A \to \mathbb{R}$ *is* monotone*, i.e., it is either strictly increasing (*$x < y \implies f(x) < f(y)$*) or strictly decreasing (*$x < y \implies f(x) > f(y)$*).*

Exercise 3.4.6. ($*$) *If a function* $f : A \to \mathbb{R}$ *is one-to-one, then we can define its* inverse function *on the range of* f *– denoted* $f(A)$ *– uniquely specified by the expression*

$$g(f(x)) = x.$$

Show that if $f : A \to \mathbb{R}$ *is continuous, then* $g : f(A) \to \mathbb{R}$ *is also continuous.*

4 Derivatives and Differential Calculus

4.1 Class Worksheet on Derivatives and Differentiability [10/8/25]

In this worksheet, we will define what the derivative of a function is, in addition to proving its basic properties. Before we get started, let's review some rates-of-change basics.

4.1.1 Warm-up: Rates of Change

Let $f : (a, b) \to \mathbb{R}$ be a function, and $x_1, x_2 \in [a, b]$ be two points.

Exercise 4.1.1. *Recall the formula for the function whose graph is the unique line which passes through the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$.*

Recall that the *difference quotient* between x_1 and x_2 is given by the following formula

$$D_f(x_1, x_2) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

The expression $D_f(x_1, x_2)$ measures the slope of the line connecting the two points $f(x_1)$ and $f(x_2)$. Concretely, it measures the *average rate of change* between the points x_1 and x_2.

Remark 4.1.2. The notation $D_f(x_1, x_2)$ is nonstandard; I picked it because I felt like it.

4.1.2 Differentiability

Key Definition 4.1.3 (Differentiability)**.**

We say a function $f : (a, b) \to \mathbb{R}$ is *differentiable at $c \in (a, b)$* if the following holds:

> "For every convergent sequence $\{x_n\}$ in (a, b) with limit c, the associated sequences
>
> $$\{D_f(x_n, c)\} = \left\{\frac{f(x_n) - f(c)}{x_n - c}\right\}$$
>
> all converge and have the same limit whenever they are defined."

Whenever this is the case, we simply write $\lim_{x\to c} D_f(c, x)$ to indicate that any choice of sequence converging to c gives the same answer.

Exercise 4.1.4. *Show that $f : (a, b) \to \mathbb{R}$ is differentiable at $c \in (a, b)$ if and only if the following holds:*

> *"For every sequence $\{h_n\}$ converging to 0 which moreover satisfies $|h_n| < |c - a|$ and $|h_n| < |c - b|$, the associated sequences*
>
> $$\{D_f(c + h_n, c)\} = \left\{\frac{f(c + h_n) - f(c)}{(c + h_n) - c}\right\}$$
>
> *all converge and have the same limit whenever they are defined."*

Exercise 4.1.5. *Show that if a function $f : (a, b) \to \mathbb{R}$ is differentiable at $c \in (a, b)$, then it is also continuous at c.* (Hint: try to show that f preserves limits of sequences converging to c. Namely, show that "$\lim_{x\to c} f(x) - f(c) = 0$")

4.1.3 Algebraic Properties of Differentiation

Whenever a function $f : (a, b) \to \mathbb{R}$ is differentiable at a point $c \in (a, b)$, we will use the notation

$$\left.\frac{df}{dx}\right|_{x=c} = \lim_{x \to c} D_f(x, c)$$

and refer to this quantity as the *derivative of f at c*.

Exercise 4.1.6 (Linearity of the Derivative)**.** *Suppose f and g are functions from (a, b) to $\mathbb{R}$ which are both differentiable at $c \in (a, b)$. Show that the following are true.*

1. *The function $(f + g)(x) = f(x) + g(x)$ is differentiable at c, and*

$$\left.\frac{d(f+g)}{dx}\right|_{x=c} = \left.\frac{df}{dx}\right|_{x=c} + \left.\frac{dg}{dx}\right|_{x=c}$$

2. *For any constant $k \in \mathbb{R}$, the function $kf(x)$ is differentiable at c, and*

$$\left.\frac{d(kf)}{dx}\right|_{x=c} = k \cdot \left.\frac{df}{dx}\right|_{x=c}$$

Exercise 4.1.7 (The Product Rule)**.** *Suppose f and g are functions from (a, b) to $\mathbb{R}$ which are both differentiable at $c \in (a, b)$. Show that the function $(f \cdot g)(x) = f(x) \cdot g(x)$ is also differentiable at c, and moreover that the following holds:*

$$\left.\frac{d(f \cdot g)}{dx}\right|_{x=c} = f(c) \cdot \left.\frac{dg}{dx}\right|_{x=c} + g(c) \cdot \left.\frac{df}{dx}\right|_{x=c}$$

Notation 4.1.8. When a function is *differentiable at every point in its domain*, we simply say the function is *differentiable*. In this case, we simply write $\frac{df}{dx}$ to denote the function whose value at c is given by

$$\frac{df}{dx}(c) = \left.\frac{df}{dx}\right|_{x=c}$$

This function is referred to as the *derivative* of f.

Exercise 4.1.9 (The Power Rule)**.** *Show that the function $f(x) = x$ is everywhere differentiable. Using this and the above, show that the function $f(x) = x^n$ is everywhere differentiable, and moreover show that*

$$\frac{d(x^n)}{dx} = nx^{n-1}$$

Warning 4.1.9.

The derivative of a continuous function, when it exists, is not necessarily itself continuous! We will see an example of the same in class.

4.2 Lecture 12: Linear Approximations, Critical Points, and the Mean Value Theorem [10/13/2025]

Learning Objectives

By the end of this lecture, we will be able to:

- Explain why maxima and minima are "critical points" of a differentiable function.
- State and prove the Mean Value Theorem
- Explain the "linear approximation" perspective on the derivative.

4.2.1 Linear approximations

Last time, I had you define what it means for a function to be *differentiable* as well as prove the basic properties of the derivative. This gave us one answer to the following question:

Question. What *is* the derivative of a function?

Our first answer: a measure of its instantaneous rate of change. Before proceeding, I want to offer us a rather different perspective on the same: that the derivative of a function is the best "linear approximation" of a function.

Let us first note the following fact.

Observation 4.2.1 (Linear functions are exactly the functions with constant derivative)**.** Let $f(x) = mx+c$ a linear function of slope m and intercept c. The slope is exactly the rate of change between any two points, which is the constant fixed value m. From the algebraic properties of the previous lecture we have that f is everywhere differentiable, and that $\frac{df}{dx}(x) = m$ the constant function at m. Thus, *linear functions have a constant rate of change.* It is not hard to convince yourself that if a function g had the derivative $\frac{dg}{dx}$ being a constant function, then g must have been linear. We will deduce this as a consequence of the mean value theorem later.

We want to take the perspective that the derivative of a function is its *best linear approximation*. Suppose a function $f : (a, b) \to \mathbb{R}$ is differentiable at a point $c \in (a, b)$. The derivative at c is computed through the following equation:

$$\left.\frac{df}{dx}\right|_{x=c} = \lim_{h\to 0} \frac{f(c+h) - f(c)}{h}$$

The *linear approximation to f at c* is the unique function specified by

$$g(c+h) = f(c) + \left.\frac{df}{dx}\right|_{x=c} \cdot h$$

Written as a function in x, we may present it as $g(x) = f(c) + \frac{df}{dx}|_c \cdot (x - c)$.

If f were a linear function, then f and g would coincide by Observation 4.2.1. However, in general, they differ. To measure their exact difference, we define the following "error function"

$$E(h) = f(c+h) - g(c+h) = f(c+h) - f(c) - \left.\frac{df}{dx}\right|_{x=c} \cdot h$$

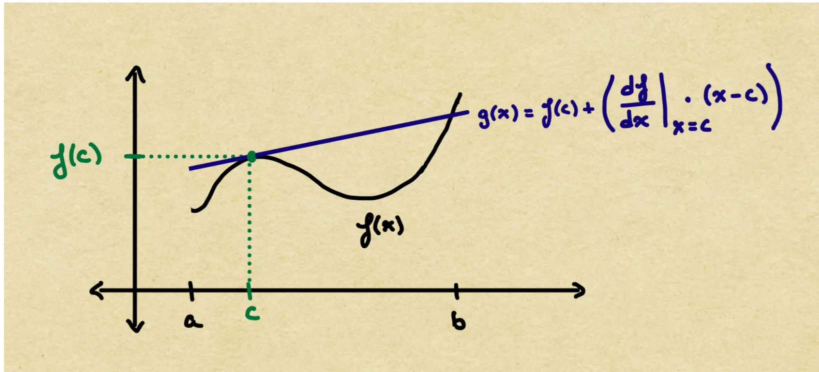

Figure 21: A depiction of the linear approximation to f at the point c.

for any given h, the error function above measures the difference between

- The value of $g(c+h)$ for g a linear function of slope exactly $\frac{df}{dx}|_c$ passing through the point $(c, f(c))$. This would be the value of $f(c+h)$ if it were actually a linear function.
- The actual value of $f(c+h)$, which may differ since f is not a linear function.

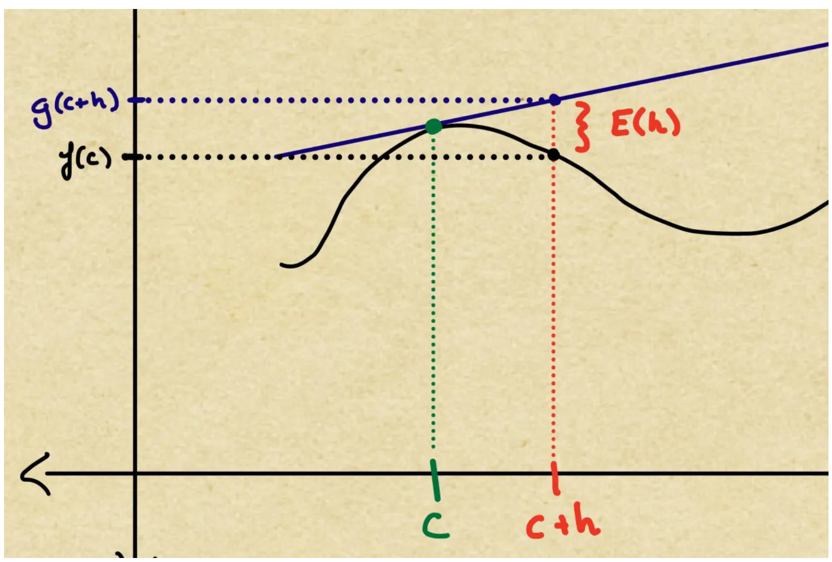

Figure 22: A depiction of the error function as the difference between a linear function of slope $\frac{df}{dx}\Big|_{x=c}$ passing through $(c, f(c))$ and the actual value of $f(c)$

To justify that this is the "best possible approximation", we will show that the error function $E(h)$ in general goes to 0 faster than any linear function would; in this way, "to first order", the function f looks like its linear approximation in a neighborhood around c. Consider the following sequence of manipulations.

$$\begin{aligned}
f(c+h) = f(c) + \left.\frac{df}{dx}\right|_{x=c} \cdot h + E(h) &\implies f(c+h) - f(x) = \left.\frac{df}{dx}\right|_{x=c} \cdot h + E(h) \\
\text{Applying } \frac{1}{h} \cdot \text{ to both sides} &\implies \frac{f(c+h)-f(x)}{h} = \left.\frac{df}{dx}\right|_{x=c} + \frac{E(h)}{h} \\
\text{Taking limits as } h \to 0 &\implies \left.\frac{df}{dx}\right|_{x=c} = \left.\frac{df}{dx}\right|_{x=c} + \lim_{h\to 0} \frac{E(h)}{h} \\
\text{Subtracting } \left.\frac{df}{dx}\right|_{x=c} \text{ from both sides} &\implies 0 = \lim_{h\to 0} \frac{E(h)}{h}
\end{aligned}$$

This last line, that $\lim_{h\to 0} \frac{E(h)}{h} = 0$, is our justification for the linear approximation being the "best possible approximation by a linear function": the error term, which measures the difference between the two, goes to 0 at a rate faster than the linear function h.

Exercise 4.2.2. *Check that* $\lim_{h\to 0} \frac{E(h)}{L(h} = 0$ *for any linear function* L *in* h. *You may assume the case where* $L(h) = h$.

Definition 4.2.3. Let $f : (a, b) \longrightarrow \mathbb{R}$ be differentiable at a point $c \in (a, b)$. The graph of the linear approximation to $f(x)$ at c, i.e., the function $g(x) = f(c) + \frac{df}{dx}\big|_{x=c} \cdot (x - c)$, is also known as the *tangent line* to $f(x)$ at the point c.

4.2.2 Critical Points, Maxima and Minima, and the Mean Value Theorem

Notation 4.2.4. Henceforth, if a function $f : (a, b) \longrightarrow \mathbb{R}$ is differentiable at every point in its domain, we will use the notation

$$f'(x) \coloneqq \frac{df}{dx} : (a, b) \longrightarrow \mathbb{R}$$

to denote the derivative of f as a function in (a, b).

For the rest of this subsection, let $f : (a, b) \longrightarrow \mathbb{R}$ be a function that is differentiable at every point in its domain. Our first result is the following.

Key Theorem 4.2.5 (Extremal Points are Critical Points)**.**

Suppose f attains a *local maximum* at some point $c \in (a, b)$. That is, suppose $\exists \epsilon > 0$ so that $\forall x \in (c - \epsilon, c + \epsilon)$, $f(x) \leq f(c)$. Then $f'(c) = 0$.

Definition 4.2.6. We call a point where $f' = 0$ *critical point* of the function f.

Proof. Consider the sequence $x_n = \{c - \frac{1}{n}$; for some sufficiently large N, this sequence is strictly between a and c. Since f is differentiable at c, we have that

$$f'(c) = \lim_{N \leq n \to \infty} \frac{f(x_n) - f(c)}{x_n - c}$$

However, since $f(c)$ is a maximum, the numerator $f(x_n) - f(c) \leq 0$, while the denominator is equal to $\frac{-1}{n}$ and also less than 0. Thus, the RHS of the above is a sequence which is ≤ 0, implying that its limit is also ≤ 0 and that $f'(c) \leq 0$.

Similarly, consider the sequence $x_n' = \{c + \frac{1}{n}\}$. Applying the same trick for this sequence, we obtain that that $f'(c) \geq 0$. It follows that $f'(c) = 0$. ☐

Remark 4.2.7. The same argument works in the case where c is a minimum; alternatively, if c is a minimum of f then it is a maximum of the function $-f$, and the linearity of the derivative yields the desired claim.

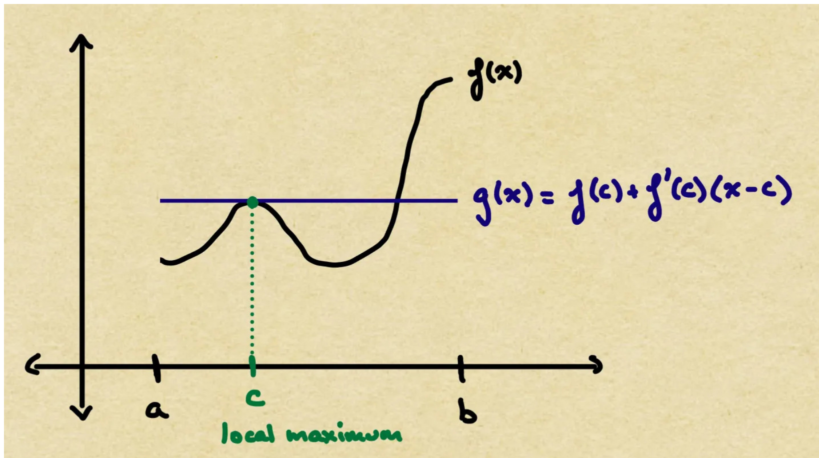

Figure 23: A depiction of the linear approximation to f at a maximal point $c \in (a, b)$

The figure above demonstrates that at a critical point, the function up to linear approximation appears to be constant. This gives an nice immediate proof of the following fact, which is the base case of a more general fact.

Theorem 4.2.8 (Rolle's Theorem)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is differentiable on (a, b). Suppose also that $f(a) = f(b) = 0$. Then there exists a point $c \in (a, b)$ so that $f'(c) = 0$, i.e., a critical point of f.*

Proof. Any differentiable function is continuous, and any continuous function on a closed interval attains both a maximum and a minimum by the Extreme Value Theorem (Theorem 3.3.6). If the maximum and minimum do not occur at a or b, we are done by the previous theorem. There is only one other case, and we write this out below.

> **Case 1, the maximum of f is strictly bigger than 0:** Then the maximum of f occurs at some point $c \in (a, b)$ since $f(a) = f(b) = 0$, and this is a critical point by the previous theorem.
>
> **Case 2, the minimum of f is strictly less than 0:** Then the minimum of f occurs at some point $c \in (a, b)$ since $f(a) = f(b) = 0$, and this is a critical point by the previous theorem.
>
> **Case 3, the maximum and minimum of f are both 0:** In this case the function is constant at 0, and the derivative is everywhere 0 so any point works.

This concludes the proof. ☐

Rolle's theorem is a special case of a much more powerful fact, itself implied by Rolle's Theorem. This will be of vital importance in the following lectures.

Key Theorem 4.2.9 Mean Value Theorem**.**

Suppose $f : [a, b] \to \mathbb{R}$ is differentiable on (a, b). Then there exists a point $c \in (a, b)$ so that

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

i.e., the instantaneous rate of change at c is equal to the average rate of change between a and b.

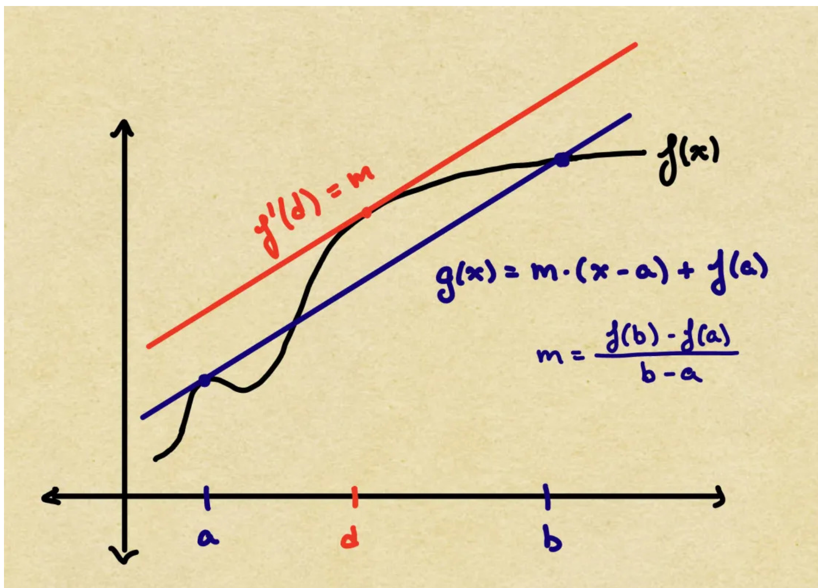

Figure 24: A depiction of the mean value theorem between a and b.

Proof. Let $m = \frac{f(b)-f(a)}{b-a}$, and consider the function $g(x) = m(x - a) + f(a)$; this is the unique linear function whose graph is the line passing through $(a, f(a))$ and $(b, f(b))$.

Then the function $(f - g)(x)$ satisfies

$$(f - g)(a) = f(a) - g(a) = f(a) - f(a) = 0$$

and similarly $(f - g)(b) = 0$. Since this is a sum of differentiable functions, it is differentiable, and Rolle's Theorem implies that there is a point $c \in (a, b)$ so that $(f - g)'(c) = 0$.

Expanding this, we find that $f'(c) - g'(c) = 0$; however, since g is linear, we have that

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

by Observation 4.2.1. It follows that

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

which was the desired result. ◻

The mean value theorem has loads of applications, not least of all the following useful fact, promised in Observation 4.2.1.

Corollary 4.2.10. *Suppose $f : [a, b] \to \mathbb{R}$ is a differentiable function on (a, b) so that $f' = m$ a constant function at some value m. Then f must be a linear function.*

Proof. Assume towards a contradiction that f is not a linear function. Consider the linear approximation of f at any point c, given by $g(x) = f(c) + m(x - c)$. Since f is not linear, there must exist some point $d \neq c \in (a, b)$ so that $f(d) \neq g(d)$, which in particular means that $f(d) - f(c) \neq g(d) - g(c)$ since $f(c) = g(c)$.

However, this implies that

$$\frac{f(d) - f(c)}{d - c} \neq \frac{g(d) - g(c)}{d - c} = m.$$

By the mean value theorem, there must exist some point L in between d and c so that $f'(L) = \frac{f(d)-f(c)}{d-c} \neq m$.

Contradiction! We assumed that f' was the constant function at m, and hence f must have been linear. ◻

4.2.3 Linear approximations after the mean value theorem

Let $f : [a, b] \to \mathbb{R}$ be a function that is differentiable on (a, b). Recall that the linear approximation of f at some point c is given by the function

$$g(x) = f(c) + f'(c)(x - c)$$

and for h sufficiently small (so that $x + h$ is in the domain of f) we have the relation

$$f(c + h) = g(c + h) + E(h) = f(c) + f'(c) \cdot h + E(h)$$

for $E(h)$ an error term to the linear approximation.

Observation 4.2.11. Given two points $c, c + h \in (a, b)$, the mean value theorem states that there is some point d in between c and $c + h$ so that

$$\frac{f(c + h) - f(c)}{h} = f'(d)$$

and in particular that $f(c + h) = f(c) + f'(d) \cdot h$

Thus, we have two expressions for $f(x + h)$ for a given h:

1. $f(c + h) = g(c + h) + E(h) = f(c) + f'(c) \cdot h + E(h)$
2. $f(c + h) = f(c) + f'(d) \cdot h$ for some d in between c and $c + h$

Expression (2) is called the *Lagrange error bound*. It says that even though there is an error term to the expression "$f(x+h) = f(c)+f'(c)\cdot h$", this is rectified if one works with a derivative of a point somewhere between x and $x + h$.

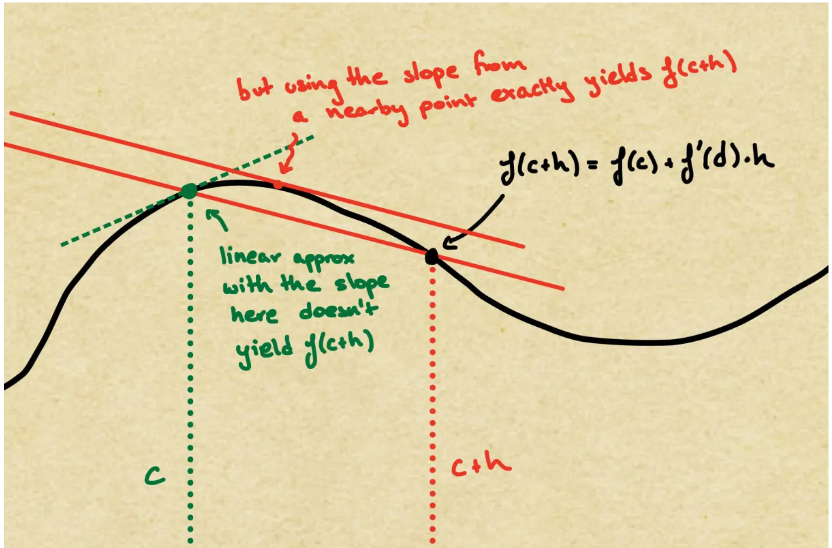

Figure 25: A depiction of the Lagrange error bound

If f' is a continuous function, then as h gets smaller, the values of $f'(d)$ become closer to $f'(c)$ for points d between c and $c+h$; this is another manifestation of the fact that the linear approximation to f is "a good approximation in a small neighborhood of c".

Remark 4.2.12. In the next class, we will explore how to use the derivative to do better than just linearly approximate: we are going to *polynomially* approximate!

4.3 Lecture 13: Taylor's Theorem and the Lagrange Error Term [10/15/25]

Learning Objectives

1. Derive the Taylor approximation to a smooth function f.
2. Prove Taylor's Theorem, and derive the Lagrange error term for the Taylor Approximation.

4.3.1 Class Activity: Linear and Quadratic Approximations

In this lecture, we will be making systematic use of the following resource: `https://www.desmos.com/calculator/xeevqhbmhw`. I recommend reopening this and fiddling with it while reviewing, and also changing the functions under consideration to see how the Taylor approximations behave.

Example 4.3.1. (Constant approximations and the Lagrange error term).

Consider a function $f : [a,b] \to \mathbb{R}$ which is differentiable on (a,b). At some point $c \in [a,b]$, the *constant approximation to f at c* is given by the function:

$$P_0 f(x) = f(c)$$

This is the "best constant function that approximates f" at c. In other words, $P_0 f$ is the unique function satisfying the following:

1. $g : [a, b] \to \mathbb{R}$ is constant
2. $g(c) = f(c)$

In the case $f(x) = \sqrt{x}$ we can see what this looks like in the following picture.

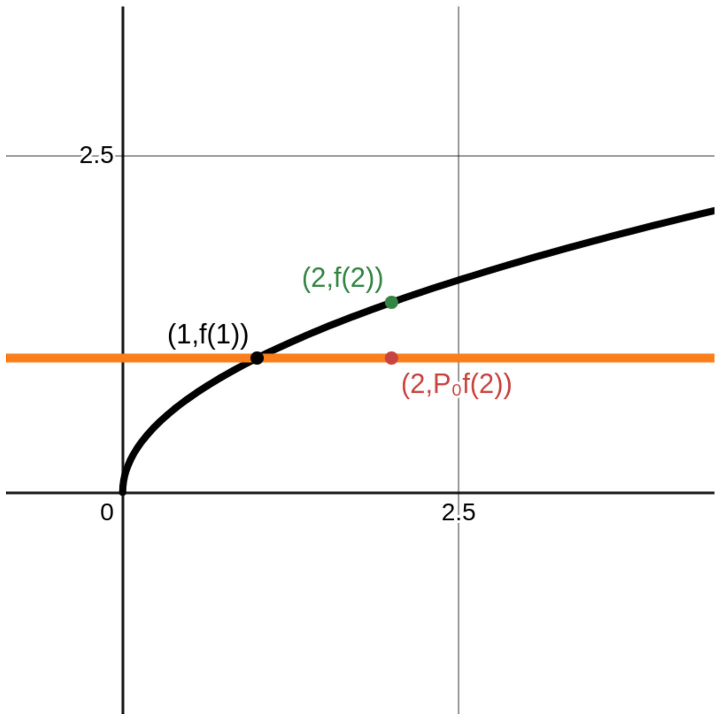

Figure 26: A depiction of the constant approximation to the function $f(x) = \sqrt{x}$ about the point $c = 1$. The constant approximation clearly agrees with f at the point 1, but differs at the point $x = 2$.

Recall Observation 4.2.11 from the previous lecture, which said that the difference between f and $P_0 f$ at some point $d \neq c \in [a, b]$ was expressible as

$$f(d) = P_0 f(d) + f'(z) \cdot (d - c) = f(c) + f'(z) \cdot (d - c)$$

for some z in between c and d. This is of course a restatement of the mean value theorem, but it indicates something strong: if we want a better approximation to f than just some constant function, it might help to incorporate $f'(c)$ in some way to our approximation.

Idea. If d is sufficiently close to c then $f'(d)$ is approximately equal to $f'(c)$, as long as f' is a continuous function.

Example 4.3.2 (Linear approximation to f at c)**.** The *linear approximation to f at c* was defined last lecture, and is given by the following function:

$$P_1 f(x) = f(c) + f'(c) \cdot (x - c)$$

This is the "best linear function approximating f" in that it satisfies the following:

1. $P_1 f(x)$ is a linear function.

2. $f(c) = P_1 f(c)$

3. $f'(c) = (P_1 f)'(c)$

In particular, if f *were* a linear function at c, then f and $P_1 f$ would just be equal functions. We depict their relationship in item 4.3.1.

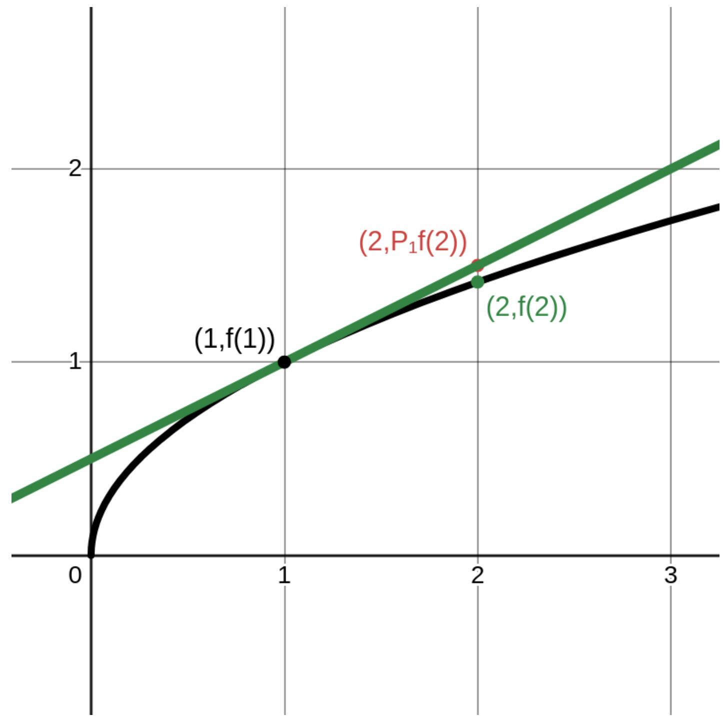

Figure 27: A depiction of the linear approximation to the function $f(x) = \sqrt{x}$ about the point $c = 1$. The constant approximation clearly agrees with f at the point 1, and has derivative equal to that of f at c, but differs at the point $x = 2$.

Recall that we arrived at the idea for considering linear approximations by considering the Lagrange error term for $f(x) - P_0(x)$. It stands to reason that the error term for the linear approximation P_1 should have something to do with a *higher-order approximation* of f. Namely, we could try to ask for a function P_2 which satisfies the following:

1. $P_2 f(x)$ is a quadratic polynomial function in $(x - c)$.

2. $f(c) = P_2 f(c)$

3. $f'(c) = (P_2 f)'(c)$

4. $f''(c) = (P_2 f)''(c)$ (as long as f'' actually exists, where we mean the derivative of f')

This perspective is /also/ suggested by the fact that if f admitted a *second derivative* f'', then the second derivative of the error term $(f - P_1 f)(x)$ is equal to $f''(x)$ (check this by taking second derivatives explicitly). Thus, the error term to the linear approximation itself seems to see *higher-order data*.

Example 4.3.3 (Quadratic approximation of f)**.** We may derive the expression for $(P_2 f)(x)$ directly from the desiderate listed above. Namely:

1. Criterion 1 means $P_2 f(x) = k_0 + k_1 \cdot (x - c) + k_2(x - c)^2$.
2. Criterion 2 forces $k_0 = f(c)$.
3. Taking derivatives, criterion 3 forces $k_1 = f'(c)$
4. Taking derivatives again, criterion 4 forces $k_2 = \frac{f''(c)}{2}$.

We depict $P_2 f(x) = f(c) + f'(c) \cdot (x - c) + \frac{f''(c)}{2} \cdot (x - c)^2$ the quadratic approximation about c, in item 4.3.1.

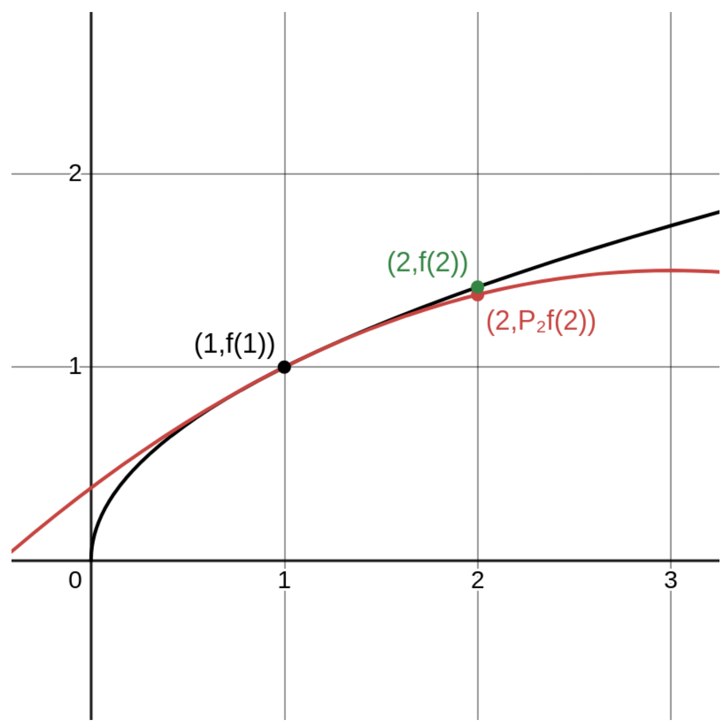

Figure 28: A depiction of the quadratic approximation to the function $f(x) = \sqrt{x}$ about the point $c = 1$. The quadratic approximation clearly agrees with f at the point 1, and has derivative equal to that of f at c, and will also have second derivative equal to that of f at c. It differs at the point $x = 2$, but it is visually a better approximation than $P_1 f$.

Discussion (Cubic and beyond)

Come up with an expression that represents the cubic approximation to f about c, based on a similar list of desiderata as above. Then try to devise what the general nth order approximation is.

4.3.2 Taylor Polynomials and Taylor's Theorem

Definition 4.3.4. We say a function $f : [a, b] \to \mathbb{R}$ is *smooth* if all iterated derivatives exist on (a, b).

Definition 4.3.5. Given a smooth function $f : [a, b] \to \mathbb{R}$, the *$n$th order Taylor polynomial* of f about c is the degree n polynomial given by

$$P_n f(x) = f(c) + f'(c) \cdot (x - c) + \frac{f''(c)}{2} \cdot (x - c)^2 + \ldots + \frac{f^{(n)}(x)}{n!} \cdot (x - c)^n$$

In summation notation, $P_n f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!} \cdot (x - c)^k$

We now turn to the component we suppressed for the linear and higher order approximations: namely, a generalized expression of the error term $f(d) - P_n f(d)$ at some point $d \neq c$ in (a, b). This will be of a very similar form to the constant error term, and is proved through an iterated application of the mean value theorem.

Key Theorem 4.3.6 (Taylor's Theorem)**.**

Let $f : [a, b] \to \mathbb{R}$ be a smooth function, and $P_n f$ the nth order Taylor polynomial of f at some $c \in (a, b)$. Then for any $d \in (a, b)$ with $d \neq c$, we have that the error term

$$f(x) - P_n f(x) = \frac{f^{(n+1)}(z)}{(n+1)!} \cdot (c - d)^{n+1}$$

for some z between c and d.

The following proof is the one described in [Slo25, §6.6], see `https://math.libretexts.org/Bookshelves/Analysis/A_Primer_of_Real_Analysis_(Sloughter)/06%3A_Derivatives/6.06%3A_Taylor's_Theorem`.

Proof. Write $P(x)$ for $P_n f(x)$. First note that $P^{(k)}(c) = f^{(k)}(c)$ for $k = 0, 1, \ldots, n$. Let

$$M = \frac{f(d) - P(d)}{(d - c)^{n+1}}.$$

Then

$$f(d) = P(d) + M(d - c)^{n+1}.$$

We need to show that

$$M = \frac{f^{(n+1)}(z)}{(n+1)!}$$

for some z between c and d. Let

$$g(x) = f(x) - P(x) - M(x - c)^{n+1}.$$

Then, for $k = 0, 1, \ldots, n$,

$$g^{(k)}(c) = f^{(k)}(c) - P^{(k)}(c) = 0.$$

Now $g(d) = 0$, so, by Rolle's theorem Theorem 4.2.8, there exists z_1 between c and d such that $g'(z_1) = 0$. Using Rolle's theorem again, we see that there exists z_2 between c and z_1 such that $g''(z_2) = 0$. Continuing for $n+1$ steps, we find z_{n+1} between c and z_n (and hence between c and d) such that $g^{(n+1)}(z_{n+1}) = 0$. Hence

$$0 = g^{(n+1)}(z_{n+1}) = f^{(n+1)}(z_{n+1}) - (n+1)!M.$$

Letting $z = z_{n+1}$, we have

$$M = \frac{f^{(n+1)}(z)}{(n+1)!},$$

as required. □

Definition 4.3.7. The term

$$\frac{f^{(n+1)}(z)}{(n+1)!} \cdot (c-d)^{n+1}$$

in the proof above is known as the *Lagrange error term* for the nth order Taylor approximation.

4.4 Homework 4 (Due Friday, October 24th)

3 out of the following 6 exercises graded for correctness, the remainder will be graded for completeness. I have marked the graded ones with an asterisk $(*)$. This homework is out of a possible **24 points**, with the graded exercises worth 6 points and the ungraded exercises worth 2 points each.

Finally, don't expect to be able to do everything in this homework immediately! I expect you to return to this once or twice with a group as we progress through the course over the next two weeks.

Exercise 4.4.1. *Show that the function $f(x) = sin(x)$ is differentiable on all of $\mathbb{R}$, then compute its derivative. (You'll need to use the formula for* sin *of the sum of angles).*

Exercise 4.4.2. $(*)$ *Do* [Abb15, Exercise 5.2.12].

Exercise 4.4.3. $(*)$ *Do part (a) of* [Abb15, Exercise 5.3.5]. *Using this, do both parts of* [Abb15, Exercise 5.3.11].

Exercise 4.4.4. *Suppose $f : [0, a] \to \mathbb{R}$ is n-times differentiable, $f(0) = f'(0) = \ldots = f^{(n-1)}(0) = 0$, and $|f^{(n)}(x)| \leq M$ for all $x \in [0, a]$. Show that*

$$|f(x)| \leq \frac{M}{n!} x^n$$

for every $x \in [0, a]$.

Exercise 4.4.5 (Some further tests for convergence)**.** $(*)$ *Prove the following for infinite series.*

1. *Prove the limit comparison test for infinite series: if $\{a_n\}$ and $\{b_n\}$ are two sequences both of which are greater than 0, and satisfy*

$$0 < \lim_{n\to\infty} \frac{a_n}{b_n} < \infty$$

   then the infinite series $\sum a_n$ converges if and only if $\sum b_n$ converges.

2. *Prove the ratio test for infinite series: if $\{a_n\}$ is any sequence satisfying*

$$\left| \lim_{n\to\infty} \frac{a_{n+1}}{a_n} \right| = r < 1$$

   then the infinite series $\sum a_n$ converges, and in fact converges absolutely.

3. *Prove the root test for infinite series: if $\{a_n\}$ is a sequence satisfying*

$$\lim_{n\to\infty} |a_n|^{\frac{1}{n}} < 1$$

   Then the series $\sum a_n$ converges. Show that if this limit is greater than 1, then the series diverges.

Exercise 4.4.6. *Show Cauchy's convolution formula: if* $\sum a_n$ *and* $\sum b_n$ *are two convergent series with limits* A *and* B*, then one has the expression*

$$\sum a_n \cdot \sum b_n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k \cdot b_{n-k} \right)$$

In particular, the series on the right converges to the limit $A \cdot B$*.*

5 Interlude: Sequences of Functions

5.1 Lecture 14: Pointwise and uniform convergence [10/20/25]

Learning Objectives

Understand how to treat with supertasks involving functions.

Last class, we discussed how to approximate smooth functions "up to nth order behaviour" using the Taylor polynomial. In this brief section, we want to discuss the extent to which we are allowed to make statements like:

$$f(x) = \lim_{n\to\infty} P_n f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!} \cdot (x-c)^k = f(c) + f'(c)\cdot(x-c) + \frac{f''(c)}{2}\cdot(x-c)^2 + \ldots$$

However, there's a concrete problem to address.

Problem 1. The series

$$P(x) := \sum_{k=0}^{\infty} \left(\frac{f^{(k)}(c)}{k!} \cdot (x-c)^k \right) = f(c) + f'(c)\cdot(x-c) + \frac{f''(c)}{2}\cdot(x-c)^2 + \ldots$$

might not converge for every value of x.

Of course, we've already dealt with Problem 1 before: this is just a collection of supertasks. Note that the series $P(x)$ depends on an input value of x, and by our previous work on supertasks we can apply some tests to decide which x-values give a well-defined sum for $P(x)$ and which don't. Let's formalize what's happening.

Definition 5.1.1. A *sequence of functions* $\{f_n\} : [a,b] \to \mathbb{R}$ is an infinite list of real-valued functions $\{f_1(x), f_2(x), f_3(x), \ldots\}$ each of which is defined on $[a,b]$.

Given a sequence of functions $\{f_n\} : [a,b] \to \mathbb{R}$, every element $x_0 \in [a,b]$ gives rise to a sequence of real numbers $\{f_n(x_0)\}_{n\in\mathbb{N}}$ just by evaluating each function on the element x_0. A first definition of convergence/Cauchy might now seem obvious.

Definition 5.1.2. Let $\{f_n\} : [a,b] \to \mathbb{R}$ be a sequence of functions.

1. We say the sequence $\{f_n\}$ *pointwise converges* to a function $f : [a,b] \to \mathbb{R}$ if for every choice of $x_0 \in [a,b]$, the sequences of *real numbers* $\{f_n(x_0)\}$ converge to the real number $f(x_0)$.
2. We say the sequence $\{f_n\}$ is *pointwise Cauchy* if for every choice of $x_0 \in [a,b]$, the sequences $\{f_n(x_0)\}$ are Cauchy sequences of real numbers.

Exercise 5.1.3. *Show that every pointwise Cauchy sequence of functions pointwise converges to a real function, and that every pointwise convergent sequence of functions is pointwise Cauchy.*

5.1.1 Pointwise doesn't accomplish everything

This might seem to solve all of our problems summarily: however, there's some issues with this definition.

Example 5.1.4 (Pointwise Cauchy sequences of functions may not have continuous limits)**.** Consider the sequence of functions $\{f_n = x^n\} : [0, 1] \to \mathbb{R}$, roughly illustrated below.

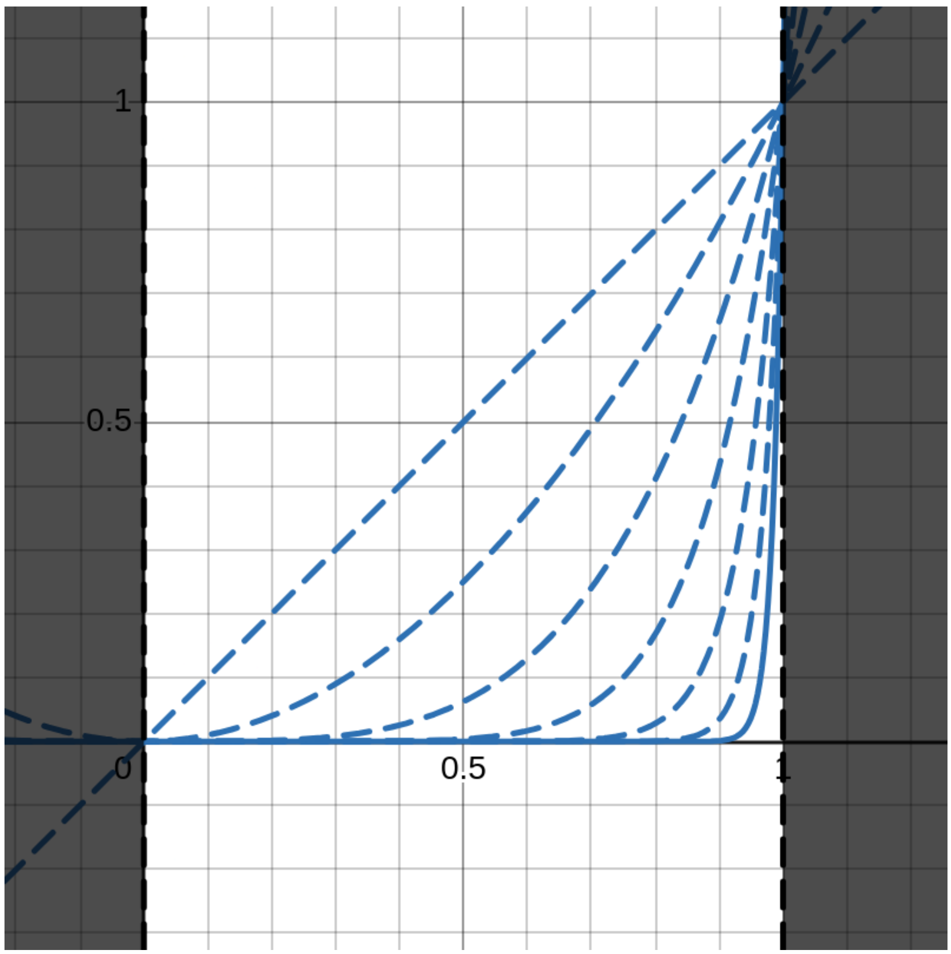

Figure 29: A depiction of the sequence of functions $\{x^n\}$ for $n = 1, 2, 4, 8, 16, 32$ and 64, on the interval $[0, 1]$

Clearly, the above sequence pointwise converges to the piecewise function

$$f = \begin{cases} 0 & 0 \le x < 1 \\ 1 & x = 1 \end{cases}$$

which you can manually convince yourself of by evaluating the Cauchy sequences yourself. Clearly, its limit is not continuous at the point 1. What exactly happened here?

Observation 5.1.5. In the sequence of functions above, the limit of the sequence $\{x^n\}$ at any point $1-\epsilon$ for $\epsilon > 0$ and the limit at 1 are different. This was enabled by the following behaviour:

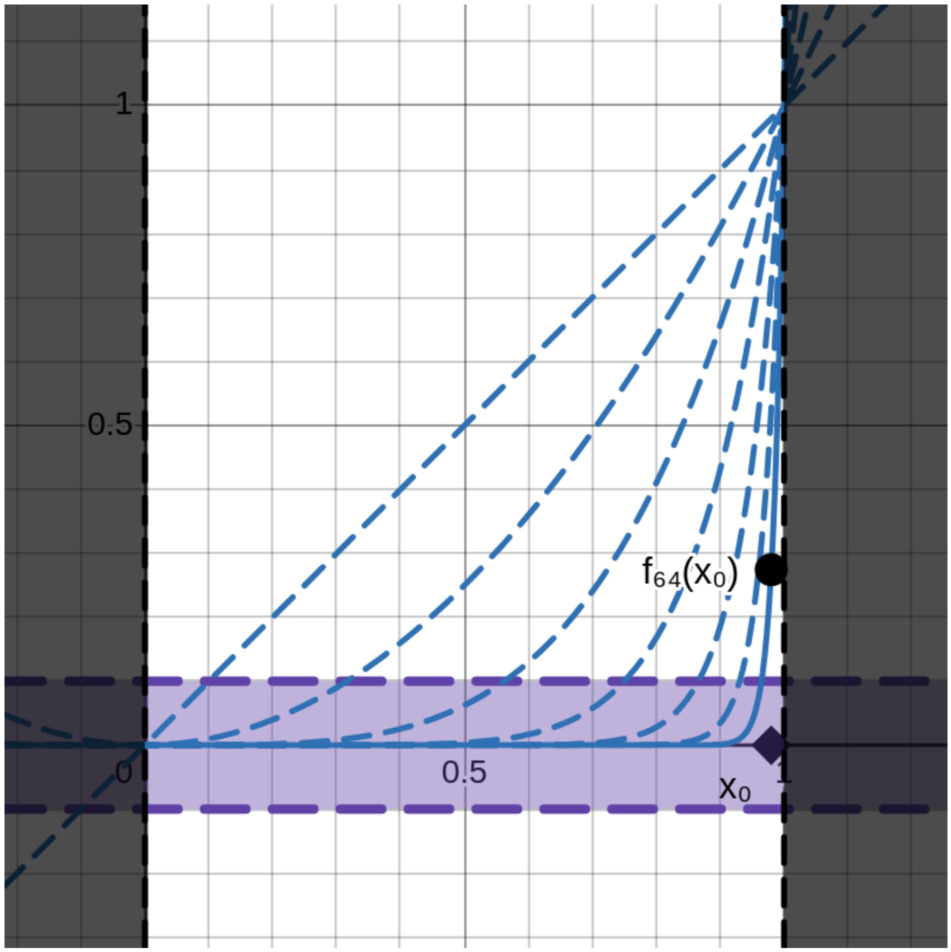

Figure 30: A depiction of the sequence of functions $\{x^n\}$ for $n = 1, 2, 4, 8, 16, 32$ and 64, on the interval $[0, 1]$. For any $\epsilon > 0$ and $N \in \mathbb{N}$, there exists some $m > N$ and $x_0 \in [0, 1]$ so that $|f_m(x_0) - y| > \epsilon$.

The problem essentially appears from the fact that even if the $\{f_n\}$ are all continuous, the sequences $f_n(1)$ may have a different limit than $f_n(1-\epsilon)$ because any function in the sequence f_n can always "sharply dip back up to 1" in order to ensure that the limit at 1 is still 1. This "sharply dipping upward" behaviour should seem very reminiscent of *uniform continuity* from the last section. We'll exclude this behaviour by enforcing a similar condition.

5.1.2 Uniform convergence

Key Definition 5.1.6 Uniform convergence.

Let $\{f_n\} : [a, b] \to \mathbb{R}$ be a sequence of functions.

1. We say that $\{f_n\}$ *uniformly converges* to a function $f : [a, b] \to \mathbb{R}$ if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ so that for every $m > N$ and $x \in [a, b]$, we have that $|f(x) - f_m(x)| < \epsilon$.
2. We say that $\{f_n\}$ is *uniformly Cauchy* if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ so that for every $m, n > N$ and $x \in [a, b]$, we have that $|f_m(x) - f_n(x)| < \epsilon$.

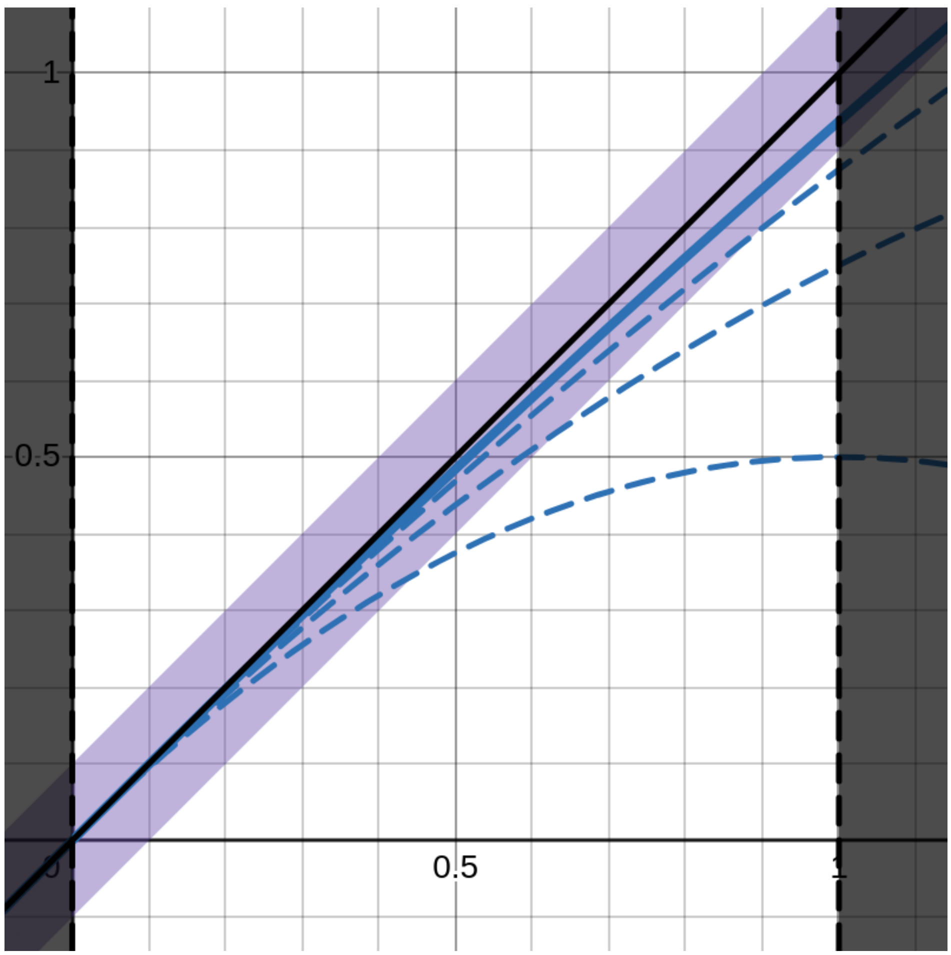

Figure 31: A depiction of the uniformly convergent sequence of functions $\{x - 2^{-n}x\}$ along with its limit $f(x) = x$. For any $\epsilon > 0$ and $N \in \mathbb{N}$, there exists some $m > N$ so that for every $x \in [0, 1]$, we have $|f_m(x) - x| > \epsilon$.

Exercise 5.1.7. *Show that any uniformly Cauchy sequence of functions is uniformly convergent to a limit, and that any uniformly convergent sequence of functions is uniformly Cauchy.*

Armed with this definition, we have the following anticipated closure property.

Key Theorem 5.1.8 (Continuous Limit Theorem)**.**

Let $\{f_n\} : [a, b] \to \mathbb{R}$ be a uniformly convergent sequence of functions with limit $f : [a, b] \to \mathbb{R}$. If every f_n is continuous, then the limit f is continuous.

Proof. We want to show continuity of f, i.e., that for any $c \in [a, b]$ and $\epsilon > 0$, there exists a $\delta > 0$ so that if $|x - c| < \delta$ then $|f(x) - f(c)| < \epsilon$. Fix $\epsilon > 0$.

For any $n \in \mathbb{N}$, we have that for any $x \in [a, b]$

$$\begin{aligned} |f(x) - f(c)| &= |f(x) - f_n(x) + f_n(x) - f_n(c) + f_n(c) - f(c)| \\ \text{(By Triangle Inequality)} \quad &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)| \end{aligned}$$

By uniform convergence of f_n, we have that there exists some $N \in \mathbb{N}$ so that $|f_n(x) - f(x)| < \frac{\epsilon}{3}$ for *every* possible choice of x in $[a, b]$. In particular, this implies that

$$\begin{aligned} |f(x) - f(c)| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| \\ &< \frac{\epsilon}{3} + |f_N(x) - f_N(c)| + \frac{\epsilon}{3} \end{aligned}$$

It remains to take care of the middle term, for which we may use the continuity of f_N; i.e., there exists a $\delta_N > 0$ so that $|x - c| < \delta_N$ implies that $|f_N(x) - f(c)| < \frac{\epsilon}{3}$. In particular setting $\delta = \delta_N$, we have showed that if $|x - c| < \delta$ then

$$|f(x) - f(c)| < \frac{\epsilon}{3} + |f_N(x) - f_N(c)| + \frac{\epsilon}{3} = \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

which is the desired result. □

5.1.3 Uniform convergence and differentiability

A variation of the previous proof will provide us with the following incredibly useful result.

Key Theorem 5.1.9 (Differentiable Limit Theorem)**.**

Let $\{f_n\} : [a, b] \to \mathbb{R}$ be a pointwise convergent sequence of functions with limit $f : [a, b] \to \mathbb{R}$. Suppose that:

1. Each f_n is differentiable.
2. The sequence of derivatives $\{f_n'\}$ is uniformly convergent with limit g.

Then the limit f is also differentiable, and moreover $f' = g$.

Proof. We would like to show the differentiability of f at any point $c \in [a, b]$. Concretely, we must show that: for every $\epsilon > 0$, there exists a $\delta > 0$ so that if $|x - c| < \delta$ then

$$\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| < \epsilon$$

For any $n \in \mathbb{N}$ and any $x \in [a, b]$, we have that

$$\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| = \left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} + \frac{f_n(x) - f_n(c)}{x - c} - f_n'(c) + f_n'(c) - g(c) \right|$$

$$\text{By Triangle Inequality } \leq \left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right| + \left| \frac{f_n(x) - f_n(c)}{x - c} - f_n'(c) \right| + |f_n'(c) - g(c)|$$

We claim that the second and third terms are entirely manageable. First, since $\{f_n'\}$ uniformly converges to g, there exists an $N \in \mathbb{N}$ so that if $m > N$ $|f_m'(c) - g(c)| < \frac{\epsilon}{3}$ irrespective of what c is.

Second, since f_m is assumed to be differentiable, there exists a $\delta_m > 0$ so that if $|x - c| < \delta_m$ then

$$\left| \frac{f_n(x) - f_n(c)}{x - c} - f_n'(c) \right| < \frac{\epsilon}{3}.$$

Altogether, we have that if $|x - c| < \delta_m$ then

$$\begin{aligned}\left| \frac{f(x) - f(c)}{x - c} - g(c) \right| &\leq \left| \frac{f(x) - f(c)}{x - c} - \frac{f_m(x) - f_m(c)}{x - c} \right| + \left| \frac{f_m(x) - f_m(c)}{x - c} - f_m'(c) \right| + |f_m'(c) - g(c)| \\ &< \left| \frac{f(x) - f(c)}{x - c} - \frac{f_m(x) - f_m(c)}{x - c} \right| + \frac{\epsilon}{3} + \frac{\epsilon}{3}\end{aligned}$$

Warning 5.1.10.

At this point we might even be done, if we had assumed that $\{f_n\}$ actually uniformly converged to f instead of just pointwise. Our plan is to redistribute:

$$\left| \frac{f(x) - f(c)}{x - c} - \frac{f_n(x) - f_n(c)}{x - c} \right| = \left| \frac{(f(x) - f_n(x)) - (f(c) - f_n(c))}{x - c} \right|$$

and if we had assumed that f_n uniformly converged to f then we could pick a large N so that for *any* choice of $x \in (c - \delta_m, c + \delta_m)$, we can make

$$|f(x) - f_m(x)| < \frac{\epsilon \cdot |x - c|}{6}$$

However, we did NOT assume uniform convergence, and this makes our work a little bit more challenging; we must show that a large choice of N works for every possible $x \in (c - \delta, c + \delta)$.

Returning to our proof, it remains to demonstrate that we may pick an $m > N$ large enough so that

$$\left| \frac{(f(x) - f_n(x)) - (f(c) - f_n(c))}{x - c} \right| < \frac{\epsilon}{3}$$

However, since f_n pointwise converges to f, it is equivalent to show that there exists an $m > N$ so that if $n_1, n_2 > m$ then

$$\left| \frac{(f_{n_1}(x) - f_{n_2}(x)) - (f_{n_1}(c) - f_{n_2}(c))}{x - c} \right| < \frac{\epsilon}{3}$$

because the result above follows by taking the limit as $n_1 \longrightarrow \infty$ and using Exercise 3.2.10. Now, using the Mean Value Theorem, we have that for any $x = x_0$ there exists some α between x_0 and c so that

$$f'_{n_1}(\alpha) - f'_{n_2}(\alpha) = \frac{(f_{n_1}(x) - f_{n_2}(x)) - (f_{n_1}(c) - f_{n_2}(c))}{x - c}$$

In particular, by the uniform convergence of $\{f'_n\}$, we may pick a sufficiently large $m > N$ so that the term on the left satisfies $|f'_{n_1}(\alpha) - f'_{n_2}(\alpha)| < \frac{\epsilon}{3}$ for every $n_1, n_2 > m$ and thus the term on the right hand side does as well. Altogether, this implies that for this choice of m, one has

$$\left| \frac{f(x) - f(c)}{x - c} - \frac{f_m(x) - f_m(c)}{x - c} \right| < \frac{\epsilon}{3}$$

and thus for $\delta = \delta_m$, we have that $|x - c| < \delta$ implies that

$$\begin{aligned} \left| \frac{f(x) - f(c)}{x - c} - g(c) \right| &< \left| \frac{f(x) - f(c)}{x - c} - \frac{f_m(x) - f_m(c)}{x - c} \right| + \frac{\epsilon}{3} + \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

finishing the claim. □

This is perhaps our hardest result so far, but the results will be incredibly rewarding both in this and in the next sections.

5.2 Lecture 15: Power series and e^x [10/22/25]

The key examples of sequences of functions that we are interested in arise from the partial sums of Taylor series. Before we discuss this particular class of series of functions, let us collect some general results on series of functions from the results of the previous lecture.

Definition 5.2.1. A *series of functions* is a sequence of functions of the form $\{s_n\} = \sum_{k=0}^{n} f_k$ for a sequence $\{f_k\} : [a, b] \longrightarrow \mathbb{R}$. Informally, we write $s = \sum_{n=0}^{\infty} f_n$.

Theorem 5.2.2 (Cauchy criterion for series of functions)**.** *Let $s = \sum_{n=0}^{\infty} f_n$ be a series of functions from $[a, b]$ to $\mathbb{R}$. Then s converges uniformly (that is, the sequence of partial sums s_n converges uniformly) on $[a, b]$ if and only if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ so that for every $m > N, k > 0$, the following sum*

$$|f_m(x) + \ldots + f_{m+k}(x)| < \epsilon$$

for any choice of $x \in [a, b]$.

Proof. This is an immediate consequence of Exercise 5.1.7. □

Corollary 5.2.3 (Weierstrass M-test)**.** *Let $s = \sum_{n=0}^{\infty} f_n$ be a series of functions from $[a, b]$ to $\mathbb{R}$, Suppose there exists a sequence $M_n > 0$ such that for each $n \in \mathbb{N}$, we have*

$$|f_n(x)| \leq M_n$$

for every $x \in [a, b]$, and moreover the series of real numbers $\sum_{n=0}^{\infty} M_n$ converges. Then the series $\sum_{n=0}^{\infty} f_n$ converges uniformly.

Proof. By the Cauchy criterion, we must show that for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ so that for every $m > N, k > 0$, the following sum

$$|f_m(x) + \ldots + f_{m+k}(x)| < \epsilon$$

for any choice of $x \in [a, b]$. However, we have that the convergence of the series $\sum_{n=0}^{\infty} M_n$ implies that for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ so that for every $m > N, k > 0$, the following sum

$$|M_m + \ldots + M_{m+k}| = M_m + \ldots + M_{m+k} < \epsilon.$$

In particular, applying the triangle inequality, we have that

$$|f_m(x) + \ldots + f_{m+k}(x)| \leq |f_m(x)| + \ldots + |f_{m+k}(x)| \leq M_m + \ldots + M_k < \epsilon$$

yielding the Cauchy criterion and hence the claim. □

5.2.1 Aside: term-by-term continuity differentiation of uniformly convergent series

The following are immediate consequences of the Continuous and Differentiable Limit Theorems.

Key Theorem 5.2.4 Term-by-term continuity theorem.

Let $s = \sum_{n=0}^{\infty} f_n$ be a series of continuous functions from $[a, b]$ to $\mathbb{R}$. If s converges uniformly, then the function $s : [a, b] \longrightarrow \mathbb{R}$ given by

$$s(x_0) = \sum_{n=0}^{\infty} f_n(x_0)$$

is a continuous function.

Key Theorem 5.2.5 Term-by-term differentiability theorem.

Let $s = \sum_{n=0}^{\infty} f_n$ be a series of differentiable functions from $[a, b]$ to $\mathbb{R}$ which converges pointwise, and such that the series of derivatives $t = \sum_{n=0}^{\infty} f_n'$ converges uniformly. Then the function

$$s(x_0) = \sum_{n=0}^{\infty} f_n(x_0)$$

is differentiable, and the derivative s' is identified with the limit of the series of derivatives t.

5.2.2 Power series

Recall that the kinds of function series that we are interested in ultimately arise from Taylor series. These form a special kind of series of functions with particularly describable behaviour.

Key Definition 5.2.6 Power Series**.**

A *power series centered around* $c \in \mathbb{R}$ is a series of functions of the following form:

$$s = \sum_{n=0}^{\infty} a_k \cdot (x - c)^n$$

for a_k some sequence of real numbers.

Convergence of power series is incredibly well behaved, as the following theorem shows.

Recollection 5.2.7. Recall the notion of *absolute convergence* from Exercise 2.5.4.

Key Theorem 5.2.8 Power series converge uniformly**.**

Let $s = \sum_{n=0}^{\infty} a_k \cdot x^n$ be a power series centered around 0. If s converges for some $x_0 \in \mathbb{R}$, then s converges absolutely on the open interval $(-x_0, x_0)$. Moreover, it converges uniformly on any closed subset $[-c, c] \subseteq (-x_0, x_0)$.

Proof. If s converges for some $x_0 \in \mathbb{R}$ then by the ratio test Exercise 4.4.5 we have that

$$\lim_{n\to\infty} \left| \frac{a_{n+1} x_0^{n+1}}{a_n x_0^n} \right| \leq 1.$$

In particular, if $|x| < |x_0|$ then we have that

$$\lim_{n\to\infty} \left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| \cdot |x| < \left| \frac{a_{n+1}}{a_n} \right| \cdot |x_0| \leq 1$$

and hence by the ratio test we have that the series $\sum_{k=0}^{\infty} |a_k x^k|$ converges for any $|x| < |x_0|$.

Now to demonstrate uniformity on closed sub-intervals: notice that if $[-c, c] \subseteq (-x_0, x_0)$ is a closed sub-interval then $|c| < |x_0|$, from which we have that $\sum_{k=0}^{\infty} |a_k x^k|$ converges. Now we may apply the Weierstrass M-test on the closed interval $[-c, c]$; for any $x \in [-c, c]$, we have that

$$|a_k x^k| < |a_k c^k|$$

for all k, and moreover that $\sum_{k=0}^{\infty} |a_k x^k|$. Letting $M_k = |a_k c^k|$, we conclude. □

Exercise 5.2.9. *Show the straightforward generalization of the above to power series centered around* c*: namely, show that if* $s = \sum_{n=0}^{\infty} a_k \cdot (x - c)^n$ *is a power series centered around* c *which converges for some* $x_0 \in \mathbb{R}$*, then* s *converges absolutely on* $(c - r, c + r)$ *where* $r = |x_0 - c|$*, and uniformly on closed sub-intervals.*

Definition 5.2.10. Given a power series $s = \sum_{n=0}^{\infty} a_k \cdot (x - c)^n$ centered around c, the largest r for which s converges on $(c - r, c + r)$ is known as its *radius of convergence*. This interval itself is known as the *interval of convergence*.

5.3 Application: e^x

Let's come up with a function f so that $f' = f$.

Observation 5.3.1. First, if f exists, it is clearly smooth on its domain: $f' = f$ so all higher derivatives exists an are equal to f.

Observation 5.3.2. Second, if f exists, it is necessarily analytic. Suppose f is defined on an interval $[a, b]$, then for any $x, c \in [a, b]$ the Lagrange error term for $(f - P_n f)(x)$ (where $P_n f$ is the nth Taylor polynomial around c) is given by:

$$\frac{f^{(n+1)}(d)}{n+1!} \cdot (x-c)^{n+1} = \frac{f(d)}{n+1!} \cdot (x-c)^{n+1}$$

for some point d between x and c. Since f is continuous on $[a, b]$, it is bounded, and we thus have that

$$\left|\frac{f(d)}{n+1!} \cdot (x-c)^{n+1}\right| \leq \left|\frac{M}{n+1!} \cdot (x-c)^{n+1}\right|$$

However, as n goes to ∞, it is clear that the sequence $\left|\frac{M}{n+1!} \cdot (x-c)^{n+1}\right|$ converges to 0, as the factorial function grows much faster than the exponential function[6]. In particular, the sequence of functions

$$\{f - P_n f\} \xrightarrow{n\to\infty} 0$$

for every x in $[a, b]$, and in particular f is uniformly approximated by its Taylor series on some closed interval in $[a, b]$.

The argument above tells us that any such function f is of the following form:

$$f(x) = f(c) + f(c) \cdot (x-c) + \frac{f(c) \cdot (x-c)^2}{2!} + \frac{f(c) \cdot (x-c)^3}{3!} + \ldots$$

Setting $c = 0$ and dividing by $f(0)$, we obtain the function

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

which is continuous and differentiable by using the same argument as in the previous observation for uniform convergence both of the series and of its derivatives. Moreover, it is continuous and differentiable *everywhere* on $\mathbb{R}$, because the above series of functions converges uniformly everywhere on $\mathbb{R}$. This gives us a full characterization of the solution $f = f'$ on any interval, and also on $\mathbb{R}$. But let's do a little bit more.

Observation 5.3.3. We claim that $\exp(x)$ as above satisfies $\exp(x + y) = \exp(x) \cdot \exp(y)$. This can be seen as follows:

[6] easy argument by just going to n large enough so that $n!$ contains many factors larger than $(x-c)^{n+1}$

$$\exp(x+y)=\sum_{n=0}^{\infty}\frac{(x+y)^n}{n!}=\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n}\binom{n}{k}\cdot\frac{x^{n-k}y^k}{n!}\right)=\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n}\frac{x^{n-k}y^k}{k!(n-k)!}\right)$$

by the binomial theorem. Similarly, by Cauchy convolution Exercise 4.4.6, we have that $f(x)\cdot f(y)$ is given by

$$\exp(x)\cdot\exp(y)=\left(\sum_{n=0}^{\infty}\frac{x^n}{n!}\right)\cdot\left(\sum_{n=0}^{\infty}\frac{y^n}{n!}\right)=\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n}\frac{x^{n-k}}{n-k!}\cdot\frac{y^k}{k!}\right)$$

by just directly applying the formula. By inspection, we see that $\exp(x+y)=\exp(x)\cdot\exp(y)$

Exercise 5.3.4. *Show that any continuous function $g:\mathbb{R}\to\mathbb{R}$ satisfying $g(x+y)=g(x)\cdot g(y)$ is given by $g(1)^x$. Start by showing that $g(n)=g(1)^n$ for natural numbers. Then show that $g(\frac{1}{n})=g(1)^{\frac{1}{n}}$. Use this to show that $g(x)=g(1)^x$ on all rational numbers, and thus must agree on all real numbers by continuity.*

It follows that the function

$$\exp(x)=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots$$

Is actually equal to

$$\exp(1)^x=\left(1+1+\frac{1^2}{2!}+\frac{x^3}{3!}+\ldots\right)^x$$

We write the letter e to denote exp(1). From the above, we have just witnessed e^x as the basic solution to the equation

$$\frac{d}{dx}e^x=e^x.$$

I believe this was Euler's original reason for happening across this function, and is the cleanest proof I know for this fact.

6 The Riemann Integral

6.1 Lecture 16: Definition of the Riemann Integral [10/29/25]

Learning Objectives

Motivate the Riemann Integral as a measure of total change. Then motivate the Fundamental Theorem of Calculus.

6.1.1 Introduction and motivation

Given a function $f : [a, b] \to \mathbb{R}$ which is differentiable on (a, b), we would like to recover the net change of the function between the values a and b utilising the derivative of the function f. Namely, if the derivative is a measure of *instantaneous* change, then performing an "infinite sum" over all of the instantaneous changes recorded by f' should in principle recover the total change of f between a and b. Let's remind ourselves of how this works for a linear function.

Example 6.1.1. Given a linear function $f(x)$ of slope m, the difference between $f(a)$ and $f(b)$ between points a and b is exactly computed by solving:

$$\frac{f(b) - f(a)}{b - a} = m \implies f(b) - f(a) = m \cdot (b - a)$$

Graphically, this is depicted by the following:

Figure 32: A depiction of the total change between a and b as a measure of the area under $f'(x) = m$

The idea behind integration is essentially the same as the idea behind differentiation; if the derivative f' indicates a good linear approximation to f at a given point c, it stands to reason that $f'(c) \cdot (c - c')$ becomes a good approximation to $f(c') - f(c)$. Indeed, the mean value theorem tells us that there is some z in between c and c' so that $f'(z) \cdot (c - c')$ is actually *equal* to $f(c) - f(c')$; in particular, this is the observation which will power the proof of our main result later in this chapter, the Fundamental Theorem of Calculus.

6.1.2 Defining the Riemann Integral

If we try to sketch out a schematic of the precise thing we're trying to do, we'll quickly notice that it looks like we're computing the area under f':

Figure 33: A depiction of process of subdividing the interval $[a, b]$ into subintervals

6.2 Lecture 17: Fundamental Theorem of Calculus I, Basic Properties of the Riemann Integral [11/3/25]

Learning Objectives

1. Prove that the integral of the derivative of a function measures total its total change: i.e., the Fundamental Theorem of Calculus I.

2. Collect other basic features of the Riemann integral, some of which are demonstrated in this week's homework.

Our first objective is to show the following.

Key Theorem 6.2.1 (Fundamental Theorem of Calculus I)**.**

Suppose that $f : [a,b] \longrightarrow \mathbb{R}$ is integrable, and $F : [a,b] \longrightarrow \mathbb{R}$ satisfies $F'(x) = f(x)$ for all $x \in [a,b]$, then

$$\int_a^b f = F(b) - F(b).$$

Proof. Let $P = \{a = x_0 < \cdots < x_n = b\} \subseteq [a,b]$ be a partition. Applying the Mean Value Theorem to F on any subinterval $[x_{i-1}, x_i]$ of P, one obtains a point $t_i \in (x_{i-1}, x_i)$ such that

$$\begin{aligned} F(x_i) - F(x_{i-1}) = F'(t_i) \cdot (x_i - x_{i-1}) \\ f(t_i) \cdot (x_i - x_{i-1}). \end{aligned}$$

Now for every interval $[x_{i-1}, x_i]$ let m_i denote the infimum of F and M_i denote the supremum of f on this interval. By definition, we learn that $m_i \leq f(t_i) \leq M_i$ for every i. It follows that we have

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

Using that $f(t_i) \cdot (x_i - x_{i-1}) = F(x_i) - F(x_{i-1})$, we learn that

$$L(f,P) \leq \sum_{i=1}^{n} F(x_i) - F(x_{i-1}) \leq U(f,P).$$

However, the middle sum telescopes to gives $F(b) - F(a)$. It follows that for every partition P, we have that

$$L(f,P) \leq F(b) - F(a) \leq U(f,P)$$

from which it follows that

$$U(f) := \sup_{P \text{ partitions}} L(f,P) \leq F(b) - F(a) \leq \inf_{P \text{ partitions}} U(f,P) =: U(f)$$

By integrability, $L(f) = U(f)$ and we conclude. $\square$

Let us now collect some basic facts about integration which will be used next time.

Theorem 6.2.2. *Assume $f : [a, b] \to \mathbb{R}$ is bounded, and let $c \in (a, b)$ be an arbitrary point. Then f is integrable on $[a, b]$ if and only if f is integrable on $[a, c]$ and on $[c, b]$. In this case, we have that*

$$\int_a^b f = \int_a^c f + \int_c^b f$$

Proof. Standard partition manipulation, see the proof of [Abb15, Theorem 7.4.1]. □

Key Theorem 6.2.3 (Algebraic Integral Theorem)**.**

Assume f and g are Riemann integrable functions from $[a, b] \to \mathbb{R}$. Then:

1. The function $f + g$ is integrable on $[a, b]$, and

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g.$$

2. Given any constant $k \in \mathbb{R}$, the function $k \cdot f$ is integrable and $\int_a^b k \cdot f = k \cdot \int_a^b f$.

3. If $m \leq f(x) \leq M$ for every $x \in [a, b]$, then

$$m \cdot (b - a) \leq \int_a^b f(x) \leq M \cdot (b - a).$$

4. If $f(x) \leq g(x)$ for every $x \in [a, b]$ then $\int_a^b f \leq \int_a^b g$.

5. The function $|f|$ given by $|f|(x) = |f(x)|$ is integrable, and

$$\left| \int_a^b f \right| \leq \int_a^b |f|$$

Proof. Exercise 6.5.2. □

6.3 Lecture 18: Fundamental Theorem of Calculus II, Integration via Taylor series [11/5/25]

Learning Objectives

1. Prove that integration produces antiderivatives, i.e., the Fundamental Theorem of Calculus II
2. Discuss the behaviour of Riemann integrability under uniform convergence
3. Use the previous feature to apply Taylor series to the computation of integrals without closed forms for the antiderivative

6.3.1 Fundamental Theorem of Calculus II

Our first objective in this section is the following:

Key Theorem 6.3.1 (Fundamental Theorem of Calculus II)**.**

Suppose that $g : [a, b] \longrightarrow \mathbb{R}$ is integrable. Define the function $G : [a, b] \longrightarrow \mathbb{R}$ via

$$G(x) = \int_a^x g(t)dt$$

Then the following hold:

1. G is a continuous function.
2. If g is continuous at a point $c \in [a, b]$, then G is differentiable at c and $G'(c) = g(c)$.

Proof. Let us first show that G is a continuous function. Namely:

(Part 1) **WTS:** For every $c \in [a, b]$, given any $\epsilon > 0$ there's a $\delta > 0$ such that $|x - c| < \delta$ implies that $|G(x) - G(c)| < \epsilon$.

Proof of claim. Assume for the moment that $x < c$, the other direction is similar with all the integration limits flipped. Expanding $G(x) - G(c)$, we obtain

$$G(x) - G(c) = \int_a^x g(t)dt - \int_a^c g(t)dt = \int_x^c g(t)dt$$

by applying Theorem 6.2.2. By the assumption that g is integrable, we have that $|g| \leq M$ for some real $M > 0$, i.e., it is bounded. Applying absolute values, we learn that

$$|G(x) - G(c)| = \left| \int_x^c g(t)dt \right| \leq \int_x^c |g(t)|dt \leq \int_x^c M\ dt = M \cdot (c - x)$$

by using the last part of the Algebraic Integral Theorem. Now fix $\epsilon > 0$. Setting $\delta = \frac{\epsilon}{M}$, we learn that if $|x - c| \leq \delta = \frac{\epsilon}{M}$, then $|G(x) - G(c)| \leq M \cdot \frac{\epsilon}{M} = \epsilon$. This gives the claim.

It remains to demonstrate that claim that G actually yields an antiderivative of g at the points where the latter is continuous.

(Part 2) **WTS:** If g is continuous at $c \in [a, b]$, then for every $\epsilon > 0$, there is a $\delta > 0$ such that if $|x - c| < \delta$ then

$$\left| \frac{G(x) - G(c)}{x - c} - g(c) \right| < \epsilon.$$

Proof of claim. Once again, we simplify:

$$\begin{aligned} \frac{G(x) - G(c)}{x - c} - g(c) &= \frac{1}{x - c} \cdot \left(\int_x^c g(t)dt \right) - g(c) \\ &= \frac{1}{x - c} \cdot \left(\int_x^c g(t)dt - g(c) \cdot (x - c) \right) \end{aligned}$$

However, note that $g(c) \cdot (x - c)$ can be regarded as the integral of the constant function at $g(c)$! Namely, we may write the above as

$$\frac{1}{x - c} \cdot \int_x^c (g(t) - g(c))\ dt.$$

Applying absolute values as in the previous claim, we learn that

$$\begin{aligned} \left| \frac{G(x) - G(c)}{x - c} - g(c) \right| &= \left| \frac{1}{x - c} \cdot \int_x^c (g(t) - g(c))\ dt \right| \\ &\leq \frac{1}{|x - c|} \cdot \int_x^c |g(t) - g(c)|\ dt \end{aligned}$$

Now fix $\epsilon > 0$. By the continuity of g at c, we have that there exists a $\delta > 0$ such that if $|t - c| < \delta$ then $|g(t) - g(c)| < \epsilon$. Choosing $x > c$ (the other direction just involves flipping all the integration limits) so that $|x - c| < \delta$, we have that

$$\frac{1}{x - c} \cdot \int_x^c |g(t) - g(c)|\ dt \leq \frac{1}{x - c} \cdot \int_x^c \epsilon\ dt = \frac{1}{x - c}\epsilon \cdot \epsilon \cdot (x - c) = \epsilon$$

Altogether, we have shown that if $|x - c| \leq \delta$ then $\left| \frac{G(x) - G(c)}{x - c} - g(c) \right| < \epsilon$ as claimed.

This finishes the proof. $\square$

6.3.2 Integral Limit Theorem

This section collects the following useful fact:

Key Theorem 6.3.2 (Integrable Limit Theorem)**.**

Assume that $\{f_n\} : [a,b] \to \mathbb{R}$ is a sequence of integrable functions which converges uniformly to some $f : [a,b] \to \mathbb{R}$. Then f is integrable and moreover

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b f$$

Proof. The proof that f is integrable is requested in Exercise 6.5.1. Using the Algebraic Integral Theorem, we have that

$$\left| \int_a^b f_n - \int_a^b f \right| = \left| \int_a^b (f_n - f) \right| \leq \int_a^b |f_n - f|.$$

Fix $\epsilon > 0$. Since $\{f_n\}$ converges uniformly to f, there exists an $N \in \mathbb{N}$ such that if $m > N$ then $|f_n(x) - f(x)| < \frac{\epsilon}{(b-a)}$ for every $n \geq N$ and $x \in [a,b]$. Thus, for $m \geq N$ we have that

$$\left| \int_a^b f_n - \int_a^b f \right| \leq \int_a^b |f_n - f| \leq \int_a^b \frac{\epsilon}{(b-a)} = \frac{\epsilon}{(b-a)} \cdot (b-a) = \epsilon.$$

It follows that the sequence of real numbers $\int_a^b f_n$ converges to $\int_a^b f$. □

6.4 Application: Estimating integrals via Taylor series

6.4.1 Working with power series

Differentiating a power series can be accomplished by term-wise differentiation, as long as the series of term-wise derivatives still uniformly converges, recall this from our aside on uniform convergence and differentiability (Lecture 15). Let's first ask you to show this latter fact.

Exercise 6.4.1. *Show that if $f : (c-\delta, c+\delta) \to \mathbb{R}$ is analytic, with the power series representation*

$$f(x) = \sum_{n=0}^{\infty} b_n \cdot (x-c)^n$$

then taking the term-wise derivatives gives a uniformly convergent power series

$$g(x) = \sum_{n=0}^{\infty} (n+1) \cdot a_{n+1} \cdot (x-c)^n.$$

(Hint: try using the ratio test at $(c+\delta)$ and $(c-\delta)$. Using this and the Differentiable Limit Theorem, show that $f'(x) = g(x)$.

In the following exercise, the trick is to differentiate n times.

Exercise 6.4.2. *Show that if $f : (c-\delta, c+\delta) \to \mathbb{R}$ is analytic, then any power series representation*

$$f(x) = \sum_{n=0}^{\infty} a_n \cdot (x-c)^n$$

*which is valid in $(c-\delta, c+\delta)$ must necessarily satisfy $a_n = \frac{f^{(n)}(c)}{n!}$. In particular, the **Taylor series is the unique power series representation**.*

The above exercise allows us to quickly compute the Taylor series of various functions without having to resort to actual differential computations.

Example 6.4.3. Recall that $\sin(x)$ is analytic everywhere around 0, and has the representation

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

It follows that $\sin(x^2)$ is analytic everywhere around 0, as the above will imply that

$$\sin(x^2) = x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \dots$$

on the entire real line. Moreover, this representation is *necessarily* the Taylor series expansion of $\sin(x^2)$ around 0, which allows for quick computations of basic facts. For example, we learn automatically that for $f = \sin(x^2)$, the coefficient of x^{10} must necessarily be $\frac{f^{(10)}(0)}{10!}$ and hence $f^{(10)}(0) = \frac{5!}{10!}$, which is a *hefty* computation otherwise.

6.4.2 Using the differentiable limit theorem to compute a gnarly integral

Let's go back to the example of $\sin(x^2)$. Trying to compute its integral directly on $[0,1]$ using the Fundamental Theorem of Calculus I leads us to a pretty thorny issue:

$$\int_0^1 \sin(x^2)\ dx$$

The above unfortunately cannot straightforwardly be solved for a closed form using u-substitution or integration by parts. We'll instead proceed by trying to approximate it. From the previous section, we have that the series of functions

$$\sin(x^2) = x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \dots$$

converges uniformly to $\sin(x^2)$ everywhere on the real line, including and especially on the closed interval $[0,1]$. This puts us in a position to use the Integrable Limit Theorem:

$$\int_0^1 \sin(x^2)\ dx = \int_0^1 (x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \dots)\ dx$$
$$= \frac{x^3}{3} - \frac{x^7}{7 \cdot 3!} + \frac{x^{11}}{11 \cdot 5!} - \frac{x^{15}}{15 \cdot 7!} + \dots \Big|_0^1$$

Remark 6.4.4. As long as the sequence of partial sums $\{s_n\}$ obtained by integration is merely pointwise convergent, then the Differentiable Limit Theorem kicks in to say that it is uniformly convergent, since the sequence $\{s_n'\}$ was assumed to be uniformly convergent. In particular, the levelwise integral above is a genuine expresssion for the anti-derivative of $\sin(x^2)$.

Thus, we learn the following:

$$\int_0^1 \sin(x^2)\ dx = \frac{x^3}{3} - \frac{x^7}{7 \cdot 3!} + \frac{x^{11}}{11 \cdot 5!} - \frac{x^{15}}{15 \cdot 7!} + \dots \Big|_0^1$$
$$= \frac{1}{3} - \frac{1}{7 \cdot 3!} + \frac{1}{11 \cdot 5!} - \frac{1}{15 \cdot 7!} + \dots \Big|_0^1$$

which is great! Of course, we're interested in actually obtained a decent estimate that doesn't require us to compute many steps before we actually obtain a serviceable answer. Here is where our work on the Lagrange error term kicks in.

6.4.3 Estimating the errors in our estimate

Let's say we pick the partial sum out to the 15th exponent. Namely, let us consider the following:

$$\frac{1}{3} - \frac{1}{7 \cdot 3!} + \frac{1}{11 \cdot 5!} - \frac{1}{15 \cdot 7!}$$

Question 6.4.4.

How good is this estimate for $\int_0^1 \sin(x^2)\ dx$?

Let's analyze this through the following steps.

Step 1: Using the Lagrange error term for $\sin(x)$

Remember that we obtained the series above as the following integral, consisting of the fourth partial sum in the integral of $\sin(x^2)$:

$$\frac{1}{3} - \frac{1}{7 \cdot 3!} + \frac{1}{11 \cdot 5!} - \frac{1}{15 \cdot 7!} = \int_0^1 (x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!})\, dx$$

This fourth partial sum was itself obtained by plugging x^2 into the seventh Taylor polynomial for $\sin(x)$:

$$P_7(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}$$

And of course, the Lagrange error term (Definition 4.3.7) says that for every $a \in [0, 1]$, there exists a $c \in [0, a]$ such that

$$\sin(x) - P_7(x) = \frac{f^{(8)}(c)}{8!} \cdot (a - 0)^8.$$

Since the eighth derivative of $\sin(x)$ is given by $\sin(x)$, which always takes values between 0 and 1 in the domain $[0, 1]$, we learn that

$$0 \leq \frac{f^{(8)}(c)}{8!} \leq \frac{1}{8!}$$

for every possible $c \in [0, 1]$. Since x^8 is always positive for any value of x, we have that for every possible choice of $a \in [0, 1]$, there is the relation

$$P_7(a) \leq \sin(a) = P_7(a) + \frac{f^{(8)}(c)}{8!}a^8 \leq P_7(a) + \frac{1}{8!}a^8$$

for some $c \in [0, 1]$. In other words,

$$P_7(x) \leq \sin(x) \leq P_7(x) + \frac{1}{8!}x^8$$

for every $x \in [0, 1]$.

This last inequality turns out to be the key:

Step 2. Bounding the integral

Precomposing with x^2 and applying the Algebraic Integral Theorem we obtain:

$$\int_0^1 P_7(x^2) \le \int_0^1 \sin(x^2) \le \int_0^1 P_7(x^2) + \frac{1}{8!}x^{16}$$

Expanding, we obtain that:

$$\int_0^1 (x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!})\,dx \le \int_0^1 \sin(x^2) \le \int_0^1 (x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!} + \frac{1}{8!}x^{16})\,dx$$

Solving on both sides, we obtain:

$$\frac{1}{3} - \frac{1}{7\cdot 3!} + \frac{1}{11\cdot 5!} - \frac{1}{15\cdot 7!} \le \int_0^1 \sin(x^2) \le \frac{1}{3} - \frac{1}{7\cdot 3!} + \frac{1}{11\cdot 5!} - \frac{1}{15\cdot 7!} + \frac{1}{17\cdot 8!}$$

Plugging this into a calculator for gratification, we obtain:

$$0.310268 \le \int_0^1 \sin(x^2) \le 0.310269$$

The fact that this is already decent approximation becomes a little less surprising once you realize that from the perspective of $\sin(x^2)$, we've used the full Taylor expansion to order 15, even though we've performed some routine manipulations for the expansion from $\sin(x)$. This is evident once you stare at the graph of this Taylor expansion:

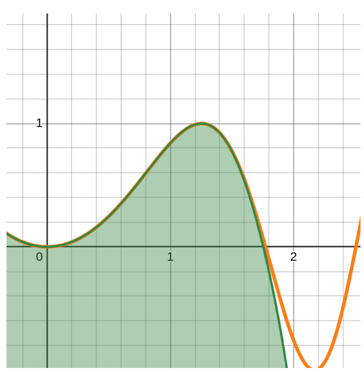

Figure 34: A depiction of the graph of $\sin(x^2)$ coupled with a shaded region under the graph of the function $x^2 - \frac{x^6}{3!} + \frac{x^{10}}{5!} - \frac{x^{14}}{7!}$

6.5 Homework 5 (Due Sunday, November 9th)

In the below, problems marked with an asterisk are graded: $(*)$ corresponds to a worth of 6 points, $(**)$ corresponds to a worth of 12. The remainder are worth 2 points and are graded for completeness. This homework is out of a possible **24 points**,.

Finally, don't expect to be able to do everything in this homework immediately! I expect you to return to this once or twice with a group as we progress through the course over the next two weeks.

Exercise 6.5.1 $(*)$**.** *Do* [Abb15, Exercise 7.2.5].

Exercise 6.5.2 $(**)$**.** *Turn in a full proof of* [Abb15, Theorem 7.4.2] *with the omitted details filled in.*

Exercise 6.5.3. *Recall that the notation* $\arcsin : [-1,1] \to \mathbb{R}$ *refers to the inverse of the function* $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to \mathbb{R}$.

1. *Show that* $\arcsin(x)$ *is differentiable, and compute its derivative. You'll find* [Abb15, Exercise 5.2.12] *useful.*
2. *Now try to compute the value of*

$$\int_0^1 \frac{1}{\sqrt{1-x^2}} dx$$

   using "u-substitution in reverse": namely, try to express the above integral as an integral in a variable t where $x = \sin t$. This technique is known as trig substitution.

3. *Find the area of a circle of radius r using the Fundamental Theorem of Calculus. Recall that a semicircle of radius r can be obtained as the graph of $\sqrt{r^2 - x^2}$.*
4. *Convince me, informally, why the volume of a 3d object V of height h can be obtained as a Riemann sum over the function*

$$\mathrm{ar} : [0,h] \to \mathbb{R}, \ \mathrm{ar}(x) = \textit{Area of the slice of } V \textit{ at height } x.$$

   Using this, find an expression for the volume of a sphere of radius r.

Exercise 6.5.4. *Prove the Mean Value Theorem for integrals: if $f : [a,b] \to \mathbb{R}$ is a continuous function, then there exists $c \in [a,b]$ such that*

$$f(c) = \frac{1}{b-a} \int_a^b f(x) dx$$

This can be thought of as saying that the function f attains its average value at some point on $[a,b]$.

Exercise 6.5.5. *Let $p > 0$ be a real number. Prove, rigorously, that the series*

$$\sum_{n=0}^{\infty} \frac{1}{n^p}$$

converges if and only if $p > 1$ and diverges otherwise by comparing it to the Riemann sum for the function $\frac{1}{x^p}$.

References

[Abb15] Stephen Abbott. *Understanding Analysis*. Springer New York, 2015. ISBN: 9781493927128. DOI: 10.1007/978-1-4939-2712-8. URL: http://dx.doi.org/10.1007/978-1-4939-2712-8. 24, 52, 67, 82, 90

[Slo25] Dan Sloughter. *A Primer of Real Analysis*. Compiled on 10/01/2025. LibreTexts, 2025. URL: https://commons.libretexts.org/book/math-22635 (visited on 10/19/2025). 65