18.100B Lecture Notes

Sanjana Das

Fall 2022

Notes for the MIT class **18.100B** (Real Analysis), taught by Tobias Colding. All errors are my responsibility.

Contents

§1 Introduction

In this class, we'll learn real analysis. We will also learn how to write mathematical proofs and prove theorems. This is typically not done in calculus. For example, you may recall the *intermediate value theorem* — if $f: [a, b] \to \mathbb{R}$ is continuous, and $f(a) < 0$ and $f(b) > 0$, then for some $a < c < b$ we have $f(c) = 0$. In calculus, you may say that a continuous function essentially means you can draw the graph with a pencil, without leaving the paper, which makes the theorem intuitively clear (if you start below the x-axis and end above it, somewhere you have to hit the x-axis). But this is not really a mathematical explanation — this isn't true over the rationals *even if* the function is continuous. So there's something special about the real numbers.

§2 The Real Numbers

Question 2.1. What is $\mathbb{R}$?

We can start with some simpler objects:

- The **natural numbers** $\mathbb{N}$ are the set $\{1, 2, 3, 4, \ldots\}$.
- The **integers** $\mathbb{Z}$ are the set $\{0, \pm 1, \pm 2, \pm 3, \ldots\}$.
- The **rational numbers** $\mathbb{Q}$ are the set $\{\frac{m}{n} \mid m \in \mathbb{Z}, n \in \mathbb{N}\}$ (we'll come back to this later).

We have $\mathbb{Q} \subseteq \mathbb{R}$, but $\mathbb{R}$ also has some additional properties. For example, $\sqrt{2}$ is in $\mathbb{R}$ but not $\mathbb{Q}$. More precisely, $\mathbb{R}$ is an *ordered field* containing all the rationals. It also has the *least upper bound* and the *greatest lower bound* properties (we won't discuss this today, but this is really what's responsible for the intermediate value theorem).

We'll now describe what these properties are.

§2.1 Fields

For now, we'll return to the rational numbers $\mathbb{Q}$, which are the numbers $\frac{m}{n}$ for $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. The rational numbers have two important operations:

1. We can *add* rational numbers:
$$\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_1 n_2 + n_1 m_2}{n_1 n_2} \in \mathbb{Q}.$$
So we have an operation $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ called *addition*.
2. We can *multiply* rational numbers:
$$\frac{m_1}{n_1} \cdot \frac{m_2}{n_2} = \frac{m_1 m_2}{n_1 n_2} \in \mathbb{Q}.$$
So we have another operation $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ called *multiplication*.

These operations have some properties. First, addition has the following properties:

1. The *commutative* property (or *abelian* property) — we have $\frac{m_1}{n_1} + \frac{m_2}{n_2} = \frac{m_2}{n_2} + \frac{m_1}{n_1}$.
2. The *associative* property — we have $\left(\frac{m_1}{n_1} + \frac{m_2}{n_2}\right) + \frac{m_3}{n_3} = \frac{m_1}{n_1} + \left(\frac{m_2}{n_2} + \frac{m_3}{n_3}\right)$.
3. There is a *neutral element* 0 — we have $0 + \frac{m_1}{n_1} = \frac{m_1}{n_1}$.

4. The existence of an *inverse* — for every rational number $\frac{m_1}{n_1}$, we have $\frac{m_1}{n_1} + \frac{-m_1}{n_1} = 0$. This means if you take any rational number, there's another rational number which can be added to the first to produce the neutral element 0.

These are the four basic properties that addition has, and from this you can deduce other properties. But we need all four of these, since if we only had three, there are examples of sets and operations where the fourth doesn't follow.

Multiplication has essentially the same properties:

5. Multiplication is commutative — $\frac{m_1}{n_1} \cdot \frac{m_2}{n_2} = \frac{m_2}{n_2} \cdot \frac{m_1}{n_1}$.
6. Multiplication is associative — $\left(\frac{m_1}{n_1} \cdot \frac{m_2}{n_2}\right) \cdot \frac{m_3}{n_3} = \frac{m_1}{n_1} \cdot \left(\frac{m_2}{n_2} \cdot \frac{m_3}{n_3}\right)$.
7. There is a neutral element 1 for multiplication — $1 \cdot \frac{m_1}{n_1} = \frac{m_1}{n_1}$.
8. Every *nonzero* rational has a multiplicative inverse — if $m \neq 0$, then there exists another rational number $\frac{n}{m}$, such that $\frac{m}{n} \cdot \frac{n}{m} = 1$. (It's possible that $\frac{n}{m}$ may not be valid according to our definition, as m may be negative; then we should take $\frac{-n}{-m}$ instead.)

Nothing here is suprising — we use these properties all the time — but we might not think of them as formal properties. The properties for addition and multiplication are exactly parallel, except that 0 doesn't have a multiplicative inverse.

In arithmetic, we're used to having four operations — $+$, $-$, $\cdot$, and $\div$. But subtraction is really the same thing as adding the inverse (and the same is true for division and multiplication).

Student Question. *In general, is the key of the inverse that it returns the neutral element?*

Answer. Yes.

Student Question. *Why did we take $\frac{-n}{-m}$?*

Answer. We defined rational numbers as $\frac{m}{n}$ with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, so we need the denominator to be positive — it's possible that $\frac{n}{m}$ is not valid, if m is negative.

Remark 2.2. The reason we exclude 0 is that if we insisted 0 had an inverse, then all the examples would be trivial.

Finally, there's one property that ties these together:

9. The *distributive law* — we have $\frac{m_1}{n_1}\left(\frac{m_2}{n_2} + \frac{m_3}{n_3}\right) = \frac{m_1}{n_1} \cdot \frac{m_2}{n_2} + \frac{m_1}{n_1} \cdot \frac{m_3}{n_3}$.

So these are the useful properties of the rational numbers. Now we'd like to formalize these by defining an abstract structure with these properties. (This course will involve some level of abstraction — this lets you make statements that hold in a greater level of generality, and is useful for isolating the key properties to make things easier to understand.)

Definition 2.3. A **field** is a set X with two operations $+$ and $\cdot$ from $X \times X \to X$. Addition has the following properties:

1. The commutative property — $x_1 + x_2 = x_2 + x_1$.
2. The associatice property — $(x_1 + x_2) + x_3 = x_1 + (x_2 + x_3)$.
3. There exists a neutral element for $+$, denoted by 0, such that $0 + X = X$.
4. The existence of inverses — for every x, there exists an element, denoted $-x$, such that $x + (-x) = 0$.

Similarly, multiplication has the following properties:

5. Commutativity — $x_1 \cdot x_2 = x_2 \cdot x_1$.
6. Associativity — $(x_1 \cdot x_2) \cdot x_3 = x_1 \cdot (x_2 \cdot x_3)$.
7. There exists a neutral element for $\cdot$, denoted by 1, such that $x \cdot 1 = x$.
8. If $x \neq 0$, there exists an element y, denoted $\frac{1}{x}$, such that $x \cdot y = 1$.

Finally, there's a property that ties these together:

9. The distributive law: $x(y + z) = xy + xz$.

We could've denoted the operations by any symbol, but we use $+$ and $\cdot$ to illustrate that they're essentially generalizations of addition and multiplication in $\mathbb{Q}$. Similarly, 0, 1, $-x$, and $\frac{1}{x}$ are all just suggestive notation; this just means there *exists* an element ($-x$ doesn't initially have any meaning).

Student Question. *Why are $+$ and $\cdot$ supplied, but 0 isn't?*

Answer. $+$ and $\cdot$ are operations, and 0 is an element in the set. Usually we single out the operations, and discuss what properties these operations have. We could've done this otherwise, but this is how it's always stated.

Student Question. *How important is closure?*

Answer. We've seen that $\mathbb{R}$ being a field isn't enough — so far this is just algebra. We won't talk about closure, but a similar notion that we *will* need is the least upper bound property, and $\mathbb{Q}$ is an example of a field which doesn't have this property (because $\sqrt{2}$ is not rational).

There is a specific notion of what it means for a field to be closed, which is something completely different (and we will not get to it in the class).

The set is closed under the *operation* because the operation goes from the set to itself (so this is given in the definition).

From these abstract properties, there's various things you can prove.

Proposition 2.4

There is a *unique* neutral element for addition.

In other words, there is a unique zero — the same is true for multiplication.

Proof. Suppose 0_1 and 0_2 are both neutral. Then $0_1 + 0_2$ must equal 0_1 (since 0_2 is neutral), and it must also equal 0_2 (since 0_1 is neutral). But now this means $0_1 = 0_2$. $\square$

Proposition 2.5
For all x, we have $0 \cdot x = 0$.

Proof. We have
$$0 \cdot x + 0 \cdot x = (0 + 0) \cdot x = 0 \cdot x$$
by distributivity. So $0 \cdot x$ has the property that $y + y = y$. But for any y with this property, we can add $-y$ to both sides, to get that
$$(y + y) + (-y) = y + (-y) = 0.$$
Since addition is associative, the left-hand side is also equal to
$$y + (y + (-y)) = y + 0 = y.$$
So then we have $y = 0$. Since $0 \cdot x$ has this property, that means $0 \cdot x = 0$. ☐

Example 2.6
$\mathbb{Q}$ is a field.

Example 2.7
$\mathbb{N}$ is *not* a field. It does have the two operations $+$ and $\cdot$, which are $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and satisfy commutativity, associativity, and distributivity. But it doesn't have a neutral element for addition (although it does have one for multiplication).

Example 2.8
$\mathbb{Z}$ is *not* a field. Addition satisfies all four properties, and multiplication satisfies the first three, but we do not have multiplicative inverses (for example, 2 doesn't have an inverse — $\frac{1}{2}$ is not in $\mathbb{Z}$).

§2.2 Ordered Sets

The real numbers also have a notion of order — we can talk about whether a real number is bigger or smaller than another.

Definition 2.9. An **ordered set** is a set X together with an ordering $<$ such that:

1. We can pair any two elements — for all $x, y \in X$, exactly one of $x < y$, $y < x$, or $x = y$ is true.
2. The transitive property — if $x < y$ and $y < z$, then $x < z$.

We may write $y < x$ as $x > y$.

Student Question. *Is there a difference between a set and a field?*

Answer. Yes — a field is a set with the two operations, which satisfy certain properties. On the other hand, sets don't have to be that special — many different things are sets.

Student Question. *Can you define an ordering on vectors?*

Answer. On $\mathbb{R}$, we have the usual ordering. But suppose we wanted an ordering of $\mathbb{R}^2$ — then there isn't really a *canonical* ordering. There *is* an ordering you could use on $\mathbb{R}^2$ — for example, the alphabetical ordering, where you first compare by the first coordinate, and if they're equal you compare by the second. This does give an ordering — for any two points in the plane, either one is bigger than the other under the ordering, or they have both coordinates equal.

§2.3 Ordered Fields

If we have a field with an ordering, we'd want the ordering and the field operations to interact.

Definition 2.10. If F is a field and $<$ is an ordering of F, then F is a **ordered field** if two additional properties hold:

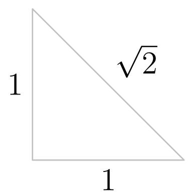
1. If $x < y$ and x, y, and z are in F, then $x + z < y + z$.
2. If $x < y$ and $z > 0$, then $xz < yz$.

Given a field and an ordering, this isn't necessarily true in general (there are fields with orderings that don't have these two properties). But if this *is* true, then we have an ordered field.

§2.4 Motivation — The Real Numbers

Our main interest is to understand the real numbers, since they're what we'll usually work with when doing analysis.

Consider a right triangle with side lengths 1 and 1; then its hypotenuse has length $\sqrt{2}$.



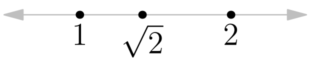
Question 2.11. Is $\sqrt{2} \in \mathbb{Q}$?

The answer is no! Suppose $\sqrt{2} = \frac{m}{n}$ for some m and n; we can assume that m and n have no common divisors. But this means $m^2 = 2n^2$.

If m isn't divisible by 2, then neither is m^2; so m must be divisible by 2. But then $m = 2m_1$, and we have $m^2 = 4m_1^2 = 2n^2$, which means $2m_1^2 = n^2$. But this means n is also divisible by 2! Then 2 is a factor of both m and n, so they have a common factor; this is a contradiction.

We certainly want to be able to talk about the side lengths of a triangle, so this means the rational numbers won't be enough.

On a calculator, we can find $\sqrt{2} = 1.414\cdots$.



So we may think of $\sqrt{2}$ in terms of its first few digits, but we really think of it as a *limit* — as we take more and more digits.

This leads to the notion of an *upper* and *lower* bound.

§2.5 Upper and Lower Bounds

Definition 2.12. Suppose we have a set S with an ordering $<$, and a subset $A \subseteq S$. Then A is **bounded from above** if there exists some $y \in S$ such that $x \leq y$ for all $x \in A$.

The notation $x \leq y$ means $x < y$ or $x = y$.

This definition makes sense — we're used to doing it on a line, but it makes sense to think about an upper bound in *any* ordering.

Example 2.13

The set $\{1, 2, 3\}$ is bounded from above, since 4, or even 3, is an upper bound.

Example 2.14

The set $\{1, 2, 3, \ldots\} = \mathbb{N}$ is *not* bounded from above.

Similarly, we can define what it means for a set to be bounded from below.

Definition 2.15. Suppose we have a set S with an ordering $<$, and a subset $A \subseteq S$. Then A is **bounded from below** if there exists some $y \in S$ such that $x \geq y$ for all $x \in A$.

Note that y doesn't have to be in A.

Example 2.16

The interval $(0, 1) \subseteq \mathbb{R}$ is bounded from above (1 is an upper bound, and so is 2) and from below (0 is a lower bound, and so is -1). But we can't find a lower bound which is *in* the set.

There's one property of $\mathbb{R}$ that isn't possible to deduce from the ones we've stated before, but is very important.

Definition 2.17. Given an ordered set $(S, <)$, a subset $A \subseteq S$ has the **least upper bound property** if there exists an upper bound y for S with the property that if z is another upper bound for S, then $y \leq z$.

Student Question. *Why do we consider other upper bounds than the least?*

Answer. It isn't always the case that there's the least upper bound. When you know that there *is* a least upper bound, you typically just look at it, but it's possible that there isn't. For example, suppose we take the set $S = \mathbb{R} \setminus \{0\}$, with the usual ordering. Then the negative numbers $\{x \mid x < 0\}$ is a subset of S. Any positive number is an upper bound for this subset. But this subset doesn't have a *least* upper bound, since 0 isn't in our set.

Another example is $\mathbb{Q}$ with the usual ordering. We will see very soon that the set $\{x > 0 \mid x^2 \leq 2\}$ has an upper bound, but it doesn't have a *least* upper bound (because $\sqrt{2}$ is not rational).

Theorem 2.18

There exists an ordered field containing $\mathbb{Q}$ with the usual ordering, with the least upper bound and greatest lower bound property.

So we take $\mathbb{Q}$, which we're familiar with, and it's contained in a larger field $\mathbb{R}$ (with the same ordering on $\mathbb{Q}$) that has the least upper bound and greatest lower bound properties.

The problem with the rational numbers was that it wasn't somehow complete — you could get arbitrarily close to $\sqrt{2}$, but $\sqrt{2}$ wasn't rational. Once you have the least upper bound property, then the field is complete — it doesn't have "gaps," like the rational numbers do.

Now that we have these properties, let's prove some more.

§2.6 The Archimedean Property

Theorem 2.19

Suppose $x \in \mathbb{R}$. Then there exists a positive integer $k \in \mathbb{N}$ with $x < k$.

Proof. Assume not. Then $\mathbb{N}$ is bounded from above, because x is an upper bound for $\mathbb{N}$. $\square$

Fact 2.20 — If $0 < x < y$, then $\frac{1}{y} < \frac{1}{x}$.

Proof. Assume not, so then $\frac{1}{x} \leq \frac{1}{y}$. Now multiplying by xy on both sides gives that $y \leq x$, contradiction. $\square$

Proposition 2.21

If $\varepsilon > 0$, there exists $k \in \mathbb{N}$ such that $0 < \frac{1}{k} < \varepsilon$.

Proof. Since $\varepsilon > 0$, then $\frac{1}{\varepsilon} > 0$, so there exists an integer $k > \frac{1}{\varepsilon}$. Now taking the inverse gives $\varepsilon > \frac{1}{k}$. $\square$

Here we used one fact that we haven't yet proved — that $\frac{1}{1/x} = x$. By definition, the inverse of x is the number y with the property that $x \cdot y = 1$. We've already observed that there's a unique such y. (If there's two, then we can consider $y_1 x y_2$, which must equal both of them.) We denote y by $\frac{1}{x}$, so then

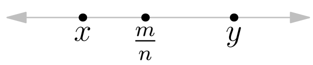
$$x \cdot \frac{1}{x} = 1.$$

But then $\frac{1}{1/x}$ is the number with the property that if we multiply it by $\frac{1}{x}$, we get 1; this is precisely x.

From the Archimedean Property, we get the following:

Theorem 2.22 (Density of the Rational Numbers)

For any two real numbers $x < y$, there is a rational number $x < \frac{m}{n} < y$.



Proof. If $x < y$, then since $\mathbb{R}$ is an ordered field we have $x + (-x) < y + (-x)$, so $0 < y - x$. Let $\varepsilon = \frac{1}{x}$. Now we know that there is a positive integer k such that $0 < \frac{1}{k} < \varepsilon$.

Now we look at

$$A = \left\{m \in \mathbb{Z} \mid \frac{m}{k} \leq x\right\} = \{m \in \mathbb{Z} \mid m \leq xk\}.$$

So A is a subset of $\mathbb{Z}$, and it's bounded from above — since xk is an upper bound.

Now take the largest such m. Then we have $\frac{m}{k} \le x$, and $\frac{m+1}{k} > x$ — we chose m to be largest, so $m+1$ is not in A. Consider the equation

$$\frac{m}{k} \le x < \frac{m}{k} + \frac{1}{k}.$$

But then we have

$$\frac{m}{k} + \frac{1}{k} \le x + \frac{1}{k} < x + \varepsilon = x + (y - x) = y,$$

so now we have

$$x < \frac{m}{k} + \frac{1}{k} < y.$$

But $\frac{m+1}{k}$ is clearly a rational number. □

Student Question. *Couldn't we go from $\frac{1}{k} < y - x$ and just add x?*

Answer. Note that x is not a rational number. But in our proof, we used the same idea — we took something smaller than the gap and repeatedly added steps of $\frac{1}{k}$. Eventually we'll get up to x, and the next step takes us past x and into the interval.

Remark 2.23. Density can be used to show uniqueness — there can't be two different ordered fields containing $\mathbb{Q}$ which have the least upper bound property.

Recall that the real numbers are an ordered field with the least upper bound and greatest lower bound property, that contains the rational numbers.

Remark 2.24. Very often, we write "there exists" and "for all." These both have mathematical abbreviations — $\exists$ and $\forall$ respectively. It is also standard to use $\exists!$ to mean that there exists a *unique* object, but this is less common.

Remark 2.25. When you write on the homework or exam, you should make sure to write enough and write clearly. Good style on the blackboard is different than good style on a paper. As an example, we may use $\Longrightarrow$ on the board to show that one statement implies the other. But when you write on paper, you should have more detail. Writing will be an important part of this course — we should be able to explain clearly and make arguments without holes.

Now we can think of $\sqrt{2}$ as the least upper bound of the set

$$A = \{x \in \mathbb{Q} \mid x > 0 \text{ and } x^2 \le 2\}.$$

We know A is bounded from above, since for example, 2 is an upper bound. Now take the *least* upper bound for A, which we call x_0.

Proposition 2.26

We have $x_0^2 = 2$.

Proof. We'll assume not, and split into two cases. We will show that $x_0^2 = 2$ in two steps — if $x_0^2 > 2$ then it can't be the *least* upper bound, and if $x_0^2 < 2$ then it can't be an upper bound.

Claim — If $x_0^2 > 2$, then there exists $h > 0$ such that $(x_0 - h)^2 > 2$ (and $x_0 - h > 0$).

Proof. We have $(x_0 - h)^2 = x_0^2 - 2x_0h + h^2$. Taking $h = \frac{1}{k}$, we can make both $2hx_0$ and h^2 as small as we want. So we can add something tiny to x^2, to get a result that's still larger than 2. ■

This means $x_0 - h$ is an upper bound for A, so x_0 cannot be the least upper bound.

Claim — If $x_0^2 < 2$, then there exists $h > 0$ such that $(x_0 + h)^2 < 2$.

Proof. We have $(x_0 + h)^2 = x_0^2 + 2x_0h + h^2$. Take h to be tiny; then we can make $2x_0h + h^2$ tiny as well, so we can squeeze in $(x_0 + h)^2$ to be less than 2 as well. ■

Now there is a rational number strictly between x_0 and $x_0 + h$, which must be in A; so then x_0 cannot be an upper bound for A. □

Student Question. *How do you show there exists h?*

Answer. We know $x_0^2 > 2$, and we want h such that $x_0^2 - 2 > 2x_0h - h^2$. We can factor out h, and write the right-hand side as $h(2x_0 - h)$. Since $x_0^2 - 2 = \varepsilon > 0$, if we didn't have $2x_0 - h$ then we could choose $h = \frac{1}{k}$. But we have $2x_0 - h > 2x_0$, so we can remove h. So we just need a slight generalization of the Archimedean property (which actually follows from the Archimedean property) — given $\varepsilon > 0$ and $x > 0$, we can find k such that $0 < \frac{x}{k} < \varepsilon$ (and this follows directly from the Archimedean property).

§2.7 Review

Recall that we defined an ordered set, a field, and an ordered field. Finally, we had the least upper bound property, and we saw the following theorem:

Theorem 2.27

There exists an ordreed field with the least upper bound property and containing $\mathbb{Q}$.

In fact, there is only one. We didn't prove this theorem; we will assume it.

For an ordered field, we've previously seen the property that if $0 < z$ and $x < y$, then $zx < zy$. Also, for any z, if $x < y$ then $x + z < y + z$.

From these axioms, we can derive some useful properties.

Fact 2.28 — If $0 < x < y$, then $\frac{1}{y} < \frac{1}{x}$.

Proof. Multiply $x < y$ by $\frac{1}{xy}$ on both sides. □

Fact 2.29 — If $0 < x < y$, then $0 < x^2 < y^2$.

Proof. Since $x > 0$, we can multiply by x to get $0 < x^2 < xy$. Then since $y > 9$, we can multiply by y to get $xy < y^2$. So by transitivity, $0 < x^2 < xy < y^2$. □

Now let's focus our attention on $\mathbb{R}$, which is the ordered field containing $\mathbb{Q}$ and with the least upper bound property.

Then if we take a nonempty subset $A \subseteq \mathbb{R}$, if A is bounded from above, then it has a least upper bound.

Definition 2.30. The least upper bound of A is denoted $\sup A$.

It's clear that the least upper bound is unique.

Now let's focus on the Archimedean property:

Theorem 2.31

For all $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $0 < 1/n < \varepsilon$.

Proof. Consider the set $A = \{n \in \mathbb{N} \mid n \leq 1/\varepsilon\}$. We want to show that A is not all of $\mathbb{N}$ — or in other words, there exists $n \in \mathbb{N}$ such that $n > 1/\varepsilon$. If we have this, then we know $1/n < \varepsilon$.

Assume for contradiction that $A = \mathbb{N}$. Then let $\alpha = \sup A$, which exists since A is bounded by $1/\varepsilon$. Then we know $n \leq \alpha$ for all n. But then $n + 1 \leq \alpha$ for all $n \in \mathbb{N}$ (since $n + 1 \in \mathbb{N}$), and therefore $n \leq \alpha - 1$ for all n. But this implies $\alpha - 1$ is also an upper bound, contradiction (because α was assumed to be the least upper bound, and $\alpha - 1 < \alpha$). □

§2.8 Writing Proofs

There are two separate parts of writing a proof — coming up with an idea, and writing it out.

Last time, we saw that we can define $\sqrt{2}$ in $\mathbb{R}$, as

$$\sqrt{2} := \sup\{\alpha > 0 \mid \alpha^2 < 2\}.$$

First we need to check this is well-defined. Let's name our set A; then to check that $\sup A$ exists, we need to show A is nonempty and bounded from above. First, A is nonempty because $1 \in A$, since $1^2 = 1 < 2$. On the other hand, we claim that 2 is an upper bound for α — we want to show that if $\alpha > 0$ and $\alpha^2 \leq 2$, then $\alpha \leq 2$. Suppose not, so $\alpha > 2$. Then $\alpha > 2 > 0$, so $\alpha^2 > 2^2 = 4$.

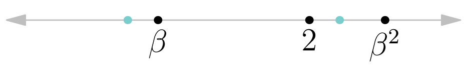
Now let $\beta = \sup A$. We want to show that $\beta > 0$ and $\beta^2 = 2$.

First, we've already seen $1 \in A$, so since β is an upper bound, $\beta \geq 1 > 0$.

To show $\beta^2 = 2$, if $\beta^2 > 2$, then consider

$$(\beta - h)^2 = \beta^2 - 2\beta h + h^2 = \beta^2 - h(2\beta - h).$$

The idea is that we want to make $h(2\beta - h)$ very small, so that $(\beta - h)^2 > 2$.



If we can construct $h > 0$ (with $h < \beta$) such that $(\beta - h)^2 > 2$, then $\beta - h$ is an upper bound for A — if we had $\alpha \geq \beta - h$, then we would have $2 < (\beta - h)^2 \leq \alpha^2$, contradicting the fact that $\alpha \in A$.

This is the idea, but now let's make it into a formal proof. First we should calculate what h should be. Recall that

$$(\beta - h)^2 = \beta^2 - h(2\beta - h),$$

so it's enough to have $h(2\beta - h) \le \frac{1}{2}(\beta^2 - 2)$. We already know that $\beta \le 2$, so $2\beta - h \le 4$. So it's enough to have $4h \le \frac{\beta^2-2}{2}$, and we can simply set

$$h = \min\left\{\frac{\beta^2 - 2}{8}, 1\right\}.$$

(The reason for 1 is that then $2\beta - h \ge 1$.)

Right now this looks somewhat ridiculous, but these side calculations are necessary to construct the rest of the proof.

Student Question. *Why do we divide by 2 — why not just take $h(2\beta - h) \le \beta^2 - 2$?*

Answer. To make the inequality strict.

Lemma 2.32

$\beta^2 \le 2$.

Proof. Suppose not; we will obtain a contradiction. So assume that $\beta^2 > 2$. Set

$$h = \min\left\{\frac{\beta^2 - 2}{8}, 1\right\},$$

and observe that $h > 0$. We see since $\beta \le 2$ that

$$\begin{aligned}(\beta - h)^2 &= \beta^2 + h^2 - 2h\beta \\ &= \beta^2 - h(2\beta - h) \\ &\ge \beta^2 - 4h \\ &\ge \beta^2 - 4\left(\frac{\beta^2 - 2}{8}\right) \\ &= 2 + \frac{1}{2}(\beta^2 - 2) \\ &> 2.\end{aligned}$$

Using this, we will see that $\beta - h$ is an upper bound for A, contradicting that β was the least upper bound for A. Namely, if $\beta - h$ is not an upper bound, then there exists some $\alpha \in A$ such that $\alpha > \beta - h > 0$, and hence $\alpha^2 > (\beta - h)^2 > 2$. This gives the desired contradiction. $\square$

Remark 2.33. Some people write Q.E.D. at the end of their proofs — or $\square$ — to indicate where the proofs end.

Let's now do this for the other bound. This is almost identical. Here we assume $\beta^2 < 2$, and look at $\beta + h$; we want to show that β was not actually an upper bound, so we want to find small h such that $(\beta + h)^2 < 2$. Now we again multiply out, so we want

$$\beta^2 + 2\beta h + h^2 < 2,$$

or equivalently

$$h(2\beta + h) < 2 - \beta^2.$$

We know $\beta \le 2$, so if we set $h \le 1$, then $2\beta + 1 \le 5$. So we can take

$$h = \min\left\{\frac{2 - \beta^2}{10}, 1\right\}.$$

Lemma 2.34
$\beta^2 \geq 2$.

Proof. Suppose not; we will obtain a contradiction. Suppose that $\beta^2 < 2$; we will show that this leads to a contradiction that β was an upper bound for A. Set

$$h = \min\left\{\frac{2-\beta^2}{10}, 1\right\},$$

and observe that by the assumption (that $\beta^2 < 2$) that $h > 0$. Now observe since $\beta \leq 2$ and $h \leq 1$ that

$$\begin{aligned}(\beta + h)^2 &= \beta^2 + h^2 + 2\beta h \\ &= \beta^2 + h(2\beta + h) \\ &\leq \beta^2 + 5h \\ &= 2 - (2 - \beta^2 - 5h) \\ &\leq 2 - \left(2 - \beta^2 - \frac{1}{2}(2-\beta^2)\right) \\ &= 2 - \frac{1}{2}(2 - \beta^2) \\ &< 2.\end{aligned}$$

This shows $\beta + h \in A$, contradicting that β was an upper bound. $\square$

The goal was to illustrate the difference between coming up with an idea and writing down a formal proof — it is an important skill to be able to write proofs, since as things become more complicated, you'll otherwise end up thinking you've proved something that has crazy implications.

§3 Sequences

§3.1 Definition of Sequences

Recall that $\mathbb{R}$ is a field containing $\mathbb{Q}$ with an ordering that makes it an ordered field. We then discussed that $\mathbb{R}$ has the least upper bound property. Using this, we could show that there was a unique element in $\mathbb{R}$ (but not in $\mathbb{Q}$) with the property that $\beta > 0$ and $\beta^2 = 2$, which we denote by $\sqrt{2}$. We also observed that $\sqrt{2} \notin \mathbb{Q}$. So this least upper bound property is sort of a kind of "completion."

We didn't prove uniqueness, but it seems sort of intuitively clear that from completeness, we must have that there could only be one such extension.

Question 3.1. How do we think about $\sqrt{2}$?

We may think of $\sqrt{2}$ as 1 (this is a very weak bound), or 1.4, or 1.41, or 1.414. If someone asked for more precision, then we'd add a few more digits. When we're thinking about it this way, we're really thinking of a *sequence*. If we're not rounding up, then we have an increasing sequence, and as we go further and further, the person we're talking to finally gets an idea of what $\sqrt{2}$ is.

Definition 3.2. A **sequence** (of numbers) is a function $a\colon \mathbb{N} \to \mathbb{R}$. In other words, for every $n \in \mathbb{N} = \{1, 2, 3, \dots\}$, we have a value $a_n \in \mathbb{R}$.

Usually we write a_n instead of $a(n)$, even though technically a sequence is a function.

Note that the ordering of a sequence is important! The sequence 1, 1.4, 1.41, 1.414, ... is a sequence, but if we reordered it, then we'd get a *different* sequence.

Example 3.3

The **binary sequence** 0, 1, 0, 1, 0, 1, ... is a sequence, where

$$a_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even.} \end{cases}$$

Remark 3.4. We don't have to start indices at 1, but we usually do.

Remark 3.5. We say that $\mathbb{N}$ begins at 1, not 0, since this is the convention used in the textbook.

Example 3.6

We can take the sequence

$$a_n = (-1)^n = \begin{cases} -1 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even} \end{cases}.$$

Example 3.7

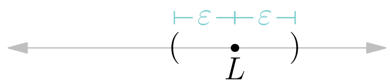
We can take the sequence $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots$ defined as

$$a_n = \left(\frac{1}{2}\right)^n.$$

§3.2 Convergence

The most important question for sequences is whether or not they converge.

Definition 3.8. For some $L \in \mathbb{R}$, a sequence a_n **converges** to L, denoted by $a_n \to L$ or $\lim_{n\to\infty} a_n = L$, if the following holds: for all $\varepsilon > 0$, there exists N such that if $n \geq N$, then $|L - a_n| < \varepsilon$.



So given any ε, we can find some N so that if we go far out for at least N terms, then all future terms of the sequence lie in this interval.

Note that if we shrink ε, we usually need to make N bigger — we need to go further out before we stabilize in that interval.

Example 3.9

Consider the sequence $a_n = (\frac{1}{2})^n$. Then $a_n \to 0$.

Proof. We want to show that given $\varepsilon > 0$, there exists N such that for all $n \geq N$,

$$\left|\left(\frac{1}{2}\right)^n - 0\right| = \left(\frac{1}{2}\right)^n < \varepsilon.$$

But if $n \geq N$, then

$$\left(\frac{1}{2}\right)^n \leq \left(\frac{1}{2}\right)^N,$$

since we're just multiplying by more factors of $\frac{1}{2}$. So we need

$$\left(\frac{1}{2}\right)^n < \varepsilon,$$

which is equivalent to saying

$$2^N > \frac{1}{\varepsilon}.$$

But this is really just the Archimedean property.

Now let's write this down formally: Given $\varepsilon > 0$, choose N such that $2^N > \frac{1}{\varepsilon}$. Then for $n \geq N$,

$$|a_n - 0| = \left(\frac{1}{2}\right)^n \leq \left(\frac{1}{2}\right)^N = \frac{1}{2^N} < \varepsilon. \qquad \square$$

Example 3.10

Consider the sequence 0, 0.9, 0.99, 0.999, . . . (if you attempt to do a calculation that should result in 1 on a calculator, you may end up with 0.999999 . . .), where

$$a_n = 0.\underbrace{9999\cdots 9}_{n}.$$

Then $a_n \to 1$.

Proof. For each n, we have

$$a_n = 1 - \frac{1}{10^{n-1}} \leq 1.$$

Then it's clear $1 \geq a_{n+1} > a_n$. Now we can write

$$|a_n - 1| = 1 - a_n = \frac{1}{10^{n-1}}.$$

Now if $n \geq N$, we have

$$\frac{1}{10^{n-1}} \leq \frac{1}{10^N - 1},$$

and similarly to the previous example, we can make this less than any ε. So $a_n \to 1$. $\square$

So if this sequence showed up on our calculator, we'd think of it as 1. But it's really a *sequence*, whose limit is 1.

So when we have a sequence, often the first question to ask is whether it converges.

Example 3.11

The binary sequence 0, 1, 0, 1, ... doesn't converge.

For example, suppose we wanted to prove it converges to 1. Then independently of how far out we go, we'll get some 0's, so the limit *can't* be 1. The same occurs for 0, or for any other number — we'll always have terms a fixed distance away.

So it's only nice sequences that converge — but a lot of questions that come up are about convergence. We'll see later that from sequences, we can build all kinds of other sequences — the most important example of this is infinite sums.

§3.3 Operations on Sequences

We can use certain operations to go from simple sequences to more complicated sequences. Then there are simple rules that address whether those more complicated sequences converge.

Suppose a_n and b_n are sequences. Then we can consider the new sequence $c_n = a_n + b_n$.

Proposition 3.12

If a_n converges to a and b_n converges to b, then $c_n = a_n + b_n$ converges to $a + b$.

Proof. Given an $\varepsilon > 0$, we want to use that a_n converges to a, we can find N such that $|a_n - a| < \varepsilon$. But we don't want to use this N — we'll actually cut the given ε in half.

So using that a_n converges to a, we have that there exists N_1 such that if $n \geq N_1$, then

$$|a_n - a| < \frac{\varepsilon}{2}.$$

Similarly, since b_n converges to b, there exists N_2 such that if $n \geq N_2$, then

$$|b_n - b| < \frac{\varepsilon}{2}.$$

Now let $N = \max\{N_1, N_2\}$. If $n \geq N$, then we have both $n \geq N_1$ and $n \geq N_2$, so we have both properties at the same time. So then

$$|(a + b) - (a_n + b_n)| = |(a - a_n) + (b - b_n)| \leq |a - a_n| + |b - b_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

(where we used the triangle inequality to split the sum into a sum of absolute values). □

Proposition 3.13

If $a_n \to a$ and $\lambda \in \mathbb{R}$, then the sequence $b_n = \lambda a_n$ converges to λa.

Proof. If $\lambda = 0$, then b_n is a sequence of all 0's, which trivially converges to 0. So we may assume $\lambda \neq 0$.

Given $\varepsilon > 0$, since $a_n \to a$, there exists an N so that if $n \geq N$, then

$$|a_n - a| < \frac{\varepsilon}{|\lambda|}.$$

(Of course, this is why we assumed $\lambda \neq 0$ — so that we could divide by it.) Now $b_n = \lambda a_n$, so for $n \geq N$ we have

$$|b_n - b| = |\lambda a_n - \lambda a| = |\lambda(a_n - a)| = |\lambda|\,|a_n - a| < |\lambda| \cdot \frac{\varepsilon}{|\lambda|} = \varepsilon.$$ □

Proposition 3.14

If $a_n \to a$ and $b_n \to b$, then the sequence $c_n = a_n b_n$ converges to ab.

Proof. This is sufficiently complicated that we need to do a bit of calculations before writing down the proof. We want to estimate $|a_n b_n - ab|$. A trick which is generally useful when we have products (for example, when proving that the product of differentiable functions is differentiable) is that these are hard to compare, but we can instead write

$$|a_n b_n - ab| = |a_n b_n - a_n b + a_n b - ab|\,.$$

Now we can look at these two terms separately — this is

$$|a_n(b_n - b) + (a_n - a)b| \leq |a_n(b_n - b)| + |(a_n - a)b| = |a_n|\,|b_n - b| + |a_n - a|\,|b|\,.$$

This looks great — $|b|$ is a constant, we can make $|a_n - a|$ as small as we want, and we can also make $|b_n - b|$ as small as we want. But we don't like $|a_n|$, because it's not fixed.

But to deal with this we can make an observation:

Claim — If $a_n \to a$, then $\{|a_n| \mid n \in \mathbb{N}\}$ is a bounded set.

Once we have this, we can take $L = \sup\{|a_n| \mid n \in \mathbb{N}\}$. Then we get

$$|a_n|\,|b_n - b| + |a_n - a|\,|b| \leq L\,|b_n - b| + |b|\,a_n - a.$$

Then we can define $M = \max\{L, |b|\} + 1$ (the $+1$ is to avoid division by 0), so that

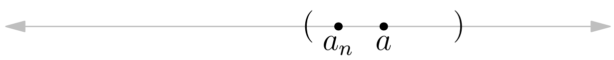
$$|a_n|\,|b_n - b| + |a_n - a|\,|b| \leq M(|b_n - b| + |a_n - a|).$$

Then this is the same as before, except we're using both ideas at the same time — given ε, there exists N_1 such that if $n \geq N_1$ then $|a_n - a| < \varepsilon/2M$, and N_2 such that if $n \geq N_2$ then $|b_n - b| < \varepsilon/2M$. Then if we take $N = \max\{N_1, N_2\}$, both these inequalities hold, and therefore

$$|a_n b_n - ab| \leq M(|b_n - b| + |a_n - a|) < M\left(\frac{\varepsilon}{2M} + \frac{\varepsilon}{2M}\right) = \varepsilon.$$

Now it remains to prove our observation.

Proof. This is intuitively obvious — as long as we go sufficiently far out, all terms are close to a.



Take N such that for all $n \geq N$, $|a_n - a| < 1$, so then $a - 1 \leq a_n \leq a + 1$. This means

$$|a_n| \leq \max\{|a - 1|\,, |a + 1|\}.$$

So if n is large, then $|a_n|$ is definitely bounded.

Now we can take all $|a_n|$ for $n \in \mathbb{N}$. We already have a bound if $n \geq N$, so there's only $N - 1$ numbers that we haven't already bounded. But that's finitely many, and any finite set has a maximum — so we can take the maximum of these extra terms, along with the bound we have for the tail, and this gives a bound for the whole set. ■

So then we are done. □

Proposition 3.15

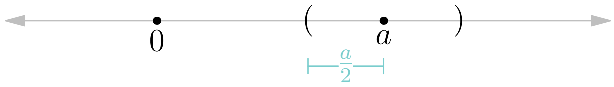
Suppose a_n is a sequence such that $a_n \neq 0$ for all $n \in \mathbb{N}$, and a_n converges to $a \neq 0$. Then the sequence $b_n = \frac{1}{a_n}$ converges to $\frac{1}{a}$.

Proof. Again, let's start by doing some side calculations — we have

$$\left|\frac{1}{a_n} - \frac{1}{a}\right| = \left|\frac{a - a_n}{a_n a}\right|.$$

The numerator is exactly what we can make as small as we want. The denominator is maybe not so great — as long as it's bounded away from 0, we're fine, but what if it becomes really small?

But now since $a_n \to a$, for N sufficiently large we can bound a_n away from 0:



We know there exists N_1 such that for all $n \geq N_1$, we have $a_n > a/2$ (or rather $|a_n| < |a/2|$, in case a is negative). Then

$$\left|\frac{a - a_n}{a_n a}\right| \leq \frac{|a - a_n|}{\frac{a}{2} \cdot a} \leq \frac{2}{a^2}\left|a - a_n\right|,$$

and the same argument as before works. $\square$

So now given two sequences, we can add them, scale them by real numbers, multiply, and divide. We can also mix up these operations — so that gives us a lot of ways to take some sequences and produce new sequences that still converge.

Example 3.16

If $a_n \to a$ and $b_n \to b$, then $a_n + \lambda b_n \to a + \lambda b$.

§3.4 Subsequences

Convergence is really important, but of course, we've seen a bunch of sequences that did not converge. Suppose we have a sequence that doesn't converge — then to understand it, we might try to look at a *subsequence*.

Example 3.17

Consider the sequence

$$a_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even,} \end{cases}$$

or the sequence

$$b_n = (-1)^n = \begin{cases} -1 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even.} \end{cases}$$

These don't converge, but they have *subsequences* — where we selectively pick out some of the elements, but maintain the ordering (and we can't pick an element more than once) — that do converge. For example, the sequence a_n is 0, 1, 0, 1, ... So we can pick out the sequences 1, 1, 1, 1, . . . or 0, 0, 0, 0, ... (In fact, in this example, any sequence of 0's and 1's is a subsequence.) Although our original sequence doesn't converge, both of these subsequences converge. The same occurs for b_n.

Example 3.18

Take the sequence $a_n = n$, which consists of 1, 2, 3, ... Then the Fibonacci numbers are a subsequence. So are the primes, or the even numbers, or odd numbers. (But we have to take these in order.) Meanwhile, something that's *not* a subsequence is 1, 4, 2, . . . — if we're at 4, we cannot go backwards (we have to maintain the ordering.) Similarly, 1, 2, 2, . . . is not a subsequence — we can't take 2 twice, since it only occurs at one spot.

Intuitively, we pick *some* of the elements of the sequence; but we have to pick them in order (with no repeated indices), and we have to pick infinitely many.

Now we'll formally define a subsequence.

Definition 3.19. A function $f\colon \mathbb{N} \to \mathbb{N}$ is (strictly) *increasing* if whenever $n_1 < n_2$, we have $f(n_1) < f(n_2)$.

Definition 3.20. A **subsequence** of a sequence a_n is a new sequence $b_n = a_{f(n)}$ for an increasing function $f\colon \mathbb{N} \to \mathbb{N}$.

It's clear that as we progress in the b_n, we go further and further out in the a_n.

Remark 3.21. We can think of subsequences as function composition — we have $a\colon \mathbb{N} \to \mathbb{R}$ and $b\colon \mathbb{N} \to \mathbb{R}$, where $b = a \circ f$.

$$\begin{tikzcd} \mathbb{N} \arrow[rr, "f"] \arrow[dr, "b"'] & & \mathbb{N} \arrow[dl, "a"] \\ & \mathbb{R} & \end{tikzcd}$$

This is the formal definition of a subsequence, but we often denote subsequences of a_n by $b_k = a_{n_k}$. Here $n_k = f(k)$.

Example 3.22 (A Non-Example)

Consider the sequence of even numbers $a_n = 2n$. Then 2, 4, 2, . . . or 2, 4, 4, . . . is *not* a subsequence.

Now we have a theorem that's trivial, but surprisingly turns out to be sometimes useful:

Theorem 3.23

Suppose a_n is a sequence. Then a_n is convergent to a if and only if all its subsequences converge to a.

Proof. We need to prove an if and only if statement. One direction, proving that if all subsequences converge to a then so does a_n, is trivial — the most trivial example of an increasing function is $f(n) = n$, which means the most trivial subsequence of a_n is a_n itself. So this follows from the fact that a_n itself is a subsequence of a_n (which means if all subsequences converge, then a_n converges).

For the other direction, assume $a_n \to a$; then we want to show that if a_{n_k} is a subsequence, then $a_{n_k} \to a$ as well.

Given $\varepsilon > 0$, we know there exists N such that if $n \geq N$, then $|a_n - a| < \varepsilon$. But now note that $n_k = f(k) \geq k$ as well, since f is increasing. So if $k \geq N$, then $n_k \geq N$ as well, which means $|a_{n_k} - a| < \varepsilon$. $\square$

This is not very deep, but it's actually a useful property.

§3.5 Monotonicity

As motivation for what will follow, we'll mention a classical example.

Example 3.24 (Zeno's Paradox)

We have a man and a turtle; the turtle has a head start, and the man wants to catch up with the turtle. This should be fairly doable. But in order to catch up with the turtle, the man first has to go half of the distance the turtle has gone. And then he has to go half the distance again. So it seems that the man can never catch up with the turtle.

This is an example of a sequence — let $a_1 = 1$, $a_2 = 1 + \frac{1}{2}$, $a_3 = 1 + \frac{1}{2} + \frac{1}{4}$, $a_4 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$, and so on —

$$a_n = 1 + \frac{1}{2} + \cdots + \frac{1}{2^{n-1}}.$$

Then a_n is a sequence. We can observe that $a_n \leq a_{n+1}$ for all n (in fact $a_n < a_{n+1}$), so this is an example of an *increasing* sequence.

Definition 3.25. A sequence a_n is **monotone increasing** if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$. Similarly, a sequence is **monotone decreasing** if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$.

Remark 3.26. The textbook defines the inequalities to be strict, and calls a sequence monotone *nondecreasing* or *nonincreasing* for the weak inequalities.

Question 3.27. When is an increasing or decreasing sequence convergent?

Theorem 3.28

If a_n is increasing, then a_n is convergent if and only if a_n is bounded.

To say that a sequence is bounded means that $\{a_n \mid n \in \mathbb{N}\}$ is bounded. Of course, the same theorem is also true for decreasing sequences, so we could restate the theorem as follows:

Theorem 3.29

If a_n is monotone, then a_n is convergent if and only if a_n is bounded.

Proof. Last time, we observed that if *any* sequence a_n is convergent, then a_n is bounded. This proves one direction immediately.

Now suppose that a_n is bounded. Then we can set

$$a = \sup\{a_n \mid n \in \mathbb{N}\}.$$

We want to show that $a_n \to a$.

First, we must have $a_n \leq a$ for all a.

On the other hand, given $\varepsilon > 0$, we know $a - \varepsilon$ is *not* an upper bound for $\{a_n\}$ (since a is the least upper bound), so there exists N with $a - \varepsilon < a_N$. But since a_n is monotone increasing, then if $n \geq N$ we have $a_n \geq a_N > a - \varepsilon$ as well. So then for all $n \geq N$ we have

$$a - \varepsilon < a_N \leq a_n \leq a,$$

which implies that

$$|a_n - a| < \varepsilon.$$

□

So if we have a monotone sequence, it's easy to determine whether it's convergent — it's convergent if and only if it's bounded.

Example 3.30

Does the sequence $a_n = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^{n-1}}$ from Zeno's paradox converge?

Solution. It suffices to check whether it is bounded. But multiplying by $1 - \frac{1}{2}$, we get

$$\left(1 - \frac{1}{2}\right) a_n = \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{n-1}}\right) - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \frac{1}{2^n}\right) = 1 - \frac{1}{2^n}.$$

This means

$$\frac{1}{2} a_n = 1 - \frac{1}{2^n} \leq 1,$$

and therefore $a_n \leq 2$ for all n.

So the sequence a_n is bounded, which means it is convergent.

But this actually proves more — this proves

$$a_n = 2\left(1 - \frac{1}{2^n}\right).$$

We can observe that $\frac{1}{2^n} \to 0$ (this is just a slight variation of the Archimedean property), and therefore using our algebraic operations from earlier, we can see that $a_n \to 2$. So in this case, we could even evaluate the limit. □

We will come back to this later on when we talk about series — this is called a *geometric series*.

§3.6 The Sandwich Principle

Sequences are often easier to understand than series, but occasionally we may have a somewhat crazy sequence.

Example 3.31

Consider the sequence

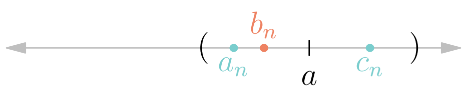
$$b_n = \frac{n^2+1}{n^2+n+1}.$$

Does b_n converge?

Theorem 3.32 (The Sandwich Principle)

Suppose we have three sequences a_n, b_n, and c_n, and $a_n \to a$ and $c_n \to a$. Suppose also that $a_n \le b_n \le c_n$ for all n. Then $b_n \to a$ as well.

Proof. Since $a_n \to a$, for all $\varepsilon > 0$ there exists N_1 such that if $n \ge N_1$, then $|a_n - a| < \varepsilon$. But since $c_n \to a$ as well, there also exists N_2 such that if $n \ge N_2$, then $|c_n - a| < \varepsilon$.



Now take $N = \max\{N_1, N_2\}$. Then we have both $|a_n - a| < \varepsilon$ and $|c_n - a| < \varepsilon$. We can write these as $a - \varepsilon < a_n < a + \varepsilon$, and $a - \varepsilon < c_n < a + \varepsilon$. But now of course, for $n \ge N$ we have

$$a - \varepsilon < a_n \le b_n \le c_n < a + \varepsilon,$$

which means $|b_n - a| < \varepsilon$ as well. □

So if one sequence is sandwiched between two other sequences, which converge to the same limit, then our original sequence must also converge to the same limit.

Solution to Example 3.31. We can write

$$b_n = \frac{n^2+n+1-n}{n^2+n+1} = 1 - \frac{n}{n^2+n+1}.$$

Now take $c_n = 1$, and $a_n = 1 - \frac{1}{n}$. Then it's clear that $c_n \to 1$ and $a_n \to 1$ (since $\frac{1}{n} \to 0$), so it suffices to show that b_n is sandwiched between them. But for any $n > 0$ we have

$$\frac{n}{n^2+n+1} = \frac{1}{n+1+\frac{1}{n}} < \frac{1}{n}$$

(since 1 and $\frac{1}{n}$ are positive). This means

$$b_n = 1 - \frac{n}{n^2+n+1} > 1 - \frac{1}{n} = a_n,$$

and of course $b_n \leq 1$ trivially (since we're subtracting something positive). $\square$

§3.7 Cauchy Sequences

In general, it's usually not super complicated to determine whether a sequence is convergent or not, but it might be very complicated to determine what the limit is. In order to prove convergence, we don't *need* to be able to evaluate the limit — an example is the monotone convergence theorem. Now we'll see another example of how we can prove convergence without finding the limit.

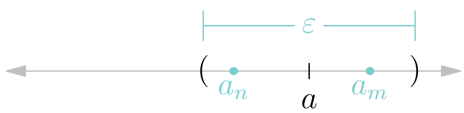
> **Definition 3.33.** A sequence a_n is a Cauchy sequence if for all $\varepsilon > 0$, there exists N such that if $n, m \geq N$, then
> $$|a_n - a_m| < \varepsilon.$$

Intuitively, a sequence is a Cauchy sequence if all elements bunch together sufficiently close when we go sufficiently far out.

> **Lemma 3.34**
>
> If a_n is convergent, then a_n is a Cauchy sequence.

The more important point is the *converse* of this lemma — if a_n is Cauchy, then it must be convergent. That's really useful, because it's often easier to prove something's a Cauchy sequence than to prove it's convergent from the definition. We'll come back to this in a later lecture.

Proof. Given $\varepsilon > 0$, since $a_n \to a$ there exists N such that $|a_n - a| < \varepsilon/2$, so then $a - \varepsilon/2 < a_n < a + \varepsilon/2$. In particular, if $m, n \geq N$ then we have $a - \varepsilon/2 < a_m, a_n < a + \varepsilon/2$.



But then a_n and a_m both lie in an interval of length ε, and their distance is at most ε. $\square$

As mentioned earlier, the reason Cauchy sequences are important is the following:

> **Theorem 3.35**
>
> Any Cauchy sequence is convergent.

We'll come back to the proof in a future lecture; it uses the fact that $\mathbb{R}$ is complete.

§3.7.1 Application to Differential Equations

This has an application to ODE (ordinary differential equations) and PDE (partial differential equations). Everything in nature — whether it's in equilibrium or evolving — satisfies some differential equations, and these differential equations are typically very complicated. In the 18th and 19th century, people were always interested in finding *exact* solutions to the differential equations, but there's few examples where this is possible — in general, we may have a very complicated equation, and it's unclear whether there even *are* solutions or how they behave.

A standard technique to find solutions uses Cauchy sequences. More precisely, it uses the *contracting mapping property.*

We want to solve some crazy differential equations, and we can't do it in one shot — we can't find an exact solution. But we can find an *approximate* solution. Then we can subtract the first solution, and get something that's a better approximation. Then we can repeat this process.

Definition 3.36. A map $T\colon \mathbb{R} \to \mathbb{R}$ is **contracting** if for all x and y,

$$|T(x) - T(y)| \leq \lambda |x - y|$$

for some fixed $0 < \lambda < 1$.

So if we take two points, then the map brings them closer together.

When solving differential equations, you'd use this property on a more fancy space than a line, but the idea is the same.

The main idea is that to solve a differential equation, we *want* to find a fixed point — a point with $T(x) = x$. We can start with anything — we take any x_0, and then we consider

$$x_0, T(x_0), T(T(x_0)), T(T(T(x_0))), \ldots.$$

This gives a sequence $a_n = T^n(x_0)$.

Proposition 3.37

The sequence a_n converges.

This is useful because we're going to later show that if $a_n \to a$, then $T(a) = a$.

Proof. We will use Theorem 3.35 — so it is enough to show that a_n is Cauchy, meaning that we want to show if m and n are sufficiently large, then $|T^m(x_0) - T^n(x_0)|$ is sufficiently small.

Let's first look at a special case — consider *consecutive* terms $T^m(x_0) - T^{m+1}(x_0)$. Then

$$\left|T^m(x_0) - T^{m+1}(x_0)\right| = \left|T(T^{m-1}(x_0)) - T(T^m(x_0))\right|.$$

But by the contracting mapping property, we have

$$\left|T^m(x_0) - T^{m+1}(x_0)\right| \leq \lambda \left|T^{m-1}(x_0) - T^m(x_0)\right|.$$

So the distance between two consecutive terms shrinks from the previous distance by a factor of λ. If we then move one step further back, we get

$$\left|T^m(x_0) - T^{m+1}(x_0)\right| \leq \lambda^2 \left|T^{m-2}(x_0) - T^{m-1}(x_0)\right|,$$

and repeating it even more times gives

$$\left|T^m(x_0) - T^{m-1}(x_0)\right| \le \lambda^{m-1} \left|T(x_0) - x_0\right|.$$

Without loss of generality, let's assume $n = m + k$ for some positive k. Now

$$|a_{m+k} - a_m| = |a_{m+k} - a_{m+k-1} + a_{m+k-1} - a_m| \le |a_{m+k} - a_{m+k-1}| + |a_{m+k-1} - a_m|$$

by the triangle inequality. Our first difference is two consecutive terms, so we can use our above calculation — we have

$$|a_{m+k} - a_{m+k-1}| \le \lambda^{m+k-2} \left|T(x_0) - x_0\right|.$$

Meanwhile, for the second term, we can split it again as

$$|a_{m+k-1} - a_m| \le |a_{m+k-1} - a_{m+k-2}| + |a_{m+k-2} - a_m|$$

by the triangle inequality. Our first term is again bounded by

$$|a_{m+k-1} - a_{m+k-2}| \le \lambda^{m+k-3} \left|T(x_0) - x_0\right|.$$

For the sake of illustration, let's imagine $\lambda = \frac{1}{2}$. Then we get $\frac{1}{2}$ to some power, and then to one less power, and so on — so we have the sum that appears in Zeno's paradox, except that all terms are multiplied by an additional factor.

In Zeno's paradox, we proved that the sequence $1 + \frac{1}{2} + \cdots$ is convergent. But then that sequence was a Cauchy sequence. That means this sequence is also a Cauchy sequence — the quantity $|a_{m+k} - a_m|$ can be made smaller than any ε, if m is sufficiently large. $\square$

Many differential equations can be set up in terms of a contracting map. So this means if we're looking for some solution (which is a function, or something complicated), then we can start with *anything*, and then look at its image, and then the image of that, and so on. And if we repeat this process, we actually get a limit!

We still need to show that the limit is actually a fixed point.

Proposition 3.38

If $a_n \to a$, then $T(a) = a$.

Proof. We have $T(a_n) = a_{n+1}$, so $T(a_n) \to a$ as well. Now if we can show that $T(a_n) \to T(a)$, we could use the fact that any sequence that's convergent has a unique limit in order to finish — then we would know that $T(a_n)$ converges to both a and $T(a)$, so we could conclude that $T(a) = a$.

But by the contracting mapping property, we have

$$|T(a_n) - T(a)| \le \lambda |a_n - a| \le |a_n - a|.$$

So since $a_n \to a$, for all ε there exists N such that $|a_n - a| < \varepsilon$, and therefore $|T(a_n) - T(a)| < \varepsilon$ as well. So this means $T(a_n) \to T(a)$ as well.

Now by the uniqueness of limits, we know $T(a) = a$. $\square$

Student Question. *What if instead of being a contraction map, we had $|T(x) - T(y)| < x - y$?*

Answer. Unfortunately, this isn't enough — we need a whole bunch of things to come together, and here they don't necessarily come together at any rate. (There exist counterexamples satisfying this equation but without the same nice properties.)

§3.8 The Bolzano–Weirstrass Theorem

Previously, we've seen the following theorem:

Theorem 3.39

If a_n is monotone, then a_n is ocnvergent if and only if a_n is bounded.

We'll now use this to prove the Bolzano–Weirstrass theorem:

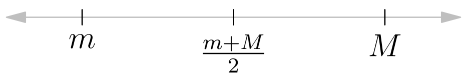
Theorem 3.40

If a_n is a bounded sequence, then a_n has a convergent subsequence.

Proof. First let's think about the idea. We know that a_n is bounded, so $\inf a_n$ and $\sup a_n$ are both real numbers; let $m = \inf a_n \leq \sup a_n = M$. Then all a_n lie in $[m, M]$.

First, set $b_1 = a_1$.

Now we can take the midway point of the interval — either there's infinitely many points in the left interval, or in the right interval. If there's infinitely many in the left interval, then we take b_2 to be the first one that comes after a_1. Suppose that's a_5.



Then we split our smaller interval into halves again — there's infinitely many points in one of the two halves, so we can take the first after a_5; suppose that's a_7.

We can keep on doing this. As we go further out, it's intuitively clear that our elements grow closer and closer together.

Now in order to prove this works, we use the sandwich principle — define two additional sequences c_n and d_n. Let $c_1 = m$ and $d_1 = M$. Then if we chose the left interval on the first step, we take $c_2 = m$ and $d_2 = (m + M)/2$. We continue in this way — we always take c_i and d_i to be the left and right endpoints of our current interval.

Then we have the property $c_1 \leq c_2 \leq c_3 \leq \cdots$, so c_n is an increasing sequence. But all the c_i are in the interval $[m, M]$, so they're all bounded above by M. So by the monotone property, c_n must converge to some limit c.

Similarly, $d_1 \geq d_2 \geq d_3 \geq \cdots$ — we either preserve the right end, or move it to the left. But all the d_i are bounded below by m; so then d_n must also converge to some limit d.

Now let's consider $|c_{n+1} - d_{n+1}|$. We can see that each time, we're taking the midpoint — so the next interval is always half the size of the previous one. So we have

$$|c_{n+1} - d_{n+1}| = \frac{1}{2} |c_n - d_n| .$$

This means $|c_2 - d_2| \leq \frac{1}{2} |c_1 - d_1|$, and more generally

$$|c_{n+1} - d_{n+1}| \leq \left(\frac{1}{2}\right)^n |c_1 - d_1| = \left(\frac{1}{2}\right)^n (M - m).$$

But we know $1/2^n \to 0$, so then $|c_{n+1} - d_{n+1}| \to 0$ as well.

Now we know $c_n \leq c \leq d \leq d_n$ — we know that the c_n are monotone nondecreasing, so $c = \sup\{c_n\}$. But we also have $c_n \leq c_{n+k} \leq d_{n+k} \leq d_n$ for all n and k. If we fix n and let k grow very large, then we see that since all $d_{n+k} \geq c_n$, we have $c_n \leq \inf\{d_n\} = d$. Then since d is an upper bound for all the c_n, it must be at least the *least* upper bound; so $c \leq d$. But then $|c - d| \leq |c_n - d_n|$, and we proved that $|c_n - d_n| \to 0$. So then we must have $c = d$.

So then c_n and d_n converge to the same limit; this means if we have any sequence $c_n \leq h_n \to d_n$, then h_n must converge to the same limit as well.

But recall that we started out with $c_1 = m$ and $d_1 = M$, and then we repeatedly defined our subsequence by taking $b_1 = a_1$, then defining b_2 to lie in the smaller interval $[c_2, d_2]$, and so on — at each step, we're choosing b_n to lie in $[c_n, d_n]$.

So then since c_n and d_n converge to the same limit, it follows by the squeezing principle that b_n converges as well. □

In fact, it turns out that our earlier theorem about Cauchy sequences is a consequence of the Bolzano–Weirstrass theorem:

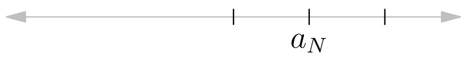
Theorem 3.41

If a_n is a Cauchy sequence, then a_n is convergent.

Recall that last time, we defined a *Cauchy sequence* — a sequence a_n such that for all $\varepsilon > 0$, there exists N such that if $m, n \geq N$, then $|a_n - a_m| < \varepsilon$. Intuitively, a sequence is Cauchy if sufficiently far out, the elements of the sequence bunch together as close as you want. Last time, we proved that if we have a convergent sequence, it must be Cauchy. But now we will prove the converse.

The converse is a property of the real numbers — later we'll see that it's a nice property to have, but it's not always true in more general settings. The reason it's true is the completeness property — it won't be true for general sets with an ordering.

Proof. Since a_n is a Cauchy sequence, then a_n must be bounded — as we saw last time, we can simply take $\varepsilon = 1$; then we know there exists N such that if $m, n \geq N$, then $|a_n - a_m| < 1$. In particular this means if we set $m = N$, then $|a_n - a_N| < 1$ for all $n \geq N$.



So from N out, everything lies in the interval $(a_N - 1, a_N + 1)$. But before N, there's only finitely many terms. So we can simply take the sup (or inf) of this interval and of the finitely many previous terms.

So a_n is bounded, and hence by the Bolzano–Weirstrass theorem, there exists a convergent subsequence — so we have $a_{n_k} \to a$ for some subsequence a_{n_k} and some a.

Lemma 3.42

If a_n is a Cauchy sequence and a_n has a convergent subsequence, then a_n converges.

Proof. Suppose $a_{n_k} \to a$. Then given $\epsilon > 0$, since a_n is Cauchy, we know there exists N_1 such that if $n, m \geq N_1$, then $|a_n - a_m| < \varepsilon/2$. But we also have that $a_{n_k} \to a$, so there exists some N_2 such that if $k \geq N_2$, then $|a_{n_k} - a| < \varepsilon/2$. Now we can combine these properties — take $N = \max\{N_1, N_2\}$. For $n \geq N$, we want to estimate $|a_n - a|$, and we can do this using the triangle inequality — we have

$$|a_n - a| \leq |a_n - a_{n_k}| + |a_{n_k} - a|.$$

Now take $k = N$, so we have $|a_{n_k} - a| \le \varepsilon/2$, and since a_{n_k} is further out than k, we also have $|a_n - a_{n_k}| < \varepsilon/2$. So then we have $|a_n - a| < \varepsilon$. ■

So since we've shown that any Cauchy sequence has a convergent subsequence, it follows that any Cauchy sequence converges. □

§3.9 Series

The fact that Cauchy sequences converge will be especially important when looking at series. Series are somewhat like sequences, but they often come up on their own.

Definition 3.43. For a sequence a_n, the sequence $s_n = a_1 + \cdots + a_n$ is called a **series**.

In a series, we have $s_{n+1} = a_1 + \cdots + a_n + a_{n+1} = s_n + a_{n+1}$ — so in other words, we're adding a term at each step. Similarly, we can see that

$$s_{n+k} = a_1 + \cdots + a_{n+k} = s_n + (a_{n+1} + \cdots + a_{n+k}).$$

So a series is really like a sequence — but it's a sequence obtained by *adding* a bunch of things together, so its basic building block is another sequence.

Notation 3.44. We use the summation notation $\sum_{i=1}^n a_i = a_1 + \cdots + a_n$.

Last class, we looked at a particular series:

Example 3.45

Consider the series $1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \cdots$ — here $a_0 = 1$, $a_1 = \frac{1}{2}$, $a_2 = \left(\frac{1}{2}\right)^2$, and more generally $a_n = \left(\frac{1}{2}\right)^n$. Then $s_n = \sum_{i=1}^n a_i = \sum_{i=1}^n \left(\frac{1}{2}\right)^n$.

Remark 3.46. We can start indices at any number, but we usually use 0 or 1.

Last time, we made a few observations: all the a_i are nonnegative, so therefore $s_{n+1} = s_n + a_{n+1} \ge s_n$. In such a case, s is convergent if and only if it is bounded. If the series *is* convergent, then we write its value as $\sum_{i=0}^\infty a_i$.

Most of the time, it's not possible to calculate this limit; but in some cases it is.

There's a few very important series:

Example 3.47

For some $\lambda \in \mathbb{R}$, the series $s_n = \sum_{i=0}^n \lambda^i$ (where $a_n = \lambda^n$) is a **geometric series**.

Usually we assume $\lambda > 0$ — if $\lambda = 0$ then the series is not very interesting. But in principle λ could also be negative.

Example 3.48

The **harmonic series** is the series $\sum_{n=1}^\infty 1/n$.

Remark 3.49. Sometimes we write the series as $\sum_{n=1}^{\infty}$ even if it's not convergent.

Example 3.50

For $\alpha > 1$, take the series $\sum_{n=1}^{\infty} 1/n^{\alpha}$.

When we have a series, the main question we're interested in is whether it's convergent. We can see how series come up — for example, they came up in an Ancient Greek paradox, and they show up whenever we want to sum infinitely many things.

For a general series, it can be very hard to determine what the limit is, but it can be much easier to determine whether it's convergent — by comparing it to series we already know.

Whenever we see a series, there's something trivial we should always check — suppose we have a series $\sum_{n=0}^{\infty} a_n$, so $s_n = a_0 + \cdots + a_n$. Then if this series is convergent, s_n must be a Cauchy sequence. But this means for all sufficiently large m and n, $|s_n - s_m|$ is arbitrarily small. In particular, we can take $m = n+1$; then $|s_n - s_{n+1}| = |a_{n+1}|$. So if the s_i are bunched together, then a_{n+1} must be small. More precisely, we have the following:

Proposition 3.51

If the series $\sum_{n=0}^{\infty} a_n$ is convergent, then $a_n \to 0$.

The converse is not true, but if a_n doesn't converge to 0, then there's no chance that the series converges. We've seen the intuition behind this already, but now let's prove this rigorously:

Proof. If $\sum_{n=0}^{\infty} a_n$ is convergent, then $s_n = a_0 + \cdots + a_n$ is also a Cauchy sequence, so then for all $\varepsilon > 0$, there exists N such that if $n, m \geq N$ then $|s_n - s_m| < \varepsilon$. Now if we take $m = n+1$, we have $|s_n - s_{n+1}| < \varepsilon$ as long as $n \geq N$. But we have $|s_n - s_{n+1}| = |a_{n+1}|$; so then we must have $|a_{n+1}| < \varepsilon$, and therefore $a_n \to 0$. □

This is a simple requirement that rules out many series from being convergent.

Now let's return to our specific examples of series.

§3.9.1 Geometric Series

We'll assume $\lambda > 0$.

Proposition 3.52

The series $\sum_{n=0}^{\infty} \lambda^n$ is convergent if $\lambda < 1$ and divergent if $\lambda \geq 1$.

Proof. First, $a_n = \lambda^n$, so if the series is convergent, we must have $\lambda^n \to 0$. It's clear that if $\lambda \geq 1$, then this is false; so the series is divergent if $\lambda \geq 1$.

Now we want to show the series is convergent if $\lambda < 1$. It turns out we can actually determine exactly what the limit is. Using the same trick we saw last class, but in more generality, we can consider

$$s_n = 1 + \lambda + \lambda^2 + \cdots + \lambda^n,$$

and then look at

$$(1-\lambda)s_n = (1 + \lambda + \cdots + \lambda^n) - (\lambda + \lambda^2 + \cdots + \lambda^{n+1}) = 1 - \lambda^{n+1},$$

since all other terms cancel out. So we have $(1-\lambda)s_n = 1-\lambda^{n+1}$, or in other words,

$$s_n = \frac{1-\lambda^{n+1}}{1-\lambda}.$$

But we have $0 < \lambda < 1$, so as $n \to \infty$, $\lambda^{n+1} \to 0$. This means

$$s_n \to \frac{1}{1-\lambda}.$$

□

In particular, in the case of $\lambda = \frac{1}{2}$, we get $\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = 2$.

Example 3.53

We've actually used series for a long time. For example, if we plug $\sqrt{2}$ into a calculator, we'd get

$$\sqrt{2} = 1.414\cdots.$$

So we can think of $\sqrt{2}$ by taking the sequence consisting of each digit — $a_1 = 1$, $a_2 = 0.4$, $a_3 = 0.01$, $a_4 = 0.004$, and so on. Then we can think of

$$\sqrt{2} = \sum_{n=0}^{\infty} a_n.$$

§3.9.2 The Harmonic Series

Question 3.54. Does the harmonic series $\sum_{n=1}^{\infty} 1/n$ converge?

We know $1/n \to 0$, so we can't determine convergence just by looking at whether $a_n \to 0$.

Proposition 3.55

The harmonic series $\sum_{n=1}^{\infty} 1/n$ is divergent.

Proof. We can write this series as

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \cdots.$$

We know that our series is convergent if and only if it's bounded. But now we can group terms: we have

$$\frac{1}{2} + \frac{1}{3} > 2 \cdot \frac{1}{4} = \frac{1}{2},$$

$$\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} > 4 \cdot \frac{1}{8} = \frac{1}{2},$$

and so on. So we'd like to group terms so that each group contributes at least $1/2$; then the series definitely can't converge.

Now let's make this rigorous.

Lemma 3.56

Let $s_n = \sum_{i=1}^{n} 1/i$; then $s_{2^n-1} \geq n/2$.

Proof. We use induction on n. In induction, we want to show that a statement holds for all $n \in \mathbb{N}$. We prove this by showing two claims:

- The induction start — the statement is true for $n = 1$;
- The induction step — if the statement holds for n, then it holds for $n + 1$.

First, when $n = 1$, we have $2^n - 1 = 1$, and $s_1 = 1 > 1/2$. So the statement is true for $n = 1$, and we have the induction start.

Now for the induction step, suppose we know $s_{2^n-1} \geq n/2$, and now consider $s_{2^{n+1}-1}$. We have

$$s_{2^{n+1}-1} = s_{2^n-1} + \frac{1}{2^n} + \frac{1}{2^n+1} + \cdots + \frac{1}{2^{n+1}-1}.$$

Now we can see that there's 2^n fractions in the sum, and the smallest one is the last — so we have

$$s_{2^{n+1}-1} \geq s_{2^n-1} + 2^n \cdot \frac{1}{2^{n+1}-1} \geq s_{2^n-1} + 2^n \cdot \frac{1}{2^{n+1}} = s_{2^n-1} + \frac{1}{2}.$$

But by the induction assumption we have $s_{2^n-1} \geq n/2$, so $s_n \geq n/2 + 1/2 = (n+1)/2$. ■

We have now shown $s_{2^n-1} \geq n/2$; this sequence is clearly not bounded, so it cannot be convergent. □

§3.10 Contracting Maps Again

Earlier, we saw the Bolzano–Weirstrass theorem, that any bounded sequence has a convergent subsequence; in particular this means that any Caucy sequence has a convergent subsequence.

Recall that earlier, we looked at *contracting maps*:

Definition 3.57. A map $T\colon [a,b] \to [a,b]$ is a **contracting map** if

$$|T(x) - T(y)| \leq L\,|x - y|$$

for all $x, y \in [a,b]$, for some fixed $L < 1$.

A few classes ago, we saw that if we start with some $x_0 \in [a,b]$ and set $a_1 = x_0$, $a_2 = T(x_0)$, $a_3 = T(T(x_0))$, and so on — in general $a_{n+1} = T(a_n)$ — then a_n is a Cauchy sequence. So then a_n converges to some limit, which we can call a_∞.

Then we used this to show that a_∞ is a fixed point of T — in other words, we have $T(a_\infty) = a_\infty$.

Example 3.58

Let f be a function $f\colon [a,b] \to [a,b]$ such that $|f'| \leq L$ for some constant $L < 1$. Then the fundamental theorem of calculus states that

$$f(x) - f(y) = \int_x^y f'(s)\,ds,$$

and taking absolute values on both sides gives

$$|f(x) - f(y)| = \left|\int_x^y f'(s)\,ds\right| \leq \int_x^y \left|f'(s)\right|\,ds \leq L\,|y - x|\,.$$

Since $L < 1$, we can see that f is then a contracting map. So this means f has a fixed point.

§3.11 Absolute Convergence

Last class, we've looked at series — for a sequence a_n of real numbers, consider the series $s_n = \sum_{i=1}^{n} a_i$.

Notation 3.59. We may write $\sum_{n=1}^{\infty} a_n$ to refer to either a series or its limit; usually it will be clear from context which one is meant.

Question 3.60. Does s_n converge?

As mentioned earlier, the first thing to consider is whether $a_n \to 0$ — if not, then the series is certainly divergent.

Definition 3.61. A series $\sum_{n=1}^{\infty} a_n$ is **absolutely convergent** if the series $\sum_{n=1}^{\infty} |a_n|$ is convergent.

Theorem 3.62

If $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then it is convergent.

Before we come to the proof, observe that $\sum_{n=1}^{\infty} |a_n|$ is a series of *nonnegative* numbers. So if we let $\tilde{s}_n = \sum_{i=1}^{n} |a_i|$, then $\tilde{s}_n$ is monotone nondecreasing. So it's convergent if and only if it's bounded.

Remark 3.63. This is a nice property because when we started with $\mathbb{R}$, we stated nine axioms (about $+$, $\cdot$, and the distributive law). But the most basic one is the commutative law — $a + b = b + a$.

But suppose we take a series $\sum_{n=1}^{\infty} a_n$. If this series is not absolutely convergent, then the ordering may matter! This seems really strange. For example, consider the **alternating series**

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n},$$

where $a_n = (-1)^n / a_n$ (it's called alternating because a_{n+1} and a_n have opposite sign — it's sometimes also called the *alternating harmonic series*). Written out explicitly, this series is

$$-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \cdots.$$

This series does converge. But imagine now that we took these infinitely many terms and reordered them. Then we could imagine taking many of the positive numbers, and then adding in one negative one. Then our sequence is kind of like the harmonic series. So we can get as far out to ∞ as we want before taking the first negative! And the first negative will set us back, but only very little. So if we keep on doing this, then the sequence will run off to ∞ — we're taking all the negative terms, but we sprinkle them in very slowly.

So if we have a series that's not absolutely convergent, then the ordering matters — in a way, the most fundamental law $a + b = b + a$ fails for infinite sums. But if we have absolute convergence, then we *can* rearrange terms (as we will see later). This is one reason why it's important to know whether a series is absolutely convergent or not.

Student Question. *Why does the alternating series converge?*

Answer. There's a simple test in general — suppose $a_n > 0$, $a_{n+1} < a_n$, and $a_n \to 0$, and we take the sequence $b_n = (-1)^n a_n$. Then the series $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty}(-1)^n a_n$ is convergent.

This is quite easy to prove. We won't state the formal proof, but intuitively — the first element will bring us some distance in the negative direction. The second will bring us in the positive direction, but not as far as where we started. The third will bring us back in the negative direction, but not as far as before. From this, we can prove that the sequence is Cauchy.

Proof of Theorem 3.62. It's enough to show that $s_n = \sum_{i=1}^{n} a_i$ is a Cauchy sequence. Consider $|s_n - s_m|$ for $n > m$. Then we have

$$|s_n - s_m| = |a_{m+1} + a_{m+2} + \cdots + a_n| \le |a_{m+1}| + |a_{m+2}| + \cdots + |a_n|$$

by the triangle inequality. But now we know that $\sum_{i=1}^{\infty} |a_i|$ is convergent. In particular, if we define $\overline{s}_n = |a_1| + \cdots + |a_n|$, then

$$|s_n - s_m| \le |a_{m+1}| + \cdots + |a_n| = |\overline{s}_n - \overline{s}_m|.$$

But since $\overline{s}_n$ is convergent, it is a Cauchy sequence, so we can make this as small as we want by taking m and n large enough; this means s_n is a Cauchy sequence as well. □

This theorem is quite important; most tests that we'll see will show that a sequence is absolutely convergent.

§3.12 Convergence of Series

Now we'll see a few tests for convergence.

§3.12.1 The Comparison Test

Theorem 3.64 (Comparison Test 1)

Consider two sequences $0 \le a_n \le b_n$, and their corresponding series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$. If $\sum_{n=1}^{\infty} b_n$ is convergent, then $\sum_{n=1}^{\infty} a_n$ is convergent as well.

This is intuitively obvious — to say that $\sum_{n=1}^{\infty} b_n$ is convergent is the same as to say that it's bounded (since a_n and b_n are nonnegative); but if $\sum_{n=1}^{\infty} b_n$ is bounded, so is $\sum_{n=1}^{\infty} a_n$. In fact, this also means

$$\sum_{n=1}^{\infty} a_n \le \sum_{n=1}^{\infty} b_n.$$

Of course, the logical negation is true as well — if $\sum_{n=1}^{\infty} a_n$ is divergent, so is $\sum_{n=1}^{\infty} b_n$.

The next comparison test is a slightly more fancy version, that's often more useful:

Theorem 3.65 (Comparison Test 2)

Suppose we have two series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$, where $0 < a_n$ and $0 < b_n$. Then if $\lim_{n\to\infty} a_n/b_n = L \ne 0$, then $\sum_{n=1}^{\infty} a_n$ is convergent if and only if $\sum_{n=1}^{\infty} b_n$ is convergent.

Proof. We'll prove a slightly more general statement:

Claim — Suppose we have $0 < a_n$ and $0 \le b_n$, and $\lim_{n\to\infty} b_n/a_n = L$. Suppose also that we know $\sum_{n=1}^{\infty} a_n$ is convergent. Then $\sum_{n=1}^{\infty} b_n$ is also convergent.

Proof. If $b_n/a_n \to L$, then b_n/a_n is bounded — so there exists M such that $|b_n/a_n| \le M$ for all n, which means $|b_n| \le M|a_n|$. But since our sequences are nonnegative, this means $b_n \le Ma_n$. Now if a_n is absolutely convergent, then Ma_n is also absolutely convergent. But then by the first comparison test, b_n is also convergent. ■

Now returning to the original theorem, if $\lim_{n\to\infty} a_n/b_n$ exists and is nonzero, then $\lim_{n\to\infty} b_n/a_n$ exists (and is equal to $1/L$). So the claim goes both ways — $\sum_{n=1}^{\infty} a_n$ is convergent if and only if $\sum_{n=1}^{\infty} b_n$ is convergent. □

Example 3.66

Does the series $\sum_{n=1}^{\infty} n/(n^2+1)$ converge?

Solution. Let $a_n = n/(n^2+1)$, and $b_n = 1/n$; then we know that $\sum_{n=1}^{\infty} b_n$ is divergent. Now we have

$$\frac{a_n}{b_n} = \frac{n^2}{n^2+1} = \frac{1}{1+\frac{1}{n^2}}.$$

Since $1/n^2 \to 0$, this means $a_n/b_n \to 1$, and therefore since $\sum_{n=1}^{\infty} b_n$ is divergent, so is $\sum_{n=1}^{\infty} a_n$. □

Example 3.67

Does the series $\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot \frac{1}{n}$ converge?

Solution. Take $a_n = \left(\frac{1}{2}\right)^n \cdot \frac{1}{n}$, and $b_n = \left(\frac{1}{2}\right)^n$. Then $0 \le a_n \le b_n$, and $\sum_{n=1}^{\infty} b_n$ converges (since it is a geometric series), so $\sum_{n=1}^{\infty} a_n$ converges as well. □

§3.12.2 The Ratio Test

If we know only one series, it should be the geometric series — it's the most important series.

Theorem 3.68 (Ratio Test)

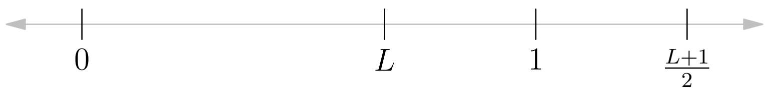
Consider a series $\sum_{n=1}^{\infty} a_n$ (where $a_n \ne 0$, but the a_n do not necessarily have to be positive), and consider the consecutive-element ratios $|a_{n+1}|/|a_n|$. If

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L$$

and $L < 1$, then $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Proof. If $|a_{n+1}|/|a_n| \to L < 1$, then there exists N such that for $n \ge N$, we have

$$\frac{|a_{n+1}|}{|a_n|} < L + \frac{1-L}{2} = \frac{L+1}{2} = \theta < 1.$$



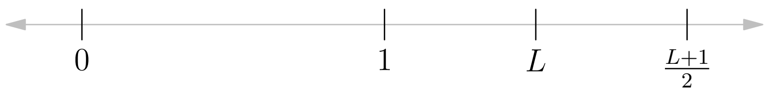
Now this means

$$|a_{n+1}| < \theta |a_n|$$

as long as $n \geq N$. But now we have $|a_{N+1}| < \theta |a_N|$, $|a_{N+2}| < \theta |a_{N+1}| < \theta^2 |a_N|$, and so on — repeating this process, we get $|a_{N+k}| < \theta^k |a_N|$ for all $k > 0$. But now we see that the entire tail of the sequence from N out is bounded above by a geometric series multiplied by a constant; so it must be convergent. □

Note that it's important that the limit is less than 1 (unlike in the second comparison test) — if we had $|a_{n+1}| / |a_n| \to L > 1$, then the sequence definitely *wouldn't* converge — there would exist N such that

$$\frac{|a_{n+1}|}{|a_n|} > \frac{L+1}{2} = \theta > 1.$$



But then $|a_{N+k}| > \theta^k |a_N|$, so the sequence would go off to ∞. (If $L = 1$, then we can't conclude anything about convergence.)

Example 3.69

Does the series $\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot n^2$ converge?

In this case, $\left(\frac{1}{2}\right)^n \to 0$ rapidly, while $n^2 \to \infty$. So it's sort of a competition between the two of them.

Solution. Let $a_n = \left(\frac{1}{2}\right)^n \cdot n^2$. Then

$$\frac{a_{n+1}}{a_n} = \frac{\left(\frac{1}{2}\right)^{n+1} \cdot (n+1)^2}{\left(\frac{1}{2}\right)^n \cdot n^2} = \frac{1}{2} \cdot \frac{(n+1)^2}{n^2}.$$

Now we can expand this as

$$\frac{1}{2} \cdot \frac{n^2 + 2n + 1}{n^2} = \frac{1}{2} \cdot \left(1 + \frac{2}{n} + \frac{1}{n^2}\right) \to \frac{1}{2}$$

(since $2/n \to 0$ and $1/n^2 \to 0$). So then $a_{n+1}/a_n \to 1/2 < 1$, which means the series is convergent by the ratio test. □

§3.13 One More Important Series

Last time, we determined that $\sum_{n=1}^{\infty} 1/n$ is divergent. We also wrote down the sequence $\sum_{n=1}^{\infty} 1/n^2$, but we didn't determine whether it's convergent or not.

Theorem 3.70

The series $\sum_{n=1}^{\infty} 1/n^2$ is convergent.

We'll prove this from first principles. But later on we'll see the *integral test*, which will show that $\sum_{n=1}^{\infty} 1/n^p$ in general (for $p > 0$) is convergent if $p > 1$ and divergent otherwise. So the harmonic series is right at the threshold of this family.

Proof. We'll use a similar argument to the one we used to prove the harmonic series is divergent. We have the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} + \cdots.$$

Now we can bunch terms together — we have

$$\frac{1}{2^2} + \frac{1}{3^2} < 2 \cdot \frac{1}{2^2} = \frac{1}{2},$$

$$\frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} < 4 \cdot \frac{1}{4^2} = \frac{1}{4},$$

and so on — we cna then take

$$\frac{1}{8^2} + \cdots + \frac{1}{15^2} < 8 \cdot \frac{1}{8^2} = \frac{1}{8},$$

and so on. But it's enough to prove that the series is bounded. And by bunching terms in this way, we've bounded the series by the geometric series, which is convergent.

Now let's do this formally, using induction. Let $s_m = \sum_{i=1}^{m} 1/i^2$. Since $1/n^2$ are nonnegative, the sequence s_m is monotone, so it's enough to show that it's bounded. We'll show that a *subsequence* is bounded:

Claim — We have $s_{2^n-1} \leq \sum_{k=1}^{n} 1/2^{k-1}$.

Proof. We use induction on n. In the base case $n = 1$, we have $2^n - 1 = 1$, and $s_1 = 1$. Meanwhile, the right-hand side is $1/2^0 = 1$, so the induction start is true.

Now for the induction step, suppose that $s_{2^n-1} \leq \sum_{k=1}^{n} 1/2^{k-1}$; now we want to look at $s_{2^{n+1}-1}$. To use the induction hypothesis, we can write

$$s_{2^{n+1}-1} = s_{2^n-1} + a_{2^n} + \cdots + a_{2^{n+1}-1}.$$

We have 2^n terms in $a_{2^n}, \ldots, a_{2^{n+1}-1}$, and the largest is the first (since the a_m are decreasing). So then

$$a_{2^n} + \cdots + a_{2^{n+1}-1} \leq a_{2^n} \cdot 2^n = 2^n \cdot \left(\frac{1}{2^n}\right)^2 = \frac{1}{2^n}.$$

Now we have

$$s_{2^{n+1}-1} \leq s_{2^n-1} + \frac{1}{2^n}.$$

Using the induction assumption, we then have

$$s_{2^{n+1}-1} \leq \sum_{k=1}^{n} \frac{1}{2^{k-1}} + \frac{1}{2^n} = \sum_{k=1}^{n+1} \frac{1}{2^{k-1}}.$$

This is exactly the claim we wanted to prove, so we are done. ■

This proves that the series s_n is bounded, and therefore convergent. □

Remark 3.71. This can be used to prove that the p-series for other $p > 1$ is also convergent, but it's somewhat tedious.

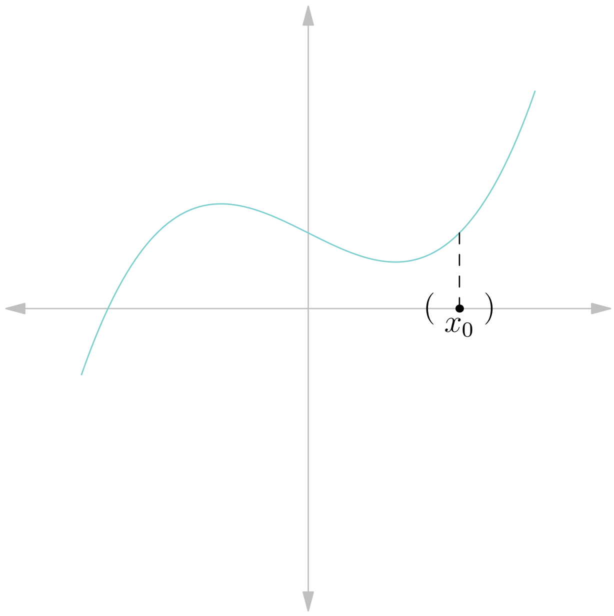
§4 The Extreme Value Theorem

The extreme value theorem is another application of Cauchy sequences; but in order to state it, we need to define continuity.

§4.1 Continuity

Definition 4.1. A function $f\colon [a,b] \to \mathbb{R}$ is continuous at some point $x_0 \in [a,b]$ if the following holds: for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$.

In words, if we're close to x_0, then the value of f should be very close to the value at x_0.



So if we have a small interval around $f(x_0)$ on the vertical axis, then we can fix a small interval around x_0 on the horizontal one sent into the vertical one.

Definition 4.2. A function $f\colon [a,b] \to \mathbb{R}$ is continuous if f is continuous at all points $x \in [a,b]$.

§4.2 The Extreme Value Theorem

Theorem 4.3

If $f\colon [a,b] \to \mathbb{R}$ is continuous, then the image of f is a bounded set, and there exists x_m and x_M such that $f(x_m) \le f(x) \le f(x_M)$ for all $x \in [a,b]$.

The **image** of f is $f([a,b]) = \{f(x) \mid x \in [a,b]\}$. So the extreme value theorem states that the image of f is bounded, and it contains a maximum and a minimum.

Proof. First, a set is bounded if it's bounded from above and below. We'll prove that $f([a,b])$ is bounded from above; the same argument can be used to show it is bounded from below.

Assume for contradiction that $f([a,b])$ is not bounded from above. Then for all $n \in \mathbb{N}$, there exists $x_n \in [a,b]$ such that $f(x_n) > n$ (otherwise n would be an upper bound). So for each integer n, we can pick some $x_n \in [a,b]$, and $f(x_n)$ is bigger than n.

Now we have a sequence $x_1, x_2, \ldots$. By the Bolzano–Weirstrass theorem, there exists a convergent subsequence x_{n_k} (since x_n is bounded — it always lies in $[a, b]$); suppose this subsequence converges to $x_\infty \in [a, b]$.

Now we know $f(x_\infty)$ is some bounded number, and f is continuous. Simply set $\varepsilon = 1$. Then we know there exists $\delta > 0$ such that if $|x - x_\infty| < \delta$, then $|f(x) - f(x_\infty)| < 1$. But the subsequence x_{n_k} converges to x, which means there exists N such that if $k \geq N$, then $|x_{n_k} - x_\infty| < \delta$.

But now for $k \geq N$ we have $|x_{n_k} - x_\infty| < \delta$, which implies $|f(x_{n_k}) - f(x_\infty)| < 1$.

But we chose x_n such that $f(x_n) > n$. So our $f(x_n)$ wander off to infinity, but they're supposed to stay within fixed distance of $f(x_\infty)$; this is a contradiction.

Now we know $f([a, b])$ is bounded from above; the same argument gives that it's bounded from below.

> **Student Question.** How do we know $f(x_\infty)$ is finite?
>
> **Answer.** It has to be a real number — $x_\infty \in [a, b]$, which means $f(x_\infty)$ is in $\mathbb{R}$.
>
> For example, consider $f(x) = 1/x$ on the interval $(0, 1]$. Even though the function goes off to ∞, at every fixed x in $(0, 1]$, $f(x)$ is still finite. But f is still unbounded.

Now we want to show that there exists $x_M \in [a, b]$ such that $f(x_M) = \sup\{f(x) \mid x \in [a, b]\}$. (The other direction is again symmetric.)

This is almost the same argument. Let $\sup\{f(x) \mid x \in [a, b]\} = M$. Now we can fix n and look at $M - 1/n$. Since this is not an upper bound for $f([a, b])$, for all n we can choose x_n such that $f(x_n) > M - 1/n$. Now again by the Bolzano–Weirstrass theorem, we know x_n has a convergent subsequence $x_{n_k} \to x_\infty$.

We claim that x_∞ is the desired x_M — we want to show that $f(x_\infty) = M = \sup\{f(x) \mid x \in [a, b]\}$.

First, we know that $f(x_\infty) \leq \sup\{f(x) \mid x \in [a, b]\}$. So we only need to show that

$$f(x_\infty) \geq \sup\{f(x) \mid x \in [a, b]\}.$$

We'll do this by showing that

$$f(x_\infty) > \sup\{f(x) \mid x \in [a, b]\} - \varepsilon$$

for all $\varepsilon > 0$. Fix some such ε. Since f is continuous at x_∞, there exists $\delta > 0$ so that if $|x - x_\infty| < \delta$, then $|f(x) - f(x_\infty)| < \varepsilon/2$. But since $x_{n_k} \to x_\infty$, there exists N such that if $k \geq N$ then $|x_{n_k} - x_\infty| < \delta$.

Now we see that if $k \geq N$, then we have $|x_{n_k} - x_\infty| < \delta$, which means

$$|f(x_{n_k}) - f(x_\infty)| < \frac{\varepsilon}{2}.$$

But now we have that $f(x_{n_k}) > M - 1/n_k$ (by the definition of the sequence). So now we have

$$f(x_\infty) > M - \frac{1}{n_k} - \frac{\varepsilon}{2}.$$

If we take k sufficiently large, then n_k is also sufficiently large. So by making n_k sufficiently large we can make $1/n_k < \varepsilon/2$, which means $f(x_\infty) > M - \varepsilon$. $\square$

§5 More About Series

Recall that for a sequence a_n of real numbers, we can look at the *series* $\sum_{n=1}^{\infty} a_n$, which may or may not be convergent — if the sequence $s_n = \sum_{i=1}^{n} a_i$ is convergent, then we say the series is convergent.

We also talked about what it means for a series to be *absolutely* convergent — we say a series $\sum_{i=1}^{\infty} a_n$ is *absolutely convergent* if the series $\sum_{i=1}^{\infty} |a_i|$ is convergent. Last time, we proved the following theorem:

Theorem 5.1

Absolute convergence implies convergence.

This is because to prove a sequence is convergent, it's enough to prove it's a Cauchy sequence. But if $\sum_{i=1}^{\infty} |a_i|$ is Cauchy, then the triangle inequality gives that the original sequence is also Cauchy.

Very often, we will test for *absolute* convergence, since that's often easier.

§5.1 Tests for Convergence

Previously, we saw two comparison tests.

1. Suppose we have two sequences $0 \leq a_n \leq b_n$. Then if $\sum_{n=1}^{\infty} b_n$ is convergent, then $\sum_{n=1}^{\infty} a_n$ is also convergent. (The negation is also useful — if $\sum_{n=1}^{\infty} a_n$ is divergent, so is $\sum_{n=1}^{\infty} b_n$.)
2. If the ratio between the two sequences is convergent — if $\lim_{n\to\infty} |a_n|/|b_n| = L$ — then if $\sum b_n$ is convergent, $\sum a_n$ is also convergent. (The converse is also true if $L \neq 0$.)

§5.1.1 The Ratio Test and Root Test

The next two tests we'll discuss are the *ratio* test and the *root* test. Both compare a given series with a geometric series — recall that the geometric series $\sum_{n=1}^{\infty} c_n$, where $c \in \mathbb{R}$, is convergent if and only if $|c| < 1$.

Theorem 5.2 (Ratio Test)

Assume $a_n \neq 0$, and suppose

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L$$

for some L. If $L < 1$, then $\sum_{n=1}^{\infty} a_n$ is absolutely convergent; meanwhile if $L > 1$, then $\sum_{n=1}^{\infty} a_n$ is divergent. If $L = 1$ then the test is inconclusive.

As a review of the proof:

Proof sketch. If $L < 1$, then we can take the midway point $L_0 = (L+1)/2 < 1$; now we know there exists N so that if $n \geq N$ then

$$\frac{|a_{N+1}|}{|a_n|} < L_0 < 1.$$

Now we have

$$|a_{n+1}| < L_0 |a_n|$$

for all $n \geq N$, so

$$|a_{n+k}| < L_0^k |a_n|$$

(by iterating this k times). Now we see that from some stage on, our series is bounded by a geometric series; so we can use the first comparison test to finish (since $L_0 < 1$) — we know that $\sum_{k=1}^{\infty} L_0^k$ converges. (Note that whether a series converges is a question about its *tail* — it doesn't really matter where we start.) More explicitly, we have $|a_{N+k}| \leq L_0^k |a_N|$. But $\sum_{i=1}^{\infty} a_i$ is convergent if and only if $\sum_{k=1}^{\infty} a_{N+k}$ is (since the second just hs finitely many terms added). If we let $b_k = a_{N+k}$, then we have $|b_k| \leq L_0^k |a_N|$. But we know $L_0^k |a_N|$ is convergent.

Of course, if $L > 1$, then our sequence a_n cannot go to 0 — the next one is a fixed multiple of the previous, where the multiple is larger than 1. So the a_n actually go off to ∞.

Meanwhile, if $L = 1$ then the test is inconclusive. This is unsurprising — we've seen that $\sum 1/n^2$ is convergent (and even absolutely convergent), where the ratio is

$$\frac{n^2}{(n+1)^2} = \frac{1}{\left(1+\frac{1}{n}\right)^2} \to 1.$$

So the ratio test is not helpful. ☐

The root test is quite similar; it again compares with a geometric series. Sometimes the root test is better, but it usually is not.

Theorem 5.3

Consider a series $\sum_{n=1}^{\infty} a_n$, and consider $\sqrt[n]{|a_n|}$ — we find the number whose nth poweri s a_n. Then if

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} < 1,$$

then (a_n) is absolutely convergent. Meanwhile, if the limit is greater than 1 then the series is divergent, and if the limit is 1 then the test is inconclusive.

Proof. Set $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L < 1$. Now again take $L_0 = (1+L)/2$, so that $L_0 < 1$ as well. Then there exists N such that if $n \geq N$ then $\sqrt[n]{|a_n|} < L_0$. This implies that $|a_n| < L_0^n$. But now we see that for $n \geq N$, our series is bounded above by a geometric series with ratio $L_0 < 1$, so this geometric series is convergent; so y the first comparison theorem, we are again done.

Similarly to before, if $L > 1$, then we can again take n so that $\sqrt[n]{a_n} > (L+1)/2$. So there exists N such that if $n \geq N$ then $\sqrt[n]{|a_n|} > L_0$. Taking the nth power of both sides, we have $|a_n| > L_0^n$. But this means $|a_n|$ wanders off to ∞, contradiction. ☐

Remark 5.4. In both situations, we can replace lim with lim sup, which we will discuss later. The proof is a trivial extension.

§6 Power Series

Definition 6.1. For a sequence a_n, its **power series** is $\sum_{n=1}^{\infty} a_n x^n$.

So for all real $x \in \mathbb{R}$, this is alseries; for differentvaleus of x we get different values.

Question 6.2. For which x is this convergent?

We could either consider the ratio test or the root test. Looking at the ratio test, we get

$$\frac{|a_{n+1}x^{n+1}|}{|a_n x^n|} = \frac{|a_{n+1}|}{|a_n|} \cdot |x|\,.$$

Now suppose that

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L \in \mathbb{R}.$$

First, if $L = 0$, then we have

$$\lim_{n\to\infty} \frac{|a_{n+1}|\, x^{n+1}}{|a_n|\, x^n}$$

is convergent, so $\sum a_n x^n$ is absolutely convergent by the ratio test.

Now suppose that $L > 0$. Then $\lim |a_n x| = |x| \cdot L$. The ratio test states the series is absolutely convergent if $|x| < 1/L$, and divergent if $|x| > L/d$.

Then $1/L$ is called the **radius of convergence**.

Example 6.3

Consider $\sum_{n=1}^{\infty} x_n/n$.

In this situation, we want to consider

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \lim_{n\to\infty} \frac{n}{n+1} = \lim_{n\to\infty} \frac{1}{1+\frac{1}{n}} = 1.$$

So we see that the radius of convergence is $1/1 = 1$.

Question 6.4. What happens at the boundary?

We know the radius of convergence is 1, so for each $-1 < x < 1$, the power series converges (we can think of the power seires as a function on $(-1, 1)$. We also know it's divergent past -1 and 1. But what happens at ± 1? In fact, at 1 we get the harmonic series, which is divergent; and at -1 we get the alternating harmonic series, which is convergent.

Now let's apply the root test to a power series. We have a power series $\sum_{n=1}^{\infty} a_n x^n$, and the root test considers

$$\sqrt[n]{|a_n x^n|} = \sqrt[n]{|a_n|} \cdot \sqrt[n]{|a_n|} \cdot |x|\,.$$

Now suppose $\lim_{n\to\infty} \sqrt[n]{|a_n|} = L$. Then:

- If $L = 0$, then by the root test, our series is absolutely convergent for all x.
- If $L > 0$, then $\lim_{n\to\infty} \sqrt[n]{|a_n|}\,|x| < 1$ if and only if $|x| < 1/L$. So if $|x| < 1/L$ then we have absolute convergence.

Again, $1/L$ is called the **radius of convergence**. The power series always converges inside and never outside, but on the boundary it's a case-by-case analysis.

§7 Continuous Functions

Recall that we say a function $f\colon [a, b] \to \mathbb{R}$ is *continuous* if it is continuous at all points. A function is continuous at one point x_0 if for all $\varepsilon > 0$, there exists some $\delta > 0$ such that if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$.

§7.1 A Useful Theorem

One thing that's often useful about continuous functions, and can be used to characterize them, is the following:

Theorem 7.1

If $f\colon [a, b] \to \mathbb{R}$ is continuous, and $x_n \to x_\infty$ (where $x_n \in [a, b]$ for all n), then $f(x_n) \to f(x_\infty)$.

Proof. For all $\varepsilon > 0$, we know there exists δ such that if $|y - x_\infty| < \delta$, then $|f(y) - f(x_\infty)| < \varepsilon$. Now given $\varepsilon > 0$, find this value $\delta > 0$. Since $x_n \to x_\infty$, there exists N such that if $n \geq N$, then $|x_n - x_\infty| < \delta$. Putting these together, if $n \geq N$, then $|x_n - x_\infty| < \delta$, which means by continuity that $|f(x_n) - f(x_\infty)| < \varepsilon$. But this is exactly what it means for $f(x_n)$ to converge to $f(x_\infty)$. $\square$

Now let's use this to prove a little lemma, that will be useful later.

Lemma 7.2

Suppose $f, g\colon \mathbb{R} \to \mathbb{R}$ are continuous, and $f = g$ on $\mathbb{Q}$. Then $f = g$ on $\mathbb{R}$.

In general, we can imagine we could take a function f that's 0 everywhere, and another function g that's 0 on $\mathbb{Q}$ and 1 on $\mathbb{R} \setminus \mathbb{Q}$ (for example, we'd have $g(\sqrt{2}) = 1$). This function g is *not* continuous, but these two functions coincide on $\mathbb{Q}$ but are not the same on $\mathbb{R}$. But this lemma states that if both functions are *continuous* and coincide on $\mathbb{Q}$, then they're the same.

Proof. Take some $x \in \mathbb{R}$. If $x \in \mathbb{Q}$, then we know $f(x) = g(x)$; so we can assume $x \in \mathbb{R} \setminus \mathbb{Q}$. Then by completeness, we know that if we take any number in $\mathbb{R}$, then there is a sequence of rational numbers that converge to it — we can look at all rational numbers $q < x$, and $\sup\{q \in \mathbb{Q} \mid q < x\}$ must equal x.

Now take a sequence $x_n \in \mathbb{Q}$ such that $x_n \to x$. Then since f is continuous, we know $f(x_n) \to f(x)$, and since g is continuous, we know $g(x_n) \to g(x)$. But $f(x_n) = g(x_n)$ for all n, since $x_n \in \mathbb{Q}$, so our two sequences are the same. But we've proven that if a sequence has a limit, then that limit is unique; this means $f(x) = g(x)$. $\square$

§7.2 Examples of Continuous Functions

Example 7.3

The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f = 1$ is continuous (as is any constant function).

Proof. For any $x, y \in \mathbb{R}$, we have $|f(x) - f(y)| = 0$. This is smaller than ε, so for any ε we can actually choose δ to be *anything*. $\square$

Example 7.4

The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ is continuous.

Proof. Fix some x_0. Then we have $|f(x_0) - f(x)| = |x_0 - x|$. So given any ε, we can choose $\delta = \varepsilon$; then if $|x_0 - x| < \varepsilon$ we have $|f(x_0) - f(x)| < \varepsilon$ as well. $\square$

Proposition 7.5

If $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ are both continuous, then $f + g$ is continuous as well.

The functions don't need to be defined over all of $\mathbb{R}$ for this to be true.

Proof. Fix some x_0; then we want to prove that $f+g$ is continuous at x_0. Now we have

$$|(f+g)(x_0)-(f+g)(x)| = |f(x_0)-f(x)+g(x_0)-g(x)| \le |f(x_0)-f(x)|+|g(x_0)-g(x)|$$

by the triangle inequality. Now we're ready to prove the statement — given $\varepsilon > 0$, since f is continuous at x_0, we can choose $\delta_1 > 0$ so that if $|x-x_0| < \delta_1$, then $|f(x)-f(x_0)| < \varepsilon/2$. Similarly, since g is continuous at x_0, we can choose $\delta_2 > 0$ so that if $|x-x_0| < \delta_2$, then $|g(x)-g(x_0)| < \varepsilon/2$. Now let $\delta = \min\{\delta_1, \delta_2\}$. Then if $|x-x_0| < \delta$, both statements hold; so

$$|(f+g)(x_0)-(f+g)(x)| < \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon. \qquad \square$$

Remark 7.6. This is very similar to our proof that $\lim_{n\to\infty}(a_n+b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$.

Proposition 7.7

If $f, g\colon \mathbb{R} \to \mathbb{R}$ are continuous, then $(fg)(x) = f(x)g(x)$ is continuous.

We can again prove this in the same way as we did for sequences:

Proof. Fix some x_0; we'll show that fg is continuous at x_0. We want to consider

$$|f(x)g(x) - f(x_0)g(x_0)|.$$

We can perform a little trick here: we can't compare these two terms, so we add something and subtract it to have terms that we can compare, to get

$$\begin{aligned} |f(x)g(x) - f(x_0)g(x_0)| &= |f(x)g(x) - f(x)g(x_0) + f(x)g(x_0) - f(x_0)g(x_0)| \\ &\le |f(x)|\,|g(x)-g(x_0)| + |g(x_0)|\,|f(x)-f(x_0)| \end{aligned}$$

by the triangle inequality. We'd now like to take x close enough to x_0 to make both terms less than $\varepsilon/2$. For the second term, this is straightforward, since $|g(x_0)|$ is a constant. On the other hand, $|f(x)|$ is not a constant. But we can deal with it in a similar way: first, since f is continuous at 0, there exists δ_1 such that if $|x-x_0| < \delta_1$, then $|f(x)-f(x_0)| < 1$. We can write $f(x) = f(x)-f(x_0)+f(x_0)$, so the triangle inequality gives that

$$|f(x)| \le |f(x)-f(x_0)| + |f(x_0)| = 1 + |f(x_0)|.$$

So then $|f(x)|$ is bounded by some constant, as long as $|x-x_0| < \delta_1$.

Now we choose δ_2 and δ_3 — we'll eventually set $\delta = \min(\delta_1, \delta_2, \delta_3)$. To choose δ_2, since g is continuous at x_0, we know given $\varepsilon > 0$, there exists $\delta_2 > 0$ such that if $|x-x_0| < \delta_2$, then

$$|g(x)-g(x_0)| < \frac{\varepsilon}{2(1+|f(x_0)|)}.$$

Meanwhile, since f is continuous at x_0, we know given $\varepsilon > 0$, there exists $\delta_3 > 0$ such that if $|x-x_0| < \delta_3$, then

$$|f(x)-f(x_0)| < \frac{\varepsilon}{2(1+|g(x_0)|)}.$$

(The added 1 is simply to avoid division by 0.)

Now if we choose $\delta = \min(\delta_1, \delta_2, \delta_3)$, then we have all three properties:

$$|f(x)-f(x_0)| < \frac{\varepsilon}{2(1+|g(x_0)|)}, \; |g(x)-g(x_0)| < \frac{\varepsilon}{2(1+|f(x_0)|)}, \text{ and } |f(x)| \le 1+|f(x_0)|.$$

Now plugging this in, we get that

$$|g(x_0)|\,|f(x) - f(x_0)| < |g(x_0)| \cdot \frac{\varepsilon}{2(1 + |g(x_0)|)} < \frac{\varepsilon}{2},$$

and similarly

$$|f(x)|\,|g(x) - g(x_0)| < (1 + |f(x_0)|) \cdot \frac{\varepsilon}{2(1 + |f(x_0)|)} = \frac{\varepsilon}{2}.$$

□

As a recap, we've proven that $f = 1$ is continuous, $f = x$ is continuous, and if f and g are continuous, then $f + g$ and fg are continuous. In fact, if f is any constant function, then it's continuous; so applying the last fact, we immediately get that for any constant c and continuous function g, the function cg is continuous.

Theorem 7.8

If f is a polynomial, then f is continuous.

Note that a polynomial is not a series — it's a *finite* sum.

Proof. Any polynomial can be built by taking constants and x, and repeatedly adding and multiplying. For instance, $x^3 = x \cdot x^2 = x \cdot x \cdot x$. So since x is continuous, then x^2 is continuous, and therefore so is x^3. Likewise, $x^2 + 5$ is continuous because x^2 and 5 are both continuous. □

Theorem 7.9

If f is continuous and $f(x) \neq 0$ for all x, then $1/f(x)$ is continuous.

This again is very similar to the fact that if we have a sequence a_n where none of the elements is 0, and the sequence converges to a nonzero limit, then their reciprocals converge to the reciprocal of the limit.

Proof. Fix x_0; then we have

$$\left|\frac{1}{f(x)} - \frac{1}{f(x_0)}\right| = \frac{|f(x_0) - f(x)|}{|f(x)f(x_0)|}.$$

Now since $f(x_0) \neq 0$, there exists $\delta_1 > 0$ such that if $|x - x_0| < \delta_1$, then

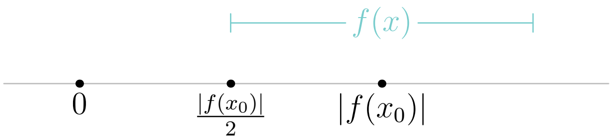
$$|f(x) - f(x_0)| < \frac{|f(x_0)|}{2}$$

(taking $\varepsilon = |f(x_0)|/2$ in the definition of continuity.) Using the triangle inequality again, this means

$$|f(x)| - |f(x_0)| < \frac{|f(x_0)|}{2},$$

which means

$$|f(x)| > \frac{|f(x_0)|}{2}.$$



Now we have

$$\left|\frac{1}{f(x) - \frac{1}{f(x_0)}}\right| = \frac{|f(x) - f(x_0)|}{|f(x)| \cdot |f(x_0)|} < \frac{2\,|f(x) - f(x_0)|}{|f(x_0)|^2}.$$

Now given ε, we can choose $\delta_2 < \delta_1$ such that if $|x - x_0| < \delta_2$ then

$$|f(x) - f(x_0)| < \frac{\varepsilon}{2} \cdot |f(x_0)|^2 .$$

Now using our previous estimate, we have

$$\left|\frac{1}{f(x)} - \frac{1}{f(x_0)}\right| < \frac{2}{|f(x_0)|^2} \cdot \frac{\varepsilon}{2}\,|f(x_0)|^2 = \varepsilon. \qquad \square$$

This is the same sort of thing we did for sequences. For a sequence, we saw that using this, from simple sequences we could construct a lot more sequences. Similarly, here all polynomials are continuous; and all rational functions ($p(x)/q(x)$ for polynomials p and q) are also continuous, as long as the polynomial in the denominator doesn't vanish.

§7.3 The Exponential Function

Now we'll define an important function that appears everywhere, the *exponential function.* (One standard place in which they appear is with radioactivity or explosive population growth — in a fixed amount of time, something may halve or double.)

Consider the power series

$$E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

where $n!$ is the function defined as $0! = 1$, $1! = 1$, $2! = 2 \cdot 1$, $3! = 3 \cdot 2 \cdot 1$, and so on — in general,

$$n! = n(n-1)(n-2)\cdots 1.$$

(Here $n \in \mathbb{N} \cup \{0\}$.) This is a power series, and when we see a power series, the first thing we can do is try to determine the radius of convergence. To do so, we can apply the ratio test: we have

$$\frac{\frac{x^{n+1}}{(n+1)!}}{\frac{x^n}{n!}} = \frac{x^{n+1}}{(n+1)!} \cdot \frac{n!}{x^n} = \frac{x^n \cdot x}{(n+1) \cdot n!} \cdot \frac{n!}{x^n} = \frac{x}{n+1}.$$

We can see that for *any* given x, this ratio converges to 0 as $n \to \infty$; so this power series is convergent everywhere.

So then we can think of $E(x)$ as a *function* of x — it's well-defined for all x.

We could define this as the exponential function. But this is not very satisfying, because we have some idea of the properties the exponential function should have. So instead, let's define the exponential function based on some of those properties.

First, we can make a few observations. In a power series, by definition we have $x^0 = 1$ (even if we plug in $x = 0$), so then $E(0) = 1$ — the first term $x^0/0!$ is defined as 1, and all future terms are clearly 0. But then $E(1)$ is some number in $\mathbb{R}$, which we define as e. Note that $e \approx 2.7\cdots$ is not rational.

Now for $k \in \mathbb{N}$, we can define

$$e^k = \underbrace{e \cdot e \cdots e}_{k \text{ times}},$$

where e^0 is defined as 1. Meanwhile, we define

$$e^{-k} = \frac{1}{e^k}.$$

Now we define

$$e^{1/k} = \sqrt[k]{e}.$$

(We've already shown that this is well-defined.) Then we can define

$$e^{p/q} = (e^{1/q})^p.$$

We can check that this also equals $(e^p)^{1/q}$, and that $e^{mp/mq} = e^{p/q}$, so this is well-defined. (If this wasn't true, then we couldn't actually define $e^{p/q}$ on rational numbers.)

Now we have defined $e^{p/q}$ for any rational number — we can think of a function $e\colon \mathbb{Q} \to \mathbb{R}$ given by $e(x) = e^x$.

But we would like e to be defined on all of $\mathbb{R}$. For example, what is $e(\sqrt{2})$? Since $\sqrt{2}$ is not rational, we can't define it in the same way. Of course, we could set it equal to anything we want, but that wouldn't be very interesting. But we expect e to have some nice properties — in particular, we would like e to be continuous.

Next time, we'll prove the following:

Theorem 7.10

The function $E\colon \mathbb{R} \to \mathbb{R}$ has the property that $E(x+y) = E(x)E(y)$.

Now we have that $E(1) = e$. This means $E(2) = E(1+1) = E(1) \cdot E(1) = e^2$.

So then E and e agree at 1 and 2, and more generally at any integer — we have

$$E(k) = E(\underbrace{1+1+\cdots}_{k}) = \underbrace{E(1)\cdots E(1)}\, k = e^k.$$

Of course, we also have $E(0) = 1 = e^0$.

We can also observe that $E(x) \cdot E(-x) = E(0) = 1$, so then $E(-x) = 1/E(x)$.

We'd now like to see that they coincide for rationals. We have

$$E(1) = E\left(\frac{1}{k} \cdot k\right) = E\left(\frac{1}{k} + \frac{1}{k} + \cdots + \frac{1}{k}\right) = E\left(\frac{1}{k}\right)^k.$$

We have $E(1) = e$. So then we have

$$E\left(\frac{1}{k}\right) = e^{1/k}.$$

(Note that $E(1/k) > 0$ — we have $E(x) = \sum_{n=0}^{\infty} x^n/n!$, so if $x > 0$ then $E(x) > 0$ as well. Then we can use the fact that positive numbers have a unique (positive) kth root.)

So then E and e coincide at all integers (both positive and negative), and on the reciprocals of integers. By combining this, we have that

$$E\left(\frac{p}{q}\right) = e\left(\frac{p}{q}\right)$$

for all rational p and q.

Next time, we'll see the following:

Theorem 7.11
E is continuous.

This holds generally — any power series is continuous inside its radius of convergence.

But we know that $E(p/q) = e^{p/q}$, and E is continuous and defined for all real numbers. We'd like to define e for all real numbers, but we'd like to define it so that it's continuous. We know that it's equal to E on $\mathbb{Q}$. But now if we take $e\colon \mathbb{Q} \to \mathbb{R}$ and then define $\widetilde{e}\colon \mathbb{R} \to \mathbb{R}$ such that $\widetilde{e}|_{\mathbb{Q}} = e$ (this notation denotes the restriction of $\widetilde{e}$ to e), then we must have $\widetilde{e} = E$ — because we proved that two continuous functions that are the same on $\mathbb{Q}$ are actually the same everywhere.

So this is one way of defining the exponential function on all numbers, not just rationals.

§8 Midterm Review

Logistical announcements about the midterm: the midterm will be in-class, 9:30–11. The topics we should review:

- Sequences:
  - Given a sequence, for example $a_n = (-1)^n + 1/n$, is it convergent? (In this example, the answer is no.)
  - Subsequences and their convergence and divergence — for example, the subsequences $a_{n_k} = a_{2k} = 1 + 1/2k$ and $b_{n_k} = a_{2k+1} = -1 + 1/(2k+1)$ are both convergent ($a_{n_k} \to 1$ and $b_{n_k} \to -1$).
  - Cauchy sequences — sequences such that if we go very far out, then they all bunch together. In $\mathbb{R}$, this is equivalent to converging (accumulating around a limit); we'll see later that this isn't true in general.
  - The Bolzano–Weirstrass theorem — if we have a bounded sequence, then it has a convergent subsequence. Using this we could prove that Cauchy sequences were convergent.

  We should do a bunch of problems — for example, from TBB or Rudin.
- Series: given a sequence a_n, we consider the infinite sum $\sum_{k=1}^{\infty} a_n$. So we're looking at a new sequence $s_n = \sum_{k=1}^{n} a_k$, and we're asking, does s_n converge?

  We may ask not only about convergence, but also about *absolute* convergence — does the series $\sum |a_n|$ converge? As we saw earlier, absolute convergence implies convergence.

  Example 8.1
  Consider the series $\sum_{n=1}^{\infty} a_n$, where $a_n = (-1)^n + 1/n$.

  The first thing we should always check is whether $a_n \to 0$. Here the answer is no, so the series is divergent.

  Example 8.2
  Consider the series $\sum_{n=1}^{\infty} 2^n/n^3$.

  Here we could consider the ratio of consecutive terms — we have

$$\frac{a_{n+1}}{a_n} = \frac{2^{n+1}n^3}{(n+1)^3 2^n} = 2\left(1 - \frac{1}{n+1}\right)^3 \to 2.$$

As $n \to \infty$, this converges to 2, so we see that $a_{n+1}/a_n \to 2$, which means very far out, the next term we add is almost twice the previous term. This means the series diverges — in fact, it fails the basic test that $a_n \to 0$. (Using the ratio test here was quite unnecessary.)

Example 8.3

Consider the series $\sum_{n=1}^{\infty} n^3/2^n$.

Now we can do the same thing — we have

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^3}{2^{n+1}} \cdot \frac{2^n}{n^3} = \frac{1}{2}\left(1 + \frac{1}{n}\right)^3 \to \frac{1}{2}.$$

Now by the *ratio test*, we have that the series converges.

Recall that the ratio test compares our series with a *geometric* series — where the next term is some fraction of the previous one.

We should know the different tests that we have covered (we have not covered all the tests, but of course the ones not covered won't be on the exam):

- The first comparison test — if $0 \le a_n \le b_n$ and $\sum b_n$ is convergent, then so is $\sum a_n$. Sometimes it is useful to take the logical negation — if $\sum a_n$ is divergent, then $\sum b_n$ is divergent as well.
- The second comparison test — if $0 < a_n, b_n$ and $\lim_{n\to\infty} b_n/a_n$ is finite and nonzero, then $\sum a_n$ is convergent if and only if $\sum b_n$ is.
- The ratio test — where we consider $\lim_{n\to\infty} |a_{n+1}| / |a_n|$. If this limit exists and equals some c, then if $c < 1$ the series is convergent, and if $c > 1$ the series is divergent (because it fails the basic test $a_n \to 0$). (If $c = 1$, the test is inconclusive.) We're essentially using the comparison principle to compare it with a geometric series.
- The root test — where we consider $\lim_{n\to\infty} \sqrt[n]{|a_n|}$. If the limit exists and equals some c, then again if $c < 1$ the series is convergent, if $c = 1$ the test is inconclusive, and if $c > 1$ the test is divergent (it again fails the basic test $a_n \to 0$).

Also, if $\sum a_n$ and $\sum b_n$ are both convergent, then $\sum(a_n + b_n)$ is also convergent, and likewise $\sum ca_n$ is convergent. We can use this to take a series and form new series that also converge — similarly to in the case of sequences.

- Power series: given a sequence a_n, we can consider the power seris $\sum a_n x^n$. For each fixed x, this gives a series, which depends not only on a_n but also x. So this defines a function — for each x, if this series converges, then we get a value. This means we can think of a power series as a function whose domain is the set of x where the series converges.

  The most important thing here is the *radius of convergence*, which can be found using the root test — we have
  $$\frac{1}{R} = \lim_{n\to\infty} \sqrt[n]{|a_n|}.$$

- We should also review what inf and sup are. One way we use them is that if we have a sequence that's monotone nondecreasing — $x_n \le x_{n+1}$ — and that's bounded by some L — then we have an increasing sequence that's bounded above. We know such a sequence converges to $\sup\{x_n \mid n \in \mathbb{N}\}$. Likewise, the same is true for inf (and nonincreasing sequences). In some cases, $\inf = \min$ (for instance, for a finite set) and likewise sometimes $\sup = \max$. (We won't be asked about the first couple of lectures, because they're just building foundations — they're not the focus of the class, since we don't really want to focus on the algebraic properties of the line. But it still may make sense to look over to them.)

- Continuous functions — what is a continuous function, and how can we create a continuous function from existing ones? For example, if f and g are continuous and λ is a constant, then $f + \lambda g$ and fg are continuous, and g/f is continuous if $f \neq 0$. This implies in particular that all polynomials are continuous.

  We should be familiar with the definition.

We should look through our notes, and for each of these five topics, we should do some problems from the book (within these topics).

We are allowed to have one of the three books for the midterm (we should not have more than one), but we are not allowed to use a computer. We will probably not need the book — if all these topics are familiar and we are comfortable about the definitions, we will not need it.

The test will have five problems, each of which may have multiple parts. Nothing will require a fantastic idea. We should have more than enough time.

§9 The Exponential Function

We defined the exponential function in two ways. One was by considering the power seires

$$E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

We saw that the radius of convergence is ∞ — the series converges for all x — so this gives us a function defined on $\mathbb{R}$.

We also defined $e = E(1) > 1$, and we looked at e^k for $k \in \mathbb{N}$; by definition

$$e^k = \underbrace{e \cdot e \cdots e}_{k}.$$

We also defined $e^{-k} = 1/e^k$. We also defined $e^{p/q}$ where $p, q \in \mathbb{Z}$ and $q \neq 0$ — one way of defining this is as

$$e^{p/q} = \underbrace{e^{1/q} \cdot e^{1/q} \cdots e^{1/q}}_{p}$$

(where $e^{1/q}$ is the qth root of e).

So now we have two functions. We also have the property that

$$e^{p_1/q_1} \cdot e^{p_2/q_2} = e^{p_1/q_1 + p_2/q_2}.$$

The next thing we want to prove is the following:

Lemma 9.1

We have $E(x+y) = E(x)E(y)$ for $x, y \in \mathbb{R}$.

The reason we want to prove this is because we already have $E(1) = e$. This then implies $E(2) = E(1+1) = E(1) \cdot E(1) = e^2$. Likewise, for every integer k, we have

$$E(k) = E(\underbrace{1 + 1 + \cdots + 1}_{k}) = \underbrace{E(1) \cdots E(1)}_{k} = e^k.$$

We can also use this rule to obtain $E(p/q)$ — we have

$$E(1) = E\left(\frac{1}{p} + \cdots + \frac{1}{p}\right) = E\left(\frac{1}{p}\right)^p,$$

which means that we have $E(1/p)^p = e$, and therefore (since $E(1/p) > 0$) $E(1/p) = e^{1/p}$.

So then just using the fact that E and e agree at 1 and both satisfy the rule $f(x+y) = f(x)f(y)$, we get that E and e coincide on $\mathbb{Q}$.

Meanwhile, we know E is a function $E\colon \mathbb{R} \to (0, \infty)$.

Proposition 9.2

The function $E\colon \mathbb{R} \to (0, \infty)$ is continuous.

Proof. We'll first show that E is continuous at 0. So given some $\varepsilon > 0$, we want to show there exists $\delta > 0$ so that if $|x| < \delta$, then $|E(x) - E(0)| < \varepsilon$.

But we have

$$E(x) - E(0) = \sum_{n=0}^{\infty} \frac{x^n}{n!} - 1 = \sum_{n=1}^{\infty} \frac{x^n}{n!}.$$

But if $|x| < \delta$, then we have

$$|E(x) - E(0)| \le \sum_{n=1}^{\infty} \frac{|x|^n}{n!} < \sum_{n=1}^{\infty} \frac{\delta^n}{n!}.$$

Now it remains to make this expression sufficiently small.

We can write this expression as

$$\sum_{n=1}^{\infty} \frac{\delta^n}{n!} = \delta \cdot \sum_{n=1}^{\infty} \frac{\delta^{n-1}}{n!} \le \delta \sum_{n=1}^{\infty} \frac{1}{n!},$$

as long as $\delta \le 1$. But the series $\sum 1/n!$ is convergent, so it equals some bounded number L; then it suffices to have $\delta L < \varepsilon$, and it's enough to take $\delta < \varepsilon/L$.

Now we have that E is continuous at 0, and we can use this along with the relationship $E(x+y) = E(x)E(y)$ to show that E is continuous everywhere else as well. Given a point x_0, we'll prove that E is continuous at x_0 — so we want to look at $|E(x) - E(x_0)|$, and we want to show that given ε, we can choose δ such that if $|x - x_0| < \delta$ then $|E(x) - E(x_0)| < \varepsilon$.

We can write $x = x_0 + (x - x_0)$, so then

$$E(x) = E(x_0 + (x - x_0)) = E(x_0) \cdot E(x - x_0).$$

This means we have

$$|E(x) - E(x_0)| = |E(x_0)E(x - x_0) - E(x_0)| = |E(x_0)| \cdot |E(x - x_0) - 1| = |E(x_0)| \cdot |E(x_0 - x) - E(0)|.$$

(In fact $E(x_0)$ is always positive, so we can ignore the absolute value sign on it.)

But x_0 is fixed, and if $|x - x_0|$ is sufficiently small, then we can make $E(x_0 - x)$ as close as we want to $E(0)$ (since E is continuous at 0). So by making δ small, we can make $|E(x_0 - x) - E(0)|$ as small as we want; and therefore we can make the entire expression as small as we want as well.

Here the basic relationship $E(x+y) = E(x)E(y)$ let us transfer information about continuity at one point to continuity everywhere. $\square$

> **Remark 9.3.** This is not necessarily the best way of proving that E is continuous — it's a consequence of a much more general fact about power series — but this is a fairly low-tech way to prove it.

Now let's return to the main thing we haven't proven yet — that $E(x+y) = E(x)E(y)$.

Proof. We first make a somewhat unrelated observation:

Theorem 9.4 (Binomial Formula)

We have

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots = \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!}a^{n-k}b^k.$$

Proof. We know that $(a+b)^n$ can be written as a sum of terms $a^{n-k}b^k$, with certain coefficients; we'd like to say what these coefficients are. We have n terms $(a+b)$ that we're multiplying out. In order to get the term $a^{n-1}b$, we have n ways to choose which term we want to get a b from; so the coefficient of $a^{n-1}b$ is n. Similarly, to find the coefficient of $a^{n-2}b^2$, we need to choose two b's from teh n different possibilities; that means the coefficient is

$$\binom{n}{2} = \frac{n!}{(n-2)!\,2!}.$$

Similarly the next coefficient is

$$\binom{n}{3} = \frac{n!}{(n-3)!\,3!},$$

and so on. ■

We're now interested in looking at $E(x)$, $E(y)$, and $E(x+y)$. We'll assume that $x, y \geq 0$.

We define $E_k(x)$ to be E chopped off after the first $k+1$ terms — so

$$E_k(x) = \sum_{n=0}^{k} \frac{x^n}{n!}.$$

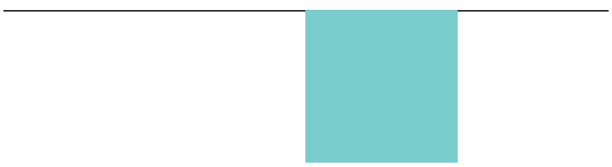
Now we can consider

$$E_k(x)E_k(y) = \sum_{n=0}^{k} \frac{x^n}{n!} \cdot \sum_{m=0}^{k} \frac{y^m}{m!}.$$

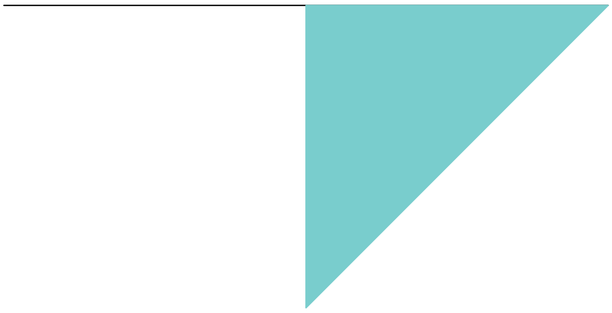
We can now write down a table:

	1	$\frac{x}{1!}$	$\frac{x^2}{2!}$	$\frac{x^3}{3!}$	$\cdots 1$	1	$\frac{x}{1!}$
$\frac{y}{1!}$	$\frac{y}{1!}$	$\frac{xy}{1!1!}$					
$\frac{y}{2!}$							

In $E_k(x)E_k(y)$, our table extends up to $x^k/k!$ and $y^k/k!$. So we're summing the terms in a square:



Now suppose that we instead take a sum of terms where the power is at most $2k$ — so then we'd be looking at a triangle. (So we're looking at terms of the form $x^m y^{2k-m}$.)

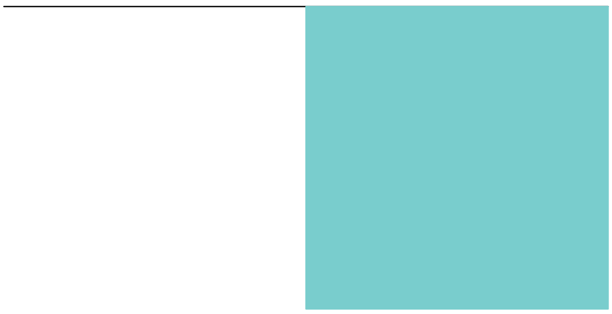


Since all terms are positive, the first sum is smaller than the second (since our triangle has more terms in it).

So then we have

$$E_k(x)E_k(y) \le \sum_{m=0}^{2k} c_{m,2k-m} x^m y^{2k-m}.$$

On the other hand, our triangle is smaller than the *larger* square that contains it:



So then

$$E_k(x)E_k(y) \le \sum_{m=0}^{2k} c_{m,2k-m} x^m y^{2k-m} \le E_{2k}(x)E_{2k}(y).$$

But using the binomial theorem, it's easy to check that the term in the middle is exactly $E_{2k}(x+y)$.

So we now have that

$$E_k(x)E_k(y) \le E_{2k}(x+y) \le E_{2k}(x)E_{2k}(y).$$

Now we have the squeezing principle — we know $E_k(x) \to E(x)$ and $E_k(y) \to E(y)$, so the left-hand side goes to $E(x)E(y)$. Similarly, the right-hand side also converges to $E(x)E(y)$. But the middle term converges to $E(x+y)$, so by the squeeze principle we have $E(x+y) = E(x)E(y)$. $\square$

§10 October 18, 2022

§10.1 The Extreme Value Theorem and Intermediate Value Theorem

Recall that if we have a function $f\colon [a,b] \to \mathbb{R}$, then we say f is continuous at a point x_0 if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$. Then we say f is *continuous* if it is continuous at all points.

For a continuous function, we have two theorems: the *extreme value theorem* (which we've discussed earlier) and the *intermediate value theorem*.

Both of these theorems use the following result:

Theorem 10.1

If $f\colon [a, b] \to \mathbb{R}$ is continuous, and $x_n \in [a, b]$ is a convergent sequence of points such that $x_n \to x_\infty$, then $f(x_n) \to f(x_\infty)$.

First, note that if $x_n \to x_\infty$ and $x_n \in [a, b]$ for all n, then we must have $x_\infty \in [a, b]$ as well, so $f(x_\infty)$ does make sense.

Proof. Since f is continuous at x_∞, given $\varepsilon > 0$ there exists $\delta > 0$ so that if $|x - x_\infty| < \delta$, then

$$|f(x) - f(x_\infty)| < \varepsilon.$$

Meanwhile, since $x_n \to x_\infty$, there exists N such that if $n \geq N$, then $|x_n - x_\infty| < \delta$. Putting these together, we now have that if $n \geq N$, then $|x_n - x_\infty| < \delta$, and by continuity $|f(x_n) - f(x_\infty)| < \varepsilon$. Since this works for any ε, then $f(x_n) \to f(x_\infty)$. □

Now we can use this to prove the extreme value theorem and the intermediate value theorem. We've already used it for the extreme value theorem, but let's do it again.

Theorem 10.2 (Extreme Value Theorem)

Suppose $f\colon [a, b] \to \mathbb{R}$ is continuous. Then there exists x_m and x_M such that

$$f(x_m) = \inf_{[a,b]} f \text{ and } f(x_M) = \sup_{[a,b]} f.$$

The proof of existence for x_m and x_M are almost the exact same, with some obvious modifications; so we will only prove it for x_M.

Proof. The first step is to prove that $f([a, b])$ is bounded from above (this notation denotes the image of f), so that $\sup f$ is a finite number. To prove this, suppose not; then for all $n \in \mathbb{N}$, there exists x_n such that $f(x_n) > n$. (Otherwise n would be an upper bound.) Now using the Bolzano–Weirstrass theorem, we know that our sequence x_n has a convergent subsequence $x_{n_k} \to x_\infty$. But we automatically have $x_\infty \in [a, b]$, so using the above theorem, we have $f(x_{n_k}) \to f(x_\infty)$. But we have $f(x_{n_k}) > n_k \geq k$ for all $k \in \mathbb{N}$, so then our terms $f(x_{n_k})$ wander off to infinity. But they're supposed to converge to something finite; this gives a contradiction.

Now we've shown that $f([a, b])$ is bounded from above. Likewise, we could conclude that $f([a, b])$ is bounded from below. So then $f([a, b])$ is bounded.

Now a very similar argument can be used to show the existence of x_M. Since $f([a, b])$ is bounded, we can let

$$s = \sup\{f(x) \mid x \in [a, b]\},$$

which is a finite number. Then for all $n \in \mathbb{N}$, we can look at $s - 1/n$. This is *not* an upper bound for our set, so then there exists some x_n such that

$$f(x_n) > s - \frac{1}{n}.$$

For each $n \in \mathbb{N}$, we can choose such an x_n; this again gives us a sequence. We again have $x_n \in [a, b]$ for all n, so by the Bolzano–Weirstrass theorem, there exists a subsequence x_{n_k} that converges to some $x_\infty \in [a, b]$. Then $f(x_{n_k}) \to f(x_\infty)$. But we also have

$$s - \frac{1}{k} \leq s - \frac{1}{n_k} < f(x_{n_k}),$$

$f(x_{n_k}) \to f(x_\infty)$, and $f(x_\infty) \leq s$. But since $s - 1/k \to s$ we must have $f(x_\infty) = s$. $\square$

Remark 10.3. On a problem set, one should not write statements such as

$$s - \frac{1}{n_k} < f(x_{n_k}) \to f(x_\infty) \leq s$$

(as written on the board). It's fine to have a line of inequalities going in the same direction; it's bad to have a line of inequalities *not* going in the same direction. This statement is somewhat questionable since we have a symbol that isn't an inequality; it's better to split it up.

Student Question. *In order to prove $f([a, b])$ is bounded, didn't we use the implicit assumption that $f(x_\infty)$ is finite?*

Answer. The function is defined $f\colon [a, b] \to \mathbb{R}$, which means it must take finite values on all of $[a, b]$ — all outputs of f must be real numbers, and ∞ is not a real number.

Student Question. *If all values of f are finite, isn't that already what we're trying to prove?*

Answer. No. For example, consider the function

$$f(x) = \begin{cases} 1/x & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

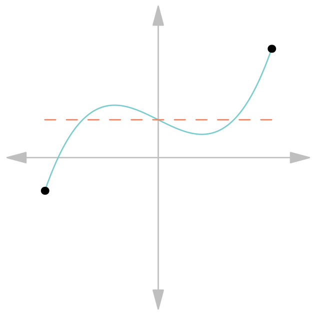
defined on the interval $[-1, 1]$. This function is not continuous, and it is not bounded. But note that this function doesn't have any infinite values — a function is not allowed to have infinite values (then it wouldn't be a function to $\mathbb{R}$).

Next we'll discuss the intermediate value theorem, which is quite important. Most functions we care about are continuous — if you change x a bit, the function won't change a lot — which means the intermediate value theorem applies.

Theorem 10.4 (Intermediate Value Theorem)

Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function. Then f achieves all values between $f(a)$ and $f(b)$ — all values between $f(a)$ and $f(b)$ are the image of some $x \in [a, b]$.

More explicitly, if $f(a) < f(b)$, then the intermediate value theorem tells us that for all $y \in [f(a), f(b)]$, there exists $x \in [a, b]$ such that $f(x) = y$. The same is true if $f(a) > f(b)$, switching the order of endpoints of the interval.



There are various related facts we'll discuss later when we discuss differentiable functions, such as the *mean value theorem*.

For simplicity, there's a minor optimization we can make. Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous, and $f(a) \leq f(b)$. Also suppose $f(a) < y < f(b)$ — it suffices to consider the strict inequalities, since otherwise we can simply use a or b. Then y is fixed, so we can define $\widetilde{f}(x) = f(x) - y$. Since we're simply subtracting a constant from f, our new function $\widetilde{f}$ is also continuous, and we have $\widetilde{f}(a) < 0$ and $\widetilde{f}(b) > 0$. All we want to show is that there exists some $x \in [a,b]$ such that $\widetilde{f}(x) = 0$ — we just want to show that 0 is achieved as the image of some x. This would mean that $f(x) - y = 0$, so $f(x) = y$, as desired.

Student Question. How do we know $\widetilde{f}(a) < 0$ and $\widetilde{f}(b) > 0$?

Answer. This is because we assumed $f(a) < y < f(b)$. So at a, we have $\widetilde{f}(a) = f(a) - y$, and we're sutracting something strictly larger than $f(a)$, giving us a negative value.

This makes the problem slightly simpler (although not by much) — we can now assume our function starts off negative and becomes strictly positive, and we want to show that it hits 0.

Student Question. *What if $f(b) < f(a)$?*

Answer. Then we'd use a symmetric argument — we'd again define $\widetilde{f}(x) = f(x) - y$, and $\widetilde{f}$ would start off negative and end up positive, and we'd still want to show that it ends up being 0 somewhere.

Note that it's important that our function is continuous!

Example 10.5

Consider the function $f\colon [-1,1] \to \mathbb{R}$ defined as

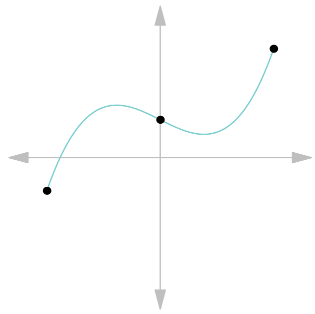
$$f(x) = \begin{cases} x & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}.$$

Then our function starts off strictly negative and ends up strictly positive, but 0 is not achieved. That's because the function isn't continuous — it's not continuous at 0.

Student Question. *What if our continuous function goes down* before *it goes up?*

Answer. This isn't a problem, as we'll see in the proof later.

Proof. We're going to divide our interval into two. At the midpoint, either the value is positive, or it's negative. If the value is negative, then we might as well move our left endpoint to the midpoint, and only consider the function defined on the right half of the interval. Similarly, if the value is positive, then we can move our right endpoint to the midpoint, and only consider the function defined on the left half of the interval.



This gives us a sequence of interval endpoints — set $a_1 = a$, $b_1 = b$. Then we have $a_2 \geq a_1$ and $b_2 \leq b_1$, where exactly one of the two values moves. In particular, we have

$$0 < b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n).$$

If we do this n times, then we accumulate a ton of factors of $1/2$, so then $(b_n - a_n) \to 0$.

But $a_n \leq b$ and $a \leq b_n$. This means a_n is an increasing sequence bounded above, so it converges to some value a_∞. Similarly, the b_n are a decreasing sequence bounded below, so they converge to some b_∞. But since $(b_n - a_n) \to 0$, we must have $a_\infty = b_\infty$.

But we have $f(a_n) < 0$ and $f(b_n) > 0$. Now since our function is continuous, and we have $f(a_n) \to f(a_\infty)$ and $f(b_n) \to f(b_\infty) = f(a_\infty)$ (since a_∞ and b_∞ are the same number). Our numbers $f(a_n)$ are all negative and our numbers $f(b_n)$ are all positive. So then their limit cannot be strictly negative or strictly positive; and therefore it must be 0. □

Student Question. What are the definitions of the sequences a_n and b_n?

Answer. We start with a continuous function f defined on $[a, b]$, and we assume $f(a) < 0$ and $f(b) > 0$. We first set $a_1 = a$ and $b_1 = b$. Then we take the midway point $(b + a)/2$. If f at the midway point is 0, then we're done. If f at the midway point is strictly positive, then we set b_2 to be the midway point and $a_2 = a_1$. Meanwhile, if it's strictly negative, then b_2 doesn't change — we keep $b_2 = b_1$ — and we instead set a_2 to be the midway point. We then keep doing this procedure.

Remark 10.6. We'll use the same reduction when proving the mean value theorem — there it's called Rolle's theorem. It's not really that important here — it just makes the proof a bit simpler.

In this case, we defined two sequences, one increasing and the other decreasing. Both are bounded, and we showed inductively that the distance between the two sequences halves at each step, which means they must converge to the same value (they're each convergent, since they're monotone and bounded). One sequence has the property that f is strictly negative, and the other that f is strictly positive; therefore the value of f at the limit cannot be either strictly negative or positive, so it must be zero.

§10.2 Midterm Information

The midterm will have five problems. Some of these problems will have parts, but they won't be enormous, and they'll be logically connected (one question will have 3 parts, and the rest will have 1 or 2). The test should be doable — Prof. Colding doesn't want us to stress too much over it. Nothing will require any great ideas.

It will be useful to go over sequences — convergence of sequences and subsequences, Cauchy sequences, what it means to be Cauchy-complete. It'll also be useful to go over series — what is a series, when does a series converge, what does it mean to be absolutely convergent, what are the basic tests (in particular, the basic test that $a_n \to 0$, the comparison test, the ratio test, and the root test). (Prof. Colding usually prefers the ratio test, since we don't have to take the nth root of something, but they're very similar.) We also want to know what continuity is; the theorem proved today, that if $x_n \to x_\infty$ and f is continuous then $f(x_n) \to f(x_\infty)$, is often useful.

We'll now discuss metric spaces (which will not be on the exam).

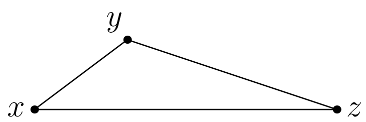
§10.3 Metric Spaces

Metric spaces are a crucial generalization of what we've talked about so far. They don't carry through *all* the properties we've seen, but they can often be useful.

Definition 10.7. A **metric space** is a set X along with a map d on $X \times X \to [0, \infty)$, a way of measuring distances between a pair of points (distances are always nonnegative). The map should satisfy the properties that:

- $d(x, y) = d(y, x)$ — the distance from x to y is the same as the distance from y to x.
- We have $d(x, y) = 0$ if and only if $x = y$.
- The triangle inequality — we have $d(x, y) + d(y, z) \geq d(x, z)$.

To visualize the triangle inequality, it essentially states that if we go from x to y and then y to z, then the distance travelled is *at least* as much as if we went directly from x to z.



Note that the distance between two points is not allowed to be infinite.

Example 10.8

Take the set $\mathbb{R}$, with distance defined as $d(x, y) = |x - y|$.

It's clear that this is symmetric and nonnegative, and that $d(x, y) = 0$ if and only if $x = y$. The triangle inequality for this space is one we've used many times before.

There's two generalizations of this.

Example 10.9 (Euclidean Distance)

Take the set $\mathbb{R}^2$, such that for two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$, we define

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

It's clear that d is nonnegative and symmetric. It's also clear that $d(x, y) = 0$ if and only if our points are the same. The triangle inequality is a bit more difficult to see, but it can be proven as well.

There's another generalization, although it's somewhat less common:

Example 10.10 (Box Distance)

Again take $\mathbb{R}^2$, and define

$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

This is again clearly symmetric, nonnegative, 0 if and only if the two pairs are the same, and this time it's easy to see that the triangle inequality holds.

Both of these examples can be generalized to $\mathbb{R}^n$ in the obvious way.

Example 10.11

Take the set of continuous functions on the interval $[a, b]$, so

$$C([a,b]) = \{f \mid f\colon [a,b] \to \mathbb{R}\},$$

with distance defined as

$$d(f,g) = \sup\{|f(x) - g(x)| \mid x \in [a,b]\}.$$

It's again clear that this is nonnegative and symmetric. If $d(f, g) = 0$ then f and g agree at all points, so are equal. Meanwhile, it's also not hard to show that it satisfies the triangle inequality.

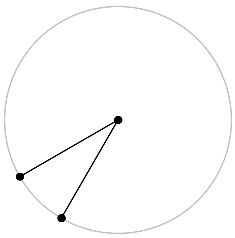
Implicit in this example is the extreme value theorem — we needed to use the fact that f and g are bounded, since otherwise $d(f, g)$ may not be defined.

Example 10.12

Take the set of points on a sphere, with distance defined as the length of the shortest path between the two points that stays on the sphere.

Example 10.13 (French Railway Metric)

Take a set of points with a 'central' point, and define the distance between two points as the distance going through the central point.



Remark 10.14. For where the name comes from, France is a country very centralized around Paris — some cities are fairly close to each other, but in order to get from one to the other via railroad, we'd have to go through Paris.

§10.3.1 Convergence in a Metric Space

Definition 10.15. Suppose (X, d) is a metric space, and $x_n \in X$ a sequence. Then we say x_n converges to $x_\infty \in X$ if for all $\varepsilon > 0$, there exists N so that if $n \geq N$ then

$$d(x_n, x_\infty) < \varepsilon.$$

Equivalently, $x_n \to x_\infty$ in X if the sequence of real numbers $d(x_n, x_\infty)$ converges to 0.

Definition 10.16. A **Cauchy sequence** in a metric space (X, d) is a sequence x_n such that for all $\varepsilon > 0$, there exists N such that if $n, m \geq N$ then

$$d(x_n, x_m) < \varepsilon.$$

Of course, in a general metric space, there's no addition or multiplication — so we can't add things together. We can talk about convergence and sequences, but we can't talk about sup — because there's no ordering. In $\mathbb{R}$, one important property we proved was the Bolzano–Weirstrass theorem, and its consequence that every Cauchy sequence is convergent. But in a general metric space, this is *not* the case — we'll get into this further next time.

Notation 10.17. As in the case of $\mathbb{R}$, if $x_n \to x_\infty$ then we write $x_\infty = \lim_{n\to\infty} x_n$.

§11 October 25, 2022

§11.1 Metric Spaces

Last class, we defined a *metric space* as a set X along with a distance d, a map $d\colon X \times X \to \mathbb{R}_{\geq 0}$ (where for x and y in X, we say $d(x, y)$ is the *distance* between x and y), with a number of properties:

- $d(x, y) = 0$ if and only if $x = y$;
- The distance is symmetric — $d(x, y) = d(y, x)$.
- The triangle inequality — $d(x, z) \leq d(x, y) + d(y, z)$.

Example 11.1

One metric space is $(\mathbb{R}, d)$, with the distance defined as $d(x, y) = |x - y|$.

This distance function is clearly nonnegative, zero if and only if $x = y$, and it's clearly symmetric; and we've already seen the triangle inequality used many times.

Example 11.2 (Euclidean Distance)

Another metric space is $(\mathbb{R}^2, d)$, where for two points $x, y \in \mathbb{R}^2$, if $x = (x_1, x_2)$ and $y = (y_1, y_2)$, then we define

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

This construction works more generally for $\mathbb{R}^n$.

The first two properties — that $d(x, y) = 0$ if and only if $x = y$, and that d is symmetric — are very easy to prove, while the triangle inequality is more challenging.

Example 11.3 (Box Metric)

Another metric space is $(\mathbb{R}^2, d)$, where for two points x and y, we define

$$d(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

This again works for $\mathbb{R}^n$. In this case, all three properties are quite easy to see.

Example 11.4

The set $\mathbb{N}$ can be made into a metric space by taking $d(n, m) = |n - m|$. This is not very interesting; a more interesting example is taking $\mathbb{N}$ with

$$d(n, m) = \left|\frac{1}{n} - \frac{1}{m}\right|.$$

§11.2 Sequences in Metric Spaces

Now we'll see how some of the concepts we've discussed earlier generalize to metric spaces. Of course, in a general metric space we don't have addition or multiplication. But we can still discuss sequences, in the same way as over $\mathbb{R}$:

Definition 11.5. A **sequence** in a metric space is a map $\mathbb{N} \to X$.

Definition 11.6. For a monotone (strictly) increasing map $f\colon \mathbb{N} \to \mathbb{N}$ (so if $k_1 < k_2$ then $f(k_1) < f(k_2)$), the sequence $x_{f(k)}$ is a **subsequence** of x_n.

This is the same definition as from earlier — we're only taking some of the terms from our sequence, but we have to take them in the same order.

§11.2.1 Cauchy Sequences

We can first generalize convergence to metric spaces:

Definition 11.7. For a metric space (X, d) and a sequence x_n, we say that x_n **converges** to x (written $x_n \to x$) if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, we have $d(x_n, x) < \varepsilon$.

Similarly, the definition of Cauchy sequences carries over as well:

Definition 11.8. A sequence x_n is a **Cauchy sequence** if for all $\varepsilon > 0$, there exists N such that for all $m, n \geq N$, we have $d(x_n, x_m) < \varepsilon$.

Again, the idea of convergence is that x_n converges to x if when we go out sufficiently far out in the sequence, the x_n bunch close together around the limit. Meanwhile, x_n is a Cauchy sequence if when we go sufficiently far out, the terms all bunch together. From this, it seems obvious that if we have a sequence that's convergent, then it must be Cauchy. We'll now prove this rigorously:

Proposition 11.9

Suppose that (X, d) is a metric space, and x_n is a sequence in X that converges to x. Then x_n is a Cauchy sequence.

Proof. Given $\varepsilon > 0$, we know that there exists N such that for all $n \geq N$, we have $d(x_n, x) < \varepsilon/2$ — this is by the fact that $x_n \to x$. Now suppose $m, n \geq N$. Then we know $d(x_n, x) < \varepsilon/2$ and $d(x_m, x) < \varepsilon/2$, which means

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$ □

So if we have a sequence in a metric space, then convergence always implies that it's a Cauchy sequence.

On $\mathbb{R}$, one basic fact we proved was the Bolzano–Weirstrass theorem. One important consequence of it was that if we have a sequence $x_n \in \mathbb{R}$ (with the standard metric), then if x_n is a Cauchy sequence, this implies x_n is convergent. The reason is that the Bolzano–Weirstrass theorem tells us that every bounded sequence has a convergent subsequence. If we now take a Cauchy sequence, it must be bounded; so then it has a convergent subsequence. But a Cauchy sequence is convergent if and only if it has a convergent sequence.

But this does *not* hold in an arbitrary metric space.

Claim — For a general metric space, there may not be a version of the Bolzano–Weirstrass theorem.

The Bolzano–Weirstrass theorem can be generalized, but only when we have a special property of the metric space. In particular, not every Cauchy sequence is convergent.

Example 11.10

Take $\mathbb{N}$ with the distance metric defined as

$$d(n, m) = \left|\frac{1}{n} - \frac{1}{m}\right|.$$

The sequence $x_n = n$ is a Cauchy sequence, but it does not converge.

Of course this sequence isn't Cauchy in $\mathbb{R}$ with the standard metric, but it *is* Cauchy in $\mathbb{N}$ with our new metric.

Remark 11.11. Note that as always, $\mathbb{N}$ does not include 0.

Proof. First, to show that x_n is Cauchy, given $\varepsilon > 0$, let N be sufficiently large so that $1/N < \varepsilon/2$ (we can find such N by the Archimedean property). Then if $n, m \geq N$, we have

$$d(x_n, x_m) = \left|\frac{1}{n} - \frac{1}{m}\right| \leq \frac{1}{n} + \frac{1}{m} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

This proves that x_n is Cauchy in this slightly odd metric space.

On the other hand, this sequence cannot converge — suppose that $x_n \to m$ for some $m \in \mathbb{N}$. Then we have

$$d(x_n, m) = d(n, m) = \left|\frac{1}{n} - \frac{1}{m}\right|.$$

This must be sufficiently small when n is sufficiently large. But we have

$$\left|\frac{1}{n} - \frac{1}{m}\right| \geq \frac{1}{m} - \frac{1}{n}$$

by the triangle inequality (since $|1/n - 1/m| + 1/n \geq 1/m$).

But if n is really large, then $1/n$ becomes really small. So then $d(x_n, m)$ is bounded away from 0 for large n (it ends up close to $1/m$), which means our distances don't go to 0, and the sequence *cannot* converge to m. $\square$

So it's possible to have a metric space with a Cauchy sequence that isn't convergent.

Another, somewhat cheap, example is the following:

Example 11.12

Consider the set $(0, 1) \subseteq \mathbb{R}$, with the usual metric $d(x, y) = |x - y|$. This is a valid metric, since it's valid on the entire line, so restricting it to $(0, 1)$ still produces a valid metric.

Now if we take any sequence $x_n \in (0, 1)$ converging to 0 — for example $x_n = 1/n$ — then x_n is a Cauchy sequence, but it can't be convergent. If it were convergent, then it would also converge (to the same value) on $\mathbb{R}$ with the usual metric; but a sequence can only converge to one value, and 0 is not in our metric space.

This example seems a bit cheap because we simply removed the limit; but both examples are important.

Student Question. *In this space $(0, 1)$, if we add in 0 and 1, then would every Cauchy sequence converge?*

Answer. Yes. We couldn't just add in one of them — if we added in 0 but not 1, then we could take a Cauchy sequence converging to 1 in $\mathbb{R}$, and it would still not converge.

But if we took the interval $[0, 1]$, now this *would* have the Bolzano–Weirstrass theorem. This is important — it's called a **compact** metric space (we'll get to compactness later), and for compact metric spaces, the Bolzano–Weirstrass theorem does hold.

Student Question. *Could you do a similar construction in the first example?*

Answer. Yes — this is called *compactification.* In some spaces, you can add a point to make it compact (although this isn't true in *all*).

So there are metric spaces where we don't have the Bolzano–Weirstrass theorem. Of course, this doesn't mean that *every* Cauchy sequence is not convergent; but it does mean there can exist Cauchy sequences that aren't convergent.

Another important example of a metric space we looked at was the space of *continuous* functions:

Example 11.13

Consider the metric space $\mathcal{C}([0, 1])$ of continuous functions $f: [0, 1] \to \mathbb{R}$, where

$$d(f, g) = \sup \{|f(x) - g(x)| \mid x \in [0, 1]\}.$$

This metric space *does* have the Bolzano–Weirstrass property, that every Cauchy sequence must be convergent — we'll prove this later.

Definition 11.14. We say that a metric space is **Cauchy complete** if every Cauchy sequence is convergent.

§11.3 Operations on Sets

We'll now discuss basic operations on sets. The reason to discuss this is in many metric spaces, the most important property is what sets are open and closed; and in order to define open and closed sets, we need to discuss set operations.

Here X will be a general set; we will not assume we have any metric on it.

Definition 11.15. For a set X, a **subset** of X, denoted $A \subseteq X$, is a set of some of the elements of X.

Any set X has two trivial subsets — the empty subset $\emptyset$ (that has no elements) and the full set X. (Of course, if X is empty then these two subsets are the same.)

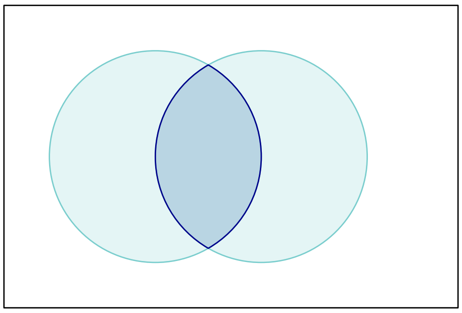
Suppose we have a subset $A \subseteq X$, and $B \subseteq A$. Then B only contains elements of A, so B only contains elements of X.

Similarly, if $A \subseteq B$ and $B \subseteq A$, then all elements of A are elements of B, and all elements of B are elements of A; this implies that $A = B$.

There's three basic operations on sets — taking *intersections*, *unions*, and *complements*.

Definition 11.16. Given a set X and two subsets $A, B \subseteq X$, their **intersection** $A \cap B$ is the set

$$A \cap B = \{x \in X \mid x \in A \text{ and } x \in B\}.$$

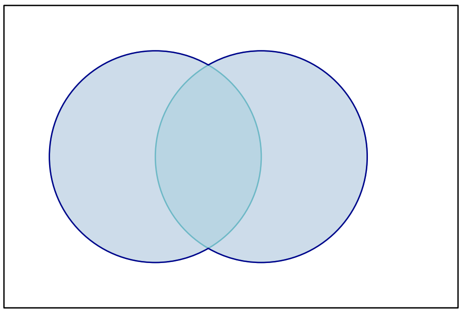


Definition 11.17. Given a set X and two subsets $A, B \subseteq X$, their **union** is the set

$$A \cup B = \{x \in X \mid x \in A \text{ or } x \in B\}.$$

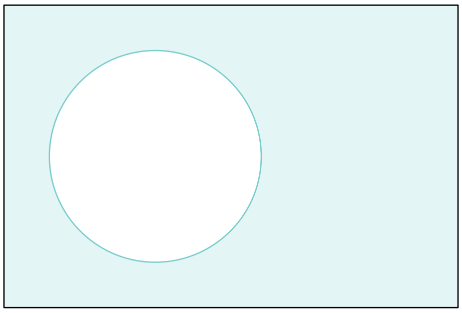
Note that elements contained in *both* A and B are also in their union (the 'or' is inclusive). In particular,

$$A \cap B \subseteq A \cup B.$$



Definition 11.18. Given a set X and a subset $A \subseteq X$, its **complement** of A, denoted as A^c or as $X \setminus A$, is the set

$$X \setminus A = \{x \in X \mid x \notin A\}.$$



We can look at the intersection of not just *two* sets, but a *bunch* of different sets. Suppose we have a collection of subsets $A_\alpha \subseteq X$ (there may be finitely many or infinitely many such sets). Then

$$\bigcap_\alpha A_\alpha = \{x \in X \mid x \in A_\alpha \text{ for all } \alpha\}.$$

Similarly, we define the union of a general collection of sets as

$$\bigcup_\alpha A_\alpha = \{x \in X \mid x \in A_\alpha \text{ for some } \alpha\}.$$

Using these three operations, we can combine them in a bunch of different ways. For example, $X \setminus (A \cup B)$ is the set of X outside of $A \cup B$, meaning that $x \notin A$ and $x \notin B$. But this means $x \in (X \setminus A) \cap (X \setminus B)$. The other direction is true as well, so then we have

$$X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B).$$

Similarly, suppose we want to describe $X \setminus (A \cap B)$ — so we're taking X and excluding the bit in both A and B. Then we can take $X \setminus A$ and $X \setminus B$; elements of the complement of the intersection must lie in *either* one of these, so we have

$$X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B).$$

Example 11.19

Consider the set $X = \mathbb{R}$, and consider the subsets $A_n = (-1/n, 1/n)$. Then we have $\bigcap_{n=1}^\infty A_n = \{0\}$, since for any nonzero number, there's some A_n not containing it.

Meanwhile, we have $\bigcap_{n=1}^\infty A_n = A_1$, since the A_n are a *nested* sequence — we have $A_1 \supseteq A_2 \supseteq \cdots$.

Example 11.20

We can also consider the sets $B_n = (-n, n)$. Now the first set is the smallest, so $\bigcap_n B_n = (-1, 1) = B_1$. Meanwhile, we have $\bigcup_n B_n = \mathbb{R}$.

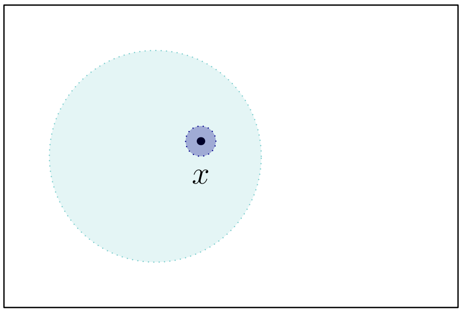
§11.3.1 Open and Closed Sets

One of the most important things about metric spaces is the notion of *open sets*.

Definition 11.21. If we have a metric space (X, d), a **metric ball** centered at x with radius $r > 0$, denoted $B_r(x)$, is the subset of X defined as

$$B_r(x) = \{y \in X \mid d(y, x) < r\}.$$

Definition 11.22. For a metric space (X, d), a subset $A \subseteq X$ is **open** if the following holds: for all $x \in A$, there exists $r > 0$ (which may depend on x) such that $B_r(x) \subseteq A$.



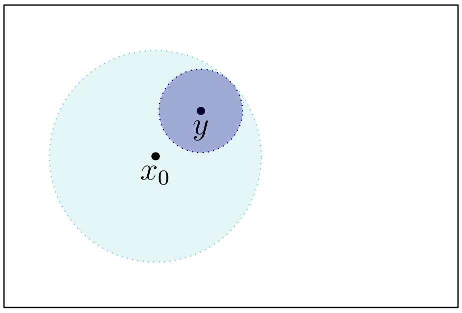
Note that to describe whether a set is open, it matters not just what the metric is, but what our set is.

Lemma 11.23

If (X, d) is a metric space and $B_R(x_0) \subseteq X$, then $B_R(x_0)$ is open.

In other words, every metric ball in a metric space is open.

Proof. We'll use the *triangle inequality.* Suppose that y is in our ball $B_R(x_0)$. Then we have $d(y, x_0) < R$.



Now the main idea is that if we look at the distance from x to y and from y to the edge of the ball, there's something 'left over.' So we can define

$$r = R - d(x, y) > 0,$$

and then consider $B_r(y)$, the set of points z with $d(y, z) < r$. Then we want to show that $z \in B_r(x_0)$ — i.e. that $d(x_0, z) < r$.

To do this, we can use the triangle inequality — we have

$$d(x_0, z) \leq d(x, y) + d(y, z) < d(x, y) + r = d(x, y) + R - d(x, y) = R.$$

So then the distance from z to the center of our giant ball is strictly less than the radius of the giant ball; this means our smaller ball lies in the larger ball. □

Definition 11.24. A **closed** set is a set whose complement is open — in a metric space (X, d), a subset $A \subseteq X$ is closed if $X \setminus A$ is open.

It may seem silly that it's necessary to have a definition for this; but sometimes we'll want to work with open sets and sometimes we'll want to work with closed sets, and it would be a bit cumbersome to not have a term for it.

Remark 11.25. Note that trivially, the empty set is open — a set is open if a condition holds for all elements in the set, but the empty set has no elements, so it's always open. Likewise, the entire set X is open as well — any ball can only contain elements in X.

But because $\emptyset$ is open, its complement X is closed. Similarly, since X is open, its complement $\emptyset$ is closed.

So X and $\emptyset$ are both open and closed. In fact, there are certain sets where X and $\emptyset$ are the only sets that are both open and closed; this is an important property, and can be used to prove that various differential equations have solutions.

Question 11.26. How do open and closed sets behave with our set operations?

Lemma 11.27

Suppose we have a metric space X, and two subsets $A, B \subseteq X$. Then $A \cap B$ and $A \cup B$ are also open.

Proof. The second statement is trivial — suppose $x \in A \cup B$. Then we must have $x \in A$ or $x \in B$; assume for simplicity that $x \in A$. Then since A is open, there exists $r > 0$ such that $B_r(x) \subseteq A$. But since $B_r(x) \subseteq A$, of course we also have $B_r(x) \subseteq A \cup B$. So the union of two open sets is open.

Now let's prove the property for intersections — this is not much harder, but it's not completely trivial. Suppose we take an element $x \in A \cap B$; this means $x \in A$ and $x \in B$. Since $x \in A$ and A is open, there exists some $r_A > 0$ such that $B_{r_A}(x) \subseteq A$. Similarly, since $x \in B$ and B is open, there exists some $r_B > 0$ such that $B_{R_B}(x) \subseteq B$. Now set $r = \min\{r_A, r_B\}$, and consider the ball $B_r(x)$. We then have

$$B_r(x) \subseteq B_{r_A}(x) \subseteq A,$$

since the two balls have the same center and $r_A \geq r$, so $B_{r_A}(x)$ can only be bigger than $B_r(x)$. Likewise, we have $B_r(x) \subseteq B_{r_B}(x) \subseteq B$. So then $B_r(x)$ is contained in both A and B, and therefore in $A \cap B$. $\square$

All we've proven here is that if we take two open sets, then their intersection and union are still open. If we look at *infinitely* many sets, then we have to be more careful.

§12 October 27, 2022

§12.1 Open and Closed Sets

Last class, we defined a *metric ball*:

Definition 12.1. For a metric space (X, d), for $x \in X$ and $r > 0$, the metric **ball** $B_r(x)$ is the set

$$B_r(x) := \{y \in X \mid d(y, x) < r\}.$$

Definition 12.2. A set $A \subseteq X$ is **bounded** if $A \subseteq B_R(x)$ for some $x \in X$ and some $R > 0$.

So a set is bounded if it lies inside some ball.

Definition 12.3. A subset $A \subseteq X$ is **open** if for all $x \in A$, there exists some $r > 0$ (possibly depending on x) such that $B_r(x) \subseteq A$.

Definition 12.4. We say that a subset $B \subseteq X$ is **closed** if its complement $X \setminus B$ is open.

There's another way to define closed subsets, and today we'll prove the equivalent definition; but this definition is nice because it can be generalized even to some spaces that *aren't* metric spaces (in some spaces it's still important to have a concept of what it means to be open and closed).

Last class, from the triangle inequality we saw that any ball $B_r(x) \subseteq X$ is an open subset. Note that it's important that our balls use a *strict* inequality — if they didn't, then this would not be the case.

We also observed that if A and B are open, then $A \cap B$ and $A \cup B$ are also open. In fact, more generally, the following is true:

Proposition 12.5

If A_α is a family of open subsets, then $\bigcup_\alpha A_\alpha$ is also open.

Proof. If we take any $x \in \bigcup_\alpha A_\alpha$, then there is some α for which $x \in A_\alpha$. But then there exists some ball $B_r(x) \subseteq A_\alpha$, and we must have $B_r(x) \subseteq \bigcup_\alpha A_\alpha$ as well. □

So *any* union of open sets is open — this is almost trivial.

Last class, we also proved that if A and B are open, then $A \cap B$ is also open — we proved this by taking the *minimum* of the radii (there's some ball with radius r_1 centered at x contained in A and another ball with radius r_2 centered at x contained in B, so the ball with radius $\min(r_1, r_2)$ is contained in both).

On the other hand, this *doesn't* generalize to infinite intersections: for open subsets A_α, it is not necessarily true that $\bigcap_\alpha A_\alpha$ is open.

Example 12.6

Take the metric space $\mathbb{R}$ with the usual metric, and let $A_n = (-1/n, 1/n)$. Then $\bigcap_n A_n = \{0\}$. But a ball around 0 is an interval $(-r, r)$, and no such ball is contained in $\{0\}$. So the intersection of our open subsets is not open.

So we can take any union, but we cannot take any intersection — but as long as we have *finitely* many subsets, the intersection is also open.

Last class, we also saw three basic operations — the intersection $A \cap B$, union $A \cup B$, and complement $X \setminus A$. Using these operations, we can translate statements about open subsets into statements about closed subsets.

Proposition 12.7

If A and B are closed, then $A \cap B$ is also closed.

Proof. Proving that $A \cap B$ is closed is the same as proving that $X \setminus (A \cap B)$ is open. But we know

$$X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B).$$

Since A is closed, $X \setminus A$ is open; since B is closed, $X \setminus B$ is open. So the complement of the intersection of two closed subsets is the union of two open subsets, and is therefore open. □

Remark 12.8. This may seem like logical operations without much content, but we'll later see that there's reasons to do this.

Proposition 12.9

If A and B are closed, then $A \cup B$ is also closed.

Proof. We need to prove that $X \setminus (A \cup B)$ is open. But we have

$$X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B).$$

If A is closed then $X \setminus A$ is open, and if B is closed then $X \setminus B$ is open. So we're looking at the intersection of two open subsets, which is also open. $\square$

The same argument can be used for the intersection of a *family* of closed subsets:

Proposition 12.10

If $A_\alpha \subseteq X$ is a (possibly infinite) family of closed subsets, then $\bigcap A_\alpha$ is also closed.

Proof. The same argument holds — the complement of the intersection is a union of open subsets, and any union of open subsets is open. $\square$

Again, it is not necessarily true that $\bigcup A_\alpha$ is closed:

Example 12.11

Consider $\mathbb{R}$ with the usual metric, and take the sets

$$A_n = \left[-1 + \frac{1}{n}, 1 - \frac{1}{n}\right].$$

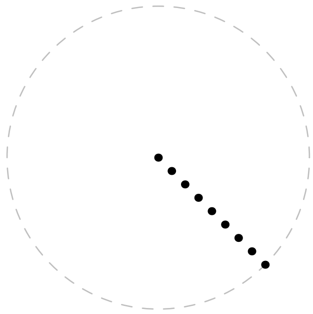
Then $\bigcup A_n = (-1, 1)$. The complement is definitely not open, because the complement contains 1 and -1, but there's no ball around 1 contained in the complement.

So then $\bigcup A_n$ is not closed.

§12.2 Limit Points

Definition 12.12. Suppose we have a metric space (X, d) and $A \subseteq X$. Suppose that $x = \lim_{n\to\infty} x_n$, where $x_n \in A$ for all n.

Note that x doesn't have to be in A — it just has to be the limit of a *sequence* lying in A.



Example 12.13

Take the metric space $\mathbb{R}$ with the usual metric, and let $A = (-1, 1)$. Then $1 \notin A$, but we can take $x_n = 1 - 1/n$; then $x_n \in A$, and $x_n \to 1$.

There is an easy observation we can make:

Proposition 12.14

Given a set A, all $x \in A$ are limit points of A.

Proof. We can take the constant sequence $x_n = x$; this lies in A, and it clearly converges to x. $\square$

Theorem 12.15

For a metric space (X, d) and a subset $A \subseteq X$, A is closed if and only if A contains all its limit points.

We know the set of limit points of A is always a (weakly) larger set than A (by the above observation); so we want to show that if A is *equal* to its set of limit points then it's closed, and vice versa.

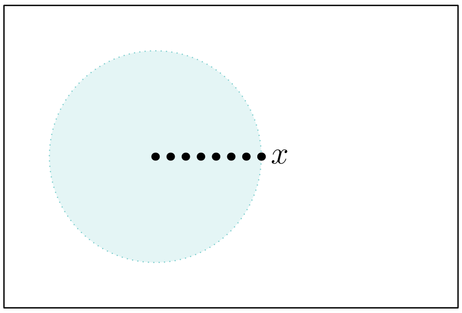
We'll prove this in two steps.

Lemma 12.16

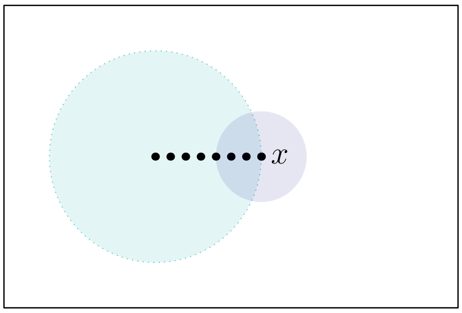
If A is closed, then A contains all its limit points.

Proof. Suppose $x \in X$ is a limit point of A. This means there is a sequence $x_n \in A$ such that $x_n \to x$; we want to show that $x \in A$.

Assume for contradiction that $x \notin A$.



If $x \notin A$, then of course $x \in X \setminus A$. But because A is closed, $X \setminus A$ is open. This means there exists $r > 0$ such that $B_r(x) \subseteq X \setminus A$.



But then this ball can't intersect A. But we claimed $x_n \to x$, and $x_n \in A$. This means that whatever r is, for n sufficiently large x_n would *have* to lie in this ball (since x_n is supposed to converge to x). But these x_n lie in A, so this is a contradiction. $\square$

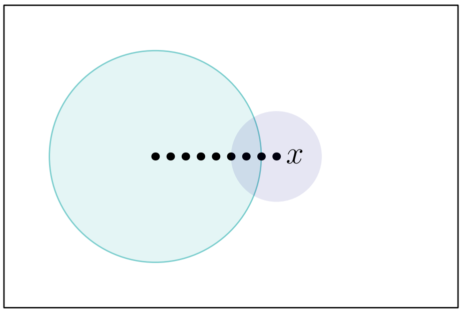
Now we want to prove the other direction.

Lemma 12.17

Suppose (X, d) is a metric space and $A \subseteq X$, and suppose that A contains all its limit points. Then A is closed.

Proof. We want to show that the complement of A is open. So suppose that x is in the complement $X \setminus A$.

Now take $n \in \mathbb{N}$ and consider the ball centered at x of radius $1/n$.



If for some n, this ball $B_{1/n}(x)$ lies in the complement $X \setminus A$, then we'd be done — we're trying to show that the complement is open, and that means we want to show that at every point there's some ball centered at that point that lies in the complement.

So we can assume for contradiction that for all n, this ball $B_{1/n}(x)$ does not lie entirely in the complement. This means the ball must intersect A — so $B_{1/n}(x) \cap A$ is nonempty.

But then for each n we can find some $x_n \in B_{1/n}(x) \cap A$. This provides a sequence; but this sequence lies in a sequence of balls with radius going to 0 centered at x, so we must have $x_n \to x$.

But A was supposed to contain all its limit points. And x_n is a sequence in A, so we must have $x \in A$. This provides the desired contradiction (since we started with some $x \notin A$). $\square$

Remark 12.18. Metric spaces are nice because it's often easy to get intuition for what's going on by drawing a picture.

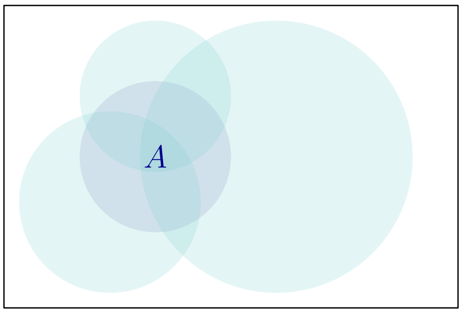
Often when we're trying to prove that a set is closed, we don't go through the route of proving its complement is open — instead we can use this equivalent property. The reason we didn't take this property as the *definition* is that it can't be generalized (while our definition can).

§12.3 Covers and Compactness

Definition 12.19. Given a metric space (X, d) and a subset $A \subseteq X$, a **cover** of A is a (not necessarily finite) collection of sets $O_\alpha \subseteq X$ such that

$$\bigcup_\alpha O_\alpha \supseteq A.$$

Note that the sets O_α may contain more elements in A; but they have to together cover all elements in A.



We often deal with covers with special properties — in particular, *open* covers.

Definition 12.20. An **open cover** is a cover consisting of open subsets.

Definition 12.21. For a metric space (X, d) and a subset $A \subseteq X$, we say that A is **compact** if for all open covers of A, there exists a *finite* subcover.

This definition may seem almost impossible to check, but there's equivalent formulations that are often quite easy to check.

Definition 12.22. Given a cover of A, a **subcover** is a collection of *some* of our sets O_α that still covers A.

So a subcover must consist only of sets in the original cover, but typically much fewer. Compactness tells us that if we have *any* cover, then we really could have used only finitely many of the sets to cover A.

This can be hard to check, but it can be quite easy to check that something is *not* compact.

Example 12.23

Let (X, d) be $\mathbb{R}$ with the usual metric, and take the set $A = (0, 1)$. Then we can take the cover $O_n = (1/n, 1)$. These sets O_n are all open, and $\bigcup_n O_n = A$. But if we only took finitely many of these O_n, then there would be some largest n we chose — so we can suppose that all the sets we took are among $O_1, \ldots, O_N$. Then we have

$$O_1 \cup O_2 \cup \cdots \cup O_N = \left(\frac{1}{N}, 1\right).$$

So there would still be small numbers that aren't covered — this means we don't have a finite subcover.

Student Question. *Could there also be another cover, though?*

Answer. Yes — if we take any metric space and any set A, then A itself is a cover; so if A is open, then A is an open cover, and of course it's a finite cover. In general, even if our set isn't compact, there may be many *finite* covers. But this doesn't mean that the set is compact — if just a single open cover doesn't have a finite subcover, then the set isn't compact.

§12.4 Bolzano–Weirstrass Theorem

One of the fundamental things about the real numbers was the property about sup — there's some sort of completeness about the real numbers. We know sup and inf were part of the ordering, but they were really convenient when we proved the Bolzano–Weirstrass theorem — that if we take any bounded set and any sequence in that bounded set, then there must be a convergent subsequence. For that, we used the sup and inf property.

The sup and inf property comes from the ordering — so that's structure that we don't have in a general metric space. But it was very convenient to have them. So we can ask how we can generalize this:

Question 12.24. What properties of a metric space guarantee that all sequences in bounded sets have a convergent subsequence?

The metric space $\mathbb{R}$ had this property, but we used its ordering to prove it — but we can wonder whether this property really has anything to do with the ordering, or whether there's some general version of it. And there is — we'll see this soon.

Lemma 12.25

Suppose (X, d) is a metric space, and $A \subseteq X$ is compact. Then A is bounded.

Proof. Fix any $x_0 \in X$ (not necessarily in A), and for each $n \in \mathbb{N}$, consider the ball $B_n(x_0)$. If for some n, A is entirely contained in this ball, then it would be bounded; so that's what we want to prove.

So suppose for contradiction that A is *not* bounded; this means that for all n, $B_n(x_0)$ does not contain A; equivalently, A intersects the complement of this ball.

But now we can consider the union of all these balls $\bigcup_n B_n(x_0)$. If we take any $x \in A$, then of course $d(x, x_0) + 1 = r$ is a finite number, and if $n > r$ then $B_n(x_0)$ will contain x. This means that $\bigcup_n B_n(x_0)$ contains A (it actually contains all of X for the same reasoning).

But then the balls $B_n(x_0)$ form an open cover of A. So since A is compact, we can find a finite subcover; this means there exists N such that

$$A \subseteq B_1(x_0) \cup B_2(x_0) \cup \cdots \cup B_N(x_0)$$

(it's possible we're including more balls than we need, but that's fine). But all these balls are contained in $B_N(x_0)$, so then

$$A \subsetneq B_N(x_0).$$

We've now proved that A is contained in a ball of sufficiently large radius. □

So if we have a compact set in a metric space, then it must be bounded. We can also show that a compact set must be *closed*:

Lemma 12.26

If (X, d) is a metric space and $A \subseteq X$ is compact, then A is closed.

Proof. It is enough to show that A contains all its limit points. So we can take some $x = \lim_{n\to\infty} x_n$, with $x_n \in A$; then we want to show that $x \in A$.

Assume for contradiction that $x \notin A$. Now for each $m \in \mathbb{N}$, consider the set

$$O_m = \left\{ y \in X \mid d(y, x) > \frac{1}{m} \right\}.$$

Claim — These sets O_m are open.

We will prove this later (it is fairly similar to an argument from last class).

Now $\bigcup_m O_m$ must contain A — if we take some $y \in A$, since $x \notin A$ we have $x \neq y$, so $d(x, y) > 0$. But this means there exists some m such that $d(x, y) > 1/m$; then y is contained in our set O_m (since O_m consists of all points with distance from x greater than $1/m$).

So then our sets O_m give a cover of A. But since A is compact, there exists a finite subcover.

But if $m_1 < m_2$, then $O_{m_1} \subseteq O_{m_2}$.

This means $A \subseteq O_m$ for some m. But we know $O_m = \{y \in X \mid d(y, x) > 1/m\}$. On the other hand x was a limit of x_n, where $x_n \in A$. This means we must have $d(x_n, x) \to 0$; but in particular, this means for n sufficiently large, $d(x_n, x) < 1/m$. This means $x_n \notin O_m$; but $x_n \in A$, contradiction.

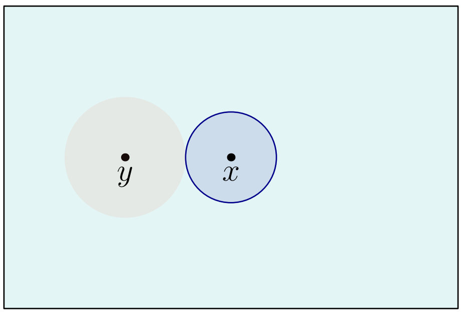
Now it remains to prove our claim that O_m are open. This follows from the triangle inequality in the same way that we proved that balls are open. We'll actually prove a slightly more general statement:

Lemma 12.27

Given a metric space (X, d) and some point $x_0 \in X$ and $r > 0$, the set

$$O = \{x \in X \mid d(x, x_0) > r\}$$

is open.



Proof. Pick some $y \in O$, so we know that $d(y, x_0) > r$. Now let $s = d(y, x_0) - r > 0$. Then we claim $B_s(y) \subseteq O$. Pick any $z \in B_s(y)$; then we want to prove that $d(z, x_0) > r$. But we have

$$d(y, x_0) \le d(y, z) + d(z, x_0),$$

which we can rearrange to get

$$d(z, x_0) \ge d(y, x_0) - d(y, z) > (r + s) - s = r,$$

as desired. ■

□

§13 November 1, 2022

§13.1 An Implication of Compactness

We've been looking at metric spaces (X, d), and last time we discussed *compactness* — X is compact if every open cover has a finite subcover. More generally, if $A \subseteq X$, then we say A is compact if every open cover of A has a finite subcover. In other words, if $A \subseteq \bigcup O_\alpha$ for a collection of open sets O_α, then $A \subseteq O_1 \cup \cdots O_n$ (so A is contained in finitely many O_i).

Recall that $O \subseteq X$ is open if and only if $X \setminus O$ is closed. And a set is closed if and only if it contains all its limit points — that's a useful way of identifying closed sets.

It's often useful to look at the *negation* of the definition of compactness — if A is compact, then we can consider the complements $U_\alpha = X \setminus O_\alpha$. Then since A is contained in the union of O_α, this means the intersection of A and $\bigcap_\alpha U_\alpha$ is empty. This gives the following:

Theorem 13.1

If (X, d) is compact and U_α are a collection of closed subsets, and $\bigcap U_\alpha$ is empty, then some finite intersection $U_1 \cap \cdots \cap U_n$ is empty.

This may seem trivial, but it is quite useful: suppose X is compact, and we have a sequence U_n of *nested* closed subsets — so $U_1 \supseteq U_2 \supseteq U_3 \supseteq \cdots$ ('nested' can also mean the opposite direction, but this isn't interesting in our situation).

Then $\bigcap_{n=1}^m U_n = U_m$, since U_m is contained in all previous sets. But X is compact, and the U_n are all closed. So if $\bigcap_{n=1}^\infty U_n = \varnothing$, then we'd have a *finite* intersection that's empty — so we'd have $\bigcap_{n=1}^N U_n = U_N = \emptyset$.

So this means if the intersection of all of the sets is empty, then from some N onwards, the sets themselves are all empty. This is a very useful property.

§13.2 Conditions for Compactness

Now suppose (X, d) is a metric space, and $A \subseteq X$ is compact. Last time, we proved that then A is closed and bounded.

For *some* metric spaces, the reverse is true — for some metric spaces X, if $A \subseteq X$ is closed and bounded, then it's also compact. But this is not *always* the case.

Example 13.2

The space $\mathbb{R}$ with the usual metric has the property that all bounded closed sets are compact.

Example 13.3

For all n, the space $\mathbb{R}^n$ with the metric

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

(where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$), called the *Euclidean distance*, has this property.

The previous example is a special case of this (taking $n = 1$). The box metric we've seen earlier also has this property; the key point is to prove that $\mathbb{R}$ has this property.

But there exist metric spaces where there are bounded closed sets that are not compact — in fact, there are many of them. We'll see a few examples of metric spaces and some bounded closed sets that are not compact. (Of course, the metric space may have other bounded closed sets that *are* compact.)

Example 13.4

Let X be the set $(0, 1)$, where $d(x, y) = |x - y|$ is the usual distance. Then X itself is not compact — we can write $X = \bigcup(1/n, 1 - 1/n)$; all of these intervals are open, and their union is all of X, but we can't have finitely many of them cover X.

A more fancy example is the following:

Example 13.5

Take the closed interval $[0, 1]$, and define X as the set of *continuous* functions on this interval — denoted

$$X = \mathcal{C}([0, 1]) = \{f \colon [0, 1] \to \mathbb{R} \mid f \text{ is continuous}\},$$

with the distance defined as

$$d(f, g) = \sup_{x \in [0,1]} |f(x) - g(x)|.$$

To see why this is well-defined, if we take two continuous functions f and g, then $h(x) = |f(x) - g(x)|$ is also a continuous function $h \colon [0, 1] \to \mathbb{R}_{\geq 0}$, and we earlier proved the *Extreme Value Theorem*, which guarantees that the sup is achieved — that there exists $x_0 \in [0, 1]$ such that $h(x_0) = \sup_{x \in [0,1]} |f(x) - g(x)|$ (and therefore the sup is also the max); then we define $h(x_0) = d(f, g)$. In particular, all our distances are finite, as they should be.

We'll now examine the following:

Question 13.6. If we have a metric space (X, d), when is it the case that any bounded sequence has a convergent subsequence?

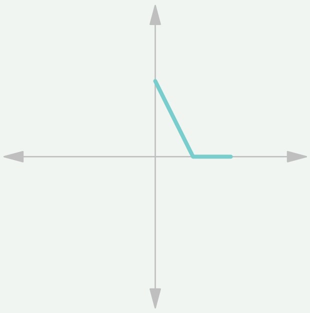
Example 13.7

We've proved earlier that this is true for $\mathbb{R}$ — that any bounded sequence has a convergent subsequence.

Remark 13.8. If you only know one metric space, it should be $\mathbb{R}$. If you're slightly more liberal with what you mean by 'one' then you should know $\mathbb{R}^n$ (in particular $\mathbb{R}^3$ and $\mathbb{R}^2$ important — if you don't travel very far, then that's the metric space we live in. If you do travel far, then we would use the 2-sphere, with the metric defined as the shortest path length on the sphere. But in a little region, the 2-sphere looks like a plane; so if you're not going up and down, then you'd think we live in $\mathbb{R}^2$). A lot of intuition you have about metric spaces comes from $\mathbb{R}$; but $\mathbb{R}$ has some properties that other metric spaces don't have. Many of these properties don't hold for *all* metric spaces, but they do hold for *some*.

Example 13.9

Our metric space $X = \mathcal{C}([0,1])$ does *not* have this property; it's very far from being compact. For example, take the sequence f_n of continuous functions on $[0,1]$, all taking values in $[0,1]$ — in other words $f_n\colon [0,1] \to [0,1]$. We can choose these functions to look as follows: they start at 1, go down linearly to $1/n$, and then become 0.

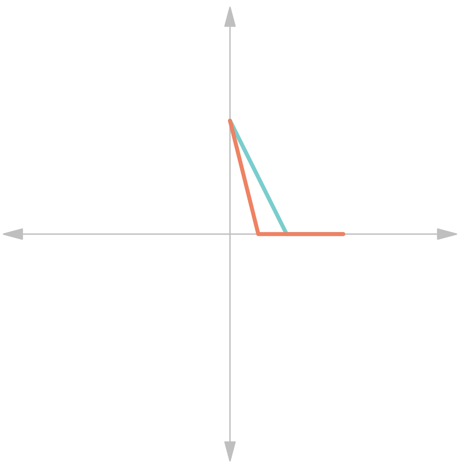


In equations,

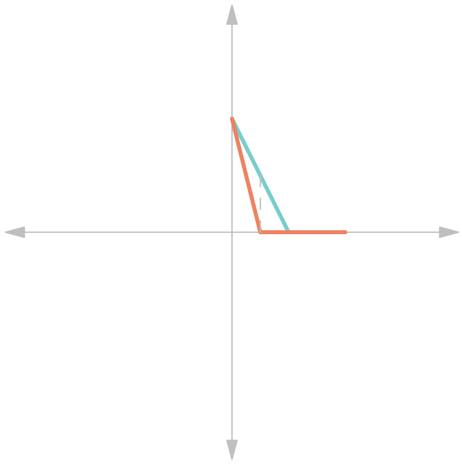
$$f_n(x) = \begin{cases} 0 & \text{if } x \geq 1/n \\ 1 - nx & \text{if } 0 \leq x \leq 1/n. \end{cases}$$

Suppose we consider

$$f_{2^{m+1}} - f_{2^m}.$$



Now we can evaluate this difference at $1/2^{m+1}$.



Then our first function vanishes, so

$$(f_{2^{m+1}} - f_{2^m})\left(\frac{1}{2^{m+1}}\right) = -f_{2^m}\left(\frac{1}{2^{m+1}}\right) = -\left(1 - \frac{2^m}{2^{m+1}}\right) = -\frac{1}{2}.$$

What this means is that if we look at the subsequence f_{2^m} of our functions f_n, then the distance between two consecutive terms is at least $1/2$.

We can do this slightly more generally — suppose we wanted to find $d(f_{2^m}, f_n)$ where $n \geq 2^{m+1}$ (so we're just further out in the sequence). Then our function f_n is only steeper than before, so the same argument gives that

$$d(f_{2^m}, f_n) \geq \frac{1}{2}.$$

This means if we look at the sequence f_{2^m} of continuous functions, then we have the property that the distance between *any* two different elements f_{2^m} and f_{2^ℓ} (with $\ell > m$) is at least $1/2$. So this sequence f_{2^m} is a sequence, but if we take any subsequence of it (so we're taking infinitely many f_{2^m} in the same order), then between any one element and all others, the distance is at least $1/2$. So there's no convergent subsequence.

So we've shown that f_{2^m} has no convergent subsequence.

Now we've looked at a fancy example of a metric space, and a sequence with the property that no subsequence is convergent. There is one more interesting observation: note that all our functions lived in a bounded subset $A \subseteq \mathbb{C}([0,1])$, where

$$A = \{f \in \mathbb{C}([0,1]) \mid 0 \leq f \leq 1\}.$$

This subset A contains all of our functions f_{2^m}.

Claim — A is bounded.

Proof. Define F to be the function that is identically 0 (so $F(x) = 0$ for all x); clearly F is in A. If we now take any function $f \in A$, we have

$$d(f, F) = \max_{x \in [0,1]} |f(x) - F(x)| = \max_{x \in [0,1]} |f(x)| \leq 1,$$

since we required that f only takes values between 0 and 1.

So we see that in $(\mathbb{C}([0,1]), d)$, the subset A is contained in $B_2(F)$ (the reason we took 2 instead of 1 is because 1 wouldn't necessarily work — $B_1(F)$ consists of functions with distance to F *strictly* less than 1 (but any radius greater than 1 would work). $\square$

To recap, we took this fancy example of a metric space, and found a sequence that lies in a bounded (and also closed) subset, but which has no convergent subsequence (because all distances between terms are at least $1/2$).

Theorem 13.10

Suppose (X, d) is a metric space, and that X is compact. Then if $x_n \in X$ is a sequence, it has a convergent subsequence.

So far, we've looked at continuous functions on the unit interval, and a subset A; if we wanted, we could just look at A with the same metric. We found a sequence in A with no convergent subsequence, so it would follow that A is not compact.

The idea of the proof comes from something we already did — but when we did it, we used some additional properties that we don't have here. We already proved that in $\mathbb{R}$ with the usual metric, if we have a bounded sequence then it has a convergent subsequence.

Proof Sketch. Suppose we have a sequence $x_n \in [0, 1]$ (what exactly the interval is doesn't matter). Then we divided the interval into two — our interval $[0, 1]$ consists of the two halves $[0, 1/2]$ and $[1/2, 1]$. If we have a sequence lying in this interval, then one of these halves must contain infinitely many terms.

So we can first set $x_{n_1} = x_1$. Then there's infinitely many of our elements in one of the two halves; let's suppose there's infinitely many in the left half. Then we take x_{n_2} to be the next element (after x_1) that lies in this half, for example x_5.

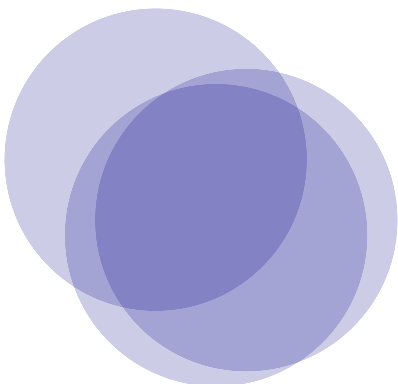
Then we restrict ourselves to this smaller interval, divide it in two, and repeat. □

This gave us a subsequence, but it used the *ordering* of $\mathbb{R}$ — and we don't have ordering in a general metric space. But we can try to extract the key points of this argument and see if we can generalize it to the setting of a compact metric space.

Proof Outline. Suppose (X, d) is compact. Suppose we start with the integer 1; then for each $x \in X$, we consider the ball $B_1(x)$. But $\bigcup_x B_1(x)$ contains all the centers of these balls, which are all our points — so $\bigcup_x B_1(x) \supseteq X$.

Then since X is compact and these open balls cover X, we must have *finitely* many balls that cover X — so we can write $B_1(x_1) \cup \cdots \cup B_1(x_n) \supseteq X$.

In the case of the real numbers, we divided our interval into *two*, and said there were infinitely many elements in one of our halves. Here we can't divide our space into two, but we have divided it into *finitely many*, which will play the same role.



Then since we have finitely many balls, there's some ball that has infinitely many terms x_n.

So to choose our subsequence, we start out by choosing $x_{n_1} = x_1$. Then we take the next element x_{n_2} to be the next element in one of the balls with infinitely many terms — and we now make sure that all elements of our subsequence lie in this ball (i.e., that for all $k > 1$ we have $x_{n_k} \in B_1(x_m)$).

Now we'd like to repeat this procedure. For that, we'll need the following fact:

Fact 13.11 — Suppose that X is compact, and $A \subseteq X$ is closed. Then A is compact.

Proof. Suppose we have an open cover of A, so $A \subseteq \bigcup_\alpha O_\alpha$. Now add one more element to our cover, namely $X \setminus A$. Since our $\bigcup_\alpha O_\alpha$ cover A, and $X \setminus A$ covers everything else, then $\bigcup_\alpha O_\alpha \cup (X \setminus A) \supseteq X$.

But then since X is compact, we can find finitely many O_α, and possibly $X \setminus A$, to cover X — so we have

$$O_1 \cup \cdots \cup O_n \cup (X \setminus A) \supseteq X.$$

But then since $X \setminus A$ doesn't help at all with covering A, then $O_1 \cup \cdots \cup O_n$ covers A; so we have found a finite subcover of A. $\square$

Previously, we also proved the following:

Fact 13.12 — In any metric space (X, d), for all $\delta > 0$ the ball $B_\delta(x)$ is open, and $\overline{B_\delta}(x) = \{y \in X \mid d(x, y) \le \delta\}$ is closed.

(This simply followed from the triangle inequality — we want to prove that the complement is open, and for every element in the complement we can use the triangle inequality to show there's a small ball around that element that's entirely contained in the complement.)

Returning to our argument, we proved our statement for $[0, 1]$ by repeatedly chopping it in halves.

To do this for a metric space, we first start with 1 and cover X with finitely many balls of radius 1 — so we can write $B = B_1(y_1) \cup \cdots \cup B_1(y_m)$. Now one of these balls (WLOG the first) has infinitely many elements of the original subsequence; so all our remaining elements of the subsequence should be chosen to be in that ball, so we guarantee that for all $k > 1$, we have $x_{n_k} \in B_1(y_1)$ for some x.

We now want to repeat this argument. This is where we now need the discussion above — we are looking at the ball $B_1(y_1)$, which is the parallel of $[0, 1/2]$ in our previous proof.

But of course we have $B_1(y_1) \subseteq \overline{B_1}(y_1)$, and $\overline{B_1}(y_1)$ is closed. But it's a subset of X, and X is compact, so then $\overline{B_1}(y_1)$ is compact as well. (The open ball $B_1(y_1)$ is not generally compact.)

Now we can repeat the same argument with $B_1(y_1)$ instead of X — we can now look at balls of radius $1/2$ around all points $y \in \overline{B_1}(y_1)$. Then we know $\bigcup_y B_{1/2}(y) \supseteq \overline{B_1}(y_1) \supseteq B_1(y_1)$. But this means *finitely many* cover our ball $\overline{B_1}(y_1)$ — so now we have

$$\overline{B_1}(y_1) \subseteq B_{1/2}(z_1) \cup \cdots \cup B_{1/2}(z_\ell).$$

So we set $x_{n_1} = x_1$, and we required that $x_{n_k} \in B_1(z)$ for all $k > 1$ (for some $k > 1$).

Student Question. How do we guarantee that this is true?

Answer. *We choose the first element randomly; then we choose the rest so that from the second one onwards, we want them all 'reasonably' bunched together (so that they're all within a ball of radius* 1*). Now we* only *consider elements of our original sequence that lie in this ball.*

Now we choose the third element so that from there on, they're bunched even *closer together (they lie in a ball of radius* $1/2$*). And so on.*

To repeat, we start with the metric space X, and we start out by looking at a collection of balls $B_1(z)$ — so then X is covered by the sets $B_1(z)$ over all $z \in X$. This means there's finitely many balls $B_1(z_1)$, ..., $B_\ell(z_\ell)$ such that

$$B_1(z_1) \cup \cdots \cup B_\ell(z_\ell) \supset X.$$

Now infinitely many terms of our sequence lie in $B_1(z_1)$.

Now when we repeat the process, we'd love to use $B_1(z_1)$ in place of the original set X. But this isn't compact, so we can't do that. But if we instead add in its boundary, then it *is* compact, and we can repeat the original process.

Now for this subsequence, we've guaranteed that our first term is just in X, from the second term onwards they all lie in a ball of radius 1 around some element z (so for $k > 1$ we have $x_{n_k} \in B_1(z)$), from the third element onwards they all lie in a ball of raidus $1/2$ arounds ome other element — so $x_{n_k} \in B_{1/2}(\tilde{z})$ for all $k > 2$. And then from the fourth element onwards they lie in a ball of radius $1/4$ around another element $\tilde{\tilde{z}}$, and so on.

It looks like we're proving that our subsequence is a Cauchy sequence — because they're bunching closer and closer together.

But now the key is that we can use the property that initially seemed trivial — the negation of the covering condition specialized to nested subsets.

We can now use this property:

> **Claim —** If (X, d) is a metric space and X is compact, and we take a sequence of nested closed subsets A_ℓ such that $A_\ell \supseteq A_{\ell+1}$, then if we also have that all A_ℓ are nonempty, then $\bigcap_\ell A_\ell$ is nonempty.

This is simply the contrapositive of our earlier statement — if we have a nested sequence of closed sets whose intersection is empty, then one of them would be empty.

Now we've constructed our sequences so that they're bunched closer and closer together. From these sets here, we can take

$$A_\ell \supseteq \overline{B}_{2^{-\ell}}(\tilde{z}).$$

(We can think of them as being equal, but we'd have to be careful because they're not necessarily nested then.)

Now to show our sequence is convergent, we want to provide a limit. We'll show that x_∞ is the intersection of all our A_ℓ (we're using compactness to prove that there's something in this intersection; in fact there must be only one element in the intersection, because the radii go to 0). $\square$

§14 November 3, 2022

§14.1 Compactness in $\mathbb{R}$

Earlier, we saw that for any metric space (X, d), if a subset $A \subseteq X$ is compact, then:

- A must be closed and bounded;
- For any *closed* subset $B \subseteq A$, B is also compact.

(We also proved that if x_n is a sequence in A, then it has a convergent subsequence x_{n_k}.)

An important property of $\mathbb{R}$ (with the standard metric) is that the *converse* of the above statement is true as well:

Theorem 14.1

If $A \subseteq \mathbb{R}$ is closed and bounded, then A is compact.

Remark 14.2. This is also true for $\mathbb{R}^n$ with the standard Euclidean metric, or with the box metric. But it isn't true for all metric spaces.

It turns out that this follows from the following fact:

Theorem 14.3

All closed intervals $[a, b] \subseteq \mathbb{R}$ are compact.

First let's see why this implies the more general claim that all closed and bounded subsets are compact.

Proof that Theorem 14.3 $\implies$ Theorem 14.1. Suppose that A is some closed and bounded subset of $\mathbb{R}$. Then since it's bounded, it's contained in some interval $[-R, R]$. Then A is a closed subset of a compact set, so it's also compact. □

Now we'll prove the theorem.

Proof of Theorem 14.3. Let O_α be a cover of $[a, b]$, so that $[a, b] \subseteq \bigcup O_\alpha$; then we want to find a finite subcover.

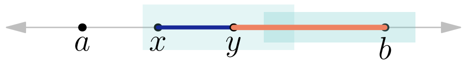
Consider the set
$$A = \{x \in [a, b] \mid [x, b] \text{ has a finite subcover}\}.$$

In other words, A is the set of x for which we can cover $[x, b]$ with finitely many of our given sets O_α.

Claim — The set A is nonempty.

Proof. We have $b \in A$ — since the O_α cover our interval $[a, b]$, there must exist some α such that O_α contains b, and therefore $[b, b]$ (which just consists of b) can be covered by a *single* set, so it can certainly be covered by finitely many. ■

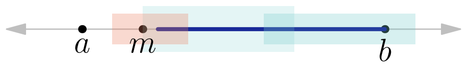
Note also that if $x \in A$, then all y with $y \geq x$ are *also* in A — the interval $[y, b]$ that we need to cover is contained in $[x, b]$, so the same finite collection covering $[x, b]$ also covers $[y, b]$.



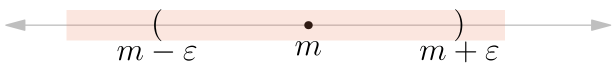
Now that we've defined A, we can consider $\inf A$, which we call m. Of course, we must have $a \leq m$.

Claim — We have $a = m$.

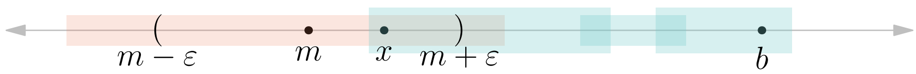
Proof. Assume for contradiction that $a < m$. Then since $m \in [a, b]$ and our sets O_α cover $[a, b]$, there must exist some set O_α such that $m \in O_\alpha$.



But since O_α is *open*, for each point in O_α there must be some ball centered at that point that's entirely contained in O_α. In particular, this means there is some $\varepsilon > 0$ such that $(m - \varepsilon, m + \varepsilon)$ is contained in O_α.



Now consider the 'midway' point $x = m + \varepsilon/2$. Then x must be in A — since $m = \inf A$, we must have some point y in A just a tiny bit to the right of m, meaning that $y \leq x$ (if we didn't have this then x would be a lower bound for A); and then this implies that $x \in A$ as well.



But since $x \in A$, the interval $[x, b]$ can be covered by finitely many O_α. Meanwhile, the interval $(m-\varepsilon, m+\varepsilon)$ is covered by one set O_α. This means we can add that set to our finite subcover to obtain a finite subcover that extends a bit to the left of m — for example, to a finite subcover of $[m - \varepsilon/2, b]$. This implies that points a bit left of m are in A, contradicting the fact that $m = \inf A$.

So we have obtained the desired contradiction, and therefore $m = a$. ■

Now to show that we can obtain a finite subcover of the entire interval $[a, b]$, we can essentially repeat the same argument. We know that a must lie in one of the open sets O_α, and therefore some interval $(a-\varepsilon, a+\varepsilon)$ must lie in that set.

But $a + \varepsilon/2$ must be in A (since $a = \inf A$, there must exist points only slightly to the right of a in A, and this again implies $a + \varepsilon/2$ is in A as well), and this means $[a + \varepsilon/2, b]$ is again contained in some finite subcover. Meanwhile $[a, a+\varepsilon/2]$ is contained in our one set O_α, so adding this one set to that finite subcover gives a finite subcover of the entire interval $[a, b]$. □

§14.2 Derivatives

We will now change topics and discuss derivatives. Suppose that $f\colon (a,b) \to \mathbb{R}$ is a function, and $x_0 \in (a,b)$.

Definition 14.4. We say that f is **differentiable** at x_0, with derivative L, if

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = L.$$

So here we're considering $x \neq x_0$, and looking at the *difference quotient*

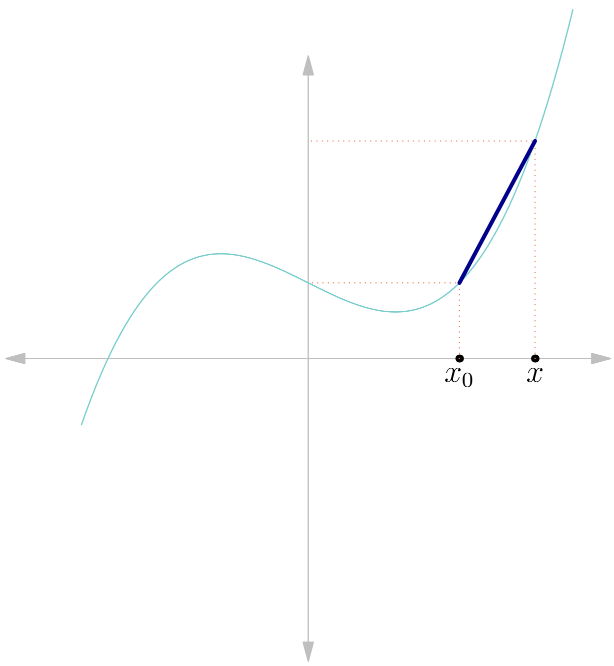
$$\frac{f(x) - f(x_0)}{x - x_0},$$

and the function is differentiable if this expression has a limit. (Note that x has to be different from x_0, as otherwise the expression does not make sense.)

We've looked at limits of *sequences* so far, but we haven't yet defined limits in this situation. This statement means that for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $|x - x_0| < \delta$ then

$$\left|\frac{f(x) - f(x_0)}{x - x_0} - L\right| < \varepsilon.$$

Remark 14.5. There are also other equivalent ways of defining this — it's equivalent to showing that for all sequences that converge to x_0, their difference quotients converge to the same L.



Note that this difference quotient is the slope of the line between $(x_0, f(x_0))$ and $(x, f(x))$.

Proposition 14.6

If $f\colon (a,b) \to \mathbb{R}$ is differentiable at x_0 with derivative L, then f is continuous at x_0.

So if f is differentiable, then it's also continuous.

Proof. Given any $\varepsilon > 0$, we want to choose δ so that that if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$. Differentiability means that we can make the expression

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - L \right|$$

as small as we want. But multiplying by $x - x_0$, we get that

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - L \right| \cdot |x - x_0| = |f(x) - f(x_0) - L(x - x_0)| \, .$$

We can make the left-hand side as small as we want; then by the triangle inequality we have

$$\begin{aligned} |f(x) - f(x_0)| &= |f(x) - f(x_0) - L(x - x_0) + L(x - x_0)| \\ &\le |f(x) - f(x_0) - L(x - x_0)| + |L(x - x_0)| \\ &\le \left| \frac{f(x) - f(x_0)}{x - x_0} - L \right| |x - x_0| + |L| \cdot |x - x_0| \, . \end{aligned}$$

By choosing x and x_0 to be close together, since L is fixed, we can make all these terms as small as we want, so we are done. $\square$

This proves that if a function is differentiable at a point, then it's also continuous at the point.

§14.3 Calculating Derivatives

Differentiable functions are nice because in general, it's easy to calculate *what* the derivative is — there's formulas for it.

Notation 14.7. We use $f'(x_0)$ to denote the derivative of f at x_0.

Proposition 14.8

If f and g are differentiable at $x_0 \in (a, b)$, then $f + g$ is also differentiable at x_0, and $(f + g)'(x_0) = f'(x_0) + g'(x_0)$.

Proposition 14.9

For any constant λ, if f is differentiable at x_0 then λf is also differentiable at x_0, and $(\lambda f)'(x_0) = \lambda f'(x_0)$.

Proposition 14.10 (Leibniz Rule)

If f and g are differentiable at x_0 then $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$.

Proposition 14.11 (Quotient Rule)

If g is never 0, then

$$\left(\frac{f}{g} \right)' = \frac{f'g - g'f}{g^2}.$$

Now let's prove these rules.

Proof of Proposition 14.8. Suppose we have two functions f and g, so we want to show

$$\left|\frac{(f+g)(x)-(f+g)(x_0)}{x-x_0}-(f'(x_0)+g'(x_0))\right|$$

is small. We can first split up the fraction as

$$\frac{f(x)-f(x_0)}{x-x_0}+\frac{g(x)-g(x_0)}{x-x_0},$$

and then collect the terms with f's and g's — so we can rewrite our expression as

$$\frac{f(x)-f(x_0)}{x-x_0}-f'(x_0)+\frac{g(x)-g(x_0)}{x-x_0}-g'(x_0).$$

Now we can use the triangle inequality, which gives that

$$\left|\frac{(f+g)(x)-(f+g)(x_0)}{x-x_0}-(f'(x_0)+g'(x_0))\right| \le \left|\frac{f(x)-f(x_0)}{x-x_0}-f'(x_0)\right|+\left|\frac{g(x)-g(x_0)}{x-x_0}-g'(x_0)\right|.$$

As long as x is close to x_0, we can guarantee that both terms are less than $\varepsilon/2$, so we are done. □

(We will not prove Proposition 14.9, as it is trivial.)

Proof of Proposition 14.10. We can use a similar idea to one we've seen a few times before. We want to consider the difference quotient

$$\frac{f(x)g(x)-f(x_0)g(x_0)}{x-x_0}.$$

We've seen expressions like this before, when looking at continuous functions and sequences, and we can do the same here — we can write this as

$$\frac{f(x)g(x)-f(x_0)g(x)+f(x_0)g(x)-f(x_0)g(x_0)}{x-x_0}.$$

Now we can split this into two fractions as

$$g(x)\cdot\frac{f(x)-f(x_0)}{x-x_0}+f(x_0)\cdot\frac{g(x)-g(x_0)}{x-x_0}.$$

As $x \to x_0$, we know that our first fraction converges to $f'(x_0)$ and our second fraction converges to $g'(x_0)$. Meanwhile $f(x_0)$ is just a constant, so the second term converges to $f(x_0)g'(x_0)$.

But for the first term, we know g is continuous, so as $x \to x_0$ we also have $g(x) \to g(x_0)$. So this means our entire expression converges to

$$f'(x_0)g(x_0)+g'(x_0)f(x_0).$$

□

Finally, instead of proving Proposition 14.11, we'll prove the special case where $f=1$. This is enough to imply the original proposition:

If $f=c$, then $f'(c)=0$ — this is because

$$\frac{f(x)-f(x_0)}{x-x_0}=0$$

for all x, and of course 0 converges to 0. So in the special case, we want to show the following:

Proposition 14.12

We have

$$\left(\frac{1}{g}\right)' = -\frac{g'}{g^2}.$$

In fact, Propositions 14.10 and 14.12 together imply Proposition 14.11 — we can think of f/g as a product

$$\frac{f}{g} = f \cdot \frac{1}{g}.$$

Then we can use the Leibniz rule, which gives

$$\left(\frac{f}{g}\right)' = f' \cdot \frac{1}{g} + f \cdot \left(\frac{1}{g}\right)'.$$

If we had 14.12, then we'd know all these derivatives — we would get

$$\frac{f'}{g} - \frac{fg'}{g^2} = \frac{f'g - fg'}{g^2}.$$

So just from the Leibniz rule and our special case, we get the quotient rule.

Proof of Proposition 14.12. We want to consider

$$\frac{\frac{1}{g(x)} - \frac{1}{g(x_0)}}{x - x_0}.$$

Obtaining a common denominator, we get

$$\frac{\frac{g(x_0)-g(x)}{g(x)g(x_0)}}{x - x_0} = \frac{g(x_0) - g(x)}{x - x_0} \cdot \frac{1}{g(x)g(x_0)}.$$

Our first term is just the difference quotient with the order reversed, so we can rewrite our expression as

$$-\frac{g(x) - g(x_0)}{x - x_0} \cdot \frac{1}{g(x)g(x_0)}.$$

Now as $x \to x_0$, our first fraction will go to $g'(x_0)$, and because g is continuous at x_0, the second term will go to $1/g(x_0)^2$. So the entire expression converges to

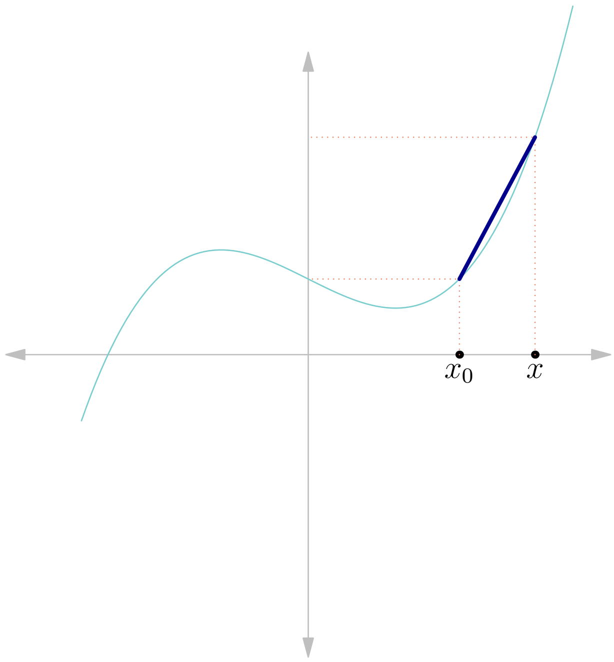
$$-\frac{g'(x_0)}{g(x_0)^2},$$

as we wanted to prove. □

§14.3.1 Some Examples

We already saw that if a function f is differentiable, then it must be continuous — if there's a jump, then the function can't be differentiable.

When we look at a difference quotient, we're considering the slope of a line:



When we're asking if f is differentiable, we want to see if these little lines all go to some line that isn't vertical (if it were vertical, then the difference quotient would go to ∞).

Example 14.13

Consider the function $x \sin(1/x)$.

First let's consider the function $\sin(1/x)$, with $x \neq 0$. This function isn't defined at 0, but we can wonder whether we could define it in a natural way at 0.

As x gets really small but positive, $1/x$ is positive and large. So as $x \to 0$ from the positive side, $1/x$ wanders off to infinity. But if $1/x$ wanders off to infinity, then sin will fluctuate very quickly — so near 0, our function $\sin(1/x)$ will fluctuate very quickly between 1 and -1. (The same will happen on the negative side.)

So it's clear that we cannot extend this to a continuous function at 0 — no matter what value we choose, close to 0 there's still values far away.

The problem here is that the amplitude of these oscillations is constant. If we multiply by x, then we're forcing the amplitude to go down; so now $x \sin(1/x)$ still has a bunch of fluctuation, but the fluctuation dies off in amplitude. So this function *is* continuous. (It's continuous everywhere except 0, but it's also continuous at 0 if we extend it to be 0 at $x = 0$.)

Question 14.14. Is the function $f(x) = x \sin(1/x)$ differentiable at 0?

Suppose we consider the difference quotient

$$\frac{f(x) - f(x_0)}{x - x_0}.$$

Our function will be nice and differentiable everywhere except 0 (by the chain rule, which we will mention soon), so we really only care about the case $x_0 = 0$, where our difference quotient becomes

$$\frac{f(x)}{x} = \sin \frac{1}{x}.$$

But $\sin\frac{1}{x}$ oscillates between -1 and 1 as $1/x \to \infty$, so it does not have a limit. So this function is continuous but not differentiable.

However, we could multiply by *another* factor of x — if we didn't have the first x then the function wasn't continuous, and with the x it was continuous but not differentiable. If we put in another x and define $g(x) = x^2 \sin(1/x)$ — so we're still killing off the amplitude, but now in a softer way — then it's easy to see that this function *is* continuous and differentiable, since we now have

$$\frac{g(x) - g(0)}{x - 0} = \frac{x^2 \sin(1/x)}{x} = x \sin\frac{1}{x} \to 0.$$

What this means is that if we look at the slopes near 0 of the lines in $x\sin(1/x)$, they don't have a limit; but if we multiply by another x, then they do have a limit (which is 0).

Example 14.15

Consider the function

$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

Of course f is continuous everywhere; and it's differentiable everywhere except possibly at 0. Meanwhile at 0 we have

$$\frac{f(x) - f(0)}{x - 0} = \frac{|x|}{x}.$$

But this expression is 1 if $x > 0$ and -1 if $x < 0$. So if we're coming from the right then the difference quotient is just 1, while if we're coming from the left then it's -1. This means the difference quotient doesn't have a limit, since it depends on which side we're coming from.

§14.3.2 The Chain Rule

The chain rule considers what happens when we take two functions and compose one with the other. We've seen earlier that if f and g are continuous, then their composition is also continuous. Now we'd like to see the same for differentiability.

Theorem 14.16

Suppose that $f\colon (a, b) \to (c, d)$ and $g\colon (c, d) \to \mathbb{R}$, and consider their composition $g \circ f\colon x \mapsto g(f(x))$. If f is differentiable at $x_0 \in (a, b)$ and g is differentiable at $f(x_0) \in (c, d)$, then $g \circ f$ is also differentiable at x_0, and

$$(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0).$$

Proof. We are considering the difference quotient

$$\frac{(g \circ f)(x) - (g \circ f)(x_0)}{x - x_0}.$$

Now we can introduce the difference quotients of f, by rewriting this as

$$\frac{g(f(x)) - g(f(x_0))}{x - x_0} \cdot \frac{f(x) - f(x_0)}{f(x) - f(x_0)}.$$

(Assume that $f(x) \neq f(x_0)$.) Then this becomes

$$\frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \cdot \frac{f(x) - f(x_0)}{x - x_0}.$$

Now we can think of $f(x)$ as a new variable y, and $f(x_0)$ as y_0. Then we can write this expression as

$$\frac{g(y)-g(y_0)}{y-y_0}\cdot\frac{f(x)-f(x_0)}{x-x_0}.$$

If y is close to y_0, then the first term converges to $g'(y_0)$, and of course the second term converges to $f'(x_0)$.

But the point is that as $x \to x_0$, since f is continuous we must have $f(x) \to f(x_0)$, so y *does* really go to y_0. So then our first term really does approach $g'(y_0)$.

(We have to also handle the case where $f(x) - f(x_0) = 0$, but that case is easy to check.) □

§15 November 8, 2022

§15.1 Finding Extrema

One useful application of differentiation is to find the maxima and minima of a function.

Question 15.1. Suppose we have a function $f\colon [a,b] \to \mathbb{R}$ such that f is differentiable. How can we find $\max f$?

This is not a priori an easy question, but there's a way to make it easier, by looking at the derivative.

Lemma 15.2

If $x_0 \in (a,b)$ is a local maximum or local minimum of f, then $f'(x_0) = 0$.

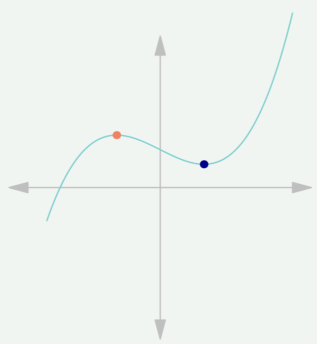
Definition 15.3. We say f has a **local maximum** at x_0 if there exists $\varepsilon > 0$ such that

$$f(x_0) = \max\{f(x) \mid x \in (x-\varepsilon, x+\varepsilon)\}.$$

So a function has a local maximum at x_0 if there exists a little interval around x_0 on which x_0 is the maximum.

Example 15.4

Suppose we have a function as follows:



Then the red point is a local maximum, and the blue point a local minimum.

Proof of Lemma. We'll prove this for local maxima. Assume that $x_0 \in (a, b)$ is a local maximum, and $f(x_0) \geq f(x)$ for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$. Now consider the difference quotient

$$\frac{f(x) - f(x_0)}{x - x_0}.$$

If $x > x_0$ (so we're on the right of x_0), then $x - x_0 > 0$, and $f(x) - f(x_0) \leq 0$, so the difference quotient is at most 0. On the other hand, if $x < x_0$ (so we're on the left of x_0), then we still have $f(x) - f(x_0) \leq 0$, but now $x - x_0 < 0$, and this means the difference quotient is at least 0.

But we have

$$f'(x_0) = \lim_{x \to 0} \frac{f(x) - f(x_0)}{x - x_0}.$$

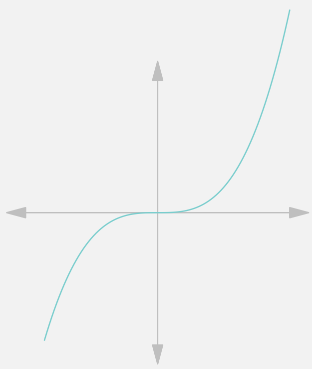
When we're approaching from the right, we know this difference quotient is nonpositive, and from the left it's nonnegative. So then the difference quotient must be 0 (since it's the limit of nonpositive as well as of nonnegative things).

(Of course, the same argument with obvious changes also gives the same statement for local minima.) □

> **Remark 15.5.** Note that it's important we could go both to the right and left — so it's important that x_0 is an interior point.

Now suppose we have a function f, and we want to find $\max f$.

In some easy situations, the maximum may be on the boundary. But typically it's not; then we want to look at for which points x_0 we have $f'(x_0) = 0$. This gives a bunch of different points, adn we can compare the values at these different points and at the endpoints to find the maximum or minimum.

> **Remark 15.6.** Suppose we take the function $f(x) = x^3$.
>
> 
>
> Then $f'(x) = 3x^2$, which means $f'(0) = 0$. But f does not have a local maximum or local minimum — which shows that the converse of our statement is not true.

§15.2 Mean Value Theorem

Theorem 15.7 (Rolle's Lemma)
Suppose $f \colon [a, b] \to \mathbb{R}$ is differentiable, and $f(a) = f(b)$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.

Proof. We know f has some value at a, and the same value at b. We can divide into cases (these cases are not mutually disjoint, but that is fine):

Case 1 (There exists some point $x_0 \in (a, b)$ such that $f(x_0) > f(a) = f(b)$). We know f is differentiable and therefore continuous, so it must achieve a maximum at some point. But that maximum must be interior, since the value at the maximum must be strictly larger than that of the endpoints. So in this case, there is an interior (global) maximum. Since this is an interior maximum, it must be a global maximum; so $f'(x) = 0$.

Case 2 (There exists $y_0 \in (a, b)$ such that $f(y_0) < f(a) = f(b)$). Similarly, there now must exist an interior minimum (since the value at the minimum is strictly less than the value at the endpoints). So at this minimum y we have $f'(y) = 0$.

Case 3 (We have $f(x) = f(a)$ for all $x \in (a, b)$). Then f is constant, which means $f'(x) = 0$ at *all* points. □

This may seem trivial, but we can reduce a more fancy and very useful statement to Rolle's theorem.

Theorem 15.8 (Mean Value Theorem)
Suppose $f\colon [a, b] \to \mathbb{R}$ is differentiable. Then there exists $c \in (a, b)$ such that

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

Proof. The main idea is to take our function and construct another function on which we can apply Rolle's lemma. We define

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a} \cdot (x - a).$$

Then we have $g(a) = f(a)$, and

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

as well. This means g is differentiable and $g(a) = g(b)$, so there exists $c \in (a, b)$ so that $g'(c) = 0$. But

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

(since this factor is a constant, and $x - a$ has derivative 1). Since we have $g'(c) = 0$, this means

$$f'(c) - \frac{f(b) - f(a)}{b - a} = 0,$$

so equivalently,

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

□

There is also a more fancy version of the mean value theorem.

Theorem 15.9 (Cauchy Mean Value Theorem)
Suppose $f, g\colon [a, b] \to \mathbb{R}$ are differentiable. Then there exists $c \in (a, b)$ such that

$$(f(b) - f(a))\, g'(c) = (g(b) - g(a))\, f'(c).$$

Remark 15.10. We could also have written this as

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

but we didn't because then we'd have to add the condition that certain expressions are nonzero.

Remark 15.11. Note that the Cauchy mean value theorem implies the usual mean value theorem — suppose $g(x) = x$. Then we have $g'(x) = 1$ for all x, so for any (differentiable) function $f\colon [a,b] \to \mathbb{R}$, the Cauchy mean value theorem on this choice of (f,g) tells us that

$$(f(b) - f(a)) \cdot 1 = (b-a) \cdot f'(c)$$

for some c, and dividing by $b-a$ gives

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

So the Cauchy mean value theorem implies the mean value theorem (by the special case where we set f to be the function we have, and g to be the identity).

Proof. If we subtract our two functions, we get that we want to show

$$(f(b) - f(a))g'(c) - (g(b) - g(a))f'(c) = 0.$$

So we can define the function

$$h(x) = (f(b) - f(a))g(x) - (g(b) - g(a))f(x).$$

Then we have

$$h'(x) = (f(b) - f(a))g'(x) - (g(b) - g(a))f'(x)$$

(since $f(b) - f(a)$ and $g(b) - g(a)$ are just constants). Meanwhile,

$$\begin{aligned} h(a) &= (f(b) - f(a))g(a) - (g(b) - g(a))f(a) \\ &= f(b)g(a) - f(a)g(a) - g(b)f(a) + g(a)f(a) \\ &= f(b)g(a) - g(b)f(a), \end{aligned}$$

Similarly, we have

$$\begin{aligned} h(b) &= (f(b) - f(a))g(b) - (g(b) - g(a))f(b) \\ &= f(b)g(b) - f(a)g(b) - g(b)f(b) + g(a)f(b) \\ &= -f(a)g(b) + g(a)f(b). \end{aligned}$$

So we can see that $h(a) = h(b)$.

Now Rolle's lemma applies to h, giving that there exists $c \in (a,b)$ such that $h'(c) = 0$. But we have a formula for the derivative of h; plugging in c gives

$$0 = h'(c) = (f(b) - f(a))g'(c) - (g(b) - g(a))f'(c).$$

But this is exactly what we wanted. □

These theorems are good for quite a lot of things; we'll now see some applications.

§15.2.1 L'Hospital's Law

Question 15.12. Suppose $f, g\colon [a, b] \to \mathbb{R}$, and we are interested in f/g; but the problem is that $g(a) = 0$, so that this quotient is not defined at a.

If $f'(a) \neq 0$, then as we approach a, the fraction will blow up. The more interesting case is when $f(a) = 0$ as well. Then does

$$\lim_{y \to a} \frac{f(x)}{g(x)}$$

exist?

Suppose $f, g\colon [a, b] \to \mathbb{R}$ are differentiable on $[a, b]$, and f' and g' are continuous on $[a, b]$.

Remark 15.13. When we said the function is differentiable on $[a, b]$ in the previous theorems, we only really needed the function to be continuous on $[a, b]$ and differentiable on (a, b). But in this theorem, we really need f and g to be differentiable on the entire interval.

We want to apply the Cauchy mean value theorem. We'll apply it not on the whole interval from a to b, but instead on the interval $[a, x]$ where $x \in [a, b]$ — since we are interested in $f(x)/g(x)$ as $x \to a$. (So we'll only think about the function on this interval, and forget what happens from a to b.)

The Cauchy mean value theorem tells us that there exists $c \in (a, x)$ such that

$$(f(x) - f(a))g'(c) = (g(x) - g(a))f'(c).$$

But $f(a)$ and $g(a)$ are both 0, so this complicated equation really just tells us that

$$f(x) \cdot g'(c) = g(x) \cdot f'(c).$$

Now dividing gives that

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}.$$

But as $x \to a$, c is squeezed between x and a, so $c \to a$ as well. If we assume $g'(a) \neq 0$ and that f' and g' are continuous, then

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

This is called L'Hospital's law.

There are other variations on this as well.

Question 15.14. Suppose $f, g\colon (a, b) \to \mathbb{R}$ are differentiable, but both f and g go to ∞ as $x \to a$. We are again interested in f/g, but at a, neither f nor g makes sense. (If only g went to ∞ and f were well-defined and continuous, then this fraction would go to 0, so this question is only interesting if *both* go to ∞ — then there's the possibility we may be able to extend f/g in a meaningful way to a.)

Example 15.15

For example, we may have the interval $(0, 1)$ and $g(x) = 1/x$.

We can again perform a very similar argument. Our functions aren't defined at a, so we can't apply the Cauchy mean value theorem on $[a, x]$ anymore. But we can instsead apply it on a slightly smaller interval

— take x_0 and some $a < x < x_0$, and apply the Cauchy mean value theorem on $[x, x_0]$. This gives that for some $c \in (x, x_0)$, we have

$$(f(x) - f(x_0))g'(c) = (g(x) - g(x_0))f'(c).$$

Since f and g blow up as $x \to a$, $g(x) - g(x_0)$ is certainly nonzero, so we have

$$\frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \frac{f'(c)}{g'(c)}.$$

Now we see that $f(x)$ and $g(x)$ both blow up as we take $x \to a$; so we can rewrite this as

$$\frac{f(x)}{g(x)}\left(\frac{1 - \frac{f(x_0)}{f(x)}}{1 - \frac{g(x_0)}{g(x)}}\right) = \frac{f'(c)}{g'(c)},$$

which means that

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)} \cdot \frac{1 - \frac{g(x_0)}{g(x)}}{1 - \frac{f(x_0)}{f(x)}}$$

for some c with $a < x < c < x_0 < b$. Now if we take $x \to a$, since x_0 is fixed and $f(x)$ and $g(x)$ blow up, then the right-hand side approaches $f'(c)/g'(c)$ — but we have to be somewhat careful because c depends on x and x_0.

Claim — Suppose that

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = L$$

(in particular, this limit exists). Then we have

$$\frac{f(x)}{g(x)} \to L.$$

To prove this, we've already found the formula

$$\frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)} \cdot \frac{1 - \frac{g(x_0)}{g(x)}}{1 - \frac{f(x_0)}{f(x)}}.$$

What this says is that if we take the interval (a, b) and fix some $x_0 \in (a, b)$, and then take some $x \in (a, x_0)$, then there exists $c \in (x, x_0)$ for which this is satisfied.

Here c depends on x. But if we take $x_0 \to a$ (which we will do eventually), then c will be squeezed between a and x_0, so it will also go to a.

To be slightly more precise, fix x_0 very close to a, and then take $x \to a$. Then in our fraction, $g(x_0)/g(x)$ and $f(x_0)/f(x)$ go to 0, so $f(x)/g(x)$ is always equal to some $f'(c)/g'(c)$ for $c \in (x, x_0)$. The problem is that c will depend on our choice of x_0 and x. But by assumption, if x_0 is close to a, then c will also be close to a, so $g'(c)/f'(c)$ is close to L.

To write this out explicitly, given $\varepsilon > 0$ there exists $\delta > 0$ such that if $|x_0 - a| < \delta$, then

$$\left|\frac{g'(c)}{f'(c)} - L\right| < \varepsilon$$

(Because if $|x_0 - a| < \delta$, then $|c - a| < \delta$ as well — because c is squeezed between a and x_0).

So then $f'(c)/g'(c)$ is virtually L (it changes a bit as x changes, but that doesn't really matter).

Student Question. *What's the order in which you're taking limits?*

Answer. Given ε, you fix δ such that when $|x_0 - a| < \delta$, this is true. Now we fix x_0 and let $x \to a$, so that our fraction goes to 1 and $f'(c)/g'(c)$ is very close to L.

§16 November 10, 2022

§16.1 Taylor Expansion

Suppose we have a function $f\colon [a,b] \to \mathbb{R}$ and a point $x_0 \in [a,b]$, and we want to understand f near x_0. We'll assume that f is differentiable, and furthermore that $f^{(k)}$ exists and is differentiable for all $k \in \mathbb{N}$ (so the derivative of f makes sense, its derivative also makes sense, and so on). In this case, we can form the *Taylor polynomials* of f:

Definition 16.1. The **Taylor polynomial** of f is the polynomial

$$P_{k-1}(x) = f(x_0) + f'(x_0)\cdot(x-x_0) + \frac{f^{(2)}(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(k-1)}(x_0)}{(k-1)!}(x-x_0)^{k-1}.$$

So then $P_{k-1}(x)$ is a polynomial in x of degree $k-1$.

Question 16.2. How well does P_{k-1} approximate f near x_0?

The advantage of the Taylor polynomial is that it's a *polynomial*, rather than a complicated function.

The key point in answering this is the following generalization of the mean value theorem:

Theorem 16.3 (Taylor expansion)

Suppose $f\colon [a,b] \to \mathbb{R}$ is a function, and $x_0 < x$ are in $[a,b]$ Then if f is k-times differentiable, then there exists $c \in (x_0, x)$ with the property that

$$f(x) = P_{k-1}(x) + \frac{f^{(k)}(c)}{k!}\cdot(x-x_0)^k.$$

Intuitively, we think of x and x_0 as being close together. This means $|x-x_0|$ is small, so we're taking something small to the kth power. This means usually $f^{(k)}(c)(x-x_0)^k/k!$ will be extremely small, which means we get a very small error term (and as k grows this error shrinks).

Proof. Fix both x (which we'll now denote as x_1 to free up the variable x for other uses) and x_0, and let M be given by

$$f(x_1) - P_{k-1}(x_1) = \frac{M}{k!}(x_1-x_0)^k.$$

(There is a unique M satisfying this equation because $x_1 - x_0 \neq 0$.) To prove Taylor expansion, we need to show that $M = f^{(k)}(c)$ for some $x_0 < c < x_1$.

To prove this, we will apply Rolle's lemma repeatedly. First, consider the function

$$g(x) = f(x) - P_{k-1}(x) - \frac{M}{k!}(x-x_0)^k.$$

Now let's think about g on the interval $[x_0, x_1]$. Since polynomials are infinitely differentiable, g is differentiable as many times as f is.

At the left endpoint, we have

$$g(x_0) = f(x_0) - P_{k-1}(x_0) - \frac{M}{k!}(x_0-x_0)^k.$$

Of course the last term is 0, and plugging into the definition of the Taylor polynomial we have $P_{k-1}(x_0) = f(x_0)$ as well, so then $g(x_0) = 0$. Meanwhile, at the right endpoint we have

$$g(x_1) = f(x_1) - P_{k-1}(x_1) - \frac{M}{k!}(x - x_0).$$

But we defined M precisely so that this expression is 0, which means $g(x_1) = 0$. This now allows us to apply Rolle's lemma — there exists $c_1 \in (x_0, x_1)$ such that $g'(c_1) = 0$.

Now we have x_0, c_1, and x_1 in that order, and we'll look at g' on the smaller interval $[x_0, c_1]$.

At the right endpoint of this interval, we constructed c_1 so that $g'(c_1) = 0$. Meanwhile, we have

$$g'(x_0) = f'(x_0) - P'_{k-1}(x_0) - \frac{M}{(k-1)!}(x_0 - x_0)^{k-1}$$

(using the fact that $k! = k(k-1)!$). (We'll assume $k > 1$; otherwise we wouldn't need this step.) The last term is 0, so

$$g'(x_0) = f'(x_0) - P'_{k-1}(x_0).$$

But looking at the Taylor polynomial again, we have

$$P'_{k-1}(x) = 0 + f'(x_0) + \frac{f''(x_0)}{1!}(x - x_0)^1 + \cdots + \frac{f^{(k-1)}(x_0)}{(k-2)!}(x - x_0)^{k-2}.$$

(Essentially we've just shifted all terms down by one step.) In particular, $P'_{k-1}(x_0) = f'(x_0)$, which means that $g'(x_0) = 0$.

So now we have seen that $g'(c_1) = 0$ and $g'(x_0) = 0$. Again by Rolle's lemma, there exists $c_2 \in (x_0, c_1)$ such that $g''(c_2) = 0$.

We can now do this again — consider the function $g^{(2)}(x)$ on the interval $[x_0, c_2]$. Then we have $g^{(2)}(x_0) = 0$, and by construction $g^{(2)}(c_2) = 0$ as well. Applying Rolle's theorem, we get that there exists $c_3 \in (x_0, c_2)$ such that $g^{(3)}(c_3) = 0$.

We now repeat this process k times. On the final step, we will find some c_k such that $g^{(k)}(c_k) = 0$. (We'll have $x_0 < c_k < c_{k-1} < \cdots < x_1$.)

But the kth derivative of g is

$$g^{(k)} = f^{(k)} - P_{k-1}^{(k)} - M$$

(we start out with $M/k! \cdot (x - x_0)^k$, and each time we take the derivative we bring a k down; so when we do it k times, we simply end up with M). Meanwhile P_{k-1} is a polynomial of degree $k - 1$; this means its kth derivative is 0. This means we have

$$0 = g^{(k)}(c_k) = f^{(k)}(c_k) - M,$$

which implies that $M = f^{(k)}(c_k)$ for some c_k in our interval — which is what we wanted to prove. □

Now we'll return to our original question:

> **Question 16.4.** Given a complicated function, how can we use the Taylor polynomial to understand it near a given point?

Example 16.5

Suppose our function is $f(x) = e^x$. We have $f'(x) = e^x$, which means $f^{(k)}(x) = e^x$ for all k. Now let's take the Taylor polynomials at 0. We know $f(0) = 1$, so all future derivatives at 0 are 1 as well, so then

$$P_{k-1}(x) = 1 + x + \frac{1}{2!} \cdot x^2 + \cdots + \frac{1}{(k-1)!} x^{k-1}.$$

This is a polynomial, so it's in principle easy to compute. But we know that

$$f(x) - P_{k-1}(x) = \frac{f^k(c)}{k!} x^k.$$

This means

$$|f(x) - P_{k-1}(x)| = \left| \frac{f^{(k)}(c)}{k!} \cdot x^k \right| = \frac{e^c}{k!} \cdot |x|^k.$$

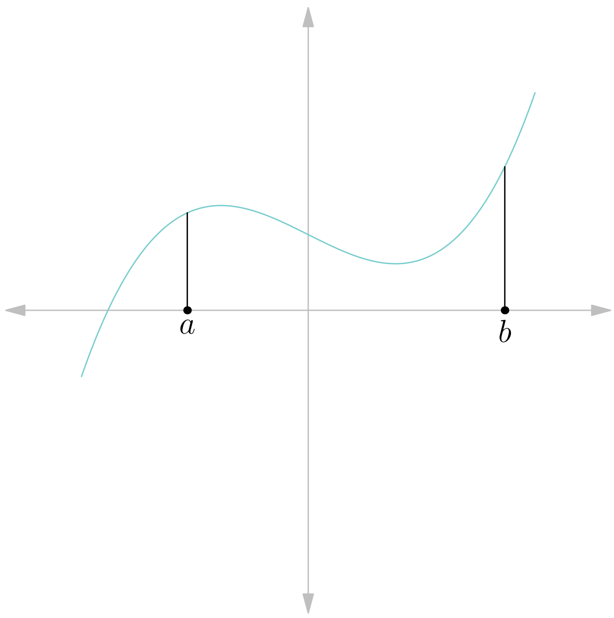
Now let's suppose $x \in (-1, 1)$, so that $c \in (-1, 1)$ as well. Then we must have $|e^c| \leq e \approx 2.7$. This means as long as $|x| < 1$, we have

$$|e^x - P_{k-1}(x)| \leq \frac{e}{k!} \cdot |x|^k.$$

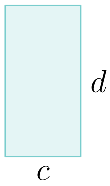
If $|x|$ is small, then $|x|^k$ is incredibly small; then the more terms we take in our Taylor approximation, the better approximation we'll get.

§16.2 Riemann Integrals

Suppose we have a function $f\colon [a, b] \to \mathbb{R}$. The idea is that we want to compute the area below the graph of f. (For now, imagine that $f > 0$, so that it's clear what we mean by 'below.')

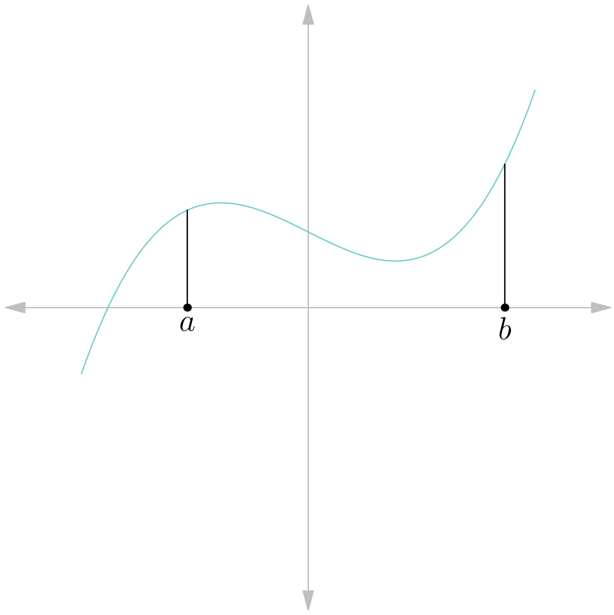


This may be complicated to compute. But suppose we instead were trying to compute the area of a *rectangle*, with side lengths c and d. Then we know its area is simply cd.

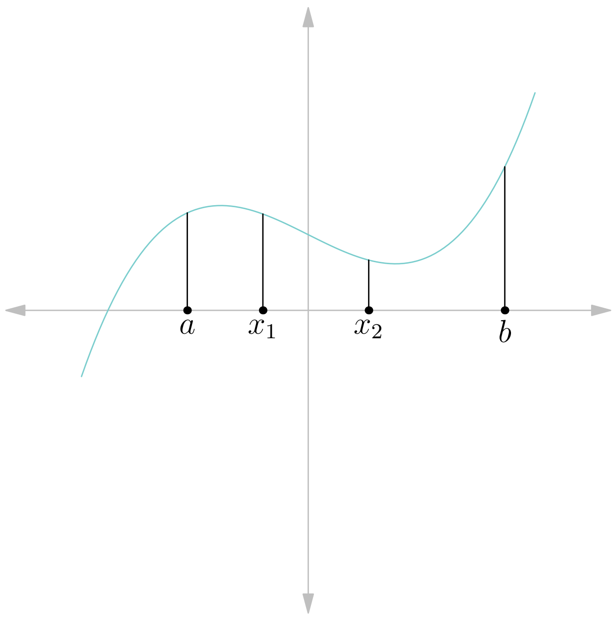


The idea is that we're going to use this to define the area below the graph of f; next class we'll show that if f is continuous, then this area is well-defined.

The main idea is that we can divide our function from a to b into a bunch of smaller intervals. The simplest thing we could do is to not divide it at all. Then we can look at its infimum; the area below the graph will certainly be larger than the area of the rectangle this produces. Meanwhile, it'll certainly be smaller than the area produced by the supremum.



To refine this, we use *partitions*. Suppose we take a number of dividing points $a = x_0 < x_1 < \cdots < x_n = b$, and we then consider $f\colon [x_i, x_{i+1}] \to \mathbb{R}$ for each smaller interval.



Then we can let $\sup_{[x_i,x_{i+1}]} f = M_i$ and $\inf_{[x_i,x_{i+1}]} f = m_i$ for each i. Then the area under our graph can be divided into these intervals as well. But the area in each part should be bounded above by M_i times the length of the interval; this means we should have

$$\text{Area}(f) \le \sum M_i(x_{i+1} - x_i).$$

SImilarly, we should have

$$\text{Area}(f) \ge \sum m_i(x_{i+1} - x_i).$$

So now the area is squeezed between

$$\sum m_i(x_{i+1} - x_i) \le \text{Area}(f) \le \sum M_i(x_{i+1} - x_i).$$

We'll prove next class that taking the sup of the left-hand side and the inf of the right-hand side over all partitions will give us the same number. Today, we'll set up some of what we need for this.

For a partition $\mathcal{P}$ given by $a = x_0 < \cdots < x_n = b$, we define

$$\mathcal{U} = \sum M_i(x_{i+1} - x_i),$$

where $M_i = \sup_{[x_i,x_{i+1}]} f$. Similarly, we define

$$\mathcal{I} = \sum m_i(x_{i+1} - x_i),$$

where $m_i = \inf_{[x_i,x_{i+1}]} f$.

There's a few things we can prove about this. First, we always have $m_i \le M_i$ (since the sup of f on an interval is certainly at least the inf). This means

$$m_i(x_{i+1} - x_i) \le M_i(x_{i+1} - x_i).$$

Summing over all i, this means

$$\sum m_i(x_{i+1} - x_i) \le \sum M_i(x_{i+1} - x_i).$$

This gives us the following trivial lemma:

Lemma 16.6

For any fixed partition $\mathcal{P}$, we have $\mathcal{I}_{\mathcal{P}} \le \mathcal{U}_{\mathcal{P}}$.

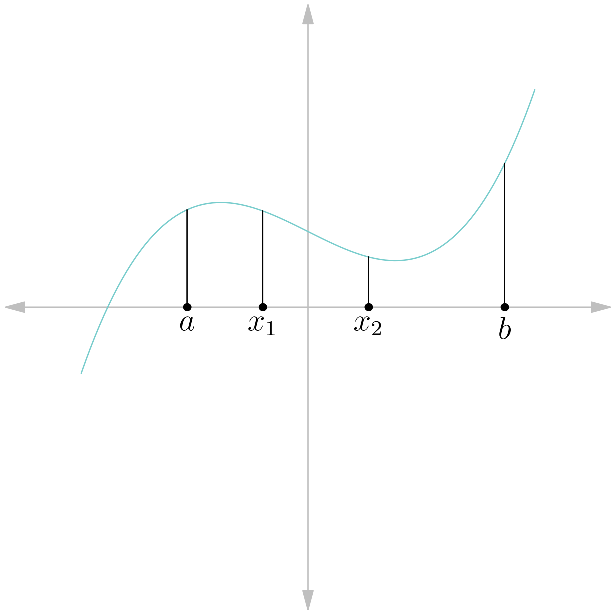
However, we want to do this over *all* partitions — because we want to make the dividing points closer and closer together, but each stage only makes sense if we have finitely many intervals.

So the next thing to do is to consider $\mathcal{I}_{\mathcal{P}_1}$ and $\mathcal{U}_{\mathcal{P}_2}$. *A priori* it might not be clear that there's any relation between them. But there should be — suppose we have two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$, with $\mathcal{P}_1$ given by $a = x_0 < x_1 < \cdots < x_n = b$ and $\mathcal{P}_2$ given by $a = y_0 < y_1 < \cdots < y_m = b$. (Note that they don't need to have the same number of dividing points.)

Definition 16.7. For two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$, the **refinement** of $\mathcal{P}_1$ and $\mathcal{P}_2$ is the partition where we include all the dividing points of both $\mathcal{P}_1$ and $\mathcal{P}_2$. (If some dividing point occurs in both, we only take it once.)

So we now have a partition $\mathcal{P}_3$ given by dividing points $a = z_0 < z_1 < \cdots < z_\ell = b$.

Claim — Suppose $\mathcal{P}$ and $\mathcal{P}^*$ are two partitions, where $\mathcal{P}^*$ is *finer* than $\mathcal{P}$ — meaning that $\mathcal{P}^*$ contains all the dividing points of $\mathcal{P}$, as well as possibly others. Then $\mathcal{I}_{\mathcal{P}} \leq \mathcal{I}_{\mathcal{P}^*}$.



When finding $\mathcal{I}_{\mathcal{P}}$, we take inf on each interval,and multiply by the length of the interval. But if we now divide this interval into sub-intervals, we're taking the length of the first interval and the minimum over that interval, and the length of the second interval and its minimum. But the minimums on these smaller intervals are *at least* the minimum on the original.

So we can see that if we take a finer partition, then $\mathcal{I}$ increases (not necessarily strictly).

Similarly, we can see that if $\mathcal{P}^*$ is finer than $\mathcal{P}$, then $\mathcal{U}_{\mathcal{P}^*} \leq \mathcal{U}_{\mathcal{P}}$.

Putting this together, we have

$$\mathcal{I}_{\mathcal{P}} \leq \mathcal{I}_{\mathcal{P}^*} \leq \mathcal{U}_{\mathcal{P}^*} \leq \mathcal{U}_{\mathcal{P}}.$$

Now returning to our two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$, we know that $\mathcal{P}_1$ has finitely many dividing points, and so does $\mathcal{P}_2$, so if we take a third partition $\mathcal{P}_3$ containing all the dividing points of both, then it's a refinement of both $\mathcal{P}_1$ and $\mathcal{P}_2$. This means we have

$$\mathcal{I}_{\mathcal{P}_1} \leq \mathcal{I}_{\mathcal{P}_3} \leq \mathcal{U}_{\mathcal{P}_3} \leq \mathcal{U}_{\mathcal{P}_2}.$$

This gives the following:

Lemma 16.8

For any two (not necessarily related) partitions, we have $\mathcal{I}_{\mathcal{P}_1} \leq \mathcal{U}_{\mathcal{P}_2}$.

(It makes sense that this *should* be true, since the left-hand side gives a lower bound for the area, and the right-hand side gives an upper bound for the area.)

Definition 16.9. For a *bounded* function $f\colon [a,b] \to \mathbb{R}$, we define the **upper Riemann integral** as

$$\inf_{\mathcal{P}} \mathcal{U}_{\mathcal{P}},$$

where the inf is taken over all partitions.

(Note that we need f to be bounded so that on each interval we can take sup and inf.)

Definition 16.10. Likewise, we define the **lower Riemann integral** as

$$\sup_{\mathcal{P}} \mathcal{I}_{\mathcal{P}}.$$

Notation 16.11. The upper Riemann integral can be denoted as $\overline{\int}_a^b f\,dx$, and the lower Riemann integral can be denoted as $\underline{\int}_a^b f\,dx$.

Definition 16.12. A bounded function f is **Riemann integrable** if $\overline{\int}_a^b f\,dx = \underline{\int}_a^b f\,dx$.

Notation 16.13. If f is Riemann integrable, then the common value is denoted as $\int_a^b f\,dx$.

Question 16.14. Which functions are integrable?

Next time, we'll prove the following theorem:

Theorem 16.15

Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous. Then f is Riemann integrable.

(We already know that all continuous functions are bounded; but in fact, they're also Riemann integrable.)

§16.2.1 Uniform Continuity

The key component of the proof will be the notion of *uniform continuity*.

Definition 16.16. We say a function $f\colon [a,b] \to \mathbb{R}$ is **uniformly continuous** if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for *all* $x_0 \in [a,b]$, if $|x - x_0| < \delta$, then $|f(x) - f(x_0)| < \varepsilon$.

This looks very similar to the definition of continuity, but the key point is that δ does *not* depend on x_0. When we say a function is continuous at x_0, we say that given x_0 and ε, there exists δ with the desired property — but δ is allowed to depend on x_0.

The key theorem to prove our result is the following theorem:

Definition 16.17. If $f\colon [a,b] \to \mathbb{R}$ is a continuous function, then f is uniformly continuous.

So any continuous function on a compact interval is uniform. It's crucial that the interval is compact:

Example 16.18

Take the function $f(x) = 1/x$ on the interval $(0,1]$. This interval is *not* compact, because 0 is not included. We know f is continuous on $(0,1]$. But given any ε, however close x and x_0 are required to be, if we move x_0 really close to 0 then we can force $1/x$ and $1/x_0$ as far apart as we want. So given *any* ε, there is no δ that works for all x_0. (If we *fixed* x_0 then for each we'd be able to find δ, since f is continuous.)

§17 November 15, 2022

§17.1 Uniform Continuity

Last time, we saw the following definition:

Definition 17.1. A function $f\colon I \to \mathbb{R}$ is **uniform continuous** if for all $\varepsilon > 0$, there exists $\delta > 0$ so that if $x, y \in I$ with $|x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$.

The distinction between uniform continuity and just continuity is that when we say a function is continuous, we mean it's continuous at all its points — so if we *fix* x_0, then for all $\varepsilon > 0$ there exists δ such that for $|y - x_0| < \delta$, $|f(y) - f(x_0)| < \varepsilon$. In particular δ is allowed to depend on x_0. Meanwhile, uniform continuity says that we can use the same δ for all x.

Example 17.2 (A Non-example)

Consider the function $f(x) = 1/x$ on $(0, 1] \to \mathbb{R}$. Then f is continuous, but it is *not* uniform continuous — suppose we have *any* ε. Then taking x and y to be really close to 0, we can make them as close together as we want while keeping $|f(x) - f(y)| = \varepsilon$.

Example 17.3 (A Non-example)

Consider the function $f(x) = x^2$ on $\mathbb{R} \to \mathbb{R}$. Then we have $f(x) - f(y) = x^2 - y^2$. If $y = x + \delta$, then

$$f(x) - f(y) = x^2 - (x + \delta)^2 = -\delta^2 - 2x\delta.$$

In particular,

$$|f(x) - f(x + \delta)| = \delta\,|\delta + 2x|\,.$$

If x becomes really large, then this difference becomes really large. So we can't use the same δ for all x to make this difference small, and therefore f is not uniform continuous.

Theorem 17.4

If $f\colon [a, b] \to \mathbb{R}$ is continuous, then f is uniform continuous.

Proof. We will prove this by contradiction. Assume f is not uniform continuous; then there exists some $\varepsilon > 0$ such that for all $\delta > 0$, there exist x and y in $[a, b]$ with $|x - y| < \delta$ such that $|f(x) - f(y)| \geq \varepsilon$. We treat ε as fixed; then we know that for all δ there exist such x and y.

Now we pick a sequence $\delta_n = 1/n$, so there exist x_n and y_n with $|x_n - y_n| < 1/n$ and $|f(x_n) - f(y_n)| \geq \varepsilon$.

Now since $[a, b]$ is compact, we know there exists a convergent subsequence $x_{n_k} \to x_\infty$. But then we have

$$|x_{n_k} - y_{n_k}| < \frac{1}{n_k} \to 0,$$

and by the triangle inequality we have

$$|x_\infty - y_{n_k}| \leq |x_\infty - x_{n_k}| + |x_{n_k} - y_{n_k}|\,.$$

The first term goes to 0 because $x_{n_k} \to x_\infty$, and the second term is bounded by $1/n_k$, which also goes to 0; so then $|x_\infty - y_{n_k}| \to 0$, and therefore $y_{n_k} \to x_\infty$ as well.

But because $[a, b]$ is compact, we have $x_\infty \in [a, b]$. So now we have a subsequence $x_{n_k} \to x_\infty$ and a corresponding subsequence $y_{n_k} \to y_\infty$, and we have that $|f(x_{n_k}) - f(y_{n_k})| \geq \varepsilon$. But the first two facts, since f is continuous, mean that $f(x_{n_k}) \to f(x_\infty)$ and $f(y_{n_k}) \to f(y_\infty)$. But this means the difference between $f(x_{n_k})$ and $f(y_{n_k})$ must go to 0, contradiction. □

So if we have a *compact* interval, then any continuous function is automatically uniformly continuous.

§17.2 Riemann Integration

Definition 17.5. Suppose $f\colon [a, b] \to \mathbb{R}$ is bounded, and $\mathcal{P}$ is a partition of $[a, b]$ by $x_0 = a < x_1 < \cdots < x_n = b$. Then we define

$$\mathcal{I}(\mathcal{P}) = \sum m_i(x_{i+1} - x_i),$$

where $m_i = \inf_{[x_i, x_{i+1}]} f$; correspondingly we define $M_i = \sup_{[x_i, x_{i+1}]} f$, and

$$\mathcal{U}(\mathcal{P}) = \sum M_i(x_{i+1} - x_i).$$

Then we always have $\mathcal{I}(\mathcal{P}) \leq \mathcal{U}(\mathcal{P})$. More generally, we can take two $\mathcal{P}_1$ and $\mathcal{P}_2$ (which may have a different number of dividing points); last class we proved that we always have

$$\mathcal{I}(\mathcal{P}_1) \leq \mathcal{U}(\mathcal{P}_2)$$

(by taking a refinement of $\mathcal{P}_1$ and $\mathcal{P}_2$). So it makes sense to take the following definitions:

Definition 17.6. We define the upper and lower integrals

$$\underline{\int_a^b} f\,dx = \sup \mathcal{I}(\mathcal{P}) \text{ and } \overline{\int_a^b} f\,dx = \inf \mathcal{U}(\mathcal{P})$$

over all partitions $\mathcal{P}$.

Question 17.7. When is the lower integral of f equal to the upper integral?

Definition 17.8. We say that f is **Riemann integrable** if $\underline{\int_a^b} f\,dx = \overline{\int_a^b} f\,dx$.

We'll now use the result we just proved about uniform continuity to prove the following theorem:

Theorem 17.9

Suppose $f\colon [a, b] \to \mathbb{R}$ is continuous. Then f is Riemann integrable.

Proof. We will use the fact that f is actually uniform continuous — so for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $|x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$.

Now let $\mathcal{P}$ be a partition with separation less than δ — i.e., that $x_{i+1} - x_i < \delta$ for all i. (So the lengths of all our tiny intervals are strictly less than δ — if δ is smaller, then this requires our partition to be finer.)

Then for any two pairs of points in this interval, we have

$$|f(x) - f(y)| < \varepsilon.$$

But on $[x_i, x_{i+1}]$, both the minimum and maximum are always achieved — so there exists x with $f(x) = m_i$, and y with $f(y) = M_i$ (with x and y in the interval $[x_i, x_{i+1}]$). So then we must have

$$M_i - m_i < \varepsilon$$

as well (since both are achieved, and for any two points in the same interval the spread is less than ε).

Now let's fix a partition which is this fine, and consider

$$0 \le \mathcal{U}(\mathcal{P}) - \mathcal{I}(\mathcal{P}) = \sum m_i(x_{i+1} - x_i) - \sum M_i(x_{i+1} - x_i).$$

We can rewrite the right-hand side as

$$\sum (M_i - m_i)(x_{i+1} - x_i).$$

But $M_i - m_i$ is bounded by ε, and $x_{i+1} - x_i$ is nonnegative; this means

$$\mathcal{U}(\mathcal{P}) - \mathcal{I}(\mathcal{P}) < \varepsilon \sum (x_{i+1} - x_i).$$

But if we take our partition and sum the lengths of all the intervals, they divide the whole interval from a to b into subintervals; this means the sum of their lengths is exactly $b - a$. So then we get

$$\mathcal{U}(\mathcal{P}) - \mathcal{I}(\mathcal{P}) < \varepsilon(b - a).$$

But ε can be as small as we want, so $\mathcal{U}(\mathcal{P}) - \mathcal{I}(\mathcal{P})$ can be as small as we want. More precisely, for any $\mathcal{P}$ chosen in this way, we have

$$\mathcal{U}(\mathcal{P}) \le \varepsilon(b - a) + \mathcal{I}(\mathcal{P}).$$

But we have $\mathcal{U}(\mathcal{P}) \ge \overline{\int_a^b} f\,dx$, since the upper integral is the inf of all $\mathcal{U}(\mathcal{P})$; and similarly we have $\mathcal{I}(\mathcal{P}) \le \underline{\int_a^b} f\,dx$; so we get

$$\overline{\int_a^b} f\,dx \le \varepsilon(b - a) + \underline{\int_a^b} f\,dx.$$

But we can take ε to be any positive real, so letting $\varepsilon \to 0$, we get that we must have

$$\overline{\int_a^b} f\,dx \le \underline{\int_a^b} f\,dx.$$

But we must have $\underline{\int_a^b} f\,dx \le \overline{\int_a^b} f\,dx$, since $\mathcal{I}(\mathcal{P}_1) \le \mathcal{U}(\mathcal{P}_2)$ for all partitions (so $\sup \mathcal{I}(\mathcal{P}) \le \inf \mathcal{I}(\mathcal{P})$); this implies they must be equal. $\square$

Initially, we had a graph and wanted to find the area below the graph. Now we see that as long as the function is continuous, it has a well-defined notion of area — so this works in surprising generality.

§17.3 Properties of Integrals

Proposition 17.10

Suppose we have a function $f\colon [a, b] \to \mathbb{R}$, and $c \in \mathbb{R}$ is a constant. Then if f is Riemann integrable, then cf is also Riemann integrable, with

$$\int_a^b cf\,dx = c\int_a^b f\,dx.$$

Proof. If $c = 0$, then the left-hand side is always 0, and it's clear that $\int_a^b 0\,dx = 0$ (since all sup's and inf's are 0). So this case is trivial and can be ignored.

We'll prove the statement when $c > 0$ (we can work out the details for $c < 0$ on our own).

Consider a partition, and look at any interval $[x_i, x_{i+1}]$. Then $m_i = \inf_{[x_i,x_{i+1}]} f$. But we have

$$\inf_{[x_i,x_{i+1}]} cf = c \inf_{[x_i,x_{i+1}]} f$$

(this is why we want to assume $c > 0$ — otherwise inf would become sup and vice versa). Similarly, we have

$$\sup_{[x_i,x_{i+1}]} cf = c \sup_{[x_i,x_{i+1}]} f.$$

Now looking at the original function, we have

$$\mathcal{I}_f(\mathcal{P}) = \sum m_i(x_{i+1} - x_i) \text{ and } \mathcal{U}_f(\mathcal{P}) = \sum M_i(x_{i+1} - x_i).$$

If we do the same for cf, then taking the same partition $\mathcal{P}$, we'll get

$$\mathcal{I}_{cf}(\mathcal{P}) = \sum cm_i(x_{i+1} - x_i),$$

and likewise

$$\mathcal{U}_{cf}(\mathcal{P}) = \sum cM_i(x_{i+1} - x_i).$$

We can factor out the c to get that

$$\mathcal{I}_{cf}(\mathcal{P}) = c\mathcal{I}_f(\mathcal{P}) \text{ and } \mathcal{U}_{cf}(\mathcal{P}) = \mathcal{U}_f(\mathcal{P}).$$

When we take the inf and sup over all partitions $\mathcal{P}$, then we simply get c times the sup and inf for f; this implies

$$\underline{\int_a^b} cf\,dx = c\underline{\int_a^b} f\,dx,$$

and the same is true for the upper integral. Since f is Riemann integrable, we have $\underline{\int_a^b} f\,dx = \overline{\int_a^b} f\,dx$; this implies the corresponding result for cf. □

Similarly, we have the following:

Proposition 17.11

If f and g are Riemann integrable, then so is $f + g$, with

$$\int_a^b (f+g)\,dx = \int_a^b f\,dx + \int_a^b g\,dx.$$

We also have another useful property:

Proposition 17.12

Suppose $f \le g$, and f and g are Riemann integrable. Then $\int_a^b f\,dx \le \int_a^b g\,dx$.

Proof. If we take the same partition and look at an interval $[x_i, x_{i+1}]$, then we have $m_i^f = \inf_{[x_i,x_{i+1}]} f \le \inf_{[x_i,x_{i+1}]} g = m_i^g$, and similarly $M_i^f \le M_i^g$ (since $f \le g$). The first statement implies $\mathcal{I}_f(\mathcal{P}) \le \mathcal{I}_g(\mathcal{P})$, and the second that $\mathcal{U}_f(\mathcal{P}) \le \mathcal{U}_g(\mathcal{P})$. Since f and g are Riemann integrable, their Riemann integrals are $\sup \mathcal{I}_f(\mathcal{P})$ and $\sup \mathcal{I}_g(\mathcal{P})$; this gives the desired result. □

One particular case of this which is quite useful is the following: suppose f is continuous (this is not strictly necessary), so that f is automatically Riemann integrable. Then $|f|$ is also continuous, and $f \leq |f|$ and $-f \leq |f|$. Then we have

$$\int_a^b f\,dx \leq \int_a^b |f|\,dx,$$

while

$$-\int_a^b f\,dx = \int_a^b (-f)\,dx \leq \int_a^b f\,dx.$$

This gives us

$$\int_a^b f\,dx \leq \int_a^b |f|\,dx \text{ and } -\int_a^b f\,dx \leq \int_a^b f\,dx.$$

Written more compactly, this gives us the following:

Proposition 17.13

If f is continuous, we have

$$\left|\int_a^b f\,dx\right| \leq \int_a^b |f|\,dx.$$

This is often useful — if we want to estimate an integral, we can bring in the values into the integration sign.

§17.4 Evaluating Integrals

Example 17.14

Let $f(x) = c$, and find $\int_a^b f\,dx$.

Solution. Then on each smaller interval we have $m_i = M_i = c$, so then

$$\mathcal{I}(\mathcal{P}) = \mathcal{U}(\mathcal{P}) = \sum c(x_{i+1} - c_i) = c(b - a).$$

So we have $\int_a^b f\,dx = c(b - a)$. □

This case is trivial, but evaluating integrals quickly gets complicated.

Example 17.15

Let $f(x) = x$, and find $\int_0^1 f\,dx$.

Solution. Now on each interval $[x_i, x_{i+1}]$, since f is increasing we have $m_i = x_i$ and $M_i = x_{i+1}$. This means

$$\mathcal{I}(\mathcal{P}) = \sum x_i(x_{i+1} - x_i) \text{ and } \mathcal{U}(\mathcal{P}) = \sum x_{i+1}(x_{i+1} - x_i).$$

It's possible to evaluate these, but it's not easy. □

As we can see, we need better tools for evaluating integrals than just using the definition. There are three tools for evaluating integrals, but the basic one is the fundamental theorem of calculus. (The other tools are *substitution* and *integration by parts*, but they all boil down to the fundamental theorem of calculus anyways.)

§17.5 The Fundamental Theorem of Calculus

The fundamental theorem of calculus is the key tool in evaluating integrals.

Theorem 17.16

Suppose $f\colon [a,b] \to \mathbb{R}$ is continuous. Then define the function $F\colon [a,b] \to \mathbb{R}$ as

$$F(x) = \int_a^x f(y)\,dy.$$

Then F is differentiable, and $F'(x) = f(x)$.

(When defining f, we take the entire interval $[a,b]$ and then only concentrate on f on the smaller interval $[a,x]$; if f is continuous, of course its restriction to the smaller interval is also continuous, so this integral is well-defined for all x.)

Proof. We want to consider the difference quotient

$$\frac{F(x_0+h)-F(x_0)}{h}.$$

For simplicity we will assume $h > 0$ (the other case is analogous). Then we have

$$\frac{F(x_0+h)-F(x_0)}{h} = \frac{\int_a^{x_0+h} f(y)\,dy - \int_a^{x_0} f(y)\,dy}{h}.$$

So we want to relate the two integrals in the numerator.

Claim — We have $\int_a^{x_0+h} f(y)\,dy = \int_a^{x_0} f(y)\,dy + \int_{x_0}^{x_0+h} f(y)\,dy$.

We'll first show that this claim implies the desired result, and prove the claim afterwards.

Assuming this claim, we can write

$$F(x_0+h) = F(x_0) + \int_{x_0}^{x_0+h} f(y)\,dy,$$

and returning to the difference quotient, this means

$$\frac{F(x_0+h)-F(x_0)}{h} = \frac{1}{h}\int_{x_0}^{x_0+h} f(y)\,dy.$$

Now we'd like to estimate the right-hand side — we can estimate it from above and below by

$$\min_{[x_0,x_0+h]} f\cdot h \le \int_{x_0}^{x_0+h} f(y)\,dy \le \max_{[x_0,x_0+h]} f\cdot h.$$

(One way to see this is that f is less than the constant function $\max_{[x_0,x_0+h]} f$ on that interval, so its integral is as well.) Dividing by h and plugging into the above equation, we get that

$$\min_{[x_0,x_0+h]} f \le \frac{F(x_0+h)-F(x_0)}{h} \le \max_{[x_0,x_0+h]} f.$$

But now as $h \to 0$, since f is continuous both the max and min go to $f(x_0)$; this means

$$\lim_{h\to 0}\frac{F(x_0+h)-F(x_0)}{h} \to f(x_0).$$

So this means F is differentiable at x_0, with $F'(x_0) = f(x_0)$. (We assumed that $h > 0$, but it's easy to check the other case as well; we only made this assumption so that we could work with intervals in the correct direction.) $\square$

We've now proven the theorem apart from the claim. There's actually a slightly more general claim that's true:

Theorem 17.17

Suppose that $f\colon [a,b] \to \mathbb{R}$ is continuous, and $c \in (a,b)$. Then

$$\int_a^b f(y)\,dy = \int_a^c f(y)\,dy + \int_c^b f(y)\,dy.$$

(We again don't actually need continuity; we just need f to be Riemann integrable.)

Proof. In order to evaluate $\int_a^b f(y)\,dy$, we are considering $\mathcal{I}_{[a,b]}(\mathcal{P})$. But we can make a refinement of this partition, where we take all the dividing points in $\mathcal{P}$ and add in c (if it's already there, then we don't do anything). Let $\mathcal{P}_c$ be this refinement.

We saw earlier that if we take a refinement, then $\mathcal{I}$ increases and $\mathcal{U}$ decreases; so we have

$$\mathcal{I}_{[a,b]}(\mathcal{P}) \le \mathcal{I}_{[a,b]}(\mathcal{P}_c).$$

But we can think of $\mathcal{P}_c$ as a union of two partitions — one of $[a,c]$ and the other of $[c,b]$ — so if we let these two parts be $\mathcal{P}_1$ and $\mathcal{P}_2$, then

$$\mathcal{I}_{[a,b]}(\mathcal{P}_c) = \mathcal{I}_{[a,c]}(\mathcal{P}_1) + \mathcal{I}_{[c,b]}(\mathcal{P}_2).$$

Now taking the sup over all partitions, we get that

$$\underline{\int_a^b} f\,dy \le \underline{\int_a^c} f\,dy + \underline{\int_c^b} f\,dy$$

for any *bounded* (not even necessarily Riemann integrable f).

On the other hand, if we want to evaluate the integrals from a to c and from c to b, then we can take a partition $\mathcal{P}_1$ and $\mathcal{P}_2$ of each, and combine them into a partition $\mathcal{P}$. Then we see

$$\mathcal{I}_{[a,b]}(\mathcal{P}) = \mathcal{I}_{[a,c]}(\mathcal{P}_1) + \mathcal{I}_{[c,b]}(\mathcal{P}_2).$$

As our partitions get finer, both partitions on the right-hand side converge to the integral; this gives the reverse inequality

$$\underline{\int_a^b} f\,dy \ge \underline{\int_a^c} f\,dy + \underline{\int_c^b} f\,dy.$$

Combining the two inequalities gives the desired result. □

The fundamental theorem of calculus is the key way we evaluate integrals.

Claim — If f is continuous and $G' = f$, then

$$\int_a^b f\,dx = G(b) - G(a).$$

Proof. We have two functions

$$F(x) = \int_a^x f(y)\,dy$$

and $G(x)$. But then we have

$$(F-G)'(x) = f(x) - f(x) = 0$$

for all x. Some time ago, we proved that if f has derivative 0, then f is constant; so here $F - G$ is constant, which means

$$(F - G)(b) = (F - G)(a).$$

We have $F(b) = \int_a^b f(y)\, dy$ and $F(a) = \int_a^a f(y)\, dy = 0$ (since we're trying to find the area below a single point, which is 0). Plugging this in, we have $\int_a^b f(y)\, dy - G(b) = 0 - G(a)$, and moving $G(b)$ over to the other side gives the desired result. $\square$

§18 November 17, 2022

§18.1 Calculating Integrals

Given a function $f\colon [a,b] \to \mathbb{R}$, we want to compute the area below its graph, or its *Riemann integral* $\int_a^b f(x)\,dx$. We saw last class that if f is continuous, then this Rieman integral is well-defined.

We've defined this integral by approximating the area from above and below using a partition; we saw that for a continuous function, as the partition becomes finer and finer, the lower and upper bounds converge to each other. So the integral *exists*, but that leaves us with the following question:

Question 18.1. How can we compute the integral?

The key to doing this is the fundamental theorem of calculus, along with the straightforward observation that

$$\int_a^b (\lambda f(x) + g(x))\,dx = \lambda \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

First let's see a few examples of the fundamental theorem of calculus in use.

Example 18.2

Suppose $f(x) = x^\alpha$ with $\alpha > 0$. Then to find $\int_a^b f(x)\,dx$, observe that if we define

$$F(x) = \frac{1}{1+\alpha} \cdot x^{1+\alpha},$$

then we have $F'(x) = f(x)$. Then the fundamental theorem of calculus tells us that

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Example 18.3

Suppose that $f(x) = e^x$. Then to find $\int_a^b f(x)\,dx$,observe that the function $F(x) = e^x$ satisfies $F'(x) = f(x)$ (the derivative of e^x is itself), and so

$$\int_a^b e^x\,dx = F(b) - F(a) = e^b - e^a.$$

There are essentially just two other methods for evaluating integrals: *substitution* and *integration by parts*. But both actually lead back to the fundamental theorem of calculus.

§18.1.1 Substitution

Suppose we have a function $f\colon [a,b] \to [c,d]$, and another function $g\colon [c,d] \to \mathbb{R}$. Then we can consider the composition $g \circ f$ (which is well-defined because all values of f are in the range of g). If f and g are differentiable, then by the chain rule $g \circ f$ is also differentiable, and

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x).$$

This is relevant to integration because if we want to calculate $\int_a^b h(x)\,dx$, and we realize that $h(x)$ can be written as $g'(f(x))f'(x)$ for some functions f and g, then we can define $H = g \circ f$, which gives us $H'(x) = h(x)$ (and therefore lets us integrate h).

Example 18.4

Suppose we want to integrate

$$\int_a^b x \cdot e^{x^2}\, dx.$$

This expression looks complicated (and if we didn't have the x it would be impossible). But we can set $f(x) = x^2$, and $g(y) = e^y$. Then we can observe that $f'(x) = 2x$ and $g'(y) = e^y$, so

$$(g \circ f)'(x) = 2x \cdot e^y = 2x \cdot e^{x^2}.$$

This means the expression we were trying to integrate actually becomes

$$\frac{1}{2} \int_a^b (g \circ f)'(x)\, dx.$$

Then using the fundamental theorem of calculus, this becomes

$$\frac{1}{2} \left((g \circ f)(b) - (g \circ f)(a)\right) = \frac{e^{b^2} - e^{a^2}}{2}.$$

As another way of thinking of this, we'll usually write $y = f(x)$, so in this example we'd write $y = x^2$, and then

$$\frac{dy}{dx} = 2x.$$

People often write this as $dy = 2x\, dx$, and then we see that $x\, dx = \frac{1}{2} \cdot dy$. So we can write our integral as

$$\int_a^b xe^x\, dx = \frac{1}{2} \int_{a^2}^{b^2} e^y\, dy.$$

(As x runs from a to b, x^2 will run from a^2 to b^2.) You can do this computation even without thinking, but it's really coming from the chain rule and fundamental theorem of calculus.

§18.2 Integration by Parts

Substitution is really useful, but it's almost always only useful to a point. Another method that's even more useful is integration by parts.

Suppose we have two functions f and G, and we're trying to calculate

$$\int_a^b f(x) \cdot G(x)\, dx.$$

(Integration by parts often applies to a *product* — a product may be set up perfectly for substitution, but if nothing else works then it often pays to integrate by parts.) The idea is to use the Leibniz formula

$$(FG)' = F'G + FG'.$$

Thinking about F' as f and G' as g, we can write this as

$$(FG)' = fG + Fg.$$

We can now attempt to integrate the Leibniz formula, to get

$$\int_a^b (FG)'\, dx = \int_a^b fG\, dx + \int_a^b Fg\, dx.$$

But applying the fundamental theorem of calculus to the left-hand side, we have

$$F(b)G(b) - F(a)G(a) = \int_a^b fG\, dx + \int_a^b Fg\, dx.$$

Notation 18.5. We often write $[F]_a^b$ to denote $F(b) - F(a)$.

Moving the last term over, we end up with

$$\int_a^b fG\,dx = [FG]_a^b - \int_a^b Fg\,dx.$$

This is surprisingly useful.

Example 18.6

Suppose we want to integrate $\int_a^b xe^x\,dx$. If we didn't have the x, this would be easy (we want to find a function whose derivative is the function we're trying to integrate — called an *antiderivative* — and the antiderivative of e^x is e^x). So we can use integration by parts to get rid of it: we define $G(x) = x$ and $F(x) = e^x$. Then $g(x) = G'(x) = 1$, and $f(x) = F'(x) = e^x$. By integration by parts, our integral then becomes

$$[FG]_a^b - \int_a^b Fg\,dx = [e^x \cdot x]_a^b - \int_a^b e^x\,dx = e^b \cdot b - e^a \cdot a - (e^b - e^a),$$

using the fundamental theorem of calculus directly on the last integral.

This is a typical example of how to use integration by parts. The key point is that we're trying to integrate something complicated — say $\int_a^b fG\,dx$ — and it would be significantly easier if instead of G, we had the derivative of G. The price we have to pay to do this is that instead of f we end up with its antiderivative F; but in cases where integration by parts works, that's not a high price. There's a slight variation of integration by parts — where taking the derivative of G and the antiderivative of f gives us something that isn't simpler, but is instead of the *same* form, so that we can get an equation.

But it's also often used when it doesn't work in just one shot — we may have to do this repeatedly.

Example 18.7

Suppose we want to integrate $\int_a^b x^n e^x\,dx$. Then we want to set $G(x) = x^n$ and $f(x) = e^x$, so letting $F(x) = e^x$ as well we end up with

$$\int_a^b x^n e^x\,dx = [FG]_a^b - \int nx^{n-1} \cdot e^x\,dx.$$

This second integral is not fantastic, but it is a little bit better than what we started with. And we can repeat this process — decreasing the exponent every time — until we finally get down to xe^x, and then to e^x.

So we can sometimes do integration by parts *repeatedly* to get our answer.

§18.3 Improper Integrals

Last time, we saw that if we're trying to integrate over an interval $[a, b]$ and we insert an intermediate point c, then we get

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

The way we thought about integrals, we had an interval which we divided into a partition. But we can also have *improper integrals.*

Suppose $f\colon [0,\infty) \to \mathbb{R}$ is continuous. Then taking $a = 0$ and b to be any positive finite number, the integral $\int_a^b f(x)\,dx$ makes sense. But we can ask the following question:

Question 18.8. Can we make sense of the integral all the way to ∞?

Definition 18.9. In this situation, the **improper integral** of f is defined as

$$\int_0^\infty f(x)\,dx = \lim_{b\to\infty} \int_0^b f(x)\,dx,$$

provided that this limit exists.

Remark 18.10. When we say that the limit exists, the limit has to be finite; we don't generally talk about limits existing if they're infinite. So if this limit is L¡ then we need that given any $\varepsilon > 0$, there exists B such that if $b \ge B$, then

$$\left| L - \int_0^b f(x)\,dx \right| < \varepsilon.$$

This does *not* mean that there *exists* a sequence of $b_i \to \infty$ such that the integral is small; it has to be small for *all* large b.

Example 18.11 (A Non-example)

Consider the function $f(x) = \sin x$, which is 2π-periodic. Then we have $\int_0^{2\pi} \sin x\,dx = 0$ (since the part above the x-axis is identical to the part below). This means

$$\int_0^{2\pi k} \sin x\,dx = 0$$

for any integer k — so there is a sequence wandering off to ∞ where the integrals are 0. But of course, if instead of going to 2π we went to 3π, then we'd get an extra piece at the top — so if we instead looked at

$$\int_0^{2\pi k + \pi} \sin x\,dx,$$

we'd get a positive fixed area (the area of one of the humps). This doesn't have limit 0, so then $\sin x$ doesn't have an improper integral to 0.

Example 18.12

Suppose $\alpha > 1$. Does the improper integral $\int_1^\infty x^{-\alpha}\,dx$ exist?

Solution. To answer this, we want to calculate

$$\int_1^b x^{-\alpha}\,dx,$$

and see whether this has a limit as $b \to \infty$. The nice thing is that it's possible to calculate this — we know that $(x^{1-\alpha})' = (1-\alpha)x^{-\alpha}$, so then

$$\left(\frac{1}{1-\alpha} \cdot x^{1-\alpha}\right)' = x^{-\alpha}.$$

This means we have

$$\int_1^b x^{-\alpha}\,dx = \left[\frac{1}{1-\alpha} x^{1-\alpha}\right]_1^b.$$

When we insert 1, we just get some number; the interesting thing is how this depends on b. We see that when we insert b, we get a constant times $b^{1-\alpha}$; and because $\alpha > 1$, we have b to a negative power. So as $b \to \infty$, this term tends to 0; this means

$$\left[\frac{1}{1-\alpha}x^{1-\alpha}\right] \to -\frac{1}{1-\alpha}.$$

So this improper integral does exist. □

Example 18.13

What happens when $\alpha = 1$ — does the improper integral $\int_1^\infty x^{-1}\,dx$ exist?

Solution. In this case, we want to calculate $\int_1^b 1/x\,dx$. We can observe that $1/x$ is the derivative of log, so

$$\int_1^b \frac{1}{x}\,dx = \log b - \log 1 = \log b.$$

But as $b \to \infty$, $\log b \to 0$, so the improper integral does not exist. □

There are variations of this as well. When we've talked about integrals, we always assumed that we started with a bounded function. But we can now suppose that we have a function that's bounded on $(0, 1]$, but isn't bounded if we extend it to 0 — for example, $1/x$.

Question 18.14. Can we make sense of the integral $\int_0^1 1/x\,dx$?

When we talked about integrals, we always started with a bounded function, and noted that if it's continuous on a compact interval then the Riemann integral exists. But this function doesn't satisfy those conditions, because it's not defined at 0; so this is also an improper integral.

In this situation, we want to find whether

$$\lim_{b\to 0}\int_b^1 \frac{1}{x}\,dx$$

exists.

More generally, this occurs when we have a function that's defined on a compact interval (and continuous), except at one of the endpoints, and we'd like to make sense of the integral. To do this, we take the limit approaching that endpoint. (In this example we have the interval $(0, 1]$, but we could also have an interval $[a, b)$, for instance. We could also have both endpoints missing, but then we'd have to take limits for both endpoints.)

In this example, we have

$$\int_b^1 \frac{1}{x}\,dx = \log 1 - \log b \to \infty$$

as $b \to 0$. So the improper integral does not exist in this case.

Example 18.15

Does the improper integral $\int_0^1 x^{-1/2}\,dx$ exist?

First, why do we have any hope that this could exist? We know that the improper integral of $1/x$ does *not* exist. But the function $1/\sqrt{x}$ still goes to ∞ near 0, but it doesn't go to ∞ as rapidly. So it's *possible* that it could exist.

Solution. For all b, we have

$$\int_b^1 \frac{1}{\sqrt{x}}\,dx = [2\sqrt{x}]_b^1 = 2 - 2\sqrt{b} \to 2$$

as $b \to 0$. So the improper integral does exist (and equals 2). $\square$

§18.4 Arc Length of Curves

Suppose we take some curve in the plane, which comes with a parametrization — for example, telling us where we are at a given time. We can write this curve using the parametrization as $\gamma(t) = (f(t), g(t))$ — so then the curve is a map $\gamma\colon [a, b] \to \mathbb{R}^2$.

Question 18.16. What is the arc length of this curve?

The idea is to approximate the curve by a 'polygon' — we take a little piece of the curve and replace it with a straight line, and we try to make these straight lines tiny. More precisely, we can take a partition $a = t_0 < \cdots < t_n = b$ of our inteerval $[a, b]$. Then on each of the smaller intervals, we replace the curve with a straight line between the corresponding endpoints. We take our partition to be finer and finer (so that the separations go to 0), and we want to see whether this gives us a well-defined length.

Suppose we have a parametrized curve $(f(t), g(t))$ where f and g are differentiable. Then taking these partitions, we get something well-defined, which is equal to

$$\int_a^b \sqrt{f'(t)^2 + g'(t)^2}\,dt.$$

(This argument actually works in greater generality.)

Example 18.17

Suppose that $\gamma(t) = (t, t^2)$, so that $f(t) = t$ and $g(t) = t^2$.

Solution. This means $f'(t) = 1$ and $g'(t) = 2t$. So then the thing we have to integrate is

$$\sqrt{f'(t)^2 + g'(t)^2} = \sqrt{1 + 4t^2}.$$

So in this case, the arc length would be

$$\int_a^b \sqrt{1 + 4t^2}.$$ $\square$

§18.5 Arcsin

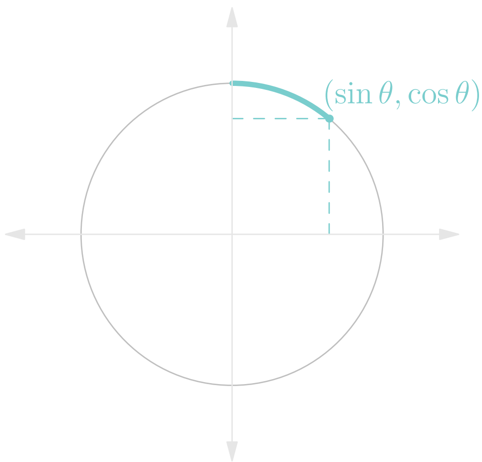
We'll now use this to define the inverse function of $\sin x$. Recall that if we have a function $f\colon A \to B$, then its inverse $f^{-1}\colon B \to A$ is a function with the property that if $f(x) = y$, then $f^{-1}(y) = x$. For this to be defined, we require that f is one-to-one and onto. (If it weren't one-to-one, then the inverse wouldn't be well-defined because there would be y corresponding to multiple values of x, which we wouldn't know which to assign, and if it weren't onto then there'd be y with nothing to assign.)

First we should define sin and cos.

Definition 18.18. For an angle θ, we define sin and cos so that the point on the unit circle at an angle θ from the x-axis is $(\cos\theta, \sin\theta)$.

The angle of the entire circle is 2π, and θ is also the length of the arc on the circle from the x-axis to our point.

Note that if we went an angle θ clockwise from the y-axis, then our point would be $(\sin\theta, \cos\theta)$ (since this is the same picture, just reflected).



Now suppose we're given x, and we want to find θ for which $x = \sin\theta$.

We can parametrize our piece of the circle by $\gamma(t) = (t, \sqrt{1-t^2})$, for $\gamma\colon [0, x] \to S^1$. Then letting $f(t) = t$ and $g(t) = \sqrt{1-t^2}$, we can see that $f'(t) = 1$, while we can calculate $g'(t)$ by the chain rule — letting $h_1(s) = \sqrt{s}$ and $h_2(r) = 1 - r^2$, we see that $h_1' = 1/2\sqrt{s}$ and $h_2' = -2r$, so

$$g'(t) = -\frac{2t}{2\sqrt{1-t^2}} = -\frac{t}{\sqrt{1-t^2}}.$$

Now we see that the arc length of this curve is

$$\int_0^x \sqrt{1 + \frac{t^2}{1-t^2}}\, dt = \int_0^x \sqrt{\frac{1-t^2+t^2}{1-t^2}}\, dt = \int_0^x \frac{1}{\sqrt{1-t^2}}\, dt.$$

But if we have $x = \sin\theta$, then θ is precisely this arclength. So that means the function

$$\arcsin(x) := \int_0^x \frac{1}{\sqrt{1-t^2}}\, dt$$

is precisely the inverse of sin.

This could also have been obtained in other ways, but one thing we should note is that all we used here was the definition of sin and cos — taking the unit circle, if we go out a distance of θ along the unit circle, then the first coordinate is cos and the second is sin. We didn't use any properties of sin and cos — in particular, we didn't use what their derivatives are. And from this, you can then conclude what the derivative of sin is (and you can do the same for cos). So this is another way of proving that the derivative of sin is cos, and the derivative of cos is $-\sin$. (This is a slightly less traditional way; the usual one is using the sum formula.)

§19 November 22, 2022

§19.1 Uniform Convergence

Suppose we have a function $f\colon I \to \mathbb{R}$ for some interval I (which may be open or closed). We have discussed what it means for f to be continuous, and to be *uniformly* continuous.

Now suppose we have a *sequence* of functions $f_n\colon I \to \mathbb{R}$. Then there are different ways of converging.

Definition 19.1. The sequence f_n converges *pointwise* to f if for all $x \in I$, the sequence $f_n(x)$ converges to $f(x)$.

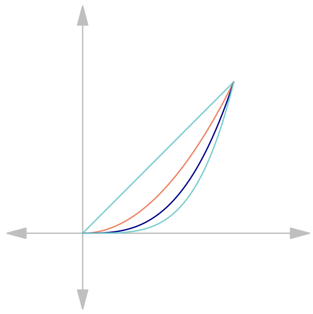
There's a canonical example of this that illustrates it doesn't give very much.

Example 19.2

Take the interval $I = [0, 1]$, with $f_n(x) = x^n$, and let

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

Then our functions look like the following:



Our functions become steeper and steeper, so if $0 \le x < 1$ is fixed, then $x^n \to 1$. But on the other hand, if $x = 1$ then of course $x^n = 1$. So this means $f_n(x) \to f(x)$ for each fixed x, and so $f_n \to f$ pointwise.

What we don't like about this picture is that we have a sequence of continuous functions which converge to something that is *not* continuous.

Question 19.3. Is there a condition that guarantees that if f_n are all continuous, so is their limit?

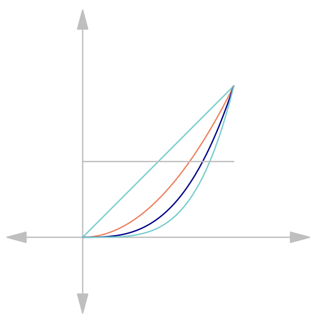
The condition that accomplishes this is *uniform* convergence. In pointwise convergence, we're fixing x. But if our functions converge at the same rate, independently of f, then we have uniform convergence.

Definition 19.4. A sequence of functions f_n converges to f *uniformly* if for all $\varepsilon > 0$, there exists N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \ge N$ and *all* x.

The point is that given ε, our value of N works for *all* x (unlike in pointwise convergence, where N can depend on x) — so N is uniform in x.

Example 19.5

Do our functions $f_n(x) = x^n$ on $[0, 1]$ converge to f uniformly?



Given n, there exists $x_n < 1$ such that $f_n(x_n) = 1/2$ — we can simply take $x_n = (1/2)^{1/n}$. But then setting $x = x_n$, we have

$$|f_n(x) - f(x)| = \left|\frac{1}{2} - 0\right| = \frac{1}{2}.$$

This shows our sequence does not converge to f uniformly — if we take $\varepsilon < 1/2$, then this expression is supposed to be strictly less than $1/2$ for n large, but at this value of x_n, it's $1/2$. So f_n does *not* converge to f uniformly (although it does converge pointwise).

Theorem 19.6

If f_n and f are functions $I \to \mathbb{R}$ and $f_n \to f$ uniformly, then if the functions f_n are all continuous, f is continuous as well.

Proof. We want to show that f is continuous, which means we want to make $|f(x) - f(x_0)|$ small. We'd like to relate this expression to the corresponding one for f_n (since that's what we have information about), so we can write it as

$$f(x) - f(x_0) = f(x) - f_n(x) + f_n(x) - f_n(x_0) + f_n(x_0) - f(x_0).$$

Then by the triangle inequality we have

$$|f(x) - f(x_0)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)|.$$

We want to show that the left-hand side is small. Because the convergence is uniform, we can make $|f(x) - f_n(x)|$ and $|f_n(x_0) - f(x_0)|$ small by making N sufficiently large; then we can use the fact that f_n is continuous to make $|f_n(x) - f_n(x_0)|$ small.

More explicitly, given any $\varepsilon > 0$, since $f_n \to f$ uniformly, there exists N such that if $n \ge N$ then

$$|f_n(y) - f(y)| < \frac{\varepsilon}{3}$$

for all $y \in I$. Now fixing $n = N$, this means the expression we wanted to make small is bounded by

$$|f(x) - f(x_0)| \le |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| < \frac{\varepsilon}{3} + |f_n(x) - f_n(x_0)| + \frac{\varepsilon}{3}.$$

But since f_N is continuous, there exists $\delta > 0$ such that if $|x - x_0| < \delta$, then $|f_N(x) - f_N(x_0)| < \varepsilon/3$. So then we get $|f(x) - f(x_0)| < \varepsilon$ (when $|x - x_0| < \delta$), as desired. □

So if our convergence is uniform, and the functions are all continuous, so is their limit. (Note that this has nothing to do with whether the interval is open or closed.)

§19.2 Weirstrass M-Test

Suppose we have a sequence $f_n \colon I \to \mathbb{R}$ of functions. Then we can consider the *sum*

$$S_n = \sum_{n=0}^{N} f_n$$

(which is also a function).

Theorem 19.7

Suppose we have a sequence $f_n : I \to \mathbb{R}$, such that $|f_n| \le M_n$ (where the M_n are constants — more explicitly, $\sup_I |f_n(x)| \le M_n$), and suppose that $\sum_{n=0}^{\infty} M_n$ converges. Then the sequence S_n converges uniformly to some function S.

In particular, if the functions f_n are all continuous, then so are the functions S_N; so if the convergence of S_N is uniform, then the function S is also continuous.

Example 19.8

Take the sequence $f_n(x) = \frac{x^n}{n!}$ on a finite interval $[-\ell, \ell]$, and consider $\sum_{n=0}^{N} \frac{x^n}{n!}$. Then we have

$$|f_n(x)| \le \frac{|x^n|}{n!} = \frac{|x|^n}{n!} \le \frac{\ell^n}{n!},$$

and we know that $\sum_{n=0}^{\infty} \frac{\ell^n}{n!}$ converges (by the ratio or root test).

We know that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, and this tells us that on any interval $[-\ell, \ell]$, the convergence $\sum_{n=0}^{N} \frac{x^n}{n!} \to e^x$ is uniform. So this tells us that e^x is also continuous.

Before we prove this, we'll make a simple observation.

Claim — Uniform convergence implies pointwise convergence.

Proof. Suppose $f_n \to f$ uniformly. Then for all $\varepsilon > 0$, for all sufficiently large n we have $|f_n(x) - f(x)| < \varepsilon$. In particular, if we *fix* any x_0 then $|f_n(x_0) - f(x_0)| < \varepsilon$ for all large n, so $f_n(x_0) \to f(x_0)$. □

Remark 19.9. As seen earlier, the converse is not true — the sequence $f_n(x) = x^n$ converges to our function f pointwise but not uniformly.

This means uniform convergence is strictly stronger than pointwise convergence.

Proof of Weirstrass M-Test. The first step is to find the function S that we want to show that our sequence converges to. Fixing x_0, we have that $|f_n(x_0)| \le M_n$, and we know that $\sum M_n$ is finite. So $S_N(x_0) = \sum_{n=0}^{N} f_n(x_0)$ is a Cauchy sequence — we have

$$|S_{N_1}(x_0) - S_{N_2}(x_0)| = \left| \sum_{n=N_1+1}^{N_2} f_n(x_0) \right| \le \sum_{n=N_1+1}^{N_2} |f_n(x_0)| \le \sum_{n=N_1}^{N_2} M_n,$$

and since $\sum_{n=0}^{\infty} M_n$ is finite, if N_1 and N_2 are sufficiently large then this sum can be made arbitrarily small. In particular, this means the sequence $S_N(x_0)$ converges to some $S(x_0)$; and therefore our functions S_N converge to some S pointwise (defining $S(x)$ to be the limit of $S(x_0)$).

So we've now found the function S that we want to show convergence to (using the fact that if $f_n \to S$ uniformly, then S must be the pointwise limit of f_n as well). Now we want to prove that

$$|S(x) - S_n(x)|$$

is small. To do so, we have

$$|S_{N_1}(x) - S_{N_2}(x)| = \left| \sum_{n=N_1+1}^{N_2} f_n(x) \right| \le \sum_{n=N_1+1}^{N_2} |f_n(x)| \le \sum_{n=N_1+1}^{N_2} M_n$$

for *all* x. Now fixing x and letting $N_2 \to \infty$, we know $S_{N_2}(x) \to S(x)$; this means that

$$|S_{N_1}(x) - S(x)| \leq \sum_{n=N_1+1}^{\infty} M_n$$

for all x. But since the series $\sum_{n=0}^{\infty} M_n$ converges, we again know that if we start sufficiently far out, we can make $\sum_{n=N_1}^{\infty} M_n$ as small as we want — so provided that $N_1 \geq N$ (where N only depends on ε) we can guarantee that $\sum_{n=N_1}^{\infty} M_n < \varepsilon$.

This proves that the convergence is uniform. □

Here the main idea was to first find the limiting function, and then use the fact that the inequalities we used when finding the limiting function were actually uniform in x.

§19.3 Power Series

Theorem 19.10

Suppose that we have a sequence of *differentiable* functions $F_n \colon [a, b] \to \mathbb{R}$ and $F \colon [a, b] \to \mathbb{R}$, such that $F_n' = f_n$ and $F' = f$. Also suppose $f_n \to f$ uniformly, and f_n are continuous. Assume also that there exists $x_0 \in I$ such that $F_n(x_0) \to F(x_0)$. Then $F_n \to F$ uniformly.

Proof. First since $f_n \to f$ uniformly, and f_n are continuous, then f is also continuous. By the fundamental theorem of calculus, we have

$$F_n(x) - F_n(x_0) = \int_{x_0}^{x} f_n(y)\, dy,$$

and likewise

$$F(x) - F(x_0) = \int_{x_0}^{x} f(y)\, dy.$$

We can then write

$$|F_n(x) - F(x)| = \left| \int_{x_0}^{x} f_n(y)\, dy - F(x_0) - \int_{x_0}^{x} f(y)\, dy + F(x_0) \right|.$$

We can now collect terms, to get

$$|F_n(x) - F(x)| = \left| F_n(x_0) - F(x_0) + \int_{x_0}^{x} f_n(y)\, dy - \int_{x_0}^{x} f(y)\, dy \right|.$$

By the triangle inequality, then

$$|F_n(x) - F(x)| \leq |F_n(x_0) - F(x_0)| + \left| \int_{x_0}^{x} f(y)\, dy - \int_{x_0}^{x} f_n(y)\, dy \right|.$$

For the first term, we know that $F_n(x_0) \to F(x_0)$, so we can make this term as small as we want by taking n large. Meanwhile for the second term, we can rewrite it as

$$\left| \int_{x_0}^{x} (f(y) - f_n(y))\, dy \right|.$$

We know that we can move the absolute value into the integral to get a bound from above, so

$$\left| \int_{x_0}^{x} (f(y) - f_n(y))\, dy \right| \leq \int_{x_0}^{x} |f(y) - f_n(y)|\, dy.$$

We also know $|f(y) - f_n(y)| \le \sup_{z\in[a,b]} |f - f_n|$, which is a constant; and our sub-interval $[x_0, x]$ can only be smaller than $[a, b]$, so then our second term is bounded above by

$$(b - a) \sup_{[a,b]} |f_n - f| .$$

This gives us

$$|F_n(x) - F(x)| \le |F_n(x_0) - F(x_0)| + (b - a) \sup_{[a,b]} |f_n - f| .$$

But now since $F_n(x_0) \to F(x_0)$, there exists N so that if $n \ge N$, then $|F_n(x_0) - F(x_0)| < \varepsilon/2$, so we can make the first term as small as we want. Meanwhile, for the second term, we know f_n converges *uniformly* to f. This means we can make $|f_n - f|$ as small as we want by making n large enough, for *all* x — so if $n \ge N'$, then we can guarantee $|f_n - f| < \varepsilon/2(b - a)$ for all x, which means

$$\sup |f_n - f| < \frac{\varepsilon}{2(b - a)}.$$

This means as long as $n \ge \max\{N, N'\}$ we have $|F_n(x) - F(x)| < \varepsilon$. $\square$

This is very useful because it allows us to differentiate power series.

Suppose we have a power series $\sum a_n x^n$, and we know for which x this power series is convergent — i.e., the radius of convergence. We can find this radius via the root or ratio test — using the root test, we know $\sum a_n x^n$ converges when

$$\limsup |a_n x^n|^{1/n} = \limsup |a_n|^{1/n} \cdot |x| < 1.$$

If $|x|$ is fixed, then this is the case exactly when

$$|x| < \frac{1}{\limsup_{n\to\infty} |a_n|^{1/n}},$$

which we call the *radius of convergence*.

Student Question. *What is the* $\limsup$*?*

Answer. In most cases, it's just the same as the limit — if $\lim_{n\to\infty} |a_n|^{1/n}$ exists, then the radius of convergence is the reciprocal of this limit. The $\limsup$ is a generalization of the limit, which we will not discuss right now; we may return to it in another lecture.

Theorem 19.11

The power series $\sum_{n=0}^{\infty} a_n x^n$ is *continuous* inside its radius of convergence. Moreover, it is differentiable (and in fact k times differentiable for all k), and its derivative is the power series $\sum_{n=1}^{\infty} n a_n x^{n-1}$.

In other words, we want to show that as long as $|x| < R$, the power series is continuous, differentiable, and the derivative is exactly what we'd expect it to be.

§20 November 29, 2022

§20.1 Limsup and Liminf

Suppose a_n is a sequence of real numbers, and let $b_n = \sup\{a_i \mid i \ge n\}$ (which may be infinite). As n becomes larger, the set we're taking the sup of grows smaller; this means $b_{n+1} \le b_n$ for all n. So the sequence b_n is monotone nonincreasing. So one of three things can happen:

(1) $b_n \to -\infty$;

(2) $b_n = \infty$ for all n;

(3) b_n convertes to some finite limit ℓ.

Definition 20.1. For a sequence a_n, with $b_n = \sup\{a_i \mid i \geq n\}$,

$$\limsup_{n\to\infty} a_n := \lim_{n\to\infty} b_n.$$

(The $\limsup$ is allowed to be $\pm\infty$.)

$\liminf$ is defined in the same way, with the obvious changes: if a_n is a sequence in $\mathbb{R}$, we consider the sequence $c_n = \inf\{a_i \mid i \geq n\}$. As n grows larger, the set being considered again grows smaller; now this means $c_{n+1} \geq c_n$ (since we're taking the inf over a smaller set). So c_{n+1} is now monotone increasing, and one of three things can happen:

(1) $c_n \to \infty$;

(2) $c_n = -\infty$ for all n;

(3) $c_n \to \ell$ for some finite ℓ.

Definition 20.2. $\liminf_{n\to\infty} a_n$ is defined as $\lim_{n\to\infty} c_n$.

Again $\liminf a_n$ is either $\pm\infty$ or ℓ.

Example 20.3

Consider the series $\sum d_n$. Then $\limsup_{n\to\infty} |d_n|^{1/n}$ is either ∞ or a finite number ℓ.

If $\limsup |d_n|^{1/n} = \ell < 1$, then $\sum d_n$ is absolutely convergent.

This is simply the root test in full generality (when we proved the root test we assumed that $|d_n|^{1/n}$ had a *limit*, but actually as long as its $\limsup$ (which is always defined) is less than 1, it still works).

Proof. First suppose that $\limsup a_n = \ell < 1$ for any sequence a_n (we'll eventually take $a_n = |d_n|^{1/n}$). Define the sequence $b_n = \sup\{a_i \mid i \geq n\}$, so that $b_n \to \ell$. This means if we take ℓ_0 with $\ell < \ell_0 < 1$, then there exists N such that if $n \geq N$, then $b_n < \ell_0$ (since $b_n \to \ell$, and $\ell_0 > \ell$).

But $b_n = \sup\{a_i \mid i \geq n\}$, which means b_n is an upper bound on a_n (since it's an upper bound on the entire set). So then $a_i \leq \ell_0$ for all $n \geq N$ as well.

Now returning to the root test, take $a_n = |d_n|^{1/n}$. We assumed that $\limsup_{n\to\infty} |d_n|^{1/n} = \ell < 1$, so this means for every $\ell < \ell_0 < 1$, there exists N such that $|d_n|^{1/n} < \ell_0$ for all $n \geq N$. But this means $|d_n| \leq \ell_0^n$.

But now we can compare the series $\sum |d_n|$ and $\sum \ell_0^n$. For n sufficiently large, we know that $|d_n| < \ell_0^n$, and $\ell_0 < 1$. So the series on the right converges, and by the comparision test (since only the tail matters, and the right-hand side ends up bigger than the left-hand side from some point on), the series on the left must converge as well. $\square$

Example 20.4

The same is true for the ratio test (we can replace lim with lim sup) — given a series $\sum d_n$, if

$$\limsup_{n\to\infty} \frac{|d_{n+1}|}{|d_n|} = \ell < 1,$$

then $\sum d_n$ is absolutely convergent.

Proof. Let $a_n = |d_{n+1}/d_n|$ (note that $a_n \geq 0$). We again have $\limsup_{n\to\infty} a_n = \ell < 1$, so again considering the sequence $b_n = \sup\{a_i \mid i \geq n\}$, this means $b_n \to \ell$. So if we take $\ell < \ell_0 < 1$, then there exists N so that if $n \geq N$, then $b_n < \ell_0$. But since $b_n = \sup\{a_i \mid i \geq n\} \geq a_n$, we have that $a_n \leq b_n < \ell_0$ for all $n \geq N$.

Now since $a_n = |d_{n+1}/d_n| < \ell_0$, we have

$$|d_{n+1}| < |d_n| \cdot \ell_0$$

as long as $n \geq N$. So first applying this to N, we have

$$|d_{N+1}| < \ell_0 |d_N|.$$

Then applying it to $N+1$, we get

$$|d_{N+2}| < \ell_0 |d_{N+1}| < \ell_0^2 |d_N|.$$

We can keep doing this to get that

$$|d_{N+m}| < \ell_0^m |d_N|$$

for all $m \geq 1$.

Now we can take our original series $\sum |d_{N+m}|$, and compare it to the series $\sum |d_N| \ell_0^m$. (We can ignore the first few terms in our original series, since the convergence doesn't depend on them.) Since the series on the right converges absolutely, the series on the left does as well. $\square$

Remark 20.5. The advantage to lim sup is that it's always well-defined — when we originally formulated the tests we assumed the relevant sequences converged, but here we don't need to assume that.

§20.2 Uniform Convergence

Question 20.6. Given a sequence of functions, when does taking the limit preserve certain nice properties? In particular, if f_n are continuous and converge to f, when can we say f is also continuous?

Suppose that $f_n : [a,b] \to \mathbb{R}$. Then there's two types of convergence that are usually used:

Definition 20.7. The sequence f_n **converges pointwise** to f if for every *fixed* $x_0 \in [a,b]$, the sequence $f_n(x_0)$ of real numbers converges to $f(x_0)$.

In other words, pointwise convergence means that given $x_0 \in [a,b]$ and $\varepsilon > 0$, there exists N such that for all $n \geq N$, $|f_n(x_0) - f(x_0)| < \varepsilon$. So not only is ε given, but x_0 is also given.

In contrast, for *uniform* convergence, we're not fixing x_0 — the only thing that's given is ε, and we need to find N which works for *all* x.

Definition 20.8. The sequence f_n **converges uniformly** to f if for every $\varepsilon > 0$, there exists N such that for *all* $x \in [a,b]$, $|f_n(x) - f(x)| < \varepsilon$.

So the key difference is that here the *same* value of N works for all x.

Remark 20.9. Another way of thinking of the condition in the definition of uniform convergence, that $|f_n(x) - f(x)| < \varepsilon$ for all x, is that

$$\sup |f_n(x) - f(x)| \leq \varepsilon$$

for all $n \geq N$ (since if $|f_n(x) - (x)| \leq \varepsilon$ for all x, the same is true for the sup). But if f and g are continuous functions on $[a, b]$ (the set of such continuous functions is denoted $\mathcal{C}([a, b])$), then there's a natural distance between f and g, given by

$$d(f, g) = \sup_{x \in [a,b]} \{|f(x) - g(x)|\}.$$

Then the statement in uniform convergence is that $d(f_n, f) \leq \varepsilon$. So another way of thinking of uniform convergence is that it's convergence in the metric space $\mathcal{C}([a, b])$ (with this metric).

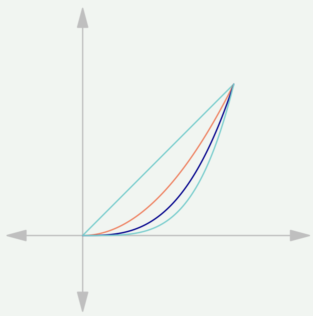
Last time, we proved the following theorem:

Theorem 20.10

If $f_n \to f$ uniformly and f_n is continuous for all n, then f is continuous.

Example 20.11 (A Non-example of Uniform Convergence)

Consider the functions $f_n \colon [0, 1] \to \mathbb{R}$ defined as $f_n(x) = x^n$.



Then the functions f_n converge *pointwise* to the function

$$f(x) = \begin{cases} 0 & x < 1 \\ 1 & x = 1. \end{cases}$$

This function is of course not continuous; that's because the convergence is not uniform.

Student Question. *Why is the convergence not uniform?*

Answer. We can show that for all n, there exists $0 < x_n < 1$ such that $f_n(x_n) = 1/2$ — our functions f_n begin at 0 and end at 1, so there must be some point where they hit $1/2$. But now $|f_n(x_n) - f(x_n)| = |f_n(x_n)| = 1/2$ (since $f(x_n) = 0$). If the convergence were uniform, then given any ε, there would exist N so that if $n \geq N$ then for all x, we'd have $|f_n(x) - f(x)| < \varepsilon$. But taking $\varepsilon = 1/2$, then there'd exist N such that $|f_n(x) - f(x)| < 1/2$ for all $n \geq N$ and for *all* x. In particular, this would have to hold for x_n, which is a contradiction.

This is a good example to keep in mind for why we need a stronger convergence than pointwise if we want the limits of our functions to have nice properties.

§20.3 Weirstrass M-Test

Last time we proved the Weirstrass M-Test, which is a test that implies uniform convergence of an infinite *sum* of sequences $\sum f_n$.

Theorem 20.12

Suppose that f_n is a sequence of functions such that for each n there exists M_n such that $|f_n| \le M_n$, and $\sum M_n < \infty$. Then the sums $\sum_{n=0}^N f_n$ uniformly converge to $\sum_{n=0}^\infty f_n$.

This can be used to prove uniform convergence of power series.

Example 20.13

Consider the power series $E(x) = \sum_{n=0}^\infty \frac{x^n}{n!}$ on some compact interval $[-\ell, \ell]$.

Suppose we look at the *finite* sums $E_i(x) = \sum_{i=0}^n \frac{x_i}{i!}$. Now let $M_i = \frac{\ell^i}{i!}$. Then we can see that

$$\left|\frac{x^i}{i!}\right| \le \frac{\ell^i}{i!}$$

for all i. But the series $\sum_{i=0}^\infty \frac{\ell^i}{i!}$ is convergent. Meanwhile, each individual f_n is continuous. So then E_n converges to E on $[-\ell, \ell]$ uniformly, and since the E_n are continuous (they are polynomials), E must also be continuous.

§20.4 Integration and Differentiation

Theorem 20.14

Suppose $f_n, f\colon [a, b] \to \mathbb{R}$ with $f_n \to f$. Then $\int_a^b f_n\,dx \to \int_a^b f\,dx$.

Proof. We want to show that $\left|\int_a^b f_n\,dx - \int_a^b f\,dx\right|$ becomes small as n becomes large. We first have

$$\left|\int_a^b f_n\,dx - \int_a^b f\,dx\right| = \left|\int_a^b (f_n - f)\,dx\right| \le \int_a^b |f_n - f|\,dx.$$

Now we can use uniform convergence — given $\varepsilon > 0$, we can find N such that for all $n \ge N$, we have

$$|f_n(x) - f(x)| < \frac{\varepsilon}{b-a}$$

for all x. But now this means

$$\int_a^b |f_n - f|\,dx \le \int_a^b \frac{\varepsilon}{b-a}\,dx = \varepsilon,$$

as desired. □

We can do a little bit better than this:

Theorem 20.15

Suppose $f_n, f\colon [a, b] \to \mathbb{R}$ with $f_n \to f$. Then the sequence $\int_a^x f_n(t)\,dt$ converges to $\int_a^x f(t)\,dt$ uniformly.

Proof. We can use the same reasoning to get that for all x,

$$\left|\int_a^x f_n(t)\,dt - \int_a^x f(t)\,dt\right| \le \int_a^x |f_n(t) - f(t)|\,dt < \frac{\varepsilon}{b-a}(x-a) \le \frac{\varepsilon}{b-a}\cdot(b-a) = \varepsilon.$$

$\square$

This allows us to do the following: suppose $f_n\colon [a,b] \to \mathbb{R}$, and $f_n(a) \to A$. Suppose also that the functions f_n are differentiable, and $f_n' \to g$ uniformly on $[a,b]$. By the fundamental calculus we have

$$f_n(x) = \int_a^x f_n'(y)\,dy + f_n(a).$$

But by the above theorem, the first term converges to $\int_a^x g(y)\,dy$, and the second term is just a number converging to A. So then $f_n \to \int_a^x g(y)\,dy + A$ uniformly.

We can state this as a theorem:

Theorem 20.16

If $f_n\colon [a,b] \to \mathbb{R}$ are functions such that:

1. $f_n(a) \to A$;
2. f_n is differentiable, and $f_n' \to g$ uniformly.

Then f_n converges to some function f uniformly, and $f' = g$.

Proof. Set $f(x) = \int_a^x g(y)\,dy + A$. Then by the fundamental theorem of calculus $f' = g$, and we saw that $f_n \to f$ uniformly. $\square$

Example 20.17

Consider the power series $\sum \frac{x^n}{n!}$. If $h_n(x) = \frac{x^n}{n!}$, then h_n is differentiable with $h_n' = \frac{x^{n-1}}{(n-1)!}$. If we then define $f_n(x) = \sum_{i=0}^n \frac{x^i}{i!}$, we have $f_n'(x) = f_{n-1}$.

Previously, by the Weirstrass M-Test, we saw that $f_n \to \exp$ uniformly on every compact interval $[-\ell, \ell]$. But then we also have $f_n' = f_{n-1} \to \exp$ uniformly (since this is the same sequence, shifted by 1).

But this gives us that exp is differentiable, and its derivative is exp — this is because here $f(x) = \int_0^x \exp(y)\,dy + 1$ (since $f_n(0) = 1$ for all n). We proved that then f_n converges to this function. But they also converge to exp, so these must be the same; so then

$$\exp(x) = \int_0^x \exp(y)\,dy + 1.$$

Taking the derivative of both sides gives the desired result.

We can do this for all power series. Suppose we have a power series $\sum_{n=0}^\infty a_n x^n$. Previously, we've seen the *radius of convergence*, which we can find using the root or ratio test — we can consider $(|a_n x^n|)^{1/n} = |a_n|^{1/n}|x| < 1$. This series is then convergent if $\limsup_{n\to\infty} |a_n|^{1/n}|x_n| < 1$, or equivalently if

$$|x| < \frac{1}{\limsup_{n\to\infty} |a_n|^{1/n}}.$$

The expression on the right-hand side is defined as the **radius of convergence**.

Now if we consider a compact interval *strictly* contained in $(-R, R)$, we know that our series converges, and in fact, the convergence is *uniform* in x.

Given a power series $\sum a_n x^n$, we could also formally form another power series where we differentiate term-by-term — giving us the new power series $\sum n a_n x^{n-1}$. If we want to, we can shift indices to write this as $\sum (m+1) a_{m+1} x^m$.

Claim — The radius of convergence of this second series is also R.

Proof. We want to compute

$$\limsup_{m\to\infty}((m+1)\,|a_{m+1}|\,|x^m|)^{1/m} = |x| \cdot \limsup_{m\to\infty}(m+1)^{1/m}\,|a_{m+1}|^{1/m}.$$

But $(m+1)^{1/m} \to 1$ as $m \to \infty$, so this is the same as

$$|x| \cdot \limsup_{m\to\infty} |a_{m+1}|^{1/m}.$$

And it also doesn't matter whether we have m or $m+1$, so this is the same as

$$|x| \cdot \limsup_{m\to\infty} |a_m|^{1/m}.$$

□

So if we start with a power series and formally take its derivative, then they have the same radius of convergence. Now we can apply what we just did — suppose we are given a power series $\sum a_n x^n$. Then we can find its radius of convergence

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{1/n}}.$$

Now suppose $\ell < R$, and we want to consider our power series on the interval $[-\ell, \ell]$. Then $\sum_{n=0}^N a_n x^n$ converges uniformly to $\sum_{n=0}^\infty a_n x^n$. But we also saw that $\sum n a_n x^{n-1}$ has the same radius of convergence, so the sums $\sum_{n=0}^N n a_n x^{n-1}$ also converge uniformly to $\sum_{n=0}^\infty n a_n x^{n-1}$.

First, since our finite sums are polynomials, they are continuous. So this lets us conclude that $\sum_{n=0}^\infty a_n x^n$ is differentiable, and its derivative is $\sum_{n=0}^\infty n a_n x^{n-1}$.

So this means inside the radius of convergence, it's easy to compute the derivative of a power series — we can simply take the derivative term-by-term.

§21 December 1, 2022

§21.1 Review of lim sup and lim inf

Last time, we defined lim sup and lim inf. To define lim sup, we consider the sequence $b_n = \sup\{a_i \mid i \geq n\}$. Then b_n is decreasing, so we can define

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} b_n$$

(the limit must exist, if we allow it to be $\pm\infty$). Likewise, to define lim inf, we consider the sequence $c_n = \{a_i \mid i \geq n\}$. This is an increasing sequence, so we can similarly define

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} c_n.$$

Theorem 21.1 (Ratio Test)
Consider a series $\sum a_n$, where $a_n \neq 0$ for all n.

(1) If we have

$$\limsup_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = \ell < 1,$$

then $\sum a_n$ is absolutely convergent.

(2) If we have

$$\liminf_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = m > 1$$

(or if the lim inf is ∞) then $\sum a_n$ is divergent.

Last time, we proved the first statement, by comparing $\sum a_n$ to a geometric series and using the comparison test. We'll now prove the second.

Proof. Consider the increasing sequence

$$b_n = \inf\left\{\frac{|a_{i+1}|}{|a_i|} \mid i \geq n\right\},$$

so we are given that $\lim_{n\to\infty} b_n = m > 1$. This means if we fix any m_0 with $1 < m_0 < m$, there exists N such that if $n \geq N$, then $b_n \geq m_0$.

Student Question. *Why are we guaranteed that such a m_0 exists?*

Answer. First, we know that the sequence b_n has a (possibly infinite) limit, since the sequence is increasing. If we then choose m_0 between 1 and its limit, since b_n must eventually get arbitrarily close to its limit, it must eventually cross m_0.

But then

$$\inf\left\{\frac{|a_{i+1}|}{|a_i|} \mid i \geq n\right\} \geq m_0,$$

and since the infimum is a lower bound, we have

$$\frac{|a_{n+1}|}{|a_n|} \geq m_0$$

for all $n \geq N$.

But now this means

$$|a_{n+1}| \geq m_0 |a_n|$$

for all $n \geq N$, for some $m_0 > 1$. Now we can iterate this inequality to get that

$$|a_{N+k}| \geq m_0^k |a_N|$$

for all $k \geq 1$. But this means $|a_{N+k}| \to \infty$ as $m \to \infty$, so the series certainly cannot converge (it fails the basic test that if a series is convergent, its elements must go to 0). $\square$

Remark 21.2. Note that here we obtained conclusions when $\limsup < 1$ or when $\liminf > 1$. But there are other possibilities, and we can't conclude anything in those cases. In particular, even if we know

$$\limsup_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} > 1,$$

it's possible that the series is not divergent — imagine our sequence occasionally jumps upwards, but usually decreases quickly. Then this doesn't tell us anything, because we can't iterate (the point in our proofs above was that we could 'snowball' by chaining inequalities, and we can't do that here).

§21.2 Convergence of Power Series

Question 21.3. Suppose we have a power series $\sum a_n x^n$. When is it convergent?

We can answer this with the ratio test — taking the ratio of consecutive terms gives us

$$\frac{|a_{n+1}x^{n+1}|}{|a_n x^n|} = \frac{|a_{n+1}|}{|a_n|} \cdot |x|.$$

When we take the $\limsup$, the constant $|x|$ factors out, and we have

$$\limsup_{n\to\infty} \frac{|a_{n+1}x^{n+1}|}{|a_n x^n|} = |x| \limsup_{n\to\infty} \frac{|a_{n+1}|}{|a_n|}.$$

If this quantity is less than 1, then we have absolute convergence.

Definition 21.4. The **radius of convergence** R of a power series is defined as

$$R = \frac{1}{\limsup_{n\to\infty} \frac{|a_{n+1}|}{|a_n|}}.$$

We can see that if $|x| < R$, then the series is absolutely convergent. In fact, we can say more:

Proposition 21.5

Suppose that $\ell < R$. Then on the interval $[-\ell, \ell]$, $\sum a_n x^n$ converges *uniformly.*

More explicitly, we can define the functions $f_N(x) = \sum_{n=0}^N a_n x^n$, and $f(x) = \sum_{n=0}^\infty a_n x^n$. Then this states that $f_n \to f$ uniformly.

Proof. For every $x \in [-\ell, \ell]$, we have

$$\limsup_{n\to\infty} \frac{|a_{n+1}x^{n+1}|}{|a_n x^n|} \le |x| \cdot \frac{1}{R} \le \frac{\ell}{R} < 1.$$

(This implies the desired result by the Weirstrass M-test, for example.) □

In particular, since each f_N is continuous (it is a polynomial), so is f.

§21.3 Differentiation of Power Series

Suppose we consider the power series obtained by formally differentiating our power series — so

$$\sum a_n x^n \rightsquigarrow \sum n a_n x^{n-1}.$$

Proposition 21.6

The radius of convergence for $\sum n a_n x^{n-1}$ is also R.

Proof. We have

$$\limsup_{n\to\infty} \left| \frac{(n+1)a_{n+1}x^n}{n a_n x^{n-1}} \right| = |x| \cdot \limsup_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} \cdot \frac{n+1}{n}.$$

But $\frac{n+1}{n} \to 1$, so then this simply equals

$$|x| \limsup_{n\to\infty} \frac{|a_{n+1}|}{a_n} = \frac{x}{R},$$

where R is the radius of convergence of the original power series. This means the radius of convergence of the new power series is R as well. □

Now suppose we are given a power series $\sum_{n=0}^{\infty} a_n x^n$ with radius of convergence R; we know this power series is continuous inside $(-R, R)$. Meanwhile, $\sum_{n=0}^{\infty} n a_n x^{n-1}$ is *also* continuous inside $(-R, R)$.

So if we think about the functions $f_N = \sum_{n=0}^{N} a_n x^n$, which have derivatives $f_N' = \sum_{n=1}^{N} n a_{n-1} x^{n-1}$. For every $\ell < R$, we know f_N' converge to $\sum_{n=1}^{\infty} n a_n x^{n-1}$ *uniformly* on $[-\ell, \ell]$. Meanwhile, f_N converges uniformly to $\sum_{n=0}^{\infty} a_n x^n$.

Previously, we proved the following result:

Theorem 21.7

Suppose that $f_N(x_0)$ converges to s for some x_0, and f_N' converges to g uniformly. Then f_N converges uniformly to the function $\int_{x_0}^{x} g \, dy + s$.

In our case, we know that f_N already converges to $\sum_{n=0}^{\infty} a_n x^n$. But the limit of f_N is unique, so

$$\int_{x_0}^{x} g \, dy + s = \sum_{n=0}^{N} a_n x^n.$$

But we know that the left-hand side is differentiable with derivative g, and so the derivative of $\sum_{n=0}^{N} a_n x^n$ exists and is g — where g is the limit of f_N', which is $\sum_{n=1}^{\infty} n a_n x^{n-1}$. This gives us the following conclusion:

Theorem 21.8

For any power series $\sum_{n=0}^{\infty}$, inside its radius of convergence, the power series is differentiable an arbitrary number of times, and the derivative can be obtained by formally differentiating each term.

(Here we've only proved it's differentiable once, but we can repeat the same argument to prove that it's differentiable twice, and so on.)

§21.4 Ordinary Differential Equations

An **ordinary differential equation** (ODE for short) is an equation with one variable involving derivatives. ('Ordinary' is as opposed to 'partial' — a partial differential equation may have several variables.)

Example 21.9

For y a function of x, suppose we are given $y' = f(x)$ for some specified function f — this is a simple example of an ODE. In this case, by the fundamental theorem of calculus we know

$$y = \int_{x_0}^{x} f(t)\,dt + y(x_0).$$

It's worth noting that here, given x_0 and y_0, there exists *exactly* one solution with $y(x_0) = y_0$.

The most basic question given a differential equation is the following:

Question 21.10. Does a solution exist, and is it unique?

(There are other, more interesting, questions to ask as well — for example, about the behavior of solutions.)

We will look at situations slightly more general than the example above, where

$$y'(x) = f(y(x)) + g(x)$$

for some functions f and g (so the derivative of y is a function of y, plus a function of x — in the example, we only had the second term).

First we'll make a few observations. If y is a solution to this equation and $y(x_0) = y_0$, then we can integrate both sides to get by the fundamental theorem of calculus that

$$y(x) = \int_{x_0}^{x} f(y(s))\,ds + \int_{x_0}^{x} g(s)\,ds + y_0.$$

This gives us an idea for how to analyze existence and uniqueness. We can define an operator T as

$$T(h) = \int_{x_0}^{x} f(h(s))\,ds + \int_{x_0}^{x} g(s)\,ds + y$$

(acting on any function h). If y solves the ODE, then $T(y) = y$.

This relates to something we saw before — given an interval I, we can think of T as an operator $T\colon \mathcal{C}(I) \to \mathcal{C}(I)$ (here $\mathcal{C}(I)$ denotes the set of continuous functions on I). Then our question becomes:

Question 21.11. Does there exist a continuous function y such that $T(y) = y$?

As we've seen before, $\mathcal{C}(I)$ is a metric space. We can define the problem more generally (forgetting about $\mathcal{C}(I)$ and working over a general metric space):

Definition 21.12. Given a metric space (X, d) and a map $T: X \to X$, a point y such that $T(y) = y$ is called a **fixed point**.

A long time ago, when we discussed metric spaces, we saw the following:

Definition 21.13. A map $T: (X, d) \to (X, d)$ is a **contracting map** if for some $\ell < 1$, for all $x, y \in X$, we have

$$d(T(x), T(y)) \leq \ell \cdot d(x, y).$$

Intuitively, *contracting* means that if we take two points, then their images end up strictly closer together than the original pair.

Theorem 21.14 (Contracting Map Property)
If (X, d) is a Cauchy complete metric space and $T: X \to X$ is a contracting map, then T has a *unique* fixed point.

Eventually we'll apply this where the metric space is the space of continuous functions on a compact interval and T is the operator defined earlier. We will try to prove that T is a contracting map; then we can apply this more general theorem to give us a unique solution to the ODE.

Recall the following definition:

Definition 21.15. A metric space is **Cauchy complete** if every Cauchy sequence is convergent.

In particular, any compact metric space is Cauchy complete, but the converse is not true (for example, $\mathbb{R}$ is not compact but is Cauchy complete).

We'll first redo the proof of this theorem.

Proof. First we will show that T has *at most* one fixed point. Assume for contradiction that x_1 and x_2 are both fixed points of T, so that $T(x_1) = x_1$ and $T(x_2) = x_2$. Then $d(T(x_1), T(x_2)) = d(x_1, x_2)$. But since T is a contracting map, we were supposed to have $d(T(x_1), T(x_2)) \leq \ell d(x_1, x_2)$. This means $d(x_1, x_2) \leq \ell d(x_1, x_2)$ for some $\ell < 1$; this is only possible if $d(x_1, x_2) = 0$, and therefore $x_1 = x_2$ (in a metric space, two points can only have distance 0 if they are the same point).

So a contracting map has *at most* one fixed point (this holds for *any* metric space, not necessarily Cauchy complete); now we will show that there *is* a fixed point (which does need Cauchy completeness).

We define a sequence in the following way: fix any $x \in X$, and define $x_0 = x$, $x_1 = T(x)$, and so on, with $x_n = T^n(x)$.

The idea is to show that x_n is a Cauchy sequence. Since our metric space is Cauchy complete, it then has a limit x_∞; we will then show that x_∞ is a fixed point.

Claim — x_n is a Cauchy sequence.

Proof. First, by the contracting map property we have

$$d(x_{n+1}, x_n) = d(T(x_n) - T(x_{n-1})) \leq \ell \cdot d(x_n, x_{n-1}).$$

Iterating this, we see that every time we bring down the index by 1 we can pop out a factor of ℓ, to obtain

$$d(x_{n+1}, x_n) \leq \ell^n d(T(x), x).$$

Now we can use the triangle inequality. If we consider two points x_n and x_{n+m}, with $n \geq N$ (and $m > 0$), by the triangle inequality we have

$$d(x_n, x_{n+m}) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + \cdots + d(x_{n+m-1}, x_{n+m}).$$

We can then apply the above bound to each consecutive distance, to obtain

$$d(x_n, x_{n+m}) \leq \ell^n d(T(x), x) + \ell^{n+1} d(T(x), x) + \cdots + \ell^{n+m-1} d(T(x), x).$$

Since $d(T(x), x)$ is a constant, we can factor it out. The remainder is a portion of a geometric series, which is convergent as $\ell < 1$. So if we take n sufficiently large, we can make this sum as small as we want. $\square$

Now we know x_n is a Cauchy sequence, so it has a limit x_∞. We now want to show x_∞ is a fixed point.

> **Claim** — If $y_n \to y$, then $T(y_n) \to T(y)$.

Proof. This essentially follows from the fact that T is continuous. To prove it explicitly, we have

$$d(T(y_n), T(y)) \leq \ell d(y_n, y).$$

So if $d(y_n, y) \to 0$, then $d(T(y_n), T(y)) \to 0$ as well. $\square$

But now using our sequence $x_n = T^n(x)$, we know $x_n \to x_\infty$, so $T(x_n) \to T(x_\infty)$ by the above claim. But $T(x_n)$ is the *same* sequence as x_n (shifted by one index), so they must have the same limit; this means we must have $x_\infty = T(x_\infty)$. $\square$

Next time we will use this to prove the existence and uniqueness of solutions to ODEs.

§22 December 6, 2022

§22.1 Solving ODEs

Last class, we started considering the ordinary differential equation

$$y'(x) = f(y(x)) + g(x).$$

We stated the following result:

Theorem 22.1 (Picard–Lindelöf Theorem)

Suppose that f is a differentiable function with $|f'| \le c$, and g a continuous function. Then given some x_0 and y_0, there exists a unique function y defined in some neighborhood of x_0, such that $y(x_0) = y_0$ and

$$y'(x) = f(y(x)) + g(x).$$

The main idea of the proof is to use the contracting mapping theorem:

Definition 22.2. For a metric space (X, d), a **contracting map** is a function $T\colon X \to X$ such that

$$d(T(x_1), T(x_2)) \le \ell d(x_1, x_2)$$

for some fixed constant $0 \le \ell < 1$.

Theorem 22.3 (Contracting Mapping Theorem)

For any Cauchy complete metric space (X, d) and contracting map T, there exists exactly one fixed point of T — i.e., there exists a unique x such that $T(x) = x$.

The main idea of the proof is to take any $x \in X$ and consider the sequence $x_n = T^n(x)$ (where $T^n(x)$ denotes T applied n times to x). We then proved that x_n converges to some point x_∞, and that x_∞ must be a fixed point. This shows the existence of at least one fixed point; meanwhile, we can't have two fixed points because the distance between them would remain the same under an application of T, but this distance is supposed to shrink.

In fact, this proof can actually give us a bound for the *rate* at which the sequence x_n converges. So if we were to try to find a fixed point of a map, we can estimate how many times we need to apply T in order to get *close* to the fixed point. (This is useful in many applications.)

Now let's use this to prove the theorem about ODEs. Choose a small interval $[-\delta + x_0, \delta + x_0]$ around x_0, and take our metric space to be $\mathcal{C}([-\delta + x_0, \delta + x_0])$ (the space of continuous functions on this interval), with metric defined by

$$d(f, g) = \sup_{x \in [-\delta + x_0, \delta + x_0]} |f(x) - g(x)|.$$

Let $I = [-\delta + x_0, \delta + x_0]$. For a function $y \in \mathcal{C}(I)$, we define the function

$$T(y) := \int_{x_0}^{x} f(y(s))\, ds + \int_{x_0}^{x} g(s)\, ds + y_0.$$

It's clear that T is a map from $\mathcal{C}(I)$ to itself. Meanwhile y is a fixed point (i.e. $T(y) = y$) if and only if

$$y = \int_{x_0}^{x} f(y(s))\, ds + \int_{x_0}^{x} g(s)\, ds + y_0.$$

But taking the derivative of both sides, this means

$$y'(x) = f(y(x)) + g(x).$$

Additionally, if we evaluate at x_0, then both integrals are 0, and we get $y(x_0) = y_0$. So a fixed point of T is indeed a solution to the ODE.

So it's enough to prove that T is a contracting map. (We can only do this for a very *small* interval around x_0 — we are going to choose δ small so that this proof works.)

Claim — There exists δ such that with $I = [-\delta + x_0, \delta + x_0]$, T is a contracting map on $\mathcal{C}(I)$.

Proof. Suppose that y_1 and y_2 are in $\mathcal{C}(I)$; then we want to compare $d(T(y_1), T(y_2))$ to $d(y_1, y_2)$. First we have

$$T(y_1) - T(y_2) = \int_{x_0}^{x} f(y_1(s))\,ds + \int_{x_0}^{x} g(s)\,ds + y_0 - \int_{x_0}^{x} f(y_2(s))\,ds - \int_{x_0}^{x} g(s)\,ds - y_0.$$

(Here $T(y_1)$ and $T(y_2)$ are both functions in x.) The latter two terms of each expression both cancel out, and we get

$$T(y_1) - T(y_2) = \int_{x_0}^{x} (f(y_1(s)) - f(y_2(s)))\,ds.$$

Now we'll make use of the fact that $|f'|$ is bounded. More generally, suppose h is any function $h\colon I \to \mathbb{R}$ and $|h'| \le c$ is bounded. Then the mean value theorem gives that

$$h(z_1) - h(z_2) = h'(z_3)(z_1 - z_2)$$

for some z_3 between z_1 and z_2. This means

$$|h(z_1) - h(z_2)| = |h'(z_3)| \cdot |z_1 - z_2| \le c\,|z_1 - z_2|\,.$$

We can now apply this to f. If we think of s as any *fixed* point, then $y_1(s)$ and $y_2(s)$ are two fixed values. Since f is differentiable and we have a bound $|f'| \le c$, we have that

$$|f(y_1(s)) - f(y_2(s))| \le c\,|y_1(s) - y_2(s)|$$

(where we're taking $z_1 = y_1(s)$ and $z_2 = y_2(s)$).

But this holds for *all* s, so then

$$|T(y_1) - T(y_2)| = \int_{x_0}^{x} (f(y_1(s)) - f(y_2(s)))\,ds \le \int_{x_0}^{x} c\,|y_1(s) - y_2(s)|\,ds.$$

We wanted to compare $d(T(y_1), T(y_2))$ and $d(y_1, y_2)$. But $d(T(y_1), T(y_2))$ is the sup of this expression over all $x \in I$, which means

$$d(T(y_1), T(y_2)) \le c \sup_{x \in I} \int_{x_0}^{x} |y_1(s) - y_2(s)|\,ds.$$

But s is *also* in the interval I, so then for each $s \in I$,

$$|y_1(s) - y_2(s)| \le \sup_{t \in I} |y_1(t) - y_2(t)| = d(y_1, y_2).$$

So our integrand $|y_1(s) - y_2(s)|$ is pointwise bounded by the constant $d(y_1, y_2)$, and $|x - x_0|$ is bounded by δ; this means

$$d(T(y_1), T(y_2)) \le c \cdot \delta \cdot d(y_1, y_2).$$

We wanted to have an inequality of the form $d(T(y_1), T(y_2)) \leq \ell d(y_1, y_2)$ for $\ell < 1$. To obtain this, we can choose $\delta = 1/2(c+1)$ (the $+1$ is there only to avoid division by 0, and is not actually important). For this choice of δ, we have

$$c\delta \leq \frac{c}{2(c+1)} \leq \frac{1}{2},$$

and taking ℓ to be this quantity gives the desired result. $\square$

As a recap, the main idea of our proof was — we wanted to solve for y in a small neighborhood of x_0. We reformulated this problem into one about a contracting map, and in order to do that, we needed to restrict the interval we were trying to find a solution for (where the interval size depends on the bound we have for f').

Remark 22.4. It's another problem in differential equations to estimate *how* small the interval needs to be, but we won't get into that.

Example 22.5

The theorem can be used to solve differential equations such as $y' = y$ (with the function f defined as $f(s) = s$), $y' = \frac{1}{y}$ (with the function f defined as $f(s) = \frac{1}{s}$), or $y' = e^{y^2}$ (with the function f defined as $f(s) = e^{s^2}$). In the first case we can solve the equation exactly, but for the last we can't — but we can still prove the *existence* (and uniqueness) of a solution. (The term $g(x)$ in the theorem isn't really important — we can handle it using the fundamental theorem of calculus.)

§22.2 Convergence Tests

We'll now discuss tests for whether a series $\sum a_n$ converges. There's two tests we talked about a lot — the *root test* and the *ratio test* (which is Prof. Colding's favorite). But there's two other major tests as well — the *alternating series test* and the *integral test*. We've already proven the alternating series test, but today we will go over it again. The integral test is one that we *haven't* yet proven, so we will do so today as well.

§22.2.1 Alternating Series Test

The *alternating series test* is interesting because it tests for convergence but not *absolute* convergence (while all the other tests do test for absolute convergence). The fundamental example is the alternating harmonic series:

Example 22.6

Does the series

$$\sum_{n=1}^{\infty} (-1)^n \cdot \frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \cdots$$

converge?

This series is *not* absolutely convergent, so it'll not pass the other tests (since those tests can only pick out whether a series is absolutely convergent). But it's an *alternating series* — terms alternate in sign, or in other words $a_{n+1}a_n < 0$ for all n. And we can prove convergence using the *alternating series test*:

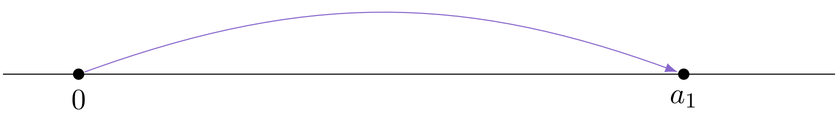
Theorem 22.7

Suppose a_n is a sequence such that:

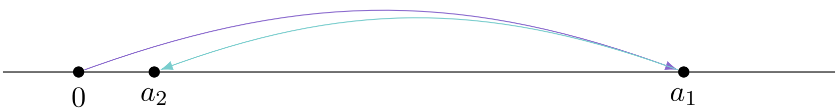
- a_n alternate in sign; i.e., $a_{n+1}a_n < 0$.
- $|a_{n+1}| < |a_n|$ for all n.
- $a_n \to 0$.

Then the series $\sum a_n$ converges.

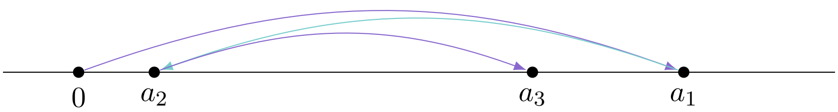
We can see intuitively why this is true by drawing a picture: we start at 0, and then we go up to a_1.



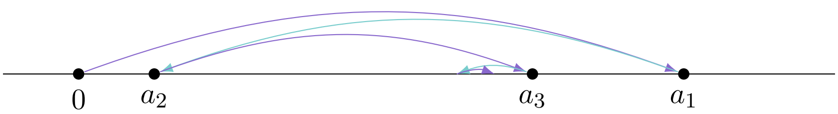
Then when we take the next step, since a_2 is negative, we're going to walk backwards. But $|a_2| < |a_1|$, so we go backwards by *less* than all the way to 0.



Then when we take the next step, we're going forwards, but less than all the way to a_1 (since a_3 is positive but $|a_3| < |a_2|$).



This continues happening. Meanwhile, the gaps between our points goes to 0, since $a_n \to 0$, so the series should converge.



Example 22.8

For the alternating harmonic series, it's clearly alternating, and $|a_n| = 1/n$ is monotone decreasing and goes to 0. So the conditions are satisfied, and by the alternating series test, the series converges.

§22.2.2 Integral Test

Example 22.9

Consider the series $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$, where $\alpha > 1$. Does this series converge?

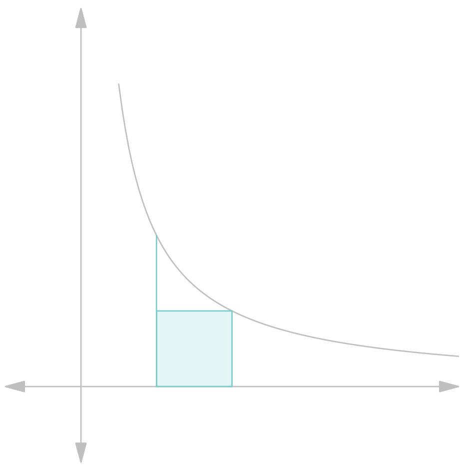
The ratio test doesn't help us here — we have

$$\lim_{n\to\infty} \frac{\frac{1}{(n+1)^\alpha}}{\frac{1}{n^\alpha}} = \lim_{n\to\infty} \left(\frac{n}{n+1}\right)^\alpha = 1,$$

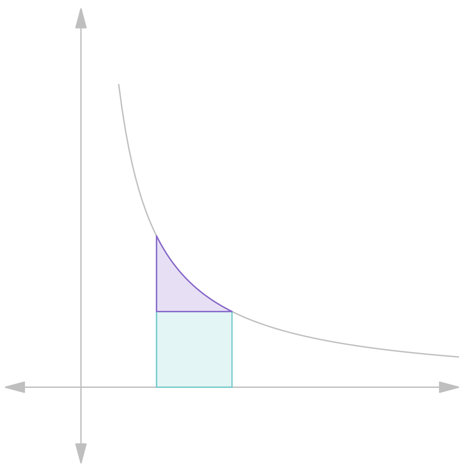
so the ratio test is inconclusive. The same would occur with the root test. So we would like another test that takes care of this situation.

The main idea is to think about the function $f(x) = x^{-\alpha}$ defined on $[1, \infty) \to \mathbb{R}$ (rather than just on the integers). Note that f is always positive, and it is decreasing, as $f'(x) = -\alpha x^{-\alpha-1} < 0$.

We can imagine the following picture. When we're computing $\frac{1}{2^\alpha}$, we can think of it as the area of a little rectangle:



But this rectangle has area *at most* the area under the entire graph on this interval:



So we can write

$$\frac{1}{2^\alpha} < \int_1^2 f(s)\,ds.$$

We can do the same for 3 to write
$$\frac{1}{3^\alpha} < \int_2^3 f(s)\,ds,$$
and so on. This tells us that for all m,
$$\sum_{n=2}^{m} \frac{1}{n^\alpha} \leq \int_1^m x^{-\alpha}\,dx.$$
But we know how to evaluate this integral — so we have
$$\sum_{n=2}^{m} \frac{1}{n^\alpha} \leq \left[\frac{x^{1-\alpha}}{1-\alpha}\right]_1^m = \frac{m^{1-\alpha}}{1-\alpha} - \frac{1}{1-\alpha}.$$
But this means
$$\sum_{n=2}^{m} \leq -\frac{1}{1-\alpha} = \frac{1}{\alpha-1}$$
(note that the first term goes to 0 as $m \to \infty$). This shows our series is *bounded*, and since all terms are positive, that means it must be convergent.

Of course, this method works more generally.

Theorem 22.10 (Integral Test)
Suppose that $f\colon [1,\infty) \to (0,\infty)$ is a *decreasing* function, and the improper integral $\int_1^\infty f(x)\,dx$ is finite. Then the series $\sum_{n=1}^\infty f(n)$ converges.

(This is actually an if and only if statement.)

§23 December 8, 2022

Today we will review the material covered in the previous classes. (The first few lectures, about the properties of $\mathbb{R}$, won't be on the exam, so we won't cover them.)

§23.1 Sequences and Limits

Definition 23.1. A sequence $x_n \in \mathbb{R}$ **converges** to a (denoted $x_n \to a$) if for all $\varepsilon > 0$ there exists N such that for all $n \geq N$, we have
$$|a - x_n| < \varepsilon.$$

Proposition 23.2 (Algebraic Operations for Sequences)
If $x_n \to x$ and $y_n \to y$, then:

- $x_n + cy_n \to x + cy$, and
- $x_n y_n \to xy$.

Another important notion regarding sequences is a *subsequence* — given a sequence, to form a subsequence we take some of its elements, in the correct order (we can't repeat elements). The formal definition is the following:

Definition 23.3. For a *strictly increasing* function $f : \mathbb{N} \to \mathbb{N}$ (i.e. whenever $m < n$ we have $f(m) < f(n)$), the sequence $x_{f(n)}$ is a subsequence of x_n.

Since f is increasing, it goes through the x's in the correct order (in particular f must be one-to-one, so we can't repeat elements).

There are two basic criteria about convergence. The first is about *monotone* sequences.

Theorem 23.4

If x_n is a monotone increasing sequence (i.e. $x_n \le x_{n+1}$ for all n), then x_n is convergent if and only if it is bounded from above, and in that case

$$\lim_{n\to\infty} x_n = \sup\{x_n \mid n \in \mathbb{N}\}.$$

Likewise, if x_n is monotone decreasing (i.e. $x_{n+1} \le x_n$), then it is convergent if and only if it is bounded below, and

$$\lim_{n\to\infty} x_n = \inf\{x_n \mid n \in \mathbb{N}\}.$$

The second useful criteria is about sequences squeezed between others.

Theorem 23.5

Suppose $a_n \le x_n \le b_n$ for all n, and $a_n \to a$ and $b_n \to a$ as well (i.e. x_n is squeezed between two sequences with the same limit). Then $x_n \to a$ as well.

§23.1.1 Cauchy Sequences

Definition 23.6. A sequence $x_n \in \mathbb{R}$ is a **Cauchy sequence** if for all $\varepsilon > 0$, there exists N such that $|x_m - x_n| < \varepsilon$ for all $m, n \ge N$.

Intuitively this states that if we go sufficiently far out, then the terms are bunched together. Meanwhile, *convergence* requires not just that the terms are bunched together, but that they're bunched together close to the limit. On $\mathbb{R}$ these two notions are equivalent:

Theorem 23.7

On $\mathbb{R}$, a sequence is convergent if and only if it is a Cauchy sequence.

This is a special property of $\mathbb{R}$. (The converse is always true — in any metric space, every convergent subsequence must be Cauchy.) The proof we saw used two facts — the Bolzano–Weirstrass theorem, and the fact that if a Cauchy sequence has a convergent subsequence, then it is itself convergent. We also used the property that every Cauchy sequence is bounded.

The latter two facts are true in general, not just in $\mathbb{R}$:

Theorem 23.8

Suppose x_n is a Cauchy sequence, and x_{n_k} a subsequence of x_n. If $x_{n_k} \to x$, then $x_n \to x$ as well.

Intuitively, the reason for this is that since x_n is a Cauchy sequence, if we go far enough out then all the x_n are very close together. But we are given that there are infinitely many of them that converge, meaning that they become very close to a certain number x. But all other x_n sufficiently far out must be close to those x_{n_k}, so they must also be close to x.

Lemma 23.9

Every Cauchy sequence is bounded.

This is simply because if x_n is Cauchy, then taking $\varepsilon = 1$, we know there exists N such that for all $n, m \geq N$, we have $|x_m - x_n| < 1$ (applying the definition of a Cauchy sequence with $\varepsilon = 1$). In particular this means $|x_n - x_N| < 1$. But x_N is a fixed number, so all such x_n are bounded. And there's only finitely many other elements ($x_1, \ldots, x_{N-1}$), which all have finite distance to x_N; this means all our points must lie within some fixed finite distance of x_n.

These two properties work in any metric space, but the Bolzano–Weirstrass theorem is particular to $\mathbb{R}$.

Theorem 23.10 (Bolzano–Weirstrass Theorem)

Every bounded sequence $x_n \in \mathbb{R}$ has a convergent subsequence.

We can use this to prove all Cauchy sequences in $\mathbb{R}$ converge — the Cauchy sequence is bounded, so it has a convergent subsequence, and therefore it must converge.

In order to prove the Bolzano–Weirstrass theorem, we constructed a subsequence squeezed between two monotone sequences — one monotone increasing, and the other monotone decreasing. Using the fact that bounded monotone sequences are convergent, we found that the two sequences had to have the same limit, and so our subsequence had to converge as well.

Definition 23.11. A metric space is **Cauchy complete** if every Cauchy sequence is convergent.

§23.2 Limsup and Liminf

Definition 23.12. For a subset $A \subseteq \mathbb{R}$, its **supremum** $\sup(A)$ is defined as the least upper bound of A — so $\sup(A)$ is an upper bound on A, and there is no smaller upper bound. Likewise its **infimum** $\inf(A)$ is the greatest lower bound of A.

Given a sequence x_n, we can then define $A_n = \{x_n, x_{n+1}, \ldots\}$, and then consider the sup and inf of *these* sets. As n becomes larger, the set A_n becomes smaller — and taking the sup over a smaller set can only give us a smaller result, so

$$\sup A_n \geq \sup A_{n+1},$$

and likewise

$$\inf A_n \leq \inf A_{n+1}.$$

This means the sequence $\sup A_n$ is decreasing, and the sequence $\inf A_n$ is increasing. (Note that here we are allowing the sup and inf to be $\pm\infty$ — we usually don't allow this, but here it's convenient to do so.)

Since both sequences are monotone, they must both have a limit, which we define as the lim sup and lim inf:

Definition 23.13. For a sequence x_n with $A_n = \{x_n, x_{n+1}, \ldots\}$, $\limsup_{n\to\infty} x_n$ is defined as $\lim_{n\to\infty} \sup\{A_n\}$, and $\liminf_{n\to\infty} x_n$ is defined as $\lim_{n\to\infty} \inf\{A_n\}$.

§23.3 Series

Given a sequence $a_n \in \mathbb{R}$, we can consider the sequence of sums $s_m = \sum_{n=0}^{m} a_n$. The key question for series is usually whether they converge.

Definition 23.14. We say that the series $\sum_{n=0}^{\infty} a_n$ converges if s_m converges (and $\sum_{n=0}^{\infty} a_n$ also denotes its limit).

There is an easy way to sometimes see that a series does *not* converge (which you should check before using tests, to avoid wasting time on series that obviously can't converge):

Proposition 23.15

If a_n converges, then $a_n \to 0$.

There are two types of convergence for a series — *convergence* (as seen above) and *absolute convergence*.

Definition 23.16. A series is **absolutely convergent** if $\sum |a_n|$ converges.

Proposition 23.17

If a series is absolutely convergent, then it is convergent.

§23.3.1 Important Examples of Series

Example 23.18

The **harmonic series** $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent.

Example 23.19

The **alternating harmonic series** $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ is convergent.

This can be proven by the alternating series test.

Probably the most important example of a series is the geometric series:

Example 23.20

For $c \in \mathbb{R}$, the series $\sum_{n=0}^{\infty} c^n$, called a **geometric series**, is convergent if and only if $|c| < 1$. In that case, we have $\sum_{n=0}^{\infty} c^n = \frac{1}{1-c}$.

If $|c| > 1$, then the geometric series $\sum c^n$ actually fails the most basic test that the terms must go to 0; so $|c| < 1$ *has* to be true for the sequence to even have a chance of converging. But in that case, it does in fact converge, and we can even calculate its limit.

Proof. The partial sums of the series are

$$s_m = 1 + c + c^2 + \cdots + c^m.$$

This means we have

$$\begin{aligned}(1-c)s_m = 1 + c + c^2 + \cdots + c^m \\ - c - c^2 - \cdots - c^m - c^{m+1}.\end{aligned}$$

Adding these together, most terms cancel out, and we are left with

$$(1-c)s_m = 1 - c^{m+1},$$

which means that

$$s_m = \frac{1-c^{m+1}}{1-c}.$$

If $|c| < 1$, then $c^{m+1} \to 0$, which means that $s_m \to \frac{1}{1-c}$. $\square$

Example 23.21

The series $\sum_{n=1}^{\infty} n^{-p}$, called the ***p*-series**, is convergent if and only if $p > 1$.

We see that when $p = 1$, this is the harmonic series and is therefore divergent; but it turns out that for any $p > 1$, the series is convergent.

This statement can be proven using the integral test (which can be used to detect both convergence and divergence); in particular, since the harmonic series is a special case, the divergence of the harmonic series can also be proven in this way.

§23.3.2 Tests for Convergence

To determine that a series is convergent, we often want to compare it with a series we already know is convergent.

The two most basic tests are the two comparison tests:

Theorem 23.22 (Comparison I)

If $\sum a_n$ and $\sum b_n$ are two series such that $b_n \neq 0$ and

$$\frac{a_n}{b_n} \to \ell$$

for some ℓ, then if $\sum b_n$ is convergent, the series $\sum a_n$ must be convergent as well.

When we use this, we'll have a series that we know is either convergent or divergent, and a series we *don't* know the convergence of, but want to compare to the known sequence to figure out. If the ratio between corresponding terms converges to a constant, then our sequence is essentially just the known sequence multiplied by a constant, which means if $\sum b_n$ converges so must $\sum a_n$.

Theorem 23.23 (Comparison II)

If $0 \le a_n \le b_n$ for all n and $\sum b_n$ is convergent, then $\sum a_n$ is convergent.

(The test is typically used to test for absolute convergence — we can replace a_n with $|a_n|$.)

Remark 23.24. Often the negation (i.e. the contrapositive) is useful as well — if $\sum a_n$ is divergent, then so is $\sum b_n$.

We used these tests to obtain two other tests, the ratio test and root test — which both come from using the two comparison tests to compare our series with a *geometric* series $\sum c^n$ (whose convergence we already know — it is absolutely convergent for $|c| < 1$).

Theorem 23.25 (Ratio test)

If $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$, then the series $\sum a_n$ is convergent.

(Prof. Colding recommends we use the ratio test when we don't know which to use.)

Theorem 23.26 (Root test)

If $\limsup_{n\to\infty} |a_n|^{1/n} < 1$, then the series $\sum a_n$ is convergent.

In the case of the ratio test, if the limit is some ℓ, then it's almost true that $|a_{n+1}| \leq \ell\,|a_n|$ (to be precise, we would have to adjust ℓ a bit). This means our series is bounded above by the geometric series with ratio $\ell < 1$, so it is convergent.

Remark 23.27. All these tests (except the alternating series test) actually imply absolute convergence.

There are two more tests we saw recently:

Theorem 23.28 (Alternating series test)

If $a_n > 0$, $a_{n+1} < a_n$, and $a_n \to 0$ for all n, then the series $\sum(-1)^n a_n$ is convergent.

Theorem 23.29 (Integral test)

If f is a monotone non-increasing function such that $f \geq 0$ and $\int_1^\infty f(s)\,ds$ is finite, then $\sum_{n=1}^\infty f(n)$ is convergent.

(If f is differentiable — which will usually be the case for the functions we see in this class — then the monotone decreasing condition can be checked by checking that $f' \leq 0$.)

This can be used to determine when the p-series $\sum_{n=1}^\infty n^{-p}$ is convergent. (It's also possible to do it without the integral test, but it's more complicated.)

§23.4 Power Series

Definition 23.30. For a sequence a_n, its **power series** is defined as the series $\sum a_n x^n$. (The power series is a function of x.)

Example 23.31

When $a_n = 1$ for all n, the corresponding power series is $\sum x^n$. This is a geometric series, so it converges if and only if $|x| < 1$, and $\sum_{n=0}^\infty x^n = \frac{1}{1-x}$.

For a power series, the first thing we usually do is determine the radius of convergence.

Definition 23.32. The **radius of convergence** of a power series $\sum a_n x^n$, denoted R, is defined such that
$$\frac{1}{R} = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|.$$

Strictly inside the radius of convergence, the power series must converge by the ratio test.

Theorem 23.33

If $f(x) = \sum a_n x^n$ is defined for $|x| < R$ (where R is the radius of convergence), then f is differentiable with $f'(x) = \sum n a_n x^n$ (the series given by differentiating term-by-term, which has the same radius of convergence).

In fact, f is differentiable an arbitrary number of times (by applying this repeatedly).

Example 23.34

The exponential function has power series $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. It has the important property that
$$\exp(x+y) = \exp(x) \cdot \exp(y).$$

§23.5 Continuous Functions

Definition 23.35. For a function $f\colon I \to \mathbb{R}$ (where $I \subseteq \mathbb{R}$ is some interval), f is **continuous** at x_0 if for all $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $|x - x_0| < \delta$, we have $|f(x) - f(x_0)| < \varepsilon$.

This intuitively states that if we're at a point x very close to x_0, then the value of the function at x must be very close to its value at x_0 as well (it can't suddenly jump to some other value).

For continuous functions on $\mathbb{R}$, there are two important properties.

Theorem 23.36 (Extreme Value Theorem)

If $f\colon [a, b] \to \mathbb{R}$ is a continuous function on a *closed* interval, then both its sup and inf are achieved — i.e., there exist x_m and x_M in $[a, b]$ such that
$$f(x_m) = \min_{[a,b]} f \text{ and } f(x_M) = \max_{[a,b]} f.$$

(There might be several places where the min or max is achieved, but they must be achieved at least *somewhere*.)

Theorem 23.37 (Intermediate Value Theorem)

If $f\colon [a, b] \to \mathbb{R}$ is continuous and $f(a) < z < f(b)$, then there must exist $c \in (a, b)$ with $f(c) = z$.

§23.6 Metric Spaces

We can think of a metric space as a generalization of $\mathbb{R}$.

Definition 23.38. A **metric space** (X, d) is a set X together with a *distance* function $d\colon X \times X \to \mathbb{R}_{\geq 0}$ such that:

- $d(x, y) = 0$ if and only if $x = y$.
- $d(x, y) = d(y, x)$ (symmetry).
- $d(x, z) \leq d(x, y) + d(y, z)$ (the triangle inequality).

Example 23.39

The space $\mathbb{R}$ with the metric $d(x, y) = |x - y|$ is a metric space.

Example 23.40

The space $\mathbb{R}^N$ with the metric

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

(called the *Euclidean distance*) is a metric space.

Example 23.41

For a closed interval $[a, b]$, the space of continuous functions on this interval — denoted $\mathcal{C}([a, b])$ — with the distance defined as

$$d(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)|$$

is a metric space.

§24 December 13, 2022

The final will have six problems. Each problem will have some sub-questions — some will have two sub-questions, one will have four (small) sub-questions. We will probably have more time on the final than the midterm.

§24.1 Continuous Functions

Intuitively, a function is continuous if it doesn't have any jumps.

Definition 24.1. A function $f\colon I \to \mathbb{R}$ is **continuous** at a point $x_0 \in I$ if for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

Definition 24.2. A function $f\colon I \to \mathbb{R}$ is continuous if it is continuous at all points $x_0 \in I$.

We also have a stronger notion of continuity — *uniform continuity*, which loosely speaking, says that we can use the same δ for all $x_0 \in I$.

Definition 24.3. A function $f\colon I \to \mathbb{R}$ is **uniformly continuous** if for all $\varepsilon > 0$, there exists $\delta > 0$ so that for all x and y with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.

Example 24.4

One function that is *not* continuous is $f\colon \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 1 & \text{if } x \le 0 \\ 0 & \text{if } x > 0. \end{cases}$$

(It's not continuous at 0, but it is continuous everywhere else.)

Example 24.5

The function $f(x) = \frac{1}{x}$ defined on $f\colon (0, \infty) \to \mathbb{R}$ is continuous, but it is not uniformly continuous — by moving very close to 0, we can find arbitrarily close points with a large difference in their image. So we can't use the same δ for all x_0.

We can produce some examples of continuous functions:

- The constant function $f(x) = c$ is continuous.
- The function $f(x) = x$ is continuous.
- If f and g are continuous, then cf, $f + g$, and fg are all continuous; if $g(x) \neq 0$ for all g, then $\frac{f}{g}$ is also continuous.

Student Question. *Does this preserve uniform continuity?*

Answer. Yes, except for the last (where we'd need to be more careful).

Theorem 24.6

If $f\colon [a, b] \to \mathbb{R}$ is a continuous function on a *compact* interval, then f is uniformly continuous.

There are two main results about continuous functions in general:

Theorem 24.7 (Extreme Value Theorem)

If $f\colon [a, b] \to \mathbb{R}$ is continuous, then $\sup f$ is achieved — in other words, there exists $x_M \in [a, b]$ such that $\sup f = f(x_M)$. Similarly, $\inf f$ is also achieved — there exists $x_m \in [a, b]$ such that $\inf f = f(x_m)$.

Theorem 24.8 (Intermediate Value Theorem)

If $f\colon [a, b] \to \mathbb{R}$ is continuous, then all values between $f(a)$ and $f(b)$ are in the image of f.

§24.2 Metric Spaces

Definition 24.9. A **metric space** is a set X along with a function $d\colon X \times X \to [0, \infty)$ (called a *distance* or *metric*) satisfying certain properties:

(1) $d(x, y) = 0$ if and only if $x = y$.

(2) $d(x, y) = d(y, x)$ (i.e., d is symmetric).

(3) $d(x, z) \le d(x, y) + d(y, z)$ (the triangle inequality).

One basic object in a metric space is *sequences* — as before, a sequence in a metric space consists of elements x_n for all $n \in \mathbb{N}$. As in $\mathbb{R}$, we can consider subsequences x_{n_k}. Similarly to in $\mathbb{R}$, we can consider when sequences are convergent, and when they are Cauchy.

Other basic objects in metric spaces are *balls*, *bounded sets*, *open* and *closed* subsets, and *compact* subsets.

§24.2.1 Examples of Metric Spaces

Example 24.10

$(\mathbb{R}, |\cdot|)$ is a metric space (with the distance function $d(x, y) = |x - y|$).

Example 24.11

$(\mathbb{R}^n, |\cdot|)$ is a metric space, under the *Euclidean distance*

$$d(x, y) = |x - y| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

Example 24.12

$(\mathbb{R}^n, |\cdot|_{\text{box}})$ is a metric space, under the *box metric*

$$d(x, y) = |x - y|_{\text{box}} = \max |x_i - y_i| .$$

(The Euclidean distance is much nicer in certain ways — in particular it is a *Hilbert space*.)

Example 24.13

$(\mathcal{C}([a, b]), d_\infty)$ is a metric space, where

$$d_\infty(f, g) = \max_{x \in [a,b]} |f(x) - g(x)| .$$

(Note that this maximum exists because of the extreme value theorem.)

§24.2.2 Convergence in Metric Spaces

Definition 24.14. A sequence $x_n \in X$ **converges** to x_∞ (denoted $x_n \to x_\infty$) if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$,

$$d(x_\infty, x_n) < \varepsilon.$$

(We can also discuss convergent subsequences, as before.)

Definition 24.15. A sequence $x_n \in X$ is a **Cauchy sequence** if for all $\varepsilon > 0$, there exists N such that for all $m, n \geq N$, $d(x_n, x_m) < \varepsilon$.

Intuitively, this says that for all ε, if we go sufficiently far out then all elements in the sequence bunch ε-close together — so the tail of the sequence bunches very closely together.

Just as in $\mathbb{R}$, a convergent subsequence is automatically also Cauchy. But the converse is not always true — this depends on the metric space.

§24.2.3 Bounded and Open and Closed Sets

Definition 24.16. For $x_0 \in X$ and $r > 0$, the **ball** of radius r centered at x_0, denoted $B_r(x_0)$, is the set of points

$$B_r(x_0) = \{x \in X \mid d(x, x_0) < r\} \subseteq X.$$

Definition 24.17. In a metric space (X, d), a set $A \subseteq X$ is **bounded** if $A \subseteq B_r(x_0)$ for some x_0 and r.

Definition 24.18. A subset $A \subseteq X$ is open if for all $x \in A$, there exists some $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$.

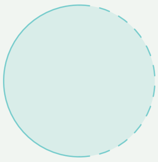
In other words, for all points in A, there needs to be a small ball centered at that point that's entirely contained in A.

Example 24.19

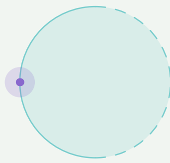
Any ball $B_r(x)$ in a metric space is open.

Example 24.20

The following set is not open:



This is because for all points on the boundary (the part on the left that the set contains), there is no ball centered at those points fully contained in the set.



Definition 24.21. A set $B \subseteq X$ is closed if $X \setminus B$ (the complement of B) is open.

Example 24.22

The set $\{x \in X \mid d(x, x_0) \geq r\}$ is closed — this is because its complement is $B_r(x_0)$, which is open.

Example 24.23

The empty set $\emptyset$ is vacuously open. Meanwhile, X is also open (since any ball is of course part of the set). This means X and $\emptyset$ are closed as well.

There is another way to characterize closed sets, in terms of limit points.

Definition 24.24. For a set $A \subseteq X$, the **limit points** of A are the points $x \in X$ such that x is the limit of a sequence $x_n \in A$.

Notation 24.25. For a set $A \subseteq X$, its set of limit points is denoted $\overline{A}$.

Note that every element in A is a limit point of A (we can take the constant sequence equal to that point) — so $A \subseteq \overline{A}$.

Theorem 24.26

A is closed if and only if $A = \overline{A}$.

Remark 24.27. You can use either as the definition. Prof. Colding uses the other one (the complement of an open set) because it generalizes — we started with $\mathbb{R}$ (which had a ton of different properties), and a lot of these generalized to general metric spaces. This was great because we could use it on other metric spaces — in particular, in the Picard–Lindelöf theorem, we used a metric space other than $\mathbb{R}$ (namely $\mathcal{C}([a, b])$). But sometimes even metric spaces aren't general enough and you want something more general — called *topological spaces*. The notion of open and closed sets carries over there, and then the first definition is the correct one.

§24.2.4 Compactness

Definition 24.28. A **cover** of A is a collection of sets O_α (with no requirements on the number of sets — there could be infinitely many and even uncountably many) such that $A \subseteq O_\alpha$. A cover is **open** if all sets O_α are open.

Definition 24.29. A set $A \subseteq X$ is **compact** if for every open cover, there exists a finite subcover.

In principle, it seems not very feasible to check that a set is compact, since we'd have to do prove this for all possible coverings. But we proved some very useful things:

Theorem 24.30

A subset of $\mathbb{R}$ (with the usual metric) is compact if and only if it is closed and bounded.

In particular, this means $[a, b]$ is always compact.

Theorem 24.31

In *any* metric space, every compact subset must be both closed and bounded.

But the converse is not true — it is true for $\mathbb{R}$ with the usual metric, but it is not true in general.

§24.2.5 Other Definitions

Definition 24.32. (X, d) is **Cauchy complete** if every Cauchy sequence is convergent.

Theorem 24.33

If (X, d) is compact, then (X, d) is Cauchy complete.

Definition 24.34. A function $f\colon X \to \mathbb{R}$ is **continuous** at a point x_0 if for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $d(x, x_0) < \delta$, then $|f(x) - f(x_0)| < \varepsilon$.

§24.2.6 Set Operations

Proposition 24.35

For any collection O_α of sets, if O_α are all open then $\bigcup O_\alpha$ is also open. Meanwhile, for any *finite* collection $O_1, \ldots, O_n$ of open sets, their intersection $\bigcap_{i=1}^n O_i$ is open.

It's not true that an intersection of infinitely many open sets is open — for example, consider $\mathbb{R}^2$ with balls of radius $\frac{1}{n}$ centered at the origin. Their union is the largest ball, but their intersection is just the origin, which is not open.

We can translate this to statements about closed sets by taking the complement:

Proposition 24.36

For any collection U_α of closed sets, their intersection $\bigcap U_\alpha$ is closed. Meanwhile, for *finitely* many closed sets $U_1, \ldots, U_n$, their union $\bigcup_{i=1}^n U_i$ is also closed.

§24.3 Differentiability

We now return to $\mathbb{R}$. (There are ways to do differentiability in greater generality, but they are more complicated.)

Definition 24.37. A function $f\colon I \to \mathbb{R}$ is **differentiable** if at x_0 if

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \ell$$

exists, and its **derivative** $f'(x_0)$ is ℓ.

Proposition 24.38

If f is differentiable, it must be continuous.

We again have basic operations that preserve differentiability:

- If f and g are differentiable, then $f + cg$ is as well, with $(f + cg)' = f' + cg'$.
- If f and g are differentiable, then so is fg, with $(fg)' = f'g + fg'$ (this is called the Leibniz rule).
- If f and g are differentiable and g is nonzero, then $\frac{f}{g}$ is also differentiable, with

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}.$$

  (This is called the quotient rule.)

(All of these apply to differentiability at a given point.)

§24.3.1 Mean Value Theorem

Theorem 24.39 (Mean Value Theorem)
If $f\colon [a, b] \to \mathbb{R}$ is differentiable, then there exists $c \in (a, b)$ such that

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

Equivalently, we can formulate this as

$$f(b) = f(a) + f'(c)(b - a)$$

for some $c \in (a, b)$.

We proved the mean value theorem from Rolle's theorem, the special case where $f(b) = f(a)$ and we need to prove there is some $c \in (a, b)$ with $f(c) = 0$.

§24.3.2 Taylor Expansion

In some sense, Taylor expansion is a generalization of the mean value theorem (stated in the second form) — it follows from applying the mean value theorem repeatedly.

Theorem 24.40
If $f\colon [a, b] \to \mathbb{R}$ is differentiable $n + 1$ times (i.e. f is differentiable, f' is differentiable, and so on, up to $f^{(n)}$ being differentiable), then there exists $c \in (a, b)$ such that

$$f(c) = f(a) + f^{(1)}(a)(b - a) + \frac{f^{(2)}(b - a)^2}{2!} + \cdots + \frac{f^{(n)}(a)}{n!}(b - a)^n + \frac{f^{(n+1)}(c)}{(n + 1)!}(b - a)^{n+1}.$$

(Only the last term involves c.)

In the case $n = 0$, this is exactly the mean value theorem; more generally this theorme follows from applying the mean value theorem repeatedly.

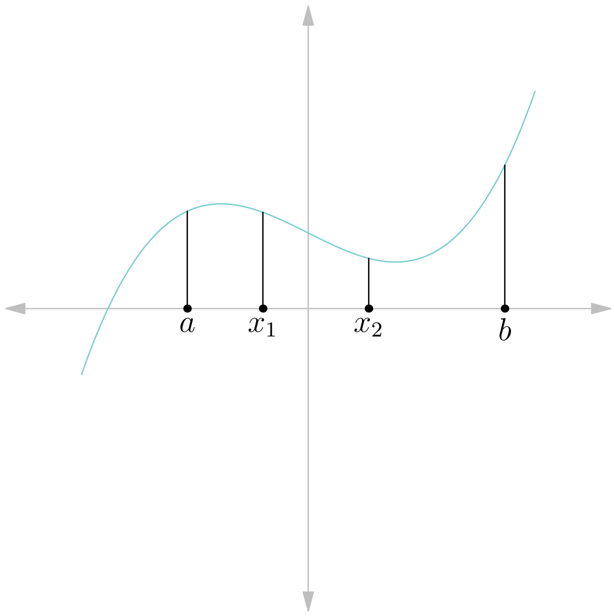
Taylor expansion is useful because the simplest functions we can think of are polynomials (it's easy to evaluate a fixed polynomial at some point); meanwhile it's not so easy to evaluate other random functions (e.g. it would be unacceptable to ask for $\sin(1.754)$ on the final). But the point is that if we know the function at a *given* point a, as well as a bunch of its derivatives, then we can use Taylor expansion to find the function *near* a — we could calculate all of the first $n + 1$ terms, and think of the last term (involving c) as our error, which it's usually possible to bound.

(We are skipping the Cauchy mean value theorem and L'Hospital's rules.)

§24.4 Integrals

There are various ways of defining integrals. The one we used was the *Riemann integral*.

The main idea is that we have a function f, and we want to figure out the area under its graph. To do this, we approximate the function from *above* and *below* by using rectangles; the integral exists if these approximations converge to the same value.



As before, we have the basic operation

$$\int_a^b (f+cg)\,dx = \int_a^b f\,dx + c\int_a^b g\,dx.$$

§24.4.1 Methods for Integration

Theorem 24.41 (Fundamental Theorem of Calculus)
If $F' = f$, then we have

$$\int_a^b f\,dx = F(b) - F(a).$$

Another way of framing the theorem is that if $F(x) = \int_a^x f\,dx$, then $F' = f$; but usually the first form is the most useful.

There's two tricks for evaluating integrals — integration by parts and substitution. (Substitution is often used when you have a messy expression and you can simplify it using substitution; integration by parts is good to try when nothing else works.)

The main idea of integration by parts is to combine the fundamental theorem of calculus with the Leibniz rule. The Leibniz rule tells us that

$$(fg)' = f'g + fg',$$

and integrating both sides gives

$$\int_a^b (fg)'\,ds = \int_a^b f'g\,ds + \int_a^b fg'\,ds.$$

We know the left-hand side by the fundamental theorem of calculus, giving the following:

Proposition 24.42 (Integration by Parts)
We have $\int_a^b f'g\,ds = f(b)g(b) - f(a)g(a) - \int_a^b fg'\,ds$.

Typically, we want to choose f and g so that the expression $g'f$ is simpler than the one we started with. It's useful if we realize that our initial expression is a product of two functions where one is the derivative of something, and we'd like to simplify the other.

Example 24.43

Suppose we want to find $\int_a^b x \cos x \, dx$. If we just had the $\cos x$, then we'd be fine; so we don't like the fact that we have the extra x.

So then we can set $g(x) = x$, so that $g'(x) = 1$ (which gets rid of the x). The price we have to pay for this is finding f such that $f'(x) = \cos x$; but this doesn't actually make the expression any more complicated, since we can take $f(x) = \sin x$.

Using integration by parts, we then have

$$\int_a^b x \cos x \, dx = [x \sin x]_a^b - \int_a^b \sin x \, dx.$$

Now we'll see substitution.

Example 24.44

Suppose we want to evaluate $\int_a^b 2xe^{x^2} \, dx$. We can realize that if we set $y = x^2$, then $\frac{dy}{dx} = 2x$. This means $2x \, dx = dy$, and so

$$\int_a^b 2xe^{x^2} = \int_{a^2}^{b^2} e^y \, dy.$$

(Note the change in boundary conditions — when x goes from a to b, y goes from a^2 to b^2.)

§24.4.2 More About Integrals

We have the following useful fact:

Proposition 24.45

We have $\left| \int_a^b f \, dx \right| \le \int_a^b |f| \, dx$.

Finally, there are some functions that are really terrible and we can't integrate them.

Example 24.46

Consider the function $f : \mathbb{R} \to \mathbb{R}$ which is 1 on rationals and 0 on irrationals. This function is not Riemann integrable (although there are other ways of definining integrals that can deal with it).

But we did see that all *continuous* functions are integrable.

§24.5 Uniform Convergence

In one of the last lectures, we saw the following:

Theorem 24.47

If f_n is a sequence of differentiable functions on an interval $[a, b]$ such that $f_n' \to g$ uniformly, and $f_n(x_0) \to \ell$, then we have

$$f_n \to \int_{x_0}^x g \, ds + \ell$$

uniformly.

Using this, we proved the following:

Theorem 24.48

Given a power series $\sum a_n x^n$, inside the radius of convergence this power series is differentiable as many times as we want, and its derivative is the expression obtained by formally differentiating each term.