NYU Honors Analysis I Recitation

Thomas Lam

Fall 2025

Contents

Remarks

These notes are an abbreviated take on what was done in recitation.

Notational conventions:

- The naturals start at 1.
- Increasing sequences need not be "strictly" increasing.
- The sequence $a_1, a_2, \ldots$ is denoted as $\{a_n\}_n$. If the starting index must be clarified, I may write $\{a_n\}_{n=1}^{\infty}$.
- $\sup_{x \in E} f(x)$ is the same as $\sup\{f(x) : x \in E\}$.
- $\sup_x f(x)$ is the same as $\sup\{f(x) : x\}$ where x is taken over the set over which $f(x)$ is defined. For example $\sup_n a_n = \sup\{a_n : n \in \mathbb{N}\}$.
- $\lim_n a_n$ is a shorthand for $\lim_{n\to\infty} a_n$. Ditto for $\limsup_n a_n$.
- log is the natural log.
- $\|f\|_\infty$ is the supremum of $|f|$ over the domain of f.
- E is usually a set. U and V are usually open sets. K is usually a compact set. C is usually a closed set.
- $\mathcal{L}_o^N$ is the *Lebesgue outer measure* on $\mathbb{R}^N$.
- $\mathcal{L}^N$ is the *Lebesgue measure* on $\mathbb{R}^N$. For example, $\mathcal{L}^2$ is "area", and $\mathcal{L}^1$ is "length".
- For a function f, we write μ_f for the Lebesgue-Stieltjes measure associated with f.

1 Sup, Inf, and Friends!

Welcome to real analysis! Real analysis is the *study of real numbers.* It's important because we actually don't understand real numbers very well. They are very unintuitive creatures! For example, many a middle schooler may think that

$$0.\overline{9} < 1.$$

But by now you probably know better! A good understanding of the real numbers is crucial for doing calculus correctly.

1.1 Exercises with sup

You should think of the "sup" as a "max". Of course, the sup isn't always obtained. For example, the supremum of the interval $[0, 1)$ is 1, but 1 isn't in the set. Thus I like to think of sup as "the max, even if there isn't a max".

To work with sup, you need to be using the raw definition: It is the least upper bound.

Example 1.1: For non-empty sets $S, T \subseteq \mathbb{R}$, define

$$S + T := \{x + y : x \in S, y \in T\}.$$

How do $\sup(S+T)$ and $\sup S + \sup T$ compare? Assume that S and T are bounded from above.

If we think of "sup" as "max" here, it's intuitive that they are equal. Let's try to prove it.

Proof. To prove an equality, we want to show $\leq$ and $\geq$.

Proof of $\leq$: How to upper bound $\sup(S + T)$? Let's just start with an element z of $S + T$. Then $z = x + y$ for $x \in S$ and $y \in T$. But

$$x \leq \sup S \qquad y \leq \sup T$$

so $z \leq \sup S + \sup T$.

This is true for all $z \in S+T$, so $\sup S+\sup T$ is an upper bound on $S+T$. So it must be at least the *least* such upper bound, which is $\sup(S+T)$. Therefore $\sup(S+T) \leq \sup S+\sup T$.

Proof of $\geq$: Now we need to upper bound $\sup S + \sup T$. Well, let's start with an $x \in S$ and a $y \in T$. I know that $x + y$ is in $S + T$, so

$$x + y \leq \sup(S + T).$$

But now how to get sups on the left side? Here is the trick: If we move the y over, then

$$x \leq \sup(S+T) - y.$$

For a fixed y, this is true for all $x \in S$. So $\sup(S+T) - y$ is an upper bound on S, and moreover needs to be at least the *least* such upper bound. So

$$\sup S \leq \sup(S+T) - y.$$

Now move the y to the other side!

$$y \leq \sup(S+T) - \sup S$$

This holds for all $y \in T$, so $\sup(S+T) - \sup S$ is an upper bound on T, and is greater than or equal to the *least* such upper bound. We conclude that

$$\sup T \leq \sup(S+T) - \sup S$$

which is what we wanted. □

Remarks:

- In general, if you know that $x \leq M$ for all $x \in S$, you can "take the sup on the left" to get $\sup S \leq M$. We did this three times in the above proof.
- Similarly, if $x \geq m$ for all $x \in S$, we can "take the inf on the left" to get $\inf S \geq m$.
- Another way to do this is to make use the following characterization of sup: $M = \sup S$ iff (1) M is an upper bound on S, and (2) for all $\varepsilon > 0$, the intersection $(M - \varepsilon, M] \cap S$ is non-empty. (This means that there are elements of S that are arbitrarily close to M, which intuitively should mean that M is the least upper bound. As an exercise you can try to prove that this characterization of supremum is equivalent to the definition of supremum.)

Let's try another example.

Example 1.2: Let a_n and b_n be two sequences, both bounded from above. How do $\sup_n(a_n + b_n)$ and $\sup_n a_n + \sup_n b_n$ compare?

This might look like the same problem, but actually no: We only get $\leq$. A counterexample which shows why $\sup_n(a_n + b_n) = \sup_n a_n + \sup_n b_n$ may not necessarily hold is given by $a_n = (-1)^n$ and $b_n = -(-1)^n$.

Let's prove that $\sup_n(a_n + b_n) \leq \sup_n a_n + \sup_n b_n$.

Proof. Let's start with some $a_n + b_n$. Then

$$a_n \leq \sup_k a_k$$

and

$$b_n \leq \sup_k b_k$$

so

$$a_n + b_n \leq \sup_k a_k + \sup_k b_k.$$

This means that $\sup_k a_k + \sup_k b_k$ is an upper bound on $a_n + b_n$, so by "taking the sup" on the left we conclude that

$$\sup_n (a_n + b_n) \leq \sup_k a_k + \sup_k b_k$$

as needed. □

Remark: Notice how I wrote $a_n \leq \sup_k a_k$ instead of $a_n \leq \sup_n a_n$. Using different letters helps a lot to prevent confusion.

1.2 The Intuition of Limsup

You should think of limsup a_n as:

- the "best upper bound on the asymptotic behavior of a_n as $N \to \infty$"
- the "upper bound on a_n **near** $n = \infty$"
- the best upper bound on the **tail** of a_n

Whereas lim is about the exact asymptotic, $\limsup$ is only an upper bound on the asymptotic (when the limit doesn't exist), whereas $\liminf$ is only a lower bound.

Limsup and liminf are really useful when we want to discuss the asymptotic behavior of a function or sequence, but the limit doesn't actually exist!

You can think of $f(x) = \sin x$ as a prototypical example for limsup and liminf. $\lim_{x\to\infty} \sin x$ does not exist. But "in the limit it's between -1 and 1", in other words,

$$-1 \leq \liminf_{x\to\infty} \sin x \leq \limsup_{x\to\infty} \sin x \leq 1.$$

In fact $\liminf_{x\to\infty} \sin x = -1$ and $\limsup_{x\to\infty} \sin x = 1$, and these are "obtained" because $\sin x = -1$ infinitely often for large x and $\sin x = 1$ infinitely often for large x.

1.2.1 Definitions

A priori the limsup is defined as

$$\limsup_{n\to\infty} a_n = \inf_n \sup_{k>n} a_k.$$

However, you should know that it is equivalent to write

$$\lim_{n\to\infty} \sup_{k>n} a_k$$

because $\{\sup_{k>n} a_k\}_n$ is monotone decreasing in n! This monotonicity is important to keep in mind.

Limsup can be defined in other contexts where I think it's easier to think about. For example, for functions,

$$\limsup_{x\to\infty} f(x) = \lim_{N\to\infty} \sup_{x>N} f(x),$$

and

$$\limsup_{x\to x_0} f(x) = \lim_{\delta\to 0^+} \sup_{0<|x-x_0|<\delta} f(x).$$

See the following Desmos visualizations:

- https://www.desmos.com/calculator/tiwrgpoa0x
- https://www.desmos.com/calculator/va3cdionyv

These visualizations should really help show you what limsup is intuitively.

1.2.2 Examples of why limsup is useful

Let's try to prove the squeeze rule!

Example 1.3: Suppose that $g_1(x) \le f(x) \le g_2(x)$, and both $\lim_{x\to x_0} g_1(x)$ and $\lim_{x\to x_0} g_2(x)$ exist and are equal to L. Prove that $\lim_{x\to x_0} f(x) = L$.

Proof. [WRONG PROOF] Just take $g_1(x) \le f(x) \le g_2(x)$ and take the limit of all three parts, to get $L \le \lim_{x\to x_0} f(x) \le L$, so $\lim_{x\to x_0} f(x) = L$. Tada? □

This is **very very very very very very very very wrong** because I don't actually know that $\lim_{x\to x_0} f(x)$ exists in the first place! So this is very bad and horrible and terrible.

...

But I *do* know that $\limsup_{x\to x_0} f(x)$ and $\liminf_{x\to x_0} f(x)$ exist. Because they *always* exist.

Proof. [Actual Proof] Taking limsup on both sides of the right inequality and liminf on both side of the left inequality, we get the following for free:

$$\liminf_{x\to x_0} g_1(x) \le \liminf_{x\to x_0} f(x) \le \limsup_{x\to x_0} f(x) \le \limsup_{x\to x_0} g_2(x)$$

Ok but, $\lim_{x\to x_0} g_1(x)$ exists, so $\liminf_{x\to x_0} g_1(x) = \lim_{x\to x_0} g_1(x) = L$... and similarly, we know that $\limsup_{x\to x_0} g_2(x) = \lim_{x\to x_0} g_2(x) = L$. So actually this is just saying that:

$$L \le \liminf_{x\to x_0} f(x) \le \limsup_{x\to x_0} f(x) \le L$$

So the liminf and limsup of f were equal, and in fact both are equal to L, so the limit exists and is L. Yay! □

Motto of the above proof: "The asymptotics of f are bounded from above by L and bounded from below by L, so the limit exists."

Here's another application. The ratio test says that if

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

then $\sum_{n=1}^{\infty} a_n$ converges. At first this seems quite intimidating if you're still getting used to limsup. But intuitively this statement is actually quite simple! In English, all it's saying is this: "If a series **eventually goes down faster than a geometric series**, then it converges!" The "eventually" part is the "lim", and the "faster" is the "sup".

In Layman's terms, a series which (eventually) converges faster than a geometric series must be convergent. That's all! The limsup formalizes this statement.

1.3 A limsup exercise

Example 1.4: Let a_n and b_n be sequences. Show that

$$\limsup_n (a_n + b_n) \le \limsup_n a_n + \limsup_n b_n$$

provided that both sides exist and are finite.

Proof. Since the limsup is the lim of a sup, let's start by working with the "inner-most operation", which is sup. Can we compare these two quantities?

$$\sup_{k>n}(a_k + b_k), \qquad \sup_{k>n} a_k + \sup_{k>n} b_k$$

It turns out we can! We proved that the correct relationship is $\leq$ in the first section. So

$$\sup_{k>n}(a_k + b_k) \leq \sup_{k>n} a_k + \sup_{k>n} b_k$$

for all n. Now we want to "send $n \to \infty$ on each side". This is safe because all the limits exist. (More precisely, we want to appeal to the fact that if $x_n \to K$ and $y_n \to L$ and $x_n \leq y_n$ for all n, then $K \leq L$. If this wasn't done in lecture, try to prove it!) So

$$\lim_{n\to\infty} \sup_{k>n}(a_k + b_k) \leq \lim_{n\to\infty} \left(\sup_{k>n} a_k + \sup_{k>n} b_k \right)$$
$$\leq \lim_{n\to\infty} \sup_{k>n} a_k + \lim_{n\to\infty} \sup_{k>n} b_k.$$

That's exactly what we wanted to prove!

□

1.4 Rigorous Write-up for the Quiz

Not done in recitation but I thought I should include this so you have a good sense of what is expected.

Theorem 1.1 (1D Cantor Intersection)

Let $\{[a_n, b_n]\}_n$ be a sequence of closed intervals such that $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ for all $n \geq 1$. Then the intersection

$$\bigcap_{n=1}^{\infty} [a_n, b_n]$$

is non-empty. Moreover, if $\lim_{n\to\infty} b_n - a_n = 0$, then the intersection has a single element.

Proof. Since $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ for all $n \geq 1$, we have that

$$a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$$

for all $n \geq 1$. In particular, $a_n \leq b_n \leq b_1$ for all n, which entails that b_1 is an upper bound on $\{a_n : n \in \mathbb{N}\}$. Thus the supremum $M := \sup_n a_n$ exists.

We claim that $M \in \bigcap_{n=1}^{\infty}[a_n, b_n]$, which will prove that the intersection is non-empty. It suffices to prove that for all n, $a_n \leq M \leq b_n$.

Fix n. Since $M = \sup_k a_k$, M is an upper bound on $\{a_k : k \in \mathbb{N}\}$, so $M \geq a_n$.

On the other hand, $a_k \leq b_n$ for all k (if $k \leq n$ then $a_k \leq a_n \leq b_n$, and if $k > n$ then $a_k \leq b_k \leq b_n$), so b_n is an upper bound on $\{a_k : k \in \mathbb{N}\}$. So b_n is $\geq$ the least such upper bound (the supremum), which is M. That is, $b_n \geq M$. This proves the claim.

Now assume that $\lim_{n\to\infty} b_n - a_n = 0$. To prove that $\bigcap_{n=1}^{\infty}[a_n, b_n]$ has exactly one element, it is sufficient to show that if $x, y \in \bigcap_{n=1}^{\infty}[a_n, b_n]$, then $x = y$.

Take such an x and y, and assume without loss of generality that $x \leq y$. Note that $a_n \leq x$ for all n, and $y \leq b_n$ for all n. So

$$|x - y| = y - x \leq b_n - a_n$$

for all n. Sending $n \to \infty$, and using the hypothesis that $b_n - a_n \to 0$, we conclude by the squeeze theorem (or by fixing $\varepsilon > 0$ or by using $\liminf$, etc.) that $|x - y| = 0$. That is, $x = y$. □

2 Series and Stuff

2.1 Comparing series and integrals

Draw a picture! (I'm too tired to reproduce a diagram in these notes, sorry! The key idea is that the series is basically a Riemann sum so draw that and the area under the curve to compare them.)

Example 2.1: Consider the series $s_n = \sum_{k=1}^{n} \frac{1}{k^2}$. Get a decent lower and upper bound on s_n. In particular can we show that $s_n \leq 2$?

Solution. By drawing a picture, we can reason that

$$\int_{k}^{k+1} \frac{1}{x^2}\,dx \leq \frac{1}{k^2} \leq \int_{k-1}^{k} \frac{1}{x^2}\,dx$$

for all k. Now let's sum this starting from $k = 2$ (we're skipping $k = 1$ because otherwise the integral on the right side explodes). This gives

$$1 + \int_{2}^{n+1} \frac{1}{x^2}\,dx \leq s_n \leq 1 + \int_{1}^{n} \frac{1}{x^2}\,dx.$$

Evaluating the integrals,

$$1.5 - \frac{1}{n+1} \leq s_n \leq 2 - \frac{1}{n}.$$

That seems like a decent bound! ■

2.2 Estimate for the Factorial

We can use the integral technique to obtain some pretty nice bounds for $n!$. First we need to turn this into a sum, so we'll instead estimate its log:

$$\log(n!) = \sum_{k=1}^{n} \log k$$

Now log is increasing and so for each k we have the estimates

$$\int_{k-1}^{k} \log x\,dx \leq \log k \leq \int_{k}^{k+1} \log x\,dx.$$

We want to sum from $k = 1$ to $k = n$, but the $k = 1$ term makes the left bound problematic. Instead we sum from $k = 2$, which is just as good because the $k = 1$ term is $\log 1 = 0$.

$$\int_{1}^{n} \log x\,dx \leq \sum_{k=2}^{n} \log k \leq \int_{2}^{n+1} \log x\,dx$$

Evaluating the integrals,

$$n \log n - n + 1 \le \log(n!) \le (n+1)\log(n+1) - 2\log 2 - n + 1.$$

Exponentiating, we end up with

$$n^n e^{-n+1} \le n! \le \frac{1}{4}(n+1)^{n+1} e^{-n+1}.$$

This suggests that $n!$ grows roughly like $n^n e^{-n}$. It turns out that the correct asymptotic is $n! \sim C n^{n+\frac{1}{2}} e^{-n}$ where the constant C is $\sqrt{2\pi}$. So, we got pretty close with a relatively elementary method!

2.3 A few exercises with series tests

Example 2.2: Find all $x \in \mathbb{R}$ for which

$$\sum_{n=1}^{\infty} \frac{nx^n}{n^2 + x^{2n}}$$

converges.

Solution. We claim that it converges for all $x \neq 1$.

If $|x| > 1$, we have that

$$\sum_{n=1}^{\infty} \frac{n|x|^n}{n^2 + x^{2n}} \le \sum_{n=1}^{\infty} \frac{n|x|^n}{x^{2n}} = \sum_{n=1}^{\infty} \frac{n}{|x|^n}$$

which converges by the ratio test, so by comparison, the series converges absolutely, and thus converges.

If $|x| < 1$, we instead write

$$\sum_{n=1}^{\infty} \frac{n|x|^n}{n^2 + x^{2n}} \le \sum_{n=1}^{\infty} \frac{n|x|^n}{n^2} = \sum_{n=1}^{\infty} \frac{|x|^n}{n} \le \sum_{n=1}^{\infty} |x|^n < \infty.$$

If $x = -1$, the series converges by the alternating series test.

If $x = 1$, then the series is

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

This is comparable to $1/n$ in the limit so this should diverge by limit comparison. Alternatively one can write something like

$$\sum_{n=1}^{\infty} \frac{n}{n^2+1} \geq \sum_{n=1}^{\infty} \frac{n}{n^2+n^2} = \frac{1}{2}\sum_{n=1}^{\infty} \frac{1}{n} = +\infty.$$

■

Example 2.3: Does $\sum_{n=1}^{\infty} \sin(1/n^2)$ converge?

Solution. Yes, by using $\sin x \leq x$ (for $x > 0$) and comparison. ■

Example 2.4: Does $\sum_{n=1}^{\infty} \frac{1}{n\log(n)^2}$ converge?

Solution. Yes by the integral test. If we want to be a bit more precise, we can write, for $n > 1000$,

$$\frac{1}{n\log(n)^2} \leq \int_{n-1}^{n} \frac{1}{x\log(x)^2}\,dx.$$

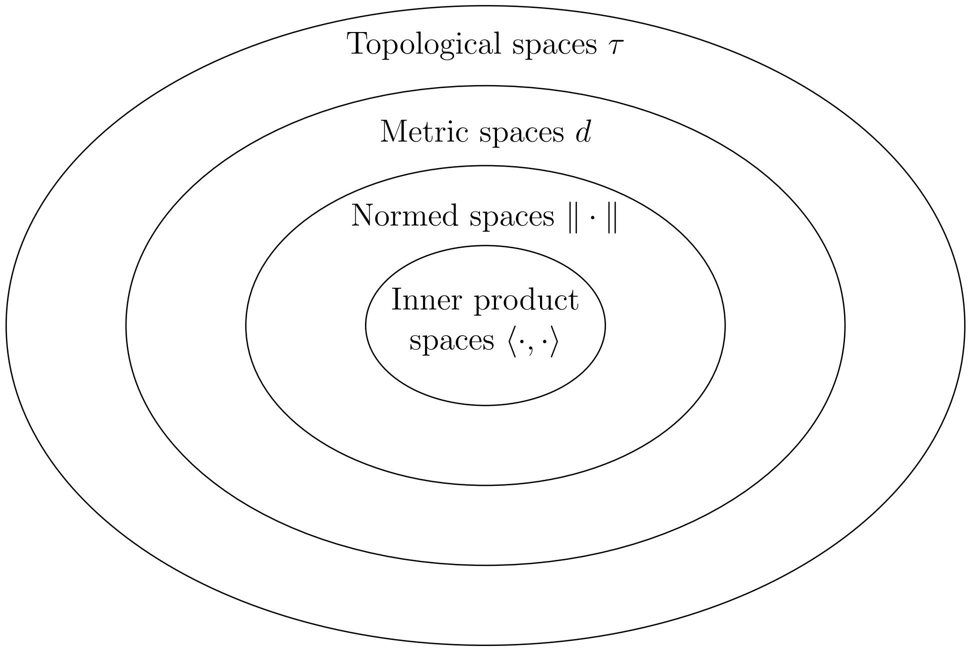
So

$$\sum_{n=10000}^{\infty} \frac{1}{n\log(n)^2} \leq \int_{9999}^{\infty} \frac{1}{x\log(x)^2}\,dx.$$

By doing calculus (u-sub with $u = \log(x)$), the integral converges, hence so does the sum. ■

3 The Analysis Hierarchy

The bulk of (undergraduate) analysis lives in the following hierarchy of spaces.



3.1 Inner Product Spaces

These are the most specialized spaces, and are quite uncommon. Some very nice things happen in inner product spaces, but I don't think we'll dive very deeply into it. You saw a bunch of it in a class called linear algebra.

Definition 3.1 (Inner Product Space)

An *inner product space* is a **vector space** X equipped with an **inner product** $\langle\cdot,\cdot\rangle$ satisfying a bunch of properties, such as $\langle tx, y\rangle = t\langle x, y\rangle$ for all $x, y \in X$ and $t \in \mathbb{R}$. (there are more conditions of course but whatever you can look it up.)

Intuitively you can think of an inner product space as "a space with a notion of **angles**". The inner product $\langle x, y\rangle$ measures how much x and y "agree".

Examples:

- Euclidean space (a.k.a. $\mathbb{R}^N$) equipped with the standard inner product, $\langle x, y\rangle := x \cdot y = \sum_j x_j y_j$.

- The sequence space l^2, consisting of all sequence of real numbers $\{x_n\}_n$ such that $\sum_n |x_n|^2 < \infty$, equipped with the inner product $\langle \{x_n\}_n, \{y_n\}_n \rangle_{l^2} := \sum_n x_n y_n$. Some work needs to be done to show that this is valid.

- The space $L^2(\mathbb{R})$ of all square-integrable "functions" $f : \mathbb{R} \to \mathbb{R}$, equipped with the inner product
$$\langle f, g \rangle_{L^2} := \int_{\mathbb{R}} f(x)g(x)\,dx.$$
(This is a lie, hence the quotes.)

3.2 Normed Spaces

Normed spaces are much more common. A lot of analysis happens here.

Definition 3.2 (Normed Space)

A *normed space* is a **vector space** X equipped with a **norm** $\|\cdot\|$ satisfying the following conditions:

- $\|x\| \geq 0$ always, with $\|x\| = 0$ if and only if $x = 0$. (also its always finite.)
- You can take out scalars: $\|tx\| = |t| \cdot \|x\|$ for all $t \in \mathbb{R}$ and $x \in X$.
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

Simply stated: A normed space is a space that has a notion of **size**.

Every inner product space is a normed space. This is because any inner product $\langle \cdot, \cdot \rangle$ induces a norm, given by
$$\|x\| := \sqrt{\langle x, x \rangle}.$$
Compare this with the identity $\|x\| = \sqrt{x \cdot x}$ in Euclidean space.

Examples

- On $\mathbb{R}$, the absolute value $|\cdot|$ is a norm.
- On $\mathbb{R}^N$, the usual norm is $\|x\| := \sqrt{\sum_j x_j^2}$. Of course, this is just given by the usual inner product as $\sqrt{x \cdot x}$.
- For $p \geq 1$, there is the L^p space, $L^p(\mathbb{R})$, of all "functions" $f : \mathbb{R} \to \mathbb{R}$ satisfying
$$\int_{\mathbb{R}} |f(x)|^p\,dx < \infty,$$

and the norm is given by $\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p \, dx\right)^{1/p}$.

Questions: (1) Why do we need to raise to the $1/p$ power? (2) Why is this a lie as written?

- There are other norms on $\mathbb{R}^N$. These are not the ones induced by the usual inner product, but they are still norms and can be useful.
  - The *taxicab norm*, $\|x\|_1 := \sum_j |x_j|$.
  - The l^∞ *norm*, $\|x\|_\infty := \max_j |x_j|$.
  - In general, for $1 \leq p < \infty$, we have the p-norm, $\|x\|_p := \left(\sum_j |x_j|^p\right)^{1/p}$. The case $p = 2$ is the standard Euclidean norm.

3.3 Metric Spaces

Metric spaces are *very* common! They arise whenever there is some notion of "distance".

Definition 3.3 (Metric Space)

A *metric space* is a **set** X equipped with a **metric** d satisfying the following conditions:

- $0 \leq d(x, y) < \infty$, with $d(x, y) = 0$ if and only if $x = y$
- Symmetry: $d(x, y) = d(y, x)$
- Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$

Note that the underlying set X no longer needs to have a vector space structure for it to qualify as a metric space. You just need "distance".

Every normed space is a metric space. If you have a norm $\|\cdot\|$, then the "distance" between x and y is given by $\|x - y\|$. You can verify that this satisfies the properties of a metric, so $d(x, y) := \|x - y\|$ is a metric for any norm $\|\cdot\|$.

Examples:

- X = the tiles of a Civilization 6 board, $d(x, y)$ = number of moves it takes to go from x to y.
- X = Warren Weaver Hall, $d(x, y)$ = time it takes to walk from x to y.
- $\mathbb{R}^N$ is a metric space induced by the standard norm (or really, any of its norms).
- X = the set of all English words, $d(x, y)$ = the number of single-letter edits you need to turn x into y.

3.3.1 Balls and Open Sets

We play with metric spaces a lot, so we've got a bunch of constructs in them that we study (which, of course, also exist in normed spaces). The first is the **ball**,

$$B(x, r) := \{y : d(x, y) < r\}.$$

Balls are important because we think of them as "neighborhoods" which surround a point x. They're also just a very convenient notation in general.

With balls, we obtain **open sets**: Sets $U \subseteq X$ such that for every $x \in U$ there is a small $r > 0$ such that $B(x, r) \subseteq U$. In layman's terms, opens sets are the sets with "wiggle room" everywhere. This intuition is why we like them: It allows us to make small changes in any direction when we're inside it.

We also get **closed sets**, which are defined to be the complements of open sets. (Closed does NOT mean "not open"!)

Then we get **compact sets**, which are defined as those sets for which every *open cover* has a finite subcover. In $\mathbb{R}^N$, these happen to be the closed and bounded sets, and therefore we can think of them as sets which are *restrictive*, and *prevent too much "change"*. An instance of this is the extreme value theorem: Any continuous function on a closed and bounded interval must have a max and min. The closed and bounded interval here is a compact set, and we see that it is preventing the continuous function on it from "exploding". More on compact sets in future recitations.

The most important types of sets are the open sets and the compact sets, for the reasons described.

3.3.2 Properties of open sets

- The union of open sets is always open (no matter how many open sets are used!).
- The **finite** intersection of open sets is open.
- $\{\}$ is open.
- The whole space is open.

3.3.3 The Shit Metric

I define the *shit metric space* as follows: It is $(\mathbb{R}^2, d_{\text{shit}})$, where

$$d_{\text{shit}}(x, y) := \begin{cases} 1, & x \neq y \\ 0, & x = y \end{cases}.$$

Verify that this is a metric!

The actual name for this is the discrete metric, but whatever. This is an important "extreme" case to keep in mind, as it can serve as a convenient counterexample to test any conjectures you have about metric spaces. We'll talk more about that in future weeks, probably.

Verify the following:

- $B_{\text{shit}}((0,0), 0.75) = \{(0,0)\}$
- $B_{\text{shit}}((0,0), 1.2) = \mathbb{R}^2$
- Every set is open.
- Every set is closed.
- Every set is bounded.
- The only compact sets are finite sets.

3.4 Topologies

Now what if there is no notion of distance? There is an even weaker structure called a *topology*: Instead of a distance, we specify what the open sets are.

Definition 3.4 (Topological space)

A *topological space* is a set X equipped with a *topology*, τ, which is a collection of sets (the "open sets") satisfying the properties listed in the *Properties of open sets* section.

It's quite abstract and it's normal to find it hard to imagine what kind of structure a topology could bring to the table. Here are some intuitions:

- Topologies specify the "neighborhoods", which therefore give some loose notion of which points are "close" to each other.
- A topology can be thought of as specifying a *notion of convergence*. For example, the topology on $\mathbb{R}$ "generated" by the intervals of the form $(a, b]$ corresponds to "convergence from below", and in some sense formalizes the concept of the left-sided limit $\lim\limits_{x \to a^-} f(x)$.

And of course, every metric space is a topological space.

4 Limits and Stuff

4.1 Be suspicious!

If your proof isn't relying on the definitions and/or theorems, that's probably a bad sign!

Problem 1: Let A and B be compact subsets of $\mathbb{R}^n$. Show that $A + B := \{a + b : a \in A, b \in B\}$ is compact.

Fake Proof: Since A and B are bounded, $A + B$ is also bounded. Since A and B are closed, $A + B$ is also closed. So by Heine-Borel, $A + B$ is compact. □

Problem 2: Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Suppose that $f'(a) > 0$ for some $a \in \mathbb{R}$, and $f'(b) < 0$ for some $b \in \mathbb{R}$. Then there exists $c \in \mathbb{R}$ such that $f'(c) = 0$.

Fake Proof: f is differentiable everywhere, so its derivative f' is continuous. So by the Intermediate Value Theorem, f' has a zero. □

Problem 3: Consider the normed vector space $C_b(\mathbb{R})$ of all bounded continuous functions on the real line, equipped with the sup norm. Then the function $F : C_b(\mathbb{R}) \to \mathbb{R}$ defined by

$$F(f) := f(5)$$

is continuous.

Fake Proof: Note that for any $f, g \in C_b(\mathbb{R})$, we have $F(f + g) = F(f) + F(g)$. Moreover, for any $t \in \mathbb{R}$ and $f \in C_b(\mathbb{R})$, we have that $F(tf) = tF(f)$. Therefore, $F : C_b(\mathbb{R}) \to \mathbb{R}$ is linear. So F is continuous. □

4.2 Limits and Continuity

Example 4.1:

$$\lim_{x \to 5} x^2$$

Solution. We claim that the limit is 25.

Fix $\varepsilon > 0$. And we'll pick $\delta =$ (TBD). Now, if $|x - 5| < \delta$, we shall show that $|x^2 - 25| < \varepsilon$.

Well, we can write

$$|x^2 - 25| = |(x - 5)(x + 5)| < \delta \cdot |x + 5|.$$

That δ will be pretty small so we just need the $|x + 5|$ to be pretty small. It should be about

10, right? Indeed,

$$|x+5| = |x-5+10| \le |x-5| + |10| < \delta + 10.$$

So provided $\boxed{\delta \le 7}$, we have that $|x+5| < 17$. Now,

$$|x^2 - 25| < \delta \cdot |x+5| < 7\delta,$$

which is $\le \varepsilon$ provided that $\boxed{\delta \le \varepsilon/7}$. Ok, so since we chose $\boxed{\delta = \min(7, \varepsilon/7)}$, both of those hold and so we're good. ■

Example 4.2:

$$\lim_{x \to 42} 1_{\mathbb{Q}}(x)$$

Solution. We claim that the limit does not exist.

Suppose the limit were L. Take $\varepsilon = 1/4$ or something. We want to show that for any $\delta > 0$, there is some x with $0 < |x - 9001| < \delta$ such that $|1_{\mathbb{Q}}(x) - L| \ge \varepsilon$.

To wit, take any $\delta > 0$. There are two cases.

- If $L \ge 1/2$, we use density of irrationals to pick $x \in (9001, 9001 + \delta)$ irrational. This gives $1_{\mathbb{Q}}(x) = 0$ so $|1_{\mathbb{Q}}(x) - L| \ge 1/2 \ge \varepsilon$.
- If $L < 1/2$, use density of rationals instead.

So the limit does not exist. ■

Example 4.3: Do Problem 3, but correctly.

Solution. We recall $F : C_b(\mathbb{R}) \to \mathbb{R}$ with $F(f) := f(5)$. Fix $f_0 \in C_b(\mathbb{R})$. We show that F is continuous at f_0.

Fix $\varepsilon > 0$. Then, for all $f \in C_b(\mathbb{R})$ with $\|f - f_0\|_\infty < \delta$, where δ shall be chosen later, we will show that $|F(f) - F(f_0)| < \varepsilon$.

Well, we want to show that $|f(5) - f_0(5)| < \varepsilon$. But

$$\delta > \|f - f_0\|_\infty = \sup_{x \in \mathbb{R}} |f(x) - f_0(x)| \ge |f(5) - f_0(5)|.$$

So we may take $\delta = \varepsilon$. ■

Example 4.4: Is $C_b(\mathbb{R})$ complete?

4.3 Exercises with Metric Spaces

Example 4.5: Let (X, d) be a metric space, and $C \subseteq K \subseteq X$ with C closed and K compact. Prove that C is compact.

This can be done using either sequences or open covers.

Example 4.6: Let (X, d) be a metric space, and $E \subseteq X$. Define the *boundary* of E as follows:

$$\partial E := \{x \in X : B(x, r) \cap E \neq \emptyset \text{ and } B(x, r) \cap E^c \neq \emptyset \ \forall r > 0\}$$

Show that ∂E is closed.

If $x \notin \partial E$ then there is $r > 0$ so that either $B(x, r) \cap E = \emptyset$ or $B(x, r) \cap E^c = \emptyset$. Without loss of generality let's suppose the former happened. Then $B(x, r) \subseteq E^c$. We claim $B(x, r) \subseteq (\partial E)^c$. Indeed, take any $y \in B(x, r)$. Then $B(y, \min(d(x, y), r - d(x, y))) \subseteq B(x, r) \subseteq E^c$ (or alternatively $B(x, r)$ is open so there's gotta be some r' so that $B(y, r') \subseteq B(x, r)$), so $B(y, \min(d(x, y), r - d(x, y))) \cap E = \emptyset$, so by definition of ∂E, $y \notin \partial E$.

5 Uniform stuff

5.1 Relative Topology

Example 5.1: Let (X, d) be a metric space, and $E \subseteq X$. Show that, for $F \subseteq E$, the following two conditions are equivalent:

1. There exists an open set $U \subseteq X$ such that $F = E \cap U$.
2. For every $x \in F$ there exists $r > 0$ such that $B(x, r) \cap E \subseteq F$.

(If either holds, we say that F is ***relatively open*** *in E, or just "open in E". Morally speaking, we are interpreting E as its own metric space, whose topology is inherited from that of X by taking all the open sets in X and intersecting them with E to form a new "restricted" topology. Think about both the conditions (1) and (2) and see which one you find more intuitive.)*

Solution. $(1 \implies 2)$ easy

$(2 \implies 1)$ for each $x \in F$ find r_x so that $B(x, r_x) \cap E \subseteq F$. Now take $U = \bigcup_{x \in F} B(x, r_x)$. This is open and you can show that $F = E \cap U$. ■

The relative topology is important because it's how you restrict a topology to a subset.

For example, we know that for $U \subseteq \mathbb{R}^n$ open, we have $f : U \to \mathbb{R}$ continuous iff $f^{-1}(V)$ open for all $V \subseteq \mathbb{R}$ open. However, this is not true if the domain is not open. Take, for example, $E = [0, \infty)$ and $f : E \to \mathbb{R}$ defined by

$$f(x) = x^2 \sin x.$$

Then $f^{-1}(\mathbb{R}) = [0, \infty)$ but $[0, \infty)$ is not open.

However, $[0, \infty)$ is *relatively open* (with respect to itself). This leads to a better, more general characterization of continuity:

$f : E \to \mathbb{R}$ is continuous iff $f^{-1}(V)$ is open in the relative topology of E, for every $V \subseteq \mathbb{R}$ open.

5.2 Uniform convergence: In the field

Example 5.2: Prove that

$$f(x) := \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^2 + |x|}$$

is a continuous function.

Motto: To show that $f_n \to f$ uniformly, find an upper bound

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

which does **not depend on** x. And such that $M_n \to 0$.

The methods used for this generalize.

Theorem 5.1 (Weierstrass-M test)

Let $f_n : E \to \mathbb{R}$, and suppose $\sum_{n=1}^{\infty} \sup_E |f_n| < \infty$. Then $\sum_{n=1}^{\infty} f_n$ converges uniformly.

Have a slightly trickier example.

Example 5.3: Prove that

$$f(x) := \sum_{n=1}^{\infty} \frac{x^{10}}{n^2 + x^2}$$

is continuous.

Key idea: suffices to show uniform convergence on compact sets.

5.3 Exploring Continuity and Completeness

Lemma 5.1

Uniformly continuous functions send Cauchy sequences to Cauchy sequences.

Proof. Let the uniformly continuous function be $f : X \to Y$, where (X, d_X) and (Y, d_Y) are metric spaces. Let $\{x_n\}_n \in X$ be Cauchy. We will show that $\{f(x_n)\}_n \in Y$ is Cauchy.

Fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that $d_Y(f(x), f(Y)) < \varepsilon$ for all $d_X(x, y) < \delta$. Since $\{x_n\}_n$ Cauchy, there exists N_δ such that $d_X(x_m, x_n) < \delta$ for all $m, n \geq N_\delta$. Now for all $m, n \geq N_\delta$,

$$d_Y(f(x_m), f(x_n)) < \varepsilon$$

due to $d_X(x_m, x_n) < \delta$ and uniform continuity. $\square$

Theorem 5.2

Let (X, d_X) and (Y, d_Y) be metric spaces, with Y complete. Let $E \subseteq X$. Let $f : X \to Y$ be a uniformly continuous function. Then there exists a unique continuous extension of f to $\overline{E}$.

Proof. I could only sketch this out in recitation, but here is the full proof.

It's not hard to see that it's unique, provided that it exists. So we just need to construct the extension and prove that it works. For each $x \in \overline{E}$, we define

$$\tilde{f}(x) := \lim_{n\to\infty} f(x_n),$$

where $x_n \in E$ is a sequence for which $x_n \to x$.

First we show that this is well-defined, because there are two problems with the definition: (1) the limit may not exist, (2) we need to ensure that the definition of $\tilde{f}(x)$ does not depend on the choice of the sequence.

(1) Let $x_0 \in \overline{E}$, and pick $x_n \to x_0$. $\{x_n\}_n$ converges so it is Cauchy. By the lemma, $\{f(x_n)\}_n$ is also Cauchy. But Y is complete, so it converges, hence the limit exists.

(2) Let $x_0 \in \overline{E}$, and pick $x_n \to x_0$. If $x'_n \in E$ is another sequence for which $x'_n \to x_0$, then we want to show $\lim_n f(x_n) = \lim_n f(x'_n)$. By the argument from (1), both limits exist. Let's say the limits are L and L', both elements of Y. Fix $\varepsilon > 0$. Then we have that $d_Y(f(x, y)) < \varepsilon$ for all $x, y \in E$ for which $d_X(x, y) < \delta$. For all n large enough, we have $d_X(x_n, x) < \delta/2$ and $d_X(y_n, x) < \delta/2$. Now $d_X(x_n, y_n) < \delta$, so

$$d_Y(f(x_n), f(y_n)) < \varepsilon.$$

Hence

$$\begin{aligned} d_Y(L, L') &\le d_Y(L, f(x_n)) + d_Y(f(x_n), f(x'_n)) + d_Y(f(x'_n), L') \\ &< \varepsilon + d_Y(L, f(x_n)) + d_Y(f(x'_n), L'), \end{aligned}$$

and sending $n \to \infty$ gives $d_Y(L, L') \le \varepsilon$. But ε was arbitrary so $L = L'$.

We've proven that $\tilde{f}$ is well-defined. Remains to show two things: (1) $\tilde{f}$ is actually an extension of f (i.e. they agree on E), and (2) $\tilde{f}$ is continuous.

(1) Easy.

(2) It is equivalent to show that $\tilde{f}$ is sequentially continuous. Take $x_0 \in \overline{E}$ and let $x_n \to x_0$, $x_n \in E$. We'll show that $\tilde{f}(x_n) \to \tilde{f}(x_0)$. Wait, but $\tilde{f}(x_n) = f(x_n)$ because $x_n \in E$, so... $\square$

6 Things are getting kinda spicy

6.1 Baire Category Theorem

The picture drawn in recitation: Can a bunch of thin things (e.g. lines and curves) come together to make a thick thing (e.g. the interior of a circle)?

- Thin things = closed sets with no interior
- Thick thing = something with an interior

Theorem 6.1 (Baire Category Theorem)

In a complete metric space, the union of countable many thin things can never be thick.

That is, if (X, d) complete, $C_n \subseteq X$ closed with empty interior, then $\bigcap_{n=1}^{\infty} C_n$ has empty interior.

That's literally it lol. I don't have very many good applications of this so that's that. The useful applications show up in functional analysis.

6.2 Ascoli-Arzela

Here is a simple instance of Ascoli-Arzela.

Theorem 6.2 (Ascoli-Arzela)

Let $f_n : [a, b] \to \mathbb{R}$ be a (continuous) sequence of functions which is

1. *uniformly bounded* (i.e. there is a constant $C > 0$ such that $|f_n(x)| \leq C$ for all x and n), and
2. *uniformly equi-continuous* (i.e. for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|f_n(x) - f_n(y)| < \varepsilon$ for all n, x, y for which $|x - y| < \delta$).

Then there exists a subsequence f_{n_k} which converges uniformly!

A quick example: Let $\mathrm{Lip}(f)$ be the best Lipschitz constant for f, i.e.

$$\mathrm{Lip}(f) := \sup_{x,y} \frac{|f(x) - f(y)|}{|x - y|}.$$

(Think of this as the "steepest slope on the graph".)

(Equivalently, $\mathrm{Lip}(f) = \inf\{L > 0 : |f(x) - f(y)| \leq L|x - y|$ *for all* $x, y\}$*. I leave it as an exercise to show that this is indeed equivalent!)*

Suppose f_n is a sequence of functions on $[1, 10]$ such that $\mathrm{Lip}(f_n) \leq 17$. Which just means that

$$|f_n(x) - f_n(y)| \leq 17|x - y|$$

for all x, y, n. Then $\{f_n\}_n$ is an example of a uniformly equi-continuous family of functions. (For any $\varepsilon > 0$ I can take $\delta = \varepsilon/17$ and this works for all x, y, and n.)

If we further assume something like, $f_n(0) = 0$ for all n, then

$$|f_n(x)| = |f_n(x) - f_n(0)| \leq 17|x - 0| \leq 170,$$

which means that f_n is uniformly bounded (by 170)! Therefore, Ascoli-Arzela applies and we get a subsequence f_{n_k} which converges uniformly.

6.3 Application of Ascoli-Arzela

Ascoli-Arzela is extremely important because it gives a notion of compactness for spaces of continuous functions. By "compactness", we philosophically mean "the technique of taking convergent subsequences".

Compactness / subsequences is incredibly useful for studying minimization problems.

Example 6.1: Use sequences to prove the extreme value theorem: If $f : K \to \mathbb{R}$ continuous and K is a compact metric space, then f has a maximum.

Proof. Let

$$M := \sup_{x \in K} f(x).$$

We want to show that the value of M is obtained. That is, we need to find x such that $f(x) = M$.

First we show that $M < \infty$ (not the important part here so I'm glossing over it). Then, this means that there exists a sequence x_n such that $f(x_n) \to M$ (i.e. we can approach the sup).

We don't know much about the sequence x_n. But we can apply compactness! This gives a subsequence x_{n_k} which converges to some $x \in [a, b]$.

By continuity,

$$f(n_k) \to f(x).$$

But $f(n_k) \to M$ so $f(x) = M$. So x obtains the maximum! ☐

In infinitely many dimensions, like in spaces of continuous functions, compactness is harder to come by, which makes it harder to take subsequences. Ascoli-Arzela is what allows us to take subsequences and make these kinds of arguments work! Here is an example.

Example 6.2 (Flappy Bird): Let $g, h : [0, 100] \to \mathbb{R}$ be functions with $g \leq h$, where we think of g being the "lower pipes" and h being the "upper pipes".
A bird starts at $(0, 0)$ and wants to end at $(100, 0)$ while dodging the pipes. That is, we are considering continuous functions $f : [0, 100] \to \mathbb{R}$ such that:

- The bird starts at $(0, 0)$, i.e. $f(0) = 0$
- The bird ends at $(100, 0)$, i.e. $f(100) = 0$
- The bird never hits the pipes, i.e. $g(x) \leq f(x) \leq h(x)$ for all $x \in [0, 100]$.

Prove that if this is possible to do, then there is an f satisfying these conditions for which $\text{Lip}(f)$ is **minimal**.
(motivation: we don't want the bird to have to fly too steeply up or down, it'll exhaust it for sure!)

Proof. Let A be the set of "legal paths" through the pipes, i.e. A = the set of all continuous $f : [0, 100] \to \mathbb{R}$ satisfying those three conditions points. The "theoretical minimum steepness", then, is

$$m := \inf_{f \in A} \text{Lip}(f).$$

The goal is to show that this infimum is actually a minimum. That is: Find an $f \in A$ such that $\text{Lip}(f) = m$.

Alright, let's start by trying to get close to m.

- Since we're given that the Flappy bird level is possible, it follows that A is non-empty.
- Moreover, $\{\text{Lip}(f) : f \in A\}$ is bounded from below by 0.
- By Week 1 stuff: Non-empty sets bounded from below have an infimum, so m exists and is finite.
- We can always get close to such an infimum: There exists $f_n \in A$ such that

$$\text{Lip}(f_n) \to m.$$

Great! We now have a sequence of Flappy bird paths $\{f_n\}_n$ whose "steepnesses" approach m. How can we procure a function whose "steepness" is *exactly* m...?

Wait hold on, here's an idea:

- Since $\text{Lip}(f_n) \to m$, we know that $\text{Lip}(f_n)$ is a bounded sequence! Let's say it's bounded by L.
- From before, we know that a sequence of functions with the same Lipschitz constant is uniformly equi-continuous!
- Moreover, since f_n is L-Lipschitz,

$$|f_n(x)| = |f_n(x) - f_n(0)| \leq L|x - 0| \leq 100L,$$

  so $\{f_n\}_n$ is uniformly bounded!

Therefore we can apply Ascoli-Arzela to find a subsequence f_{n_k} which converges uniformly to some $f : [0, 100] \to \mathbb{R}$.

Is this limit, f, a valid Flappy bird path?

- Since $f_{n_k} \to f$ uniformly, this convergence is also true pointwise. So, $f_{n_k}(0) \to f(0)$ and $f_{n_k}(100) \to f(100)$. This tells us that $f(0) = 0$ and $f(100) = 0$.
- Does f dodge the pipes? We know that

$$g(x) \leq f_{n_k}(x) \leq h(x)$$

  for all x and k. By pointwise convergence, we can send $k \to \infty$ to find that

$$g(x) \leq f(x) \leq h(x)$$

  for all x, which means that f dodges the pipes as well.
- Uniform convergence preserves continuity, so f is continuous.

So f is indeed a valid Flappy birth path (i.e. $f \in A$).

Finally, does f actually achieve the theoretical minimum steepness? Well, we know that

$$|f_{n_k}(x) - f_{n_k}(y)| \leq \text{Lip}(f_{n_k})|x - y|$$

because that's what Lip means. What happens when we send $k \to \infty$?

By pointwise convergence, the left side just converges to $|f(x) - f(y)|$. As for the right side: By how I chose $\{f_n\}_n$ at the start, I know that $\text{Lip}(f_{n_k}) \to m$! Thus,

$$|f(x) - f(y)| \leq m|x - y|.$$

So m is *at least* the steepness of f, $\text{Lip}(f)$. But m is the infimum of all possible steepnesses, i.e. $m \leq \text{Lip}(f)$. So $m = \text{Lip}(f)$, meaning that we've achieved the minimum possible steepness. □

7 Taylor Spam

7.1 L'Hopital sucks ngl

Example 7.1: Compute
$$\lim_{x\to\infty} \frac{x}{3x+\sin x}.$$

Why is applying L'Hopital illegal here?

7.2 Setting up Taylor

Definition 7.1

Fix some function g. We say that a function f is $o(g(x))$, as $x \to x_0$, if:

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = 0$$

Essentially, we use $o(g(x))$ as a "placeholder" for some function/expression that vanishes when divided by $g(x)$ (and sending $x \to x_0$).

Examples (as $x \to 0$):

- x^2 is $o(x)$ as $x \to 0$. That's because $\frac{x^2}{x} \to 0$ as $x \to 0$.
- x is NOT $o(x)$.
- $\sin(x)$ is $o(1)$. That's because $\sin(x)/1 \to 0$.
- $\sin(x)$ is NOT $o(x)$. Recall that $\sin(x)/x \to 1$, not 0.
- We can write $\frac{o(x^5)}{x^2}$ as $o(x^3)$. This is because
$$\frac{o(x^5)/x^2}{x^3} = \frac{o(x^5)}{x^5} \to 0$$
by definition of $o(x^5)$.
- We can write $o(o(x))$ as $o(x)$. (Why?)

Theorem 7.1 (Taylor)

Let $f : [a, b] \to \mathbb{R}$ be n-times differentiable. Then:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x - x_0)^2}{2} + \ldots + f^{(n)}(x_0)\frac{(x - x_0)^n}{n!} + o((x - x_0)^n)$$

Remember, by definition of little-o, this $o((x - x_0)^n)$ thing is a placeholder for an expression that satisfies the key property $\lim_{x\to x_0} \frac{o((x - x_0)^n)}{(x - x_0)^n} = 0$.

Key Point: Taylor expansion around x_0 is used to get an *approximation* to f near $x = x_0$. The remainder term $o((x - x_0)^n)$ tells you how good this approximation is.

Example: We have

$$\sin x = 0 + x + 0 \cdot x^2 - \frac{1}{6}x^3 + 0 \cdot x^4 + o(x^4).$$

This means that $\sin(x)$ is approximately $x - \frac{1}{6}x^3$ near $x = 0$. This approximation is *very good*: The difference goes to 0 at a rate faster than x^4. For example, at $x = 0.1$, we could expect that the error of this approximation may be around 0.0001.

7.3 Taylor: In the Field

The Taylor expansion

$$f(x) = f(a) + f'(a)(x - a) + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + o((x - a)^n)$$

gives an *order-n approximation near $x = a$.* The more terms you write out, the better the approximation.

Example 7.2:

$$\lim_{x\to 0} \frac{\sin(x) - x}{x^3}$$

Solution. Write $\sin x = x - x^3/6 + o(x^4)$. This is an EQUALITY. I can replace $\sin x$ with this "approximation" and you can't stop me.

$$\lim_{x\to 0} \frac{\sin x - x}{x^3} = \lim_{x\to 0} \frac{x^3/6 + o(x^4)}{x^3} = \lim_{x\to 0} \frac{1}{6} + \frac{o(x^4)}{x^3}$$

But $\frac{o(x^4)}{x^3} \to 0$ because $x \to 0$ and $\frac{o(x^4)}{x^4} \to 0$ so their product $\frac{o(x^4)}{x^4}$ goes to 0 too. So the limit is $\boxed{1/6}$. ■

Example 7.3:

$$\lim_{x\to 0}\left(\frac{\sin x}{x}\right)^{1/x}$$

Solution. Write $\sin x = x + o(x^2)$ (that's all we need). Then:

$$\lim_{x\to 0}\left(\frac{\sin x}{x}\right)^{1/x} = \exp\left(\lim_{x\to 0}\frac{\log(\sin(x)/x)}{x}\right)$$

$$= \exp\left(\lim_{x\to 0}\frac{\log((x+o(x^2)/x)}{x}\right)$$

Using $o(x^2)/x = o(x)$:

$$= \exp\left(\lim_{x\to 0}\frac{\log(1+o(x))}{x}\right)$$

Write $\log(1+y) = y + o(y)$. Taking $y = o(x)$:

$$= \exp\left(\lim_{x\to 0}\frac{o(x)+o(o(x))}{x}\right)$$

But as we remarked a while back, $o(o(x)) = o(x)$. Moreover clearly we have $o(x) + o(x) = o(x)$. Thus this is:

$$= \exp\left(\lim_{x\to 0}\frac{o(x)}{x}\right)$$

Apply definition of little-o:

$$= \exp(0) = \boxed{1}$$

■

Example 7.4:

$$\lim_{x\to\infty}\sqrt{x+1}-\sqrt{x}$$

Solution. It is true that

$$\sqrt{1+x} = 1 + \frac{1}{2}x + o(x).$$

However it is nonsense to apply Taylor expansion immediately. This is because Taylor expansion at $x = 0$ is used to approximate the function near $x = 0$. Whereas, the limit here is concerned with $x = +\infty$.

Instead we write the limit as

$$\lim_{x\to 0^+}\sqrt{\frac{1}{x}+1}-\sqrt{\frac{1}{x}}.$$

Now we are in position to apply Taylor. Rewriting,

$$= \lim_{x \to 0^+} \frac{\sqrt{1+x} - 1}{\sqrt{x}}.$$

Applying Taylor,

$$= \lim_{x \to 0^+} \frac{1 + \frac{1}{2}x + o(x) - 1}{\sqrt{x}} = \lim_{x \to 0^+} \frac{1}{2}\sqrt{x} + \frac{o(x)}{\sqrt{x}}.$$

We have $o(x)/\sqrt{x} \to 0$ (because we can write it as $o(x)/x \cdot \sqrt{x}$) so the limit evaluates to 0. ■

The method here also shows that the exponent α for which $\lim_{x\to\infty} x^\alpha(\sqrt{x+1} - \sqrt{x})$ exists and is non-zero, is $\alpha = 0.5$.

Taylor is also useful for determining the convergence of some sums and integrals, and ascertaining differentiability of certain contrived functions, but I will not discuss this here.

7.4 More applications of uniform convergence

Uniform convergence preserves some regularity. Some sample facts:

- If f_n continuous and $f_n \to f$ uniformly, then f continuous.
- If $f_n \to f$ pointwise and $f_n' \to g$ uniformly, then f is differentiable and $f' = g$. In other words,
$$\lim_{n\to\infty} \frac{d}{dx} f_n(x) = \frac{d}{dx} \lim_{n\to\infty} f_n(x).$$
- If f_n continuous and $\sum_{n=1}^\infty f_n$ converges uniformly, then $\sum_{n=1}^\infty f_n$ is continuous.
- (Weierstrass M-test) If $\sum_{n=1}^\infty \|f_n\|_\infty < \infty$ then $\sum_{n=1}^\infty f_n$ converges uniformly.
- If f_n differentiable, $\sum_{n=1}^\infty f_n$ converges, and $\sum_{n=1}^\infty f_n'$ converges uniformly, then $\sum_{n=1}^\infty f_n$ is differentiable. In other words,
$$\frac{d}{dx} \sum_{n=1}^\infty f_n = \sum_{n=1}^\infty \frac{d}{dx} f(x).$$

Example 7.5: Let

$$f(x) := \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^3 + x^2}.$$

- Does the sum converge pointwise? Does it converge uniformly?
- Is f continuous?
- Is f differentiable?

Proof.

- For a fixed x we have
$$\sum_{n=1}^{\infty} \left| \frac{\sin(nx)}{n^3 + x^2} \right| \leq \sum_{n=1}^{\infty} \frac{1}{n^3} < \infty$$
so we have absolute convergence of the series. Since x was arbitrary we deduce pointwise convergence.

- Actually,
$$\sum_{n=1}^{\infty} \left\| \frac{\sin(nx)}{n^3 + x^2} \right\|_{\infty} \leq \sum_{n=1}^{\infty} \frac{1}{n^3} < \infty$$
so we have uniform convergence by the M-test.

- f is continuous because the sum converges uniformly, and the partial sums are all continuous.

- We claim that f is differentiable. By the theorem about differentiability under uniform convergence, it suffices to prove that
$$\sum_{n=1}^{\infty} \frac{d}{dx} \frac{\sin(nx)}{n^3 + x^2}$$
converges uniformly. Evaluating the derivative, this is
$$\sum_{n=1}^{\infty} \frac{n\cos(nx)}{n^3 + x^2} - \frac{2x\sin(nx)}{(n^3 + x^2)^2}.$$
To show this converges uniformly, we apply the M-test.
$$\sum_{n=1}^{\infty} \left\| \frac{n\cos(nx)}{n^3 + x^2} - \frac{2x\sin(nx)}{(n^3 + x^2)^2} \right\|_{\infty} \leq \sum_{n=1}^{\infty} \left\| \frac{n\cos(nx)}{n^3 + x^2} \right\|_{\infty} + \left\| \frac{2x\sin(nx)}{(n^3 + x^2)^2} \right\|_{\infty}$$
$$\leq \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{n=1}^{\infty} \left\| \frac{2x\sin(nx)}{(n^3 + x^2)^2} \right\|_{\infty}$$
So just need to show that $\sum_{n=1}^{\infty} \left\| \frac{2x\sin(nx)}{(n^3+x^2)^2} \right\|_{\infty} < \infty$.

- This is true but it is not that easy to show. Instead let's cheat by instead showing that the original series is differentiable over $[-M, M]$, where M is fixed (but arbitrarily large). Then we just need the uniform convergence over $[-M, M]$. This makes the proof easier:

$$\sum_{n=1}^{\infty} \left\| \frac{2x \sin(nx)}{(n^3 + x^2)^2} \right\|_{\infty} \leq \sum_{n=1}^{\infty} \frac{2M}{n^6} < \infty$$

  So we have uniform convergence, hence by the theorem f is differentiable on $[-M, M]$. But M can be as large as we want, so f is differentiable on $\mathbb{R}$.

□

8 Uhoh

8.1 Midterm Problems

8.1.1 Bounding sums

You can't do this:

$$\sum_{n=1}^{\infty}(-1)^n x^n \leq \sum_{n=1}^{\infty}(-1)^n |x|^n$$

Similarly you can't do this:

$$\sum_{n=1}^{\infty}(-1)^n \frac{1}{n^2} \leq \sum_{n=1}^{\infty}(-1)^n \frac{1}{n}$$

A good rule of thumb is to shove in an absolute value into the summand before you write inequalities.

8.1.2 Theorem does not apply, therefore the conclusion of the theorem is false?

For example, let's say f_n continuous but $\sum f_n$ does *not* converge uniformly. This does NOT mean that $\sum f_n$ is not continuous!

(See: Question 4b on the exam, where people said that lack of uniform convergence implies f not differentiable...)

8.1.3 Local uniform convergence does not imply uniform convergence

If $f_n \to f$ uniformly on each $[a, b] \subseteq (0, 1)$, this does not imply that $f_n \to f$ uniformly on $(0, 1)$.

If you know that f_n is continuous on $(0, 1)$ and want to show that f is continuous $(0, 1)$, it *would* be sufficient to have that $f_n \to f$ uniformly on $(0, 1)$. But this may not be the case. At times, we may only be able to show that $f_n \to f$ uniformly on $[\delta, 1 - \delta]$ for each $\delta > 0$. Although we can't conclude that $f_n \to f$ uniformly on $(0, 1)$, we actually can argue that f is continuous.

This is because, as $f_n \to f$ uniformly on $[\delta, 1 - \delta]$ for each $\delta > 0$, it follows that f is continuous on $[\delta, 1 - \delta]$ for each $\delta > 0$. This means that f is continuous on $(0, 1)$ since δ can be made arbitrarily small.

8.1.4 More exercises

Example 8.1: Let

$$f(x) = \sum_{n=1}^{\infty} \sqrt{n} e^{-nx}.$$

- Show that the series converges uniformly over $x \in [a, \infty)$ for any $a > 0$.
- Is f differentiable on $(0, \infty)$? If so, write the series for the derivative f'.

Solution. We apply the M-test. Fix $a > 0$. Then

$$\sum_{n=1}^{\infty} \sup_{x \in [a,\infty)} \left| \sqrt{n} e^{-nx} \right| = \sum_{n=1}^{\infty} \sqrt{n} e^{-na}.$$

This is $< \infty$ because, by the ratio test,

$$\limsup_{n \to \infty} \frac{\sqrt{n+1} e^{-n+1a}}{\sqrt{n} e^{-na}} = \limsup_{n \to \infty} \sqrt{\frac{n+1}{n}} e^{-a} = e^{-a} < 1.$$

So the M-test applies and, therefore, we have uniform convergence of the series over $[a, \infty)$.

f is differentiable on $(0, \infty)$. To show this, it suffices to prove that f is differentiable on $[a, \infty)$ for every $a > 0$. By a theorem, this will be true if we prove that

- $\sum_{n=1}^{\infty} \sqrt{n} e^{-nx}$ converges uniformly on $[a, \infty)$, and
- $\sum_{n=1}^{\infty} \frac{d}{dx} \sqrt{n} e^{-nx}$ converges uniformly on $[a, \infty)$.

We have already shown the first bullet point. For the second, we compute

$$\sum_{n=1}^{\infty} \frac{d}{dx} \sqrt{n} e^{-nx} = \sum_{n=1}^{\infty} -n\sqrt{n} e^{-x}.$$

To show this converges uniformly on $[a, \infty)$, we apply the M-test. ■

Example 8.2: Does $\sum_{n=2}^{\infty} (-1)^n \frac{x^n}{\log n}$ converge uniformly on $x \in (0, 1]$?

Solution. As a first attempt, we may try the M-test. We have

$$\sup_{x \in (0,1]} \left| (-1)^n \frac{x^n}{\log n} \right| = \frac{1}{\log n},$$

and $\sum_{n=2}^{\infty} \frac{1}{\log n} = +\infty$, so the M-test fails.

This does not mean that the answer is "no"! M-test succeeding implies uniform convergence, but this doesn't mean that M-test failing implies no uniform convergence.

It turns out that we *do* have uniform convergence. To show this, we must do it "by hand", i.e. we must show that

$$\sup_{x\in(0,1]} \left| \sum_{n=2}^{N} (-1)^n \frac{x^n}{\log n} \right| \to 0$$

as $N \to \infty$.

There's a theorem that applies here.

Theorem 8.1 (Alternating Series with Error Bound)

Let $a_n \to 0$, $a_n \geq 0$ be decreasing. Then

- (Alternating Series Test) $\sum_{n=1}^{\infty} (-1)^n a_n$ converges.
- (Error Bound) Let $S = \sum_{n=1}^{\infty} (-1)^n a_n$. Then

$$\left| S - \sum_{n=1}^{N} (-1)^n a_n \right| \leq a_{N+1}.$$

Applying this we find that

$$\left| \sum_{n=2}^{N} (-1)^n \frac{x^n}{\log n} \right| \leq \frac{x^{N+1}}{\log(N+1)},$$

hence

$$\sup_{x\in(0,1]} \left| \sum_{n=2}^{N} (-1)^n \frac{x^n}{\log n} \right| \leq \frac{1}{\log(N+1)}$$

which clearly $\to 0$ as $N \to \infty$. ■

8.2 Exercises with Taylor

Example 8.3: Find all $p, q \in \mathbb{R}$ for which

$$\int_0^{\infty} \frac{(x - \sin x)^q}{x^p}\, dx$$

exists and is finite.

Solution. Let $\delta > 0$ be a small number which we'll choose later. It suffices to get that both

$$I := \int_{100}^{\infty} \frac{(x - \sin x)^q}{x^p}\, dx$$

and

$$II := \int_{0}^{\delta} \frac{(x - \sin x)^q}{x^p}\, dx$$

exist. Note that $\int_{\delta}^{100} \frac{(x-\sin x)^q}{x^p}\, dx$ exists and is finite (why?) so we don't need to worry about it.

For I, we note that for all $x > 100$, the integrand is positive, and

$$\frac{x}{2} \leq x - \sin x \leq 2x$$

for all such x. Hence

$$(1/2)^q \int_{100}^{\infty} x^{p-q}\, dx \leq \int_{100}^{\infty} \frac{(x - \sin x)^q}{x^p}\, dx \leq 2^q \int_{100}^{\infty} x^{q-p}\, dx.$$

From this we see that $\int_{100}^{\infty} \frac{(x-\sin x)^q}{x^p}\, dx$ converges iff $\int_{100}^{\infty} x^{q-p}\, dx$ converges, which is when $q - p < -1$.

For II, we have

$$\sin x = x - \frac{x^3}{6} + o(x^4).$$

So

$$x - \sin x = \frac{x^3}{6} + o(x^4)$$

and hence, for all x small enough (say, $0 < x < \delta \ll 1$; here we choose δ), we have

$$0.01x^3 \leq \frac{x^3}{6} - 0.1x^4 \leq x - \sin x \leq \frac{x^3}{6} + 0.1x^4 \leq 0.3x^3.$$

So

$$0.01^q \int_0^{\delta} x^{3+q-p}\, dx \leq \int_0^{\delta} \frac{(x - \sin x)^q}{x^p}\, dx \leq 0.3^q \int_0^{\delta} x^{3+q-p}\, dx.$$

Since the integrand is ≥ 0 *(why is this important?)*, we deduce that the integral II converges iff $\int_0^{\delta} x^{3+q-p}\, dx$ converges, which occurs exactly when $3 + q - p > -1$.

To summarize, our two conditions are $q + 1 < p$ and $4 + q > p$. Thus the original integral converges iff $q + 1 < p < q + 4$. ■

9 Domination

9.1 The limit theorems

Theorem 9.1 (Monotone Convergence)

Let I be any interval, let $f_n : I \to \mathbb{R}$ be **non-negative**, *increasing in* n, and **Riemann integrable**. Let the pointwise limit be f and suppose moreover that f is **Riemann integrable**. Then

$$\lim_{n\to\infty} \int_I f_n \, dx = \int_I f \, dx.$$

Theorem 9.2 (Dominated Convergence)

Let I be any interval, let $f_n : I \to \mathbb{R}$ be **Riemann integrable**, and converge pointwise to some $f : I \to \mathbb{R}$ which is also **Riemann integrable**. If there exists a Riemann-integrable function $g : I \to \mathbb{R}$ for which

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

for all n and x, then

$$\lim_{n\to\infty} \int_I f_n \, dx = \int_I f \, dx.$$

Notes:

- The bounded convergence theorem and uniform convergence theorem are both directly implied by dominated convergence. So monotone convergence and dominated convergence are technically the only two limit theorems you need to know (*for now*).
- In both theorems there is this annoying additional caveat that we must **assume that the limit f is Riemann integrable**. Ideally we'd like to not check for this. Unfortunately, for the Riemann integral we must check for this (why?).

  This is one (minor) reason as to why the Lebesgue integral is better: If you replace "Riemann" with "Lebesgue" in both of the monotone convergence and dominated convergence theorems, then you no longer need to assume this!

  If I forget to check for integrability of the limit, this is why.

Quick note about notation: *I'll be using the partial derivative notation $\frac{\partial}{\partial x}$ at times. It means the same thing as $\frac{d}{dx}$. You're treating all other variables as constant and differentiating in x. We tend to write $\frac{\partial}{\partial x}$ when there are multiple variables. For example $\frac{d}{dx}f(x, y)$ is a bit strange but $\frac{\partial}{\partial x}f(x, y)$ is the gold standard for notation. It's nothing scary I promise.*

9.2 Examples

Example 9.1: Compute the limit

$$\lim_{n\to\infty}\int_0^2 (2x-x^2)^n\,dx.$$

Solution. Let $f_n(x) = (2x-x^2)^n\,dx$. We recognize that $0 < 2x - x^2 < 1$ for all $x \in (0,1)$, thus

$$\lim_{n\to\infty} f_n(x) = 0$$

for all $x \in (0,2)$. Thus $f_n \to f$ pointwise, where $f \equiv 0$. We now must show that the functions $\{f_n\}_n$ are dominated by some integrable function g. Indeed, we can just take $g(x) = 1$, which is integrable over $(0,2)$. So we may swap the limit and integral to get

$$\lim_{n\to\infty}\int_0^2 (2x-x^2)^n\,dx = \int_0^2 \lim_{n\to\infty}(2x-x^2)^n\,dx = \int_0^2 0\,dx = 0.$$

■

Example 9.2: Compute the limit

$$\lim_{t\to 0}\int_{-\infty}^{\infty}\frac{\cos(xt)}{1+x^2}\,dx.$$

Note that, although the convergence theorems are stated for sequences of functions and limits of the form $\lim_{n\to\infty}$, they work equally well for functions parametrized by a continuous variable, such as in the above example, provided that it's done properly. We will stick to sequences for the time being for safety, but it should not be too hard to show the following:

Theorem 9.3 (Dominated convergence for continuous variable)

Let $f(x,t) : I \times T \to \mathbb{R}$ be a function of two variables, with I and T both intervals of $\mathbb{R}$. Let $t_0 \in T$. Suppose that the integral $\int_I f(x,t)\,dx$ exists for all $t \in T$, and that the pointwise limit $\lim_{t\to t_0} f(x,t)$ exists and is integrable. Then, if there exists $\delta > 0$ and a function g which is integrable over I with

$$|f(x,t)| \leq g(x)$$

for all x and for all t with $|t - t_0| < \delta$, then

$$\lim_{t\to t_0}\int_I f(x,t)\,dt = \int_I \lim_{t\to t_0} f(x,t)\,dt.$$

This theorem, which seems to be unstated in typical sources, is quite natural to me. However when written out it may be a bit wordy, so let's generally stick with sequences for the time being.

9.3 Regularity of functions defined by an integral

Example 9.3: Let
$$f(x) = \int_0^1 \sin(x^2 + y^2)\, dy.$$
Is f continuous? Is it differentiable?

Solution. We claim f is differentiable. But as a warm-up let's show that f is continuous. Fix $x_0 \in \mathbb{R}$ and we shall show continuity at x_0 by showing that if $x_n \to x_0$, then
$$\lim_{n\to\infty} f(x_n) = f(x_0).$$
In other words, we want to show that
$$\lim_{n\to\infty} \int_0^1 \sin(x_n^2 + y^2)\, dy = \int_0^1 \sin(x_0^2 + y^2)\, dy.$$
This means that we want to justify an interchange of integrals!

If we let $h_n(t) = \sin(x_n^2 + y^2)$, we have that h_n is dominated by 1, which is integrable over $[0, 1]$. Moreover h_n converges pointwise. So we may conclude by dominated convergence.

How about differentiability? This is a bit trickier, since now we want to interchange an integral with a *derivative*. That is, we want to argue that $\frac{d}{dx}\int_0^1 \sin(x^2 + y^2)\, dy$ exists, with
$$\frac{d}{dx}\int_0^1 \sin(x^2 + y^2)\, dy = \int_0^1 \frac{\partial}{\partial x}\sin(x^2 + y^2)\, dy.$$
However, note that derivatives *are* limits, so what we're really trying to show is that, for every $x_0 \in \mathbb{R}$, we have
$$\lim_{x\to x_0} \frac{\int_0^1 \sin(x^2 + y^2)\, dy - \int_0^1 \sin(x_0^2 + y^2)\, dy}{x - x_0} \overset{?}{=} \int_0^1 \lim_{x\to x_0} \frac{\sin(x^2 + y^2) - \sin(x_0^2 + y^2)}{x - x_0}\, dy,$$
or
$$\lim_{x\to x_0} \int_0^1 \frac{\sin(x^2 + y^2) - \sin(x_0^2 + y^2)}{x - x_0}\, dy \overset{?}{=} \int_0^1 \lim_{x\to x_0} \frac{\sin(x^2 + y^2) - \sin(x_0^2 + y^2)}{x - x_0}\, dy.$$
So, we just need to verify an interchange of limits! To justify this one, we note by the MVT that for any $x \neq x_0$ there is some ξ_x in between x and x_0 such that
$$\frac{\sin(x^2 + y^2) - \sin(x_0^2 + y^2)}{x - x_0} = 2\xi_x \cos(\xi_x^2 + t^2).$$
Thus for all $x \in (x_0 - 1, x_0 + 1)$ we have the uniform bound
$$\left|\frac{\sin(x^2 + y^2) - \sin(x_0^2 + y^2)}{x - x_0}\right| \leq 2(|x_0| + 1),$$

which allows us to use dominated convergence to win. ∎

The methods we used here generalize. Try to prove the following theorem by yourself.

Theorem 9.4

Let $f(x,t) : I \times T \to \mathbb{R}$ be a function of two variables, with I and T both intervals. Suppose that there exists a function $g(t) : T \to \mathbb{R}$ which dominates $\partial_x f$, i.e.

$$\left|\frac{\partial f}{\partial x}(x,t)\right| \leq g(t)$$

for all $x \in I$ and $t \in T$. Then

$$\frac{d}{dx}\int_T f(x,t)\,dt = \int_T \frac{\partial f}{\partial x}(x,t)\,dt$$

for all x. (Here the left side should be interpreted as the derivative of $y \mapsto \int_T f(y,t)\,dt$, evaluated at $y = x$.)

Example 9.4 (Not done in recitation): The Gamma function $\Gamma(z)$ is defined by the integral

$$\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}\,dt$$

for $z > 0$. Prove that Γ is continuous.

Proof. Fix $z_0 > 0$ and we will show that Γ is continuous at z_0. It suffices to show that if we take a sequence $z_n \to z_0$, then

$$\lim_{n\to\infty}\int_0^\infty t^{z-1}e^{-t}\,dt = \int_0^\infty t^{z_0-1}e^{-t}\,dt$$

which, in turn, reduces to justifying the interchange of the limit and integral above.

To do this, we shall dominate the functions $f_n(t) := t^{z_n-1}e^{-t}$. Because of how convergence works, we may assume that

$$0 < z_0 - \delta < z_n < z_0 + \delta$$

for all n (why?). Thanks to this, we get that

$$t^{z_n-1} \leq t^{(z_0+\delta)-1} \text{ when } t > 1,$$

$$t^{z_n-1} \leq t^{(z_0-\delta)-1} \text{ when } t < 1.$$

Therefore,

$$|f_n(t)| \leq g(t),$$

where

$$g(t) := \begin{cases} t^{z_0+\delta-1}e^{-t}, & t \leq 1 \\ t^{z_0-\delta-1}e^{-t}, & t > 1 \end{cases}.$$

This is integrable (why?), so dominated convergence applies! □

9.4 Weird Tricks

I don't expect you to be able to do these, I'm presenting this technique purely for fun. :)

Example 9.5: Compute
$$\int_0^1 \frac{x^{17}-1}{\log x}\,dx$$

Solution. Let $f(t) = \int_0^1 \frac{x^t-1}{\log x}\,dx$. We claim that f is differentiable and that

$$f'(t) \stackrel{?}{=} \int_0^1 \frac{\partial}{\partial t}\frac{x^t-1}{\log x}\,dx.$$

To show this, fix t_0. Then we will show that

$$\lim_{t\to t_0} \frac{f(t)-f(t_0)}{t-t_0}$$

exists, which we can write as

$$\lim_{t\to t_0} \int_0^1 \frac{x^t - x^{t_0}}{(t-t_0)\log x}\,dx.$$

We now want to dominate the integrand here by some integrable $g(x)$. To do this, note by the MVT that

$$\frac{x^t - x^{t_0}}{t-t_0} = \log x \cdot x^{c_t}$$

for some c_t in between t and t_0. So

$$\left|\frac{x^t - x^{t_0}}{t-t_0}\right| \le |\log x| \cdot |x^{c_t}| \le |\log x|,$$

and therefore,

$$\left|\frac{x^t - x^{t_0}}{(t-t_0)\log x}\right| \le 1,$$

which is integrable over $(0,1)$. So the dominated convergence applies. *(technically you should be taking a sequence $t_n \to t_0$ first but whatever.)*

So the claim is proven, i.e. we now have justified that

$$f'(t) = \int_0^1 \frac{\partial}{\partial t}\frac{x^t-1}{\log x}\,dx = \int_0^1 x^t\,dx = \frac{1}{t+1}.$$

Noting that $f(0) = 0$, we conclude that f satisfies the following differential equation:

$$\begin{cases} f'(t) = \frac{1}{t+1} \\ f(0) = 0 \end{cases}$$

The solution is given by $f(t) = \log(1+t)$. We conclude that the answer is $f(17) = \boxed{\log(18)}$. Voila. ■

You can try this one on your own.

Example 9.6: Evaluate the integral

$$\int_{-\infty}^{\infty} e^{-x^2} \cos(2x)\, dx.$$

(Hint: consider $e^{-x^2}\cos(2xt)$... integration by parts... differential equation...)

10 The Birth of Measure Theory

10.1 The Problem of Area

What does "area" mean? Intuitively we expect area to be a function which assigns numbers to sets in a way that measures their "size". Some properties we expect:

(a) $\text{Area}(E) \geq 0$ for all sets E

(b) $\text{Area}(\emptyset) = 0$

(c) (Additivity) $\text{Area}(E \cup F) = \text{Area}(E) + \text{Area}(F)$ whenever E and F are disjoint

(d) (Behavior under "limits") If $E_1 \subseteq E_2 \subseteq E_3 \subseteq \dots$, and $E_\infty := \bigcup_{n=1}^{\infty} E_n$, then

$$\text{Area}(E_\infty) = \lim_{n\to\infty} \text{Area}(E_n).$$

Any function $\mu : \mathcal{F} \to [0, \infty]$ which satisfies these rules is called a *measure*.

Definition 10.1 (Measure)

A *measure* is a function $\mu : \mathcal{F} \to [0, \infty]$ such that:

- $\mu(\emptyset) = 0$
- $\mu(E \cup F) = \mu(E) + \mu(F)$ whenever E and F are disjoint
- $\mu(E_\infty) = \lim_{n\to\infty} \mu(E_n)$ whenever $\{E_n\}_n$ is an increasing sequence of sets, and $E_\infty = \bigcup_{n=1}^{\infty} E_n$

Here, the "domain" of μ, $\mathcal{F}$, is a *sigma-algebra*.

What I want you to know is that "a measure is something which satisfies the rules that area and volume should satisfy".

A sigma algebra is basically just a set of sets which has some structure. The rules for a sigma algebra include containing the empty set, closure under complement, and closure under countable union. You don't need to be that familiar with the definition until you take advanced probability courses.

10.2 Examples of Measures

10.2.1 "Area"

Area, in $\mathbb{R}^2$, is a measure (once we've properly defined it, which is not quite yet).

10.2.2 "Volume", "Length"

The notions of volume and length, in $\mathbb{R}^3$ and $\mathbb{R}$ respectively, are measures (once we've properly defined them).

10.2.3 Counting measure

The function $\mu(S) := |S|$ is a very stupid example of a measure, where $|S|$ denotes the cardinality of S.

10.2.4 Mass

The function $m(E)$ which measures the amount of mass in a region E of space is a measure. This is what physicists mean when they write "dm".

10.2.5 Probability

$\mathbb{P}$ is a measure.

10.2.6 Measure induced by a function

Let $f \geq 0$ be an integrable function on $\mathbb{R}$. Then μ given by

$$\mu(E) := \int_E f(x)\,dx$$

is a measure (on an appropriate sigma algebra). This is an extremely common occurrence for how measures pop up. You can think of the function f as a "weight": The larger $f(x)$ is, the more weight it puts at x.

10.3 How to build a measure...?

Let's try to define the "Area" function now. It seems not obvious how to define it! Let's follow these steps.

Step 1: Simple shapes

Not sure how to find the area of a circle. But maybe I can start by defining the area of really easy shapes.

Definition 10.2 (Measure of a rectangle)

The (elementary) *measure* of a rectangle $R = [a, b] \times [c, d]$ is given by $(b - a)(d - c)$. It is denoted as $\operatorname{meas} R$.
In general we can define the measure of a rectangle $R = I_1 \times I_2 \times \cdots \times I_n = \mathbb{R}^n$ *as* $|I_1| \cdot |I_2| \cdot \ldots |I_n|$ *for any intervals* $\{I_i\}_{i=1}^n$*, where* $|I_i|$ *is the length of interval* I_i*.*

Step 2: Cover anything else with simple shapes

How can we use this to define the area of other things? We could cover them with smaller and smaller rectangles and sum them up!

Definition 10.3 (Lebesgue Outer Measure)

The *Lebesgue Outer Measure* of a set $E \subseteq \mathbb{R}^n$ is given by

$$\mathcal{L}_o^n(E) := \inf \left\{ \sum_{i=1}^{\infty} \operatorname{meas}(R_i) : \{R_i\}_{i=1}^{\infty} \text{ is a sequence of rectangles such that } \bigcup_{i=1}^{\infty} R_i \supseteq E \right\}.$$

The $\mathcal{L}$ stands for "Lebesgue", the n is dimension, and the subscript o stands for "outer". (Spoiler alert: this is *not* a measure, which is why it's called an *outer measure*.)

Is this good definition of "area" (when $n = 2$) and "volume" (when $n = 3$)? Does it satisfy the rules a measure needs to satisfy?

10.4 The Problem of Totality: Vitali Ruins Everything

There is one more rule that you'd expect area to satisfy.

(e) (Totality) All sets have an area.

However, $\mathcal{L}_o^N$ does **not** satisfy this.

Theorem 10.1 (Vitali ruins the day)

There exists some really ugly disjoint sets $E, F \subseteq \mathbb{R}$ such that $\mathcal{L}^1_o(E \cup F) \neq \mathcal{L}^1_o(E) + \mathcal{L}^1_o(F)$.

These two "rules" are, therefore, in contradiction:

(c) (Additivity) $\text{Area}(E \cup F) = \text{Area}(E) + \text{Area}(F)$ whenever E and F are disjoint

(e) (Totality) All sets have an area.

Which one do we give up on? Additivity is simply too important to give up, so we give up totality.

10.5 How to build a measure (for real this time)

Step 1: Simple shapes

For a certain class of simple shapes (like rectangles), define their measure.

Step 2: Cover anything else with simple shapes

For any E, compute its *outer measure* by covering it with smaller and smaller simple shapes.

This gives a way to "measure" all sets.

Step 3 (NEW!): Throw out bad sets

We toss out any sets that could break the additivity condition, leaving us with a *sigma algebra* $\mathcal{F}$ of *measureable* sets, which serves as the "domain" of our measure.

I'm glossing over how exactly this works. For Lebesgue measure, $\mathcal{F}$ generally will contain most sets that you can think of. More on that later.

Let's look at a few examples.

10.5.1 Lebesgue Measure

This is the standard notion of length, area, volume, etc.

1. (Simple shapes) We can define the measure $\operatorname{meas} R$ of any rectangle $R \subseteq \mathbb{R}^N$.

2. (Coverings) We can now define the outer measure via

$$\mathcal{L}_o^N(E) := \inf \left\{ \sum_n \operatorname{meas} R_n : \bigcup_n R_n \supseteq E \right\}.$$

3. (Throw out bad sets) Let $\mathcal{F}$ be the sigma algebra of *Lebesgue-measurable sets* (more on what that looks like). Then the restriction of $\mathcal{L}_o^N$ to $\mathcal{F}$ is the **Lebesgue measure**, $\mathcal{L}^N$ on $\mathbb{R}^N$.

10.5.2 Lebesgue-Stieltjes Measure

In Calculus you may recall writing dg for a differentiable function g, and then writing $\int f\, dg$. This is kinda nonsense unless done correctly.

When g is differentiable, $\int f\, dg$ just means $\int f g'\, dx$, which makes "sense" since $g' = \frac{dg}{dx}$ and then we "rearrange" this to get $g'\, dx = dg$. But when g is not differentiable, such as in an integral like

$$\int_{[0,3]} x^2\, d\lfloor x \rfloor,$$

then we can't do that, so we must think of a different way to interpret this. It turns out that for any increasing g, we can think of the differential dg as *integrating with respect to a measure.* This measure is called the Lebesgue-Stieltjes measure. Here is how we build it.

Let g be increasing and right-continuous.

1. (Simple shapes) The simple shapes are intervals of the form $(a, b]$. Their measure shall be

   $$\operatorname{meas}(a, b] := g(b) - g(a).$$

   think of integrals as Riemann sums, and think of this as redefining what the bases of the rectangles are. the more g increases, the larger we define the bases to be.

2. (Coverings) For any set $E \subseteq \mathbb{R}$, we can define its outer measure,

   $$\mu_g^*(E) := \inf \left\{ \sum_{n=1}^{\infty} \operatorname{meas}(a_n, b_n] : \bigcup_{n=1}^{\infty} (a_n, b_n] \supseteq E \right\}.$$

3. (Throw out bad sets) Restrict μ_g^* to a sigma-algebra called the *Borel sigma-algebra*, $\mathcal{B}(\mathbb{R})$, which you can intuitively think of as all sets you can make using open sets.

   This leaves us with a measure $\mu_g : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ called the **Lebesgue-Stieltjes measure** associated with g.

We will explore this measure a little more in the next section of these notes.

10.6 Lebesgue Measurable Sets

In the construction of the Lebesgue measure, we tossed out a bunch of bad sets to end up with $\mathcal{F}$, the sigma-algebra of *Lebesgue-measurable sets*. What are those, precisely?

I'll give you two ways to tell if a set $E \subseteq \mathbb{R}^N$ is (Lebesgue) measurable. (Sometimes just called "measurable".)

10.6.1 Cookie Cutting Condition

E is measurable if and only if it can be used as a good cookie cutter. In other words, we have that

$$\mathcal{L}_o^N(A) = \mathcal{L}_o^N(A \cap E) + \mathcal{L}_o^N(A \cap E^c)$$

for all sets A.

(See also: "Cartheodory cutting condition")

This can be kinda hard to work with though.

10.6.2 Outer Regularity

E is measurable if and only if for all $\varepsilon > 0$ there exists an open $U \supseteq E$ such that

$$\mathcal{L}_o^N(U \setminus E) < \varepsilon.$$

Intuitively, open sets are nice, and so a measurable set is one that's close enough to a nice set.

Consequences of this are:

- All open sets are measurable.
- All closed sets are measurable.
- All Borel sets are measurable (anything you can make using countably many open sets).

That's pretty good. Here's an even better condition.

10.6.3 Not Trying To Break Math

Write down a set. It's measurable.

It's REALLY HARD to make a set that isn't measurable. The only way to do it is to use the axiom of choice and non-constructively define a horrible set. So if you can even write down the set explicitly then it has to be measurable. Generally all sets are measurable as long as you're not trying to break math.

There's also a notion of a *measurable function*, which is a function which interacts nicely with measurable sets and, therefore, can be reasoned with nicely. I won't bother defining this but it's not that bad.

10.7 Lebesgue Integration

Once we have a measure, we can define what it means to integrate. The way we do this is to first define the integral of *simple functions*. I will do this all on $\mathbb{R}$ and for the Lebesgue measure $\mathcal{L}^1$ (i.e. "length") so that you can imagine it better, but everything I'm about to write works for *any* measure.

Definition 10.4 (Simple function)

A *simple function* is a function of the form

$$s(x) = \sum_{k=1}^{n} c_k 1_{E_k}$$

for $c_k > 0$ and E_k measurable.
The *integral* of $s(x)$ is then

$$\int_{\mathbb{R}} s(x)\, dx := \sum_{k=1}^{n} c_k \mathcal{L}^1(E_k).$$

Think of simple functions as "Riemann sums", or "a bunch of rectangles". We take $c_k > 0$ because we don't want to deal with negatives just yet.

Next, for any function $f \geq 0$, we can try to approximate with simple functions. And that's basically how you get the Lebesgue integral.

Definition 10.5 (Lebesgue Integral (non-negative functions))

Let $f \geq 0$ (be measurable). Then

$$\int_{\mathbb{R}} f(x)\,dx := \sup\left\{\int_{\mathbb{R}} s(x)\,dx : s \text{ is simple, and } 0 \leq s \leq f\right\}.$$

When f can be both positive and negative, we split it into two parts.

Definition 10.6 (Lebesgue Integral)

Take a function f. Split it into the positive part $f^+ := \max(0, f) \geq 0$ and the negative part $f^- := \max(0, -f) \geq 0$. Then the integral of f is

$$\int_{\mathbb{R}} f\,dx := \int_{\mathbb{R}} f_+\,dx - \int_{\mathbb{R}} f_-\,dx$$

provided that the right side isn't $\infty - \infty$.
If the right side is a finite number then we say f is *integrable*.
The integral

$$\int_E f(x)\,dx$$

simply means $\int_{\mathbb{R}} f(x) \cdot 1_E(x)\,dx$.
(If you're using a different measure, you should be very wary of writing $\int_a^b$. Do you see why?)

This leads to the "Big Three" convergence theorems.

Theorem 10.2 (Monotone convergence theorem)

Let $0 \leq f_n \leq f$, such that $f_n \nearrow f$. Then

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,dx = \int_{\mathbb{R}} f(x)\,dx.$$

Theorem 10.3 (Fatou's Lemma)

Let $f_n \geq 0$. Then

$$\int_{\mathbb{R}} \liminf_{n\to\infty} f_n(x)\,dx \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,dx.$$

Theorem 10.4 (Dominated Convergence)

Let f_n be a sequence of functions converging pointwise to some f, where $\{f_n\}_n$ is *dominated* by an integrable g. That is, $|f_n| \leq g$ and g is integrable. Then

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,dx = \int_{\mathbb{R}} f(x)\,dx.$$

I'm not expecting you to know how to prove these, but I want you to know that dominated convergence is a real thing that exists, and is extremely useful.

10.8 Examples of integrals with respect to other measures

10.8.1 Doubled lengths

It's not hard to see that

$$\int_0^1 x^2\,dx = \frac{1}{3}.$$

Now, let μ be the measure which is twice the Lebesgue measure, so $\mu(E) = 2\mathcal{L}^1(E)$. Then what is

$$\int_0^1 x^2\,d\mu?$$

Intuitively, the lengths of the bases of all "rectangles" will have doubled, so it makes sense that the area under the curve has doubled. The answer is $\frac{2}{3}$.

10.8.2 Counting measure

It's not hard to see that

$$\int_1^\infty \frac{1}{x^2}\,dx = 1.$$

Now, let μ be the number of elements shared with $\mathbb{N}$, i.e.

$$\mu(E) := |\mathbb{N} \cap E|.$$

For example $\mu([2.5, 5.5]) = |\{3, 4, 5\}| = 3$. What is

$$\int_1^\infty \frac{1}{x^2}\,d\mu?$$

Before we answer this we should be more precise with what $\int_1^\infty$ means. Is 1 included or excluded? Let's just say that what we meant to write was

$$\int_{[1,\infty)} \frac{1}{x^2}\,d\mu.$$

Now, since $\mu(\{n\}) = 1$ for all $n \in \mathbb{N}$, intuitively μ is putting a "mass of 1" on every natural number, and everything else is worthless. So it turns out that

$$\int_{[1,\infty)} \frac{1}{x^2}\, d\mu = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

11 Apparently this is the last recitation

i literally found out at the end of recitation

11.1 The Dirac Delta "Function"

Physicists say that the Dirac Delta $\delta_a(x)$ is a "function" for which:

- $\delta_a(x) = 0$ for all $x \neq a$
- $\int_{\mathbb{R}} \delta_a(x)\, dx = 1$.

After measure theory, we know that this is BS: A function which is 0 "almost everywhere" must integrate to 0. So what is δ_a?

It turns out that it's not a function. Rather, it is a *measure*.

Definition 11.1 (Dirac Delta)

The Dirac Delta at $a \in \mathbb{R}$ is a measure δ_a defined as follows:

$$\delta_a(E) := \begin{cases} 1, & a \in E \\ 0, & a \notin E \end{cases}$$

In other words, δ_a assigns a mass of 1 to the point $x = a$, and everything else is worthless.

You can check from the definition of Lebesgue integral that for any f,

$$\int_{\mathbb{R}} f(x)\, d\delta_a = f(a).$$

This makes sense: Only $x = a$ matters.

Playing with this a bit more, we can create a funny measure called the *Dirac Comb*,

$$\mathcal{C} := \sum_{n \in \mathbb{Z}} \delta_n.$$

The measure $\mathcal{C}$ assigns a mass of 1 to every integer, and everything else is worthless. We can also define this explicitly as

$$\mathcal{C}(E) = |E \cap \mathbb{Z}|,$$

so $\mathcal{C}(E)$ counts the number of integers inside E.

You can convince yourself that

$$\int_{\mathbb{R}} f(x)\, d\mathcal{C} = \sum_{n=-\infty}^{\infty} f(n).$$

Therefore, all sums are integrals, and thus any theorems which apply to Lebesgue integrals can also be used to reason about sums.

11.2 Back to Lebesgue-Stieltjes

We learned about Riemann-Stieltjes but it sucks. Let's think about Lebesgue-Stieltjes instead. We saw in the last section how it can be defined using the 3-step recipe.

Let's look at some explicit examples now. Let $g(x) = 1_{[0,\infty)}(x)$. Then g is increasing and right-continuous, so we can use it as an integrator. What is the associated Lebesgue-Stieltjes measure, μ_g?

Claim: $\mu_g = \delta_0$

Proof. *(Not done in recitation)* It suffices to prove that $\mu_g(\{0\}) = 1$ and $\mu_g(\mathbb{R} \setminus \{0\}) = 0$, or alternatively, that $\mu_g(\{0\}) = \mu_g(\mathbb{R}) = 1$.

Those sets are all Borel, so we just need to compute the outer measures $\mu_g^*(\{0\})$ and $\mu_g^*(\mathbb{R} \setminus \{0\})$.

First we prove $\mu_g^*(\{0\}) = 1$. The half-open interval $(-10, 10]$ covers $\{0\}$, so

$$\mu_g^*(\{0\}) \leq \text{meas}(-10, 10] = g(10) - g(-10) = 1 - 0 = 1.$$

On the other hand, any covering of $\{0\}$ with half-open intervals must involve some $(a, b]$ which contains 0, so that $a < 0 \leq b$. Thus

$$\mu_g^*(\{0\}) = \inf\{\dots\} \geq \text{meas}(a, b] = \text{meas}(a, b] = g(b) - g(a) = 1 - 0 = 1.$$

Now we prove $\mu_g^*(\mathbb{R}) = 1$. Since $\mathbb{R} \supseteq \{0\}$, we have $\mu_g^*(\mathbb{R}) \geq \mu_g^*(\{0\}) = 1$. For the other inequality, we just note that $\bigcup_{n\in\mathbb{Z}}(n, n+1]$ covers $\mathbb{R}$, and you can verify that $\sum_{n\in\mathbb{Z}} \text{meas}(n, n+1] = \text{meas}(0, 1] = 1$. □

It therefore follows that for all f,

$$\int_{\mathbb{R}} f(x)\, d\mu_g = f(0).$$

When written as a Stieltjes integral, we can think of this as saying that

$$\int_{\mathbb{R}} f(x)\, dg = \int_{\mathbb{R}} f(x) g'(x)\, dx = \int_{\mathbb{R}} f(x)\delta_0(x)\, dx \text{ (not rigorous)}$$

so that, in a sense, δ_0 can be thought of as the derivative of the step function $g(x)$. Intuitively this does make sense! If g has a derivative, it would surely be something that's "infinite" at $x = 0$.

Similarly, if we take $g(x) = \lfloor x \rfloor$, then the associated Lebesgue-Stieltjes measure μ_g is the Dirac comb $\mathcal{C} = \sum_{n\in\mathbb{Z}} \delta_n$! We therefore can say that the derivative of $\lfloor x \rfloor$ is $\mathcal{C}$, in a sense.

Example 11.1: Let $f(x) = 1_{[0,\infty)}(x)$. Then as you agonized over during homework, the *Riemann*-Stieltjes integral $\int_{\mathbb{R}} f\, df$ does not exist. This is very annoying and I'll ask you to not stress over this technicality so that you can study literally anything else.
Anyways, the *Lebesgue*-Stieltjes integral $\int_{\mathbb{R}} f\, df$ certainly exists. As a Lebesgue integral, we're really interpreting this integral as $\int_{\mathbb{R}} f\, d\mu_f$ where μ_f is the Lebesgue-Stieltjes measure associated with f. And, as we've shown, μ_f is just the Dirac delta (at 0)! Thus

$$\int_{\mathbb{R}} f\, df = \int_{\mathbb{R}} f\, \delta_0 = f(0) = \boxed{1}.$$

11.3 Bounded Variation

We can now (kinda) calculate integrals of the form $\int f\, dg$ where g is increasing and right-continuous. But what if g is not necessarily increasing? For example, can we make sense of the integral

$$\int_{-1}^{1} x^3\, d|x|?$$

If we try to associate a Lebesgue-Stieltjes measure with $|x|$ in the way we've been doing, we notice a problem that comes from the decreasing part: $\text{meas}(-1, 0] = |0| - |-1| = -1$, which is really bad because measures should not spit out negative values!

This suggests that the associated Lebesgue-Stieltjes measure, $\mu_{|}x|$, is not quite a measure. Rather, it is a *signed measure*, which satisfies some different properties but ultimately is defined quite similarly to measures. Signed measures are a bit annoying until one takes a course in measure theory and proves that every signed measure can be written as the difference between two measures!

$$\mu = \mu_+ - \mu_-$$

Let's put that aside for now and try to figure out we could compute $\int f\, dg$ for g not increasing. It turns out that when g has *bounded variation*, we can make sense of this.

Definition 11.2

The *variation* of g over an interval I is basically how much it goes up and down over I. More formally,

$$\operatorname*{Var}_I g = \sup\left\{\sum_{i=1}^{n} |g(x_i) - g(x_{i-1})| : x_0 < x_1 < \cdots < x_n, \{x_i\}_{i=0}^{n} \in I\right\}.$$

For example, $\operatorname{Var}_{[-1,1]} x^2 = 4$.
If $\operatorname{Var}_I g < \infty$ then we say g has *bounded (pointwise) variation.*

It turns out that if g has bounded variation then it can be written as the difference between two increasing functions *(fun exercise!)*:

$$g = g^+ - g^-$$

For example, $|x| = \max(x, 0) - \min(x, 0)$. Now to interpret $\int f\, dg$, we can write

$$\int f\, dg = \int f\, d(g^+ - g^-) = \int f\, dg^+ - \int f\, dg^-,$$

where we can now evaluate both of those integrals on the right!

The astute reader may have realized that this is the analogue of the decomposition of signed measures. Indeed, the associated Lebesgue-Stieltjes measures decompose in this way.

Example 11.2: Let's evaluate $\int_{-1}^{1} x^3\, d|x|$. From the decomposition $|x| = \max(x, 0) - \min(x, 0)$, we have

$$\int_{-1}^{1} x^3\, d|x| = \int_{-1}^{1} x^3\, d\max(x, 0) - \int_{-1}^{1} x^3\, d\min(x, 0).$$

We see that

$$\max(x, 0) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases},$$

so the associated Lebesgue-Stieltjes measure is the normal length measure on $(0, \infty)$, and no mass is assigned to $(-\infty, 0)$. It follows that

$$\int_{-1}^{1} x^3\, d\max(x, 0) = 0 + \int_{0}^{1} x^3\, dx = \frac{1}{4}.$$

Similarly, we can find that $\int_{-1}^{1} x^3\, d\min(x, 0) = \int_{-1}^{0} x^3\, dx = \frac{-1}{4}$. So

$$\int_{-1}^{1} x^3\, d|x| = \frac{1}{4} - -\frac{1}{4} = \frac{1}{2}.$$

11.4 Cantor Set

There are a few more things in measure theory that I think you guys should probably know. And they're all bad.

Question 1: Is every set $E \subseteq \mathbb{R}$ with zero (Lebesgue) measure necessarily at most countably infinite?

Unfortunately no, because of the Cantor set.

Ok, let's prepare the second question. A lot of the examples of the Lebesgue-Stieltjes measures on $\mathbb{R}$ that we've talked about so far are either "very nice" or "jumpy".

By "very nice", I mean stuff like dx^2 or $d|x|$, in which it ended up being the case that we could write

$$\int f\, dg = \int fh\, dx$$

for some function h. ($dx^2 = 2x\, dx$, $d|x| = \operatorname{sgn} x\, dx$

By "jumpy", I mean measures that are *purely atomic*, such as $d1_{[0,\infty)}$ or $d\lfloor x \rfloor$.

Are all Lebesgue-Stieltjes measures a combination of these guys?

Question 2: If a measure μ has no atoms (i.e. sets $\{x\}$ of positive measure), must it be correspond to a function in the sense that $d\mu = f\, dx$ for some f?

Unfortuntaely no. There's something called the Cantor function. It is continuous, so its associated Lebesgue-Stieltjes measure has no atoms. Moreover its derivative is zero *almost everywhere*. And yet, the associated Lebesgue-Stieltjes measure is quite non-trivial.

For a bit more on the Cantor set, you can see my notes for the 21-235 course at CMU (section 4, "Cantor Ruins the Day"). I don't expect you to be that familiar with Cantor stuff, just wanted to give you guys a glimpse of it.

12 Appendix

I wrote this up but I could never find a good time to talk about it lol.

12.1 All Norms on $\mathbb{R}^N$ are equivalent

Theorem 12.1

Every norm on the vector space $\mathbb{R}^N$ generates the same topology.

This is huge. Some sample consequences of this:

- To prove that some $K \subseteq \mathbb{R}^N$ is compact, you can show that it is compact with respect to the l^1 norm.
- If a sequence $x_n \in \mathbb{R}^N$ converges with respect to the usual norm, then it converges with respect to the taxicab norm.
- $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous at $(x_0, y_0) \in \mathbb{R}^2$ iff for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x,y) - f(x_0,y_0)| < \varepsilon$ for all x, y for which $|x - x_0| < \delta$ and $|y - y_0| < \delta$ (i.e. $\|(x,y) - (x_0,y_0)\|_\infty < \delta$).

This is a reflection of the principle that *compactness, convergence, and continuity are topological properties.*

Let's prove the theorem now.

Proof. Let $\|\cdot\|_\infty$ be the l^∞ norm on $\mathbb{R}^N$, i.e.

$$\|(x_1, \ldots, x_N)\|_\infty := \max_{1 \le k \le N} |x_k|,$$

and let $\|\cdot\|_\sharp$ be another norm on $\mathbb{R}^N$.

CLAIM: *There exists constants $c, C > 0$ such that*

$$c\|x\|_\infty \le \|x\|_\sharp \le C\|x\|_\infty$$

for all $x \in \mathbb{R}^N$.

Before we prove this claim, let me explain why this claim proves the theorem. This is because if U is open with respect to $\|\cdot\|_\sharp$, then for every point $x \in U$, we can find a ball $B_\sharp(x,r) \subseteq U$. But by the claim, $B_\infty(x, cr) \subseteq B_\sharp(x,r)$. In particular, $B_\infty(x, r/C) \subseteq U$. So this proves that every $x \in U$ has an open $\|\cdot\|_\infty$-ball around it that's contained in U, hence

U is open with respect to $\|\cdot\|_\infty$. Similar reasoning shows that any set which is open with respect to $\|\cdot\|_\infty$ must be open with respect to $\|\cdot\|_\sharp$. So the topologies that these norms induce must be the same.

Now let's prove the claim. Observe first that

$$\|x =: (x_1, \dots, x_N)\|_\sharp = \left\| \sum_{k=1}^N x_k e_k \right\|_\sharp \le \sum_{k=1}^N |x_k| \cdot \|e_k\|_\sharp$$

$$\le \|x\|_\infty \sum_{k=1}^N \|e_k\|_\sharp.$$

So we have $\|x\|_\sharp \le C\|x\|_\infty$ for all x, where the constant is $C = \sum_{k=1}^N \|e_k\|_\sharp$.

To show the other side of the inequality, observe that since $\|x\|_\sharp \le C\|x\|_\infty$, we have that the function $f(x) := \|x\|_\sharp$ is continuous with respect to $\|\cdot\|_\infty$ (why?). Now let S be the unit sphere with respect to $\|\cdot\|_\infty$, i.e.

$$S := \{x \in \mathbb{R}^N : \|x\|_\infty = 1\}.$$

S is closed with respect to the usual norm (why?) and bounded, so it is compact with respect to the usual norm. I leave it to you to show that the usual norm and $\|\cdot\|_\infty$ are equivalent (can you show that there are constants A and B so that $A\|\cdot\|_\infty \le \|\cdot\| \le B\|\cdot\|_\infty$?) hence S is compact with respect to $\|\cdot\|_\infty$.

Therefore f obtains a minimum c on S. So $\|x\|_\sharp \ge c$ for all $\|x\|_\infty = 1$. By scaling, we conclude that $\|x\|_\sharp \ge c\|x\|_\infty$ for all $x \in \mathbb{R}^N$. □

12.2 ODE existence and uniqueness

Remember this theorem?

Theorem 12.2 (Banach Fixed Point Theorem)

Let (X, d) be a complete metric space, and let $f : X \to X$ be a *contraction*, i.e.

$$d(f(x), f(y)) \le Ld(x, y)$$

for all $x, y \in X$, where $0 \le L < 1$.
Then there exists a unique fixed point of f, i.e. $x_0 \in X$ such that $f(x_0) = x_0$.

You proved it in homework a long while ago by picking some point x and then showing that the sequence $x, f(x), f(f(x)), \dots$ is Cauchy.

When you're picturing this theorem you're probably thinking of X as like a subset of Euclidean space. What's really nice is that this can apply to other contexts! It's incredible that this theorem works seamlessly in like, spaces of function, and the classic example of this is one first proofs in analysis that truly captivated me.

So here's something called a differential equation:

$$\begin{cases} u'(t) = f(u(t), t) \\ u(0) = u_0 \end{cases}$$

Here $f(x,t) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a given function, and $u_0 \in \mathbb{R}$ is a given constant. We are trying to solve for u. For example, we can easily solve the differential equation

$$\begin{cases} u'(t) = t^2 \\ u(0) = 0 \end{cases}$$

as $u(t) = \frac{1}{3}t^3$. (Here the f is $f(x,t) = t^2$.) And

$$\begin{cases} u'(t) = u(t) \\ u(0) = 1 \end{cases}$$

solves as $u(t) = e^t$. (Here $f(x,t) = x$).

What differential equations have a solution? Here's an amazing result.

Theorem 12.3 (Cauchy-Lipschitz-Picard-Lindelöf, simplified slightly)

Suppos f is *Lipschitz in space*, i.e. there is a constant $L \geq 0$ such that

$$|f(x,t) - f(y,t)| \leq L|x - y|$$

for all $x, y, t \in \mathbb{R}$. Then there exists a solution $u : [0,T] \to \mathbb{R}$ to the differential equation, for some amount of time T.

Proof. We choose $T = \frac{1}{2L}$, which will make sense later. First, notice that to solve the differential equation, we can take $u'(t) = f(u(t), t)$ and integrate both sides to get

$$u(t) = u_0 + \int_0^t f(u(s), s)\, ds. \qquad (*)$$

It suffices to find a u which solves this "integral form" of the equation.

Let's turn this into a problem about finding a fixed point... What metric space should we work in? Let's take it to be $X = C([0,T])$, which is a complete metric space. *(Recall that $C([0,T])$ is equipped with the sup norm $\|\cdot\|_\infty$.)* Our goal is to find a $u \in C([0,T])$ solving $(*)$.

Now we need to come up with a function $F : C([0,T]) \to C([0,T])$, which is related to $(*)$... how about we take $F(u)$ to be the right side of $(*)$? That is, for a function $u : [0,T] \to \mathbb{R}$, we take $F(u)$ to be the function

$$t \mapsto u_0 + \int_0^t f(u(s), s)\, ds.$$

Then $(*)$ reads as $u = F(u)$, meaning that we need a **fixed point** of F...

Wait, we'd be done if F were a contraction! Calculating,

$$\begin{aligned}\|F(u) - F(v)\|_\infty = \sup_{0 \le t \le T} \left| \int_0^t f(u(s), s) - f(v(s), s)\, ds \right| &\le \int_0^T |f(u(s), s) - f(v(s), s)|\, ds \\ &\le L \int_0^T |u(s) - v(s)|\, ds \\ &\le L \int_0^T \|u - v\|_\infty\, ds = LT\|u - v\|_\infty = \frac{1}{2}\|u - v\|_\infty.\end{aligned}$$

So F is a contraction, and we've nuked the problem by using the Banach Fixed Point Theorem. I think that's really cool. □

If you look up the theorem you'll see that the proof is more complicated. This is because everyone else considers a more general f, whose domain isn't necessarily on all of $\mathbb{R}$, so a bit more care is needed to set things up. It's basically the same proof though.