MATH 205A: Real Analysis I

Lecturer: Professor Lenya Ryzhik

Notes by: Andrew Lin

Autumn 2022

Introduction

Professor Ryzhik (who we should call Lenya) and Zhenyuan Zhang (the CA for the course) are the course staff for this quarter. Office hours, class materials, lecture notes, and problem sets can be found on Canvas. Homework will be assigned biweekly, and there will also be a midterm (but probably no final).

There are many texts on real analysis, and Royden's "Real Analysis," Rudin's "Real and Complex Analysis," and Tao's "An Epsilon of Room" (volume I) may be useful references. On the other hand, there is a book by Evans and Gariepy ("Measure Theory and Fine Properties of Functions") which is very dry but extremely correct. If we get to Fourier analysis, there is a nice book "Fourier Analysis" by Duoandikoetxea and Pinsky's "Introduction to Fourier Analysis and Wavelets," which is nice to read, often not correct, but correctable. None of these are required texts, but some of them might be useful for us, and they might be available from the library.

We'll start with measure theory and integration, then maybe Fourier analysis and Brownian motion, and perhaps cover the maximal function and harmonic analysis if we have time.

1 September 29, 2022

We'll begin with the **Lebesgue measure**, whose goal is to generalize the length of an interval or the volume of a cube to more general sets. Such a "measure" m should satisfy the following properties:

1. $m(E)$ should be defined for all sets E and always be nonnegative.
2. The measure of any interval I (open or closed) should be its length.
3. If sets E_i are disjoint, then $m\left(\bigcup_{i=1}^{\infty} E_n\right) = \sum_{j=1}^{\infty} m(E_n)$.
4. For any x, we should have $m(E + x) = m(E)$ (translation invariance).

Unfortunately, it is not actually possible for all of these to hold. Specifically, let $\oplus$ be addition modulo 1, so that for any $x, y \in [0, 1)$ we have

$$x \oplus y = \begin{cases} x + y & x + y < 1, \\ x + y - 1 & x + y \geq 1. \end{cases}$$

It turns out that we still have translation invariance with $\oplus$ (on a circle) if we assume properties (1), (3), and (4). Indeed, for any $E \subset [0, 1)$, we can let $E_1 = [0, 1 - y) \cap E$ and $E_2 = [1 - y, 1] \cap E$. Then $E_1 \oplus y = E_1 + y$ and

$E_2 \oplus y = E_2 + (y-1)$, so

$$m(E \oplus y) = m(E_1 \oplus y) + m(E_2 \oplus y) = m(E_1 + y) + m(E_2 + y - 1) = m(E_1) + m(E_2) = m(E).$$

(Basically we just keep track of which part of E crossed over 1 and shift it back down by 1 by ordinary translation-invariance.) So now define the equivalence relation $x \sim y$ if $x = y \oplus q$ for some rational number q (we can check that this is indeed an equivalence relation), and thus we can split $[0,1)$ into equivalence classes. So if P is a set which contains one element from each equivalence class (note that we use the axiom of choice here), then we can enumerate the countable set $\mathbb{Q} \cap [0,1)$ and define $P_j = P \oplus q_j$. These P_j are pairwise disjoint subsets of $[0,1)$ (because whenever $p_1 \oplus q_1 = p_2 \oplus q_2$, we must have $p_1 \sim p_2$, and each P_j only has one element per equivalence class), and in fact $[0,1)$ is the union of the P_js. Thus we have $m([0,1)) = \sum_{j=1}^{\infty} m(P_j) = \sum_{j=1}^{\infty} m(P)$ by translation invariance, so if $m(P) = 0$ then $m([0,1) = 0$ and if $m(P) > 0$ then $m([0,1)) = \infty$. In either case, we break property (2).

So we need to drop one of the four assumptions – we don't want to give up (2) if we want to generalize length, and (3) and (4) should hold in $\mathbb{R}^n$ by "physical common sense." So the point is that **not every set should have a measure**.

We'll begin by defining the **outer Lebesgue measure**, which we can define for all sets but will not satisfy property (3).

Definition 1

Let $E \subseteq \mathbb{R}$ be any subset. The **outer measure** $m^*(E)$ of E is defined by

$$m^*(E) = \inf_{\text{covers of } E \text{ by a countable union of open intervals } I_n} \left\{ \sum \ell(I_n) \right\}$$

In other words, we approximate E from above and we want to do so in the smallest way possible.

As an exercise, we can ask what happens if we only allow covers with finitely many intervals. We can check that $m^*(S) = 0$ for any countable set S. Also, notice that $A \subseteq B \implies m^*(A) \le m^*(B)$ because anything that covers B also covers A.

Proposition 2

Let I be an (open, closed, or semi-closed) interval. Then $m^*(I) = \ell(I)$.

Proof. First of all, notice that the endpoints do not matter, because any of the intervals $(a,b), [a,b], [a,b), (a,b]$ are contained in $(a-\varepsilon, b+\varepsilon)$. So $m^*(I) \le b - a + 2\varepsilon$ for any $\varepsilon > 0$, which means $m^*(I) \le b - a = \ell(I)$.

On the other hand, suppose we have a **closed** interval $I = [a,b] \subseteq \bigcup_{j=1}^{\infty} I_j$ for open intervals I_j. By compactness, there is a finite subcover such that $I \subseteq \bigcup_{k=1}^{n} I_{j_k}$. Thus

$$[a,b] \subseteq \bigcup_{k=1}^{N} I_j \implies |b-a| \le \sum_{k=1}^{N} \ell(I_{j_k}) \le \sum_{j=1}^{\infty} \ell(I_j)$$

(somewhat painful exercise but doable). So taking the infimum over all possible coverings, we indeed see that $\ell(I) \le m^*(I)$. Combining these inequalities gives us the desired result for closed intervals. Finally, for the **other kinds of intervals**, $m^*([a+\varepsilon, b-\varepsilon]) = b - a - 2\varepsilon$, so $m^*((a,b))$ (or the same for semi-closed intervals) is at least $m^*([a+\varepsilon, b-\varepsilon]) \ge b - a - 2\varepsilon$, which again gives the desired result when combined with $m^*(I) \le \ell(I)$. $\square$

Proposition 3

The outer measure is countably **sub**additive, meaning that

$$m^*\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} m^*(A_n).$$

Proof. If $m^*(A_k) = \infty$ for some k, then the result automatically holds. Otherwise, suppose $m^*(A_k) < \infty$ for all k, and choose intervals I_{kj} such that A_k is covered by $\bigcup_{j=1}^{\infty} I_{kj}$ for all k, with $\sum_{j=1}^{\infty} \ell(I_{kj}) \le m^*(A_k) + \frac{\varepsilon}{2^k}$. (We can do this by the definition of outer measure.) Then $\bigcup A_k$ is covered by the entire set of I_{kj}s, so

$$m^*\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{k,j} \ell(I_{kj}) \le \sum_k \left(m^*(A_k) + \frac{\varepsilon}{2^k}\right) = \sum m^*(A_k) + \varepsilon,$$

and taking $\varepsilon \to 0$ yields the result. $\square$

(This is one way to prove that $m^*(S) = 0$ if S is countable, because the measure of a single point is zero.) It turns out that we can get equality instead of inequality under some special conditions, though. Recall that the distance between two sets is

$$\operatorname{dist}(A, B) = \inf\{x - y : x \in A, y \in B\}.$$

Lemma 4

If $\operatorname{dist}(E, F) > 0$, then $m^*(E \cup F) = m^*(E) + m^*(F)$.

Proof. Choose δ so that $|x - y| \ge \delta$ for all $x \in E$ and $y \in F$. We claim that

$$m^*(A) = \inf_{A \subseteq \cup I_j, |I_j| < \delta} \sum \ell(I_j).$$

In other words, we can always choose our intervals to be of length at most δ by splitting up our intervals into open intervals such that we only gain at most an $\frac{\varepsilon}{2^j}$ measure. So now if we let $E \cup F$ be covered by a set of intervals with each interval having length at most $\frac{\delta}{10}$, then I_j cannot intersect both E and F (because it has length less than δ), so we can let I_j' be the intervals intersecting E and I_j'' be the intervals intersecting F. So $m^*(E) \le \sum \ell(I_j')$ and $m^*(F) \le \sum \ell(I_j'')$, and we can choose the I_js so that $\sum \ell(I_j)$ (the sum of the lengths of all of the intervals) is at most $m^*(E \cup F) + \varepsilon$. Thus $m^*(E \cup F) + \varepsilon \ge \sum \ell(I_j') + \sum \ell(I_j'') \ge m^*(E) + m^*(F)$, so $m^*(E \cup F) \ge m^*(E) + m^*(F)$, and combining this with the previous result implies equality. $\square$

We can generalize outer Lebesgue measure to $\mathbb{R}^n$ by replacing open intervals with open boxes (and requiring that the measure of $B = (a_1, b_1) \times \cdots \times (a_n, b_n)$ to be its volume $|B| = (b_1 - a_1) \cdots (b_n - a_n)$. The advantage of using boxes over open balls is the following: call two boxes "almost disjoint" if their interiors are disjoint. Then we can check that for a finite union E of pairwise almost disjoint boxes B_j, we have $m^*(E) = \sum_{j=1}^{N} |B_j|$, so if E is instead a countable union of pairwise almost disjoint boxes, then $m^*(E) = \sum_{j=1}^{\infty} |B_j|$ as well. Indeed, $E \subseteq \bigcup_{j=1}^{\infty} B_j \implies m^*(E) \le \sum_{j=1}^{\infty} |B_j|$, but since $E \supseteq \bigcup_{j=1}^{N} B_j$ for any N, $m^*(E) \ge \sum_{j=1}^{N} |B_j|$. Since this holds for any N, we must indeed have equality for the countable sum.

The point now is that any open set in $\mathbb{R}$ is a finite or countable disjoint union of open intervals, but this does not hold in $\mathbb{R}^n$ for $n \ge 2$. Instead, we use the notion of **dyadic boxes** – take the unit lattice in $\mathbb{R}^n$ and repeatedly break each box into 2^n pieces. Then every dyadic box is contained in exactly one box from the previous iteration, so for any open set $\mathcal{O}$ and any $x \in \mathcal{O}$ we can draw a ball around x contained in $\mathcal{O}$, which must contain a dyadic box. Choosing the dyadic box to contain x (our boxes are closed), $\mathcal{O}$ is then a union of dyadic boxes because it's the union of the dyadic boxes across all x. Call a dyadic box $B \in \mathcal{O}$ **maximal** if it is not contained in any other dyadic box also contained in $\mathcal{O}$ (in other words, double the box and see if it is still completely within $\mathcal{O}$). Now it is nice that any two maximal dyadic boxes are almost disjoint (since any two dyadic boxes that intersect must have one contained in another), so any open set is a union of almost pairwise disjoint pairwise (dyadic, but that doesn't matter any more) boxes, and there are only countably many such boxes.

This is all we'll say about the outer measure for now – we'll now talk about **Lebesgue measurability**, which tells us which sets we can actually define a measure on.

Definition 5

A set F is $\mathcal{G}_\delta$ if it is an intersection of a countable collection of open sets (notice that this does not need to be an open set).

Proposition 6

For any set $A \subseteq \mathbb{R}^n$, the following are true:

1. There is an open set $O \supset A$ such that $m^*(A) + \varepsilon \geq m^*(O)$ (so we can approximate from above with open sets),
2. There is a $\mathcal{G}_\delta$ set $G \supset A$ such that $m^*(A) = m^*(G)$,
3. $m^*(A) = \inf\{m^*(O) : O \text{ open and } O \supseteq A\}$.

Proof. (1) is achieved by covering A with a union of open boxes, and (3) is the definition of the outer measure. For (2), take $O_j \supseteq A$ so that $m^*(O_j) \leq m^*(A) + \frac{\varepsilon}{2^j}$ (in this case we could have just taken any numbers going down to zero), and we can take $G = \bigcap_{j=1}^\infty O_j$. Then $G \supseteq A$ and $m^*(G) \leq m(O_j)$ for each j, so $m^*(G) \leq m^*(A) + \frac{\varepsilon}{2^j}$ for all j, meaning that $m^*(G) \leq m^*(A)$. Since G contains A, we must have equality. □

Definition 7

A set A is **Lebesgue measurable** if for any $\varepsilon > 0$, there exists an open set $O \supseteq A$ such that $m^*(O \setminus A) < \varepsilon$.

Notice that this condition is not the same as $m^*(A) + \varepsilon \geq m^*(O)$. Indeed, just because we only need ε more measure to cover O than A does not mean that $O \setminus A$ can be covered with ε measure, especially if A is a terrible mess. So measurability kind of requires us to be able to cover both A and $O \setminus A$ nicely, and next time, we'll discuss a more abstract sense of measurability due to Caratheodory.

2 October 4, 2022

Last lecture, we defined the outer measure of a set $E \subset \mathbb{R}$ to be

$$m^*(E) = \inf_{E \subseteq \cup I_k} \sum \ell(I_k)$$

and similarly making the definition with boxes in $\mathbb{R}^n$. We defined E to be **(Lebesgue) measurable** if there is an open set $O \supseteq E$ such that $m^*(O \setminus E) < \varepsilon$ for any $\varepsilon > 0$. We can now check a few properties:

- If E_j are measurable sets, then $\cup_{j=1}^\infty E_j$ is also measurable (by covering each E_j by an open set O_j with $m^*(O_j \setminus E_j) \le \frac{\varepsilon}{2^j}$ and taking the union of the O_js).
- Any open set is measurable (by taking itself), and any closed set is also measurable. Indeed, let E be a closed set and find an open set O containing E of outer measure at most $m^*(E) + \varepsilon$, noticing that $O \setminus E$ is open and thus an at most countable union of almost disjoint boxes B_j). Then any finite subset of these boxes is a positive distance away from E (since we have closed sets that do not intersect), so outer measure is additive and $m^*(E) + \sum_{j=1}^N |B_j| \le m^*(O) \le m^*(E) + \varepsilon$. Thus $m^*(O \setminus E) \le \sum_{j=1}^\infty |B_j| \le \varepsilon$ as well.
- The complement of a measurable set is also measurable. Indeed, if O_n is an open set containing E with $m^*(O_n \setminus E) < \frac{1}{n}$ with complement F_n, then $E^c = F_n \cup (O_n \setminus E)$. Then $m^*(E^c \setminus F_n) < \frac{1}{n}$, so if we look at $F = \bigcup_{n=1}^\infty F_n$, then $m^*(E^c \setminus F) \le m^*(E^c \setminus F_n) < \frac{1}{n}$ for all n, and thus $m^*(E^c \setminus F) = 0$. Thus $E^c \setminus F$ is measurable (any set of outer measure zero is measurable because we can just cover it with an open set of arbitrarily small measure ε to satisfy the definition), and F is also measurable because each F_n is closed and thus measurable. Thus E^c (their union) is measurable, as desired.
- If E_n are measurable, then the intersection $\bigcap_{n=1}^\infty E_n$ is measurable. Indeed, this is just some manipulation:

$$E = \bigcap_{n=1}^\infty E_n \implies E^c = \bigcup_{n=1}^\infty E_n^c,$$

and since each E_n is measurable, so is each E_n^c, and so is their union E^c, and so is its complement E.

We thus have some nice closure properties for the set of measurable sets:

Definition 8

A collection of sets $\mathcal{F}$ is an **algebra** if it is closed under complements and finite unions (equivalently, for any $A, B \in \mathcal{F}$, we have $A^c, A \cup B \in \mathcal{F}$). $\mathcal{F}$ is a **σ-algebra** if it is also closed under countable unions.

Everything we've proven in the list above gives us the following result:

Theorem 9

The collection of all Lebesgue measurable sets forms a σ-algebra.

There is another important σ-algebra used by probabilists (in the sense that they consider the σ-algebra "generated by" some fundamental collection):

Definition 10

The **Borel σ-algebra** is the smallest σ-algebra that contains all open sets.

It will turn out that not all Lebesgue measurable sets are Borel – this may not be surprising because the example of a non-Lebesgue-measurable-set (the stuff we "need to remove") has nothing to do with open sets, but it may be surprising because Lebesgue measurable sets have to do with approximations **by** open sets.

It's important to note that there is an alternative definition of measurability that we may also encounter:

Definition 11 (Caratheodory definition of measurability)

A set E is **C-measurable** if for each (not necessarily measurable) $B \subseteq \mathbb{R}^n$, we have $m^*(B) = m^*(B \cap E) + m^*(B \cap E^c)$.

With this definition, we directly see that E is C-measurable if and only if E^c is C-measurable. Notice that we always have the inequality $m^*(B) \le m^*(B \cap E) + m^*(B \cap E^c)$ (we can cover the left-hand side by covering the two terms on the right-hand side separately), so we only need to check the other inequality, which we can do by checking that $m^*(B) \ge m^*(B \cap E) + m^*(B \cap E^c) - \varepsilon$ for any $\varepsilon > 0$. We can now check a few properties under C-measurability (the goal is ultimately to prove that we again have a σ-algebra):

- If $m^*(E) = 0$, then E is C-measurable. This is easier to prove than with the Lebesgue measurability definition: notice that $m^*(B \cap E) = 0$, so measurability just requires $m^*(B) \ge m^*(B \cap E^c)$, which is true.
- Any "corner" $A = \{x_1 > a_1, x_2 > a_2, \cdots, x_d > a_d\} \subseteq \mathbb{R}^n$ is C-measurable. Indeed, for any $B \subseteq \mathbb{R}^d$, the inequality we need to check is automatically satisfied if $m^*(B) = \infty$. Otherwise, $m^*(B)$ is finite, so for any $\varepsilon > 0$ there is a countable collection of almost disjoint closed boxes D_n such that $B \subseteq \cup_n D_n$ and $m^*(B) + \varepsilon > \sum |D_n|$. Now if we consider
$$A_n = \left\{x_1 \ge a_1 + \frac{\varepsilon}{2^n}, \cdots, x_n \ge a_n + \frac{\varepsilon}{2^n}\right\}$$
and define the "intersection boxes" $D_n' = D_n \cap A$ and $D_n'' = D_n \cap A_n^c$ (notice that A_n^c is bigger than A^c, and notice that we're intersecting each box with a different "modified corner" A_n^c). We want to say that these add up to D_n plus only a small amount. Without loss of generality, we'll assume that all sides of D_n are of size at most 1, meaning that the intersection of D_n' and D_n'' is at most $\frac{\varepsilon}{2^n}$ (because one of the sides of the boxes has length at most $\frac{\varepsilon}{2^n}$ by definition of A_n). Then taking the sets $B_1 = A \cap B \subseteq \cup_n D_n'$ and $B_2 = A^c \cap B \subseteq \cup_n D_n''$, we find
$$m^*(B_1) + m^*(B_2) \le \sum_n |D_n'| + \sum_n |D_n'| \le \sum_n |D_n| + \frac{\varepsilon}{2^n} \le m^*(B) + 2\varepsilon.$$
Well, we've actually cheated a bit here, because D_n'' isn't actually a box if it intersects A near the "actual corner." So there should really be a constant C_d factor in the $\frac{\varepsilon}{2^n}$ argument (depending on the dimension d – it's the maximum number of boxes needed to split up such an intersection, but we can just take 2^d), but otherwise the argument works.
- If E_1 and E_2 are C-measurable, then so is $E_1 \cup E_2$. Indeed,
$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c) = m^*(A \cap E_1) + m^*((A \cap E_1^c) \cap E_2) + m^*((A \cap E_1^c) \cap E_2^c)$$
by measurability of E_2. Now the last term can be written as $(A \cap E_1^c) \cap E_2^c = A \cap (E_1 \cup E_2)^c$ (because both sides are "in A but not E_1 or E_2), and $A \cap (E_1 \cup E_2) = (A \cap E_1) \cup ((A \cap E_1^c) \cap E_2)$, so by subadditivity of the outer measure we indeed have
$$m^*(A) \ge m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c).$$
- Thus, if $E_1, \cdots, E_n$ are pairwise disjoint C-measurable sets, then $m^*(A \cap (\bigcup_{i=1}^n E_i)) = \sum_{i=1}^n m^*(A \cap E_i)$. (For example, this gives us additivity of outer measure for C-measurable sets if we take A to be the whole space.) Indeed, this can be done by induction; the base case $n = 1$ is clear, and assuming that it works for $(n-1)$ we have (by C-measurability of E_n)
$$m^*\left(A \cap \bigcup_{i=1}^n E_i\right) = m^*\left(\left(A \cap \bigcup_{i=1}^n E_i\right) \cap E_n\right) + m^*\left(\left(A \cap \bigcup_{i=1}^n E_i\right) \cap E_n^c\right).$$
The first term is just $m^*(A \cap E_n)$ because all of the E_is are disjoint, and the second term is $m^*\left(A \cap \bigcup_{i=1}^{n-1} E_i\right)$ (because we can't intersect E_n if we're in E_n^c). Applying the inductive hypothesis to this last term yields the result.

- Next, suppose $E_1, E_2, \cdots$ are C-measurable and $E = \bigcup_{k=1}^{\infty} E_k$. Define $\tilde{E}_k = E_k \setminus \bigcup_{j<k} \tilde{E}_j$, so that $\tilde{E}_1 = E_1$, and then the union of $\tilde{E}_1$ through $\tilde{E}_k$ is the same as the union of E_1 through E_k but constructed inductively to be disjoint. Now we know that the sets $F_n = \bigcup_{k=1}^{n} \tilde{E}_k$ are C-measurable (because it's obtained from a series of manipulations of finite unions and complements of the E_ks). Thus for any set $A \subseteq \mathbb{R}^n$, we have

$$m^*(A) = m^*(A \cap F_n) + m^*(A \cap F_n^c) \geq m^*(A \cap F_n) + m^*(A \cap E^c)$$

because $F_n \subset E$ implies that $E^c \subset F_n^c$. But now applying the previous bullet point, we get

$$m^*(A) \geq \sum_{k=1}^{n} m^*(A \cap \tilde{E}_k) + m^*(A \cap E^c),$$

and taking $n \to \infty$ yields

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap \tilde{E}_k) + m^*(A \cap E^c).$$

Since E is the union of the $\tilde{E}_k$s, subadditivity gives us $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c)$, so the countable union E is indeed measurable.

Putting this all together gives us our result:

Theorem 12

The collection of C-measurable sets is a σ-algebra.

From this, we see that **any open set is C-measurable**. This is because corners are C-measurable, and the σ-algebra containing all open corners also contains all closed boxes, and the countable unions of closed boxes are the open sets. And we get a stronger result:

Proposition 13

A set is C-measurable if and only if it is Lebesgue measurable.

Proof. Let E be a Lebesgue measurable set, and suppose we have any $A \subseteq \mathbb{R}^n$. To show that $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c)$, take an open set $O \supseteq E$ such that $m^*(O \setminus E) < \varepsilon$. But because we know that O is C-measurable (since it's open), we have

$$\begin{aligned} m^*(A) &= m^*(A \cap O) + m^*(A \cap O^c) \\ &\geq m^*(A \cap E) + m^*(A \cap O^c) \\ &\geq m^*(A \cap E) + (m^*(A \cap E^c) - \varepsilon) \end{aligned}$$

because we can apply subadditivity to $A \cap E^c = (A \cap (O \setminus E)) \cup (A \cap O^c)$, and the outer measure of $A \cap (O \setminus E)$ is at most the measure of $O \setminus E$, which is ε. Since ε is arbitrary, we indeed get $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c)$ as desired.

On the other hand, suppose E is C-measurable. Then

$$E = \bigcup_{n=1}^{\infty} E \cap \{||x|| \leq n\},$$

so it suffices to show that any bounded C-measurable set is Lebesgue measurable (since we know Lebesgue measurability is preserved under countable intersection). Bounded sets have finite outer measure, so there is an open set $O \supseteq E$

such that $m^*(O) \le m^*(E) + \varepsilon$. But then $O = E \cup (O \setminus E)$, and because E is C-measurable by assumption and O is C-measurable because it is open, we have (by additivity of outer measure for C-measurable sets) $m^*(O) = m^*(E) + m^*(O \setminus E)$. But this means that $m^*(O \setminus E) = m^*(O) - m^*(E) \le \varepsilon$, so E is Lebesgue measurable. ☐

Notice that the only "analysis" part of this argument open sets are C-measurable – the rest is basically manipulating sets. And now we'll generalize and explain the general construction of a measure:

Definition 14

Let X be a set. A mapping $\mu^* : 2^X \to \mathbb{R}$ is an **outer measure** on X if $\mu^*(\varnothing) = 0$, $\mu^*(A) \le \sum_{k=1}^{\infty} \mu^*(A_k)$ whenever $A \subseteq \cup_{k=1}^{\infty} A_k$ (subadditivity), and it is **finite** if $\mu^*(X) < \infty$.

For example, the Lebesgue measure (this is the same as the Lebesgue outer measure), counting measure (counting the number of elements in a set), and delta measure (where $\delta(A) = 1$ if $0 \in A$ and 0 otherwise) are all outer measures. And if B is Lebesgue measurable for some fixed set B, we can restrict μ^* to B and define the measure $\mu_B^*(A) = \mu^*(A \cap B)$. The Caratheodory definition is then powerful because all of this algebraic manipulation still goes through for general measures:

Definition 15

A set E is **μ-measurable** if for every $A \subseteq X$, we have $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$.

And the same argument we just made for the Lebesgue measure carries over directly:

Theorem 16

For any outer measure μ^*, the set of μ-measurable sets forms a σ-algebra.

3 October 6, 2022

Last lecture, we defined an **outer measure** on a set X to be a map $\mu^* : 2^X \to \mathbb{R}$ satisfying $\mu^*(\varnothing) = 0$ and countable subadditivity $\mu^*\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} \mu^*(A_k)$. (In the Lebesgue case, "outer" makes sense because we're approximating our sets "from the outside," but there isn't any other explanation of the name for now.) We then said that a set E is **μ-measurable** if $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$ for any A (this is the Caratheodory definition of measurability), and then we can write $\mu(E) = \mu^*(E)$ if E is measurable. Our argument (working with various set theory manipulations) then showed that the set of μ-measurable sets is a σ-algebra.

Remark 17. *The concepts of "outer measure" and "measure" should be thought of as basically one and the same: we define the outer measure, and then we use the definition of measurability to decide which sets are measurable and write $\mu(E)$ instead of $\mu^*(E)$ for measurable E.*

Proposition 18

Let $E_1, E_2, \cdots$ be pairwise disjoint measurable sets. Then we have countable **additivity**, meaning that

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k).$$

(The proof is the same as for the Lebesgue measure.)

Proposition 19

Let E_j be measurable sets forming a nested sequence where $E_{j+1} \subseteq E_j$ for all j, and assume that $\mu(E_1) < \infty$. Then $\mu\left(\bigcap_j E_j\right) = \lim_{j\to\infty} \mu(E_j)$.

This result is "continuity from above:" if we have smaller and smaller measurable sets nested inside each other, then the measure of the limit is the limit of the measures. To explain why the finiteness of E_1 matters, we should be careful about measures "escaping off to infinity:" for example, if $E_j = (j, \infty)$, then the intersection is the empty set but each E_j has infinite measure.

Proof. Define $F_n = E_n \setminus E_{n+1}$ for all n, and let $E = \bigcap_{j=1}^{\infty} E_j$ be the intersection of the sets. (The F_ns are disjoint and measurable.) Notice that being in E_1 but not E means that we're in some smallest E_n, so

$$E_1 \setminus E = \bigcup_{j=1}^{\infty} F_j \implies \mu(E_1 \setminus E) = \sum_{j=1}^{\infty} \mu(F_j),$$

which simplifies (by a telescoping sum) to

$$\mu(E_1) - \mu(E) = \sum_{j=1}^{\infty} \mu(E_j) - \mu(E_{j+1}) = \lim_{j\to\infty} \mu(E_1) - E_{j+1},$$

and canceling out the $\mu(E_1)$s (here we use finiteness) yields the result. □

It turns out that when we nest in the other direction, there's no trouble with "losing measure" in the same way:

Proposition 20

Suppose E_k are measurable sets with $E_{k+1} \supseteq E_k$ for all k. Then $\lim_{n\to\infty} \mu(E_n) = \mu\left(\bigcup_{n=1}^{\infty} E_n\right)$.

In particular, $\mu(E_n)$ must increase to infinity if the union of the sets has infinite measure.

Proof. Similarly to before, we have the telescoping sum

$$\mu(E_{k+1}) = \mu(E_1) + \sum_{j=1}^{k} \mu(E_{j+1}) - \mu(E_j) = \mu(E_1) + \sum_{j=1}^{k} \mu(E_{j+1} \setminus E_j)$$

for all k. Taking $k \to \infty$, we find that

$$\lim_{k\to\infty} \mu(E_k) = \lim_{k\to\infty} \left(\mu(E_1) + \sum_{j=1}^{k} \mu(E_{j+1} \setminus E_j)\right) = \mu(E_1) + \sum_{j=1}^{\infty} \mu(E_{j+1} \setminus E_j),$$

and now by countable additivity we find that

$$\lim_{k\to\infty} \mu(E_k) = \mu\left(E_1 \cup \bigcup_{j=1}^{\infty} E_{j+1} \setminus E_j\right) = \mu\left(\bigcup_{j=1}^{\infty} E_j\right).$$

□

Definition 21

A measure μ^* on $\mathbb{R}^n$ is **Borel** if every Borel set is measurable. μ^* is **Borel regular** if it is Borel and for any set A, there is a Borel set $B \supseteq A$ with $\mu^*(A) = \mu(B)$. Finally, μ^* is **Radon** if it is Borel regular and for every compact set K (closed, thus measurable) we have $\mu(K) < \infty$.

Proposition 22

Let μ be a Borel regular measure. Then if A is a measurable set with $\mu(A) < \infty$, then $\nu = \mu|_A$ (the measure restricted to A) is a Radon measure.

Note here that A is not necessarily Borel, but we should prove this first when A is a Borel set and use Borel regularity to restrict to B instead of A. (This is a routine check that we can work out on our own.)

Proposition 23

Let μ be a Borel regular measure, and suppose we have (not necessarily measurable) sets E_j satisfying $E_j \subseteq E_{j+1}$. Then $\lim_{j\to\infty} \mu^*(E_j) = \mu^*\left(\bigcup_{i=1}^\infty E_j\right)$.

Proof. By definition, we can find Borel (measurable) sets $B_k \supseteq E_k$ such that $\mu(B_k) = \mu^*(E_k)$. These B_ks may not be increasing, but we can define the sets $C_k = \bigcap_{j=k}^\infty B_j$ and notice that $E_k \subseteq E_j \subseteq B_j$ for all $j > k$, meaning that $E_k \subseteq C_k$. But also $\mu^*(E_k) = \mu(B_k) \geq \mu(C_k)$, and C_k is an increasing sequence (because we're taking an intersection over fewer sets), so

$$\boxed{\liminf_{k\to\infty} \mu^*(E_k)} \geq \lim_{k\to\infty} \mu(C_k) = \mu\left(\bigcup_{k=1}^\infty C_k\right) \geq \boxed{\mu^*\left(\bigcup_{k=1}^\infty E_k\right)}$$

where the inequalities come from $\mu^*(E_k) \geq \mu(C_k)$, the fact that C_k are measurable, and that $E_k \subseteq C_k$, respectively. But we also have $\boxed{\limsup_{k\to\infty} \mu^*(E_k) \leq \sum_{j=1}^\infty \mu^*(E_j)}$ (because E_k is always contained in the union), so combining these gives the result. □

Theorem 24

Let μ be a Radon measure. Then we have the following properties:

1. For each set A, we have $\mu^*(A) = \inf\{\mu(U) : U \supseteq A, U \text{ open}\}$, like for the Lebesgue measure.
2. For each μ-measurable set A, we have $\mu^*(A) = \mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}$ (in other words, our measure is **inner regular**).

We'll prove these facts first for Borel sets:

Lemma 25

Let μ be a Borel measure on $\mathbb{R}^n$ and B be a Borel set. Then we have the following:

1. If $\mu(B) < \infty$, then for all $\varepsilon > 0$ there is a closed set $C \subseteq B$ with $\mu(B \setminus C) < \varepsilon$.
2. If μ is Radon, then for all $\varepsilon > 0$ there is an open set B such that $\mu(U \setminus B) < \varepsilon$.

Proof. For (1), if $\mu(B)$ is finite, then $\nu = \mu|_B$ is Radon by Proposition 22. Let $\mathcal{F}$ be the set of μ-measurable sets A such that for any $\varepsilon > 0$, there is a closed set $C \subseteq A$ with $\nu(A \setminus C) < \varepsilon$ – we want to show that $\mathcal{F}$ contains all Borel sets. We will do that by showing that it's a σ-algebra that contains all closed sets.

For the first step, $\mathcal{F}$ indeed contains all closed sets, because we can just take $C = A$ in that case. Next, notice that the countable intersection A of sets $A_j \in \mathcal{F}$ is also in $\mathcal{F}$. Indeed, take $C_j \subseteq A_j$ closed with $\mu(A_j \setminus C_j) < \frac{\varepsilon}{2^j}$. Then $\bigcap_{j=1}^{\infty} C_j \subseteq \bigcap_{j=1}^{\infty} A_j$, so letting C be the intersection of the C_js, we have

$$\nu(A \setminus C) \le \nu\left(\bigcup_{j=1}^{\infty} A_j \setminus C_j\right) \le \sum \nu(A_j \setminus C_j) < \varepsilon.$$

We may wish to repeat this argument with the union, but $\bigcup_{j=1}^{\infty} C_j$ may not be closed, so we should look at $\bigcup_{j=1}^{m} C_j$ for a large enough m. Similar manipulations tell us that

$$\lim_{m\to\infty} \nu\left(A \setminus \left(\bigcup_{j=1}^{m} C_j\right)\right) = \nu\left(A \setminus \bigcup_{j=1}^{\infty} C_j\right) \le \nu\left(\bigcup_{j=1}^{\infty} A_j \setminus C_j\right) \le \sum_{j=1}^{\infty} \nu(A_j \setminus C_j) < \varepsilon,$$

where we use that ν is finite, meaning we don't have to worry about infinite measures when applying Proposition 19. So we can choose m large enough so that $\nu\left(A \setminus \bigcup_{j=1}^{m} C_j\right) < \varepsilon$ and let our set be $C = \bigcup_{j=1}^{m} C_j$.

Next, let $\mathcal{G}$ be the collection of all sets A in $\mathcal{F}$ such that A^c is also in $\mathcal{F}$. Then $\mathcal{G}$ contains all open sets (because they are countable unions of closed dyadic boxes and the complement of an open set is closed), and it is a σ-algebra. Indeed, it is closed under complements by definition, and for any sets $A_1, A_2, \dots \in \mathcal{G}$ we know that $\bigcup_{k=1}^{\infty} A_k$ is in $\mathcal{F}$ (we showed closure under infinite unions above) and $\left(\bigcup_{k=1}^{\infty} A_k\right)^c = \bigcap_{k=1}^{\infty} A_k^c$ is also in $\mathcal{F}$ (we showed closure under intersections and $A_k^c \in \mathcal{F}$ because $A_k \in \mathcal{G}$). Thus $\bigcup_{k=1}^{\infty} A_k \in \mathcal{G}$. This means $\mathcal{G}$ satisfies the conditions to be a σ-algebra containing closed sets, so it contains all Borel sets. Thus $\mathcal{F} \supseteq \mathcal{G}$ must contain all Borel sets as well, as desired.

For part (2), let B be a Borel set. If $\mu(B^c)$ is finite, then we can find some closed set $C \subseteq B^c$ with $\mu(B^c \setminus C) < \varepsilon$ by part (1). But then $\mu(C^c \setminus B) = \mu(B^c \setminus C) < \varepsilon$ (draw a Venn diagram), so the open set C^c does the job. On the other hand, if $\mu(B^c) = \infty$, let $B(0, m)$ be the ball of radius m centered at the origin, and let $B_m^c = B^c \cap B(0, m)$. Then for each integer m, we can find $C_m \subseteq B(0, m) \cap B^c$ such that $\mu((B(0, m) \cap B^c) \setminus C_m) < \frac{\varepsilon}{2^m}$. Denoting $Q_m = B(0, m) \cap B$, we have $B = \bigcup_{m=1}^{\infty} Q_m$ and $\mu(C_m^c \setminus Q_m) = \mu(B(0, m) \cap B^c \setminus C_m) < \frac{\varepsilon}{2^m}$ (by the same Venn diagram as before). Taking the union of these open sets C_m^cs, we get an open set O, and $\mu(O \setminus B) < \sum \frac{\varepsilon}{2^m} = \varepsilon$, as desired. □

Proof of Theorem 24. For part (1), if $\mu^*(A) = \infty$, then the inequality automatically holds. Otherwise, since μ is Borel regular, there is some Borel set $B \supseteq A$ with $\mu^*(B) = \mu(A)$. Thus by part (2) of Lemma 25 applied to B, there is some $U \supseteq B$ with $\mu(U) < \mu(B) + \varepsilon$. Thus we've found $U \supset A$ with $\mu^*(U) < \mu^*(A) + \varepsilon$, which implies the result.

Finally, for part (2), if $\mu(A) < \infty$, we know that $\nu = \mu|_A$ is a Radon measure, so applying part (1) to ν gives us an open set $U \supseteq A^c$ with $\mu(U \setminus A^c) = \nu(U) < \varepsilon$. But then we can use the closed set $C = U^c$ to get $\nu(A \setminus C) = \nu(U \setminus A^c) < \varepsilon$. Since $A \setminus C \subseteq A$, we know that $\nu(A \setminus C) = \mu(A \setminus C) < \varepsilon$, as desired. Meanwhile, if A is an infinite measure, we do the same trick with intersecting with larger and larger annuli / balls. And for compactness, we need to make sure to use $C \cap B(0, m)$ for some larger enough m (such that the measure of A outside the ball is small enough). □

Essentially, what we're saying is that we have something similar to the construction of the Lebesgue measure – for **any** measure μ, we can start by defining μ on the open sets U and then define μ^* based on that.

4 October 11, 2022

We'll start today by discussing **measurable functions** (and soon we'll be able to actually start doing analysis). Continuous functions are those where preimages of an open sets are open, but open sets don't form a σ-algebra and thus the limit of a sequence of continuous functions don't need to be continuous (in fact it can be seriously discontinuous). This is bad, because Riemann integration is only built for functions whose discontinuities form a set of measure zero, and even those are hard to work with under limits. So we want to build something that works better for integration theory.

Proposition 26

For any function $f(x)$, the following are equivalent:

1. For any $\alpha \in \mathbb{R}$, the set $\{x : f(x) > \alpha\}$ is μ-measurable.
2. For any $\alpha \in \mathbb{R}$, the set $\{x : f(x) \geq \alpha\}$ is μ-measurable.
3. For any $\alpha \in \mathbb{R}$, the set $\{x : f(x) \leq \alpha\}$ is μ-measurable.
4. For any $\alpha \in \mathbb{R}$, the set $\{x : f(x) < \alpha\}$ is μ-measurable.

Proof. We see that (1) and (3) are equivalent, as are (2) and (4), because the complement of a measurable set is always measurable. Also, notice that $\{x : f(x) > \alpha\} = \bigcup_{n=1}^{\infty}\{x : f(x) \geq \alpha + \frac{1}{n}\}$, so (2) implies (1). Similarly we have (3) implies (4), and all of these together (2 implies 1 implies 3 implies 4 implies 2) mean that any of (1), (2), (3), (4) imply the other. $\square$

Definition 27

Let X be a measure space and Y be a topological space. We say that $f : X \to Y$ is ***μ*-measurable** (also simply **measurable**) if for every open set $O \subset Y$, $f^{-1}(O)$ is μ-measurable.

Proposition 28

Suppose $f, g : X \to \mathbb{R}$ are μ-measurable functions. Then for any $c \in \mathbb{R}$, $cf, f + c, f + g$, and fg are all μ-measurable.

Proof. We see that $\{f(x) + c > \alpha\} = \{f(x) > \alpha - c\}$ is always measurable, and something similar works for cf (though we need to write out separate cases for c positive, negative, or zero). Next, notice that

$$\{f(x) + g(x) < \alpha\} = \bigcup_{q \in \mathbb{Q}} \{f(x) < \alpha - q\} \cap \{g(x) < q\}.$$

(The idea here is that if $f(x)+g(x) < \alpha$, we can always pick a rational number q close enough to $g(x)$.) Since the right-hand side is measurable, so is the left-hand side. Finally, notice that $\{f^2(x) > \alpha\} = \{f(x) > \sqrt{\alpha}\} \cup \{f(x) < -\sqrt{\alpha}\}$, so f^2 is measurable, and thus $\frac{1}{4}((f+g)^2 - (f-g)^2) = fg$ is measurable as well (given that f, g are measurable). $\square$

All of these properties that we've just mentioned should be expected of a nice class of functions.

Theorem 29

If $f_1, f_2, \cdots, f_n$ are all μ-measurable, then so are the functions $g_n(x) = \sup_{1 \le j \le n} f_j(x)$, $q_n(x) = \inf_{1 \le j \le n} f_j(x)$, $g(x) = \sup_n f_n(x)$, $q(x) = \inf f_n(x)$, $s(x) = \limsup_n f_n(x)$, and $w(x) = \liminf_n f_n(x)$.

Proof. All of these are set theory checks (they're general enough that the proofs cannot be too complicated): notice that

$$\{g_n(x) > \alpha\} = \bigcup_{j=1}^{n} \{f_j(x) > \alpha\}, \quad \{g(x) > \alpha\} = \bigcup_{j=1}^{\infty} \{f_j(x) > \alpha\}$$

and the right-hand sides are measurable so the left-hand sides are as well. A similar argument works for $q_n(x)$ and $q(x)$. For the limsup and liminf, we have to be slightly more clever and write

$$s(x) = \limsup_n f_n(x) = \inf_n \left(\sup_{k \ge n} f_k(x) \right),$$

and the inner term on the right is measurable, so the whole infimum is as well. A very similar argument works for $w(x)$ too. $\square$

This result is extremely useful in real analysis, and it's important to note that everything started with the Caratheodory definition (which makes it easy to check that the set of measurable sets forms a σ-algebra). So it's important that we started with the right definition, and we'll see that this definition is useful when we get into integration too.

Definition 30

A measurable function $f(x)$ is **simple** if it takes at most countably many values, meaning that it can be written as

$$f(x) = \sum_j \alpha_j 1\{x \in A_j\},$$

where $1\{x \in A_j\}$ is the indicator function of a measurable set A_j (also denoted 1_{A_j}) and the A_js are pairwise disjoint.

Theorem 31

Let $f(x)$ be μ-measurable, and suppose $f(x) \ge 0$ for μ-almost-every x (so except for a set of measure zero). Then there are μ-measurable sets A_k such that $f(x) = \sum_{k=1}^{\infty} \frac{1}{k} 1\{x \in A_k\}$.

In particular, truncating this sum at some finite n means that $f(x)$ can only take on at most 2^n values, so we can approximate any μ-measurable function with simple functions by breaking up into the 2^n disjoint sets on which f takes on those values.

Proof sketch. We can essentially imagine "raising a water level" by 1, then $\frac{1}{2}$, then $\frac{1}{3}$, and so on, and adding as much to each set A_k to keep the water below the function.. Start with our function f, and let A_1 be the set of values where $f(x) \ge 1$. Then define A_2 to be the set of values where $f(x) - 1\{x \in A_1\} \ge \frac{1}{2}$, A_3 to be the set of values where $f(x) - 1\{x \in A_1\} - \frac{1}{2} \cdot 1\{x \in A_2\} \ge \frac{1}{3}$, and so on. Notice that for all k, we have

$$f(x) \ge \sum_{j=1}^{k} \frac{1}{j} 1\{x \in A_j\}$$

inductively by the definition of A_j, so taking the limit yields

$$f(x) \geq \sum_{j=1}^{\infty} \frac{1}{j} 1\{x \in A_j\}.$$

But if this inequality were strict, then we must have $x \in A_j$ for all j (because otherwise the A_js would make $f(x) - \sum_j \frac{1}{j} 1\{x \in A_j\}$ arbitrarily close to zero), meaning that $f(x) \geq \sum_{j=1}^{\infty} \frac{1}{j} = \infty$. $\square$

Notice that all that's important with our $\frac{1}{k}$ factor is that it goes to zero but the sum $\sum_k \frac{1}{k}$ diverges, so we can replace it with any other sequence α_k with $\alpha_k \to 0$ and $\sum \alpha_k = \infty$.

We'll next discuss **extension of a continuous function** from a compact set:

Problem 32

Let $K \subset \mathbb{R}^n$ be a compact set, and let f be continuous on K. Can we extend f to all of $\mathbb{R}^n$, and do we need to increase the supremum norm of f when we do so?

(Here, we should not think of "compact" as being particularly nice, because closed sets can look pretty ugly. This is a common kind of problem in analysis, in which we want to gain some additional structure without losing too much of what we already have.) The answer turns out to be as nice as possible:

Theorem 33

Let $K \subset \mathbb{R}^n$ be compact, and let $f : K \to \mathbb{R}$ be continuous. Then there exists a continuous function $\overline{f} : \mathbb{R}^n \to \mathbb{R}$, such that $\overline{f}(x) = f(x)$ on K and $\sup_{x \in \mathbb{R}^n} |\overline{f}(x)| = \sup_{x \in K} |f(x)|$.

Proof. We need to extend $f(x)$ to the open set $U = K^c$ so that if $x \in U$ and $y \in K$ is near x, we have $f(x)$ close to $f(y)$. Let $s_1, s_2, \cdots$ be a dense set in K (for example, take a finite cover of K using balls of radius $\frac{1}{2^n}$ for each $n \geq 0$, and pick a point in each ball), and the idea will be to define a weight $w_j(x)$ such that $\sum_{j=1}^{\infty} w_j(x) = 1$ for all $x \in U$. Then we can define $\overline{f}(x) = \sum_{j=1}^{\infty} w_j(x) f(s_j)$ – specifically, we wish to take weighted averages of values of f on s_j, and for s_j close to x we weight further.

To do this mathematically, consider the function

$$u_s(x) = \max\left[2 - \frac{|x - s|}{\text{dist}(x, K)}, 0\right].$$

We can check that $0 \leq u_s(x) \leq 1$ for all $x \in U$ and $s \in K$ because $\text{dist}(x, K) \leq |x - s|$, and $u_s(x) = 0$ if $|x - s| \geq 2 \,\text{dist}(x, K)$. Notice also that $u_s(x) \to 1$ (uniformly) as $x \to \infty$ and s is kept fixed. Also, this function is continuous for any s, so we can define

$$\sigma(x) = \sum_{j=1}^{\infty} \frac{1}{2^j} u_{s_j}(x),$$

which is also continuous by the Weierstrass test. Additionally, notice that $\sigma(x) > 0$ for all $x \in U$, because there is a point $y \in K$ where $\text{dist}(x, K) = \text{dist}(x, y)$, and there is some s_j sufficiently close to y that makes $u_{s_j}(x) \neq 0$. And with that we can finally define weights by renormalizing as

$$v_j(x) = \frac{1}{2^j} \frac{u_{s_j}(x)}{\sigma(x)}.$$

This is always nonnegative, and we have $\sum_{j=1}^{\infty} v_j(x) = 1$ for all $x \in U$. So then we can define

$$\overline{f}(x) = \sum_{j=1}^{\infty} v_j(x) f(s_j)$$

for all $x \in U$, and keep $\overline{f}(x) = f(x)$ for all $x \in K$. We check that this function has the desired properties: for any $x \in U$, we can find a ball around U not intersecting K, so $\overline{f}(x) = \frac{1}{\sigma(x)} \sum_{j=1}^{\infty} \frac{f(s_j)}{2^j} u_{s_j}(x)$. But σ and u_{s_j} are all continuous, and $0 \le u_{s_j}(x) \le 1$ and $|f(s_j)| \le M$ for all j, where $M = \sup_{y \in K} |f(y)|$. Thus our sum is continuous (again by Weierstrass) and bounded by $\frac{1}{\sigma(x)} \sum_{j=1}^{\infty} \frac{M}{2^j} u_{s_j}(x) = \frac{1}{\sigma(x)} M(\sigma(x)) = M$. Thus this function is continuous on U and bounded by the same sup norm.

Finally, we must check that this function is continuous on K. For any $x \in K$ and $y \in U$, our goal is to compare $f(x)$ with $\overline{f}(y)$ and show that $\overline{f}(y)$ is close to $f(x)$ (using that points contributing to the sum for y are also close to x). To make this more precise, note that $u_s(y) = 0$ if $|y - s| \ge 2 \operatorname{dist}(y, K)$. So for any $x \in K$, we can find $\delta > 0$ such that $|f(x) - f(z)| < \varepsilon$ if $|x - z| < \delta$. Now for any $y \in U$ with $|x - y| < \frac{\delta}{100}$, notice that if $|x - s_k| \ge \delta$, then $|y - s_k| \ge \frac{\delta}{4}$ by the triangle inequality (because s_k is far from x, but y is close to x). But this means $|y - s_k| > 2 \operatorname{dist}(y, K)$, meaning that $u_{s_k}(y) = 0$. Thus

$$\overline{f}(y) = \sum_j v_j(x) f(s_j) = \sum_{j : |s_j - x| < \delta} v_j(x) f(s_j)$$

is only a weighted average over points at most δ away from x, while we can write $f(x)$ in the same way because $\sum v_j(y) = 1$ for any $y \in U$:

$$f(x) = \sum_{i=1}^{\infty} v_j(y) f(x) = \sum_{|s_j - x| < \delta} v_j(y) f(x).$$

If we subtract these two expressions, we see that

$$|\overline{f}(y) - f(x)| \le \sum_{|s_j - x| < \delta} v_j(y) |f(x) - f(s_j)| < \frac{\varepsilon}{100} \sum v_j(y) = \frac{\varepsilon}{100} < \varepsilon.$$

So indeed we find that for any $\varepsilon > 0$, if $|x - y| < \frac{\delta}{100}$, then $|f(x) - \overline{f}(y)| < \varepsilon$. So $\overline{f}$ is continuous on K as well (because $f(y) \to f(x)$ for y within K and also within U), completing the proof. $\square$

One useful thing to ask is if $f(x)$ is Lipschitz on K (meaning that $|f(x) - f(y)| \le C|x - y|$ for all $x, y \in K$), whether $\overline{f}$ constructed in this way is also Lipschitz, and if so whether the Lipschitz constant needs to change.

Theorem 34 (Lusin)

Let μ be a Borel-regular measure and let $f : \mathbb{R}^n \to \mathbb{R}$ be μ-measurable. Let $A \subseteq \mathbb{R}^n$ be μ-measurable with $\mu(A) < \infty$. Then for any $\varepsilon > 0$, there is some compact set $K_\varepsilon \subset A$ such that f is continuous on K_ε and $\mu(A \setminus K_\varepsilon) < \varepsilon$.

In other words, any measurable function restricts to a continuous function on almost all of A. In particular, for any f, we can look at the function $\overline{f}$ formed by extending f_{K_ε} to all of A using Theorem 33, yielding the following corollary:

Corollary 35

Let A be μ-measurable and $f : A \to \mathbb{R}$ be $\mu|_A$-measurable with $\mu(A) < \infty$. Then there is a continuous function $\overline{f} : \mathbb{R}^n \to \mathbb{R}$ and a compact set $K_\varepsilon \subseteq A$ such that $\mu(A \setminus K_\varepsilon) < \varepsilon$ and f agrees with $\overline{f}(x)$ on all of K_ε.

Proof of Theorem 34. We'll construct a family of continuous functions f_p such that $f_p(x) \to f(x)$ uniformly on K_ε. Define the sets $B_{p,j} = \left[\frac{j}{2^p}, \frac{j+1}{2^p}\right)$ for $j \in \mathbb{Z}$ and $p \in \mathbb{N}$, and let $A_{p,j} = A \cap f^{-1}(B_{p,j})$ and $A = \bigcup_{j=-\infty}^{\infty} A_{p,j}$. Then we have compact sets $K_{p,j} \subseteq A_{p,j}$ such that $\mu(A_{p,j} \setminus K_{p,j}) < \frac{\varepsilon}{10 \cdot 2^{p+|j|}}$. This means that when we sum over j, we find

$$\mu\left(A \setminus \bigcup_{j \in \mathbb{Z}} K_{p,j}\right) < \frac{\varepsilon}{2 \cdot 2^p}$$

for each p, and thus we can pick $N(p)$ so that $\mu\left(A \setminus \bigcup_{|j| \le N_p} K_{p,j}\right) < \frac{\varepsilon}{2^p}$ as well (here we use continuity from above, in particular the fact that $\mu(A)$ is finite). Then we get a compact set $K_\varepsilon = \bigcup_{|j| \le N(p)} K_{p,j}$, but the $K_{p,j}$s are pairwise disjoint for any fixed p because each one is contained within $A_{p,j}$, which are contained within preimages of distinct intervals $B_{p,j}$.

Now for each p, define $\overline{f}_p(x)$ to be $\frac{j}{2^p}$ on $K_{p,j}$, meaning that we're defining the function on the compact set $K_{\varepsilon,p} = \bigcup_{|j| \le N(p)} K_{p,j}$. Each of these $\overline{f}$s are continuous (because they're constant on disjoint compact sets), and $|\overline{f}_p(x) - f(x)| < \frac{1}{2^p}$ for all $x \in K_{\varepsilon_p}$. Thus if we look at the functions $\overline{f}_p$ on all of $\overline{K_\varepsilon} = \bigcap_{p=1}^{\infty} K_{\varepsilon,p}$, we have continuous functions $\overline{f}_p$ defined on K_ε which uniformly converge to f, meaning that f is continuous on K_ε. Since we lose measure $\frac{\varepsilon}{2^p}$ for each p, $\mu(A \setminus K_\varepsilon) < \varepsilon$, as desired. □

5 October 13, 2022

We'll start today with a result similar to Lusin's theorem:

Theorem 36 (Egorov)
Let μ be a measure on $\mathbb{R}^m$, and let $f_n : \mathbb{R}^m \to \mathbb{R}$ be μ-measurable. Assume that $\mu(A) < \infty$ and $f_n \to g$ μ-almost-everywhere on A. Then for all $\varepsilon > 0$, there is a set $B_\varepsilon \subseteq A$ such that $f_n \to g$ **uniformly** on B_ε and $\mu(A \setminus B_\varepsilon) < \varepsilon$.

This theorem fails if we do not have the assumption $\mu(A) < \infty$: for example, consider the functions $f_n : \mathbb{R} \to \mathbb{R}$ given by $f_n(x) = \frac{x}{n}$, which converge to 0 but do not converge uniformly on unbounded set.

Proof. The idea is to "throw out the sets" which are not yet good enough to be uniformly convergent: define

$$C_{ij} = \bigcup_{k=j}^{\infty} \left\{x \in A : |f_k(x) - g(x)| > \frac{1}{2^i}\right\}.$$

Here, i is the "scale of closeness," and not being in C_{ij} means that once we're at the jth function, we're within $\frac{1}{2^i}$ of the limit. Furthermore, $\bigcap_{j=1}^{\infty} \mu(C_{ij}) = 0$ for any i, because a point x can only be in that set if $f_n(x)$ does not converge to $g(x)$ (because for any j we can find some $k > j$ such that $f_k(x)$ is more than $\frac{1}{2^i}$ away from $g(x)$) . Thus for each i, there is some N_i such that $\mu(C_{i,N_i}) < \frac{\varepsilon}{2^i}$.

Then for all $x \in A \setminus C_{i,N_i}$, we have that $|f_n(x) - g(x)| < \frac{1}{2^i}$ once $n \ge N_i$, so $f_n \to g$ uniformly on the set $B = A \setminus \bigcup_{i=1}^{\infty} C_{i,N_i}$ (because for any ε we can find an i, yielding an N_i that uniformly works for all x), and $\mu(A \setminus B) < \sum_i \frac{\varepsilon}{2^i} = \varepsilon$, as desired. □

Notice that the structure of $\mathbb{R}^m$ is not so important here – we don't need compactness or anything – so this result is really a statement about measure spaces in general.

We'll now mention an alternative notion of convergence which is useful in probability theory:

Example 37

Consider a sequence of functions on $[0,1]$, where f_n is the indicator function of an interval of length $\frac{1}{n}$. Then for large n, the "probability" of f_n being zero gets larger and larger, but f_n doesn't converge pointwise to 0 anywhere if we make the intervals line up back-to-back (and loop around $[0,1]$), since $f_n(x)$ will be 1 infinitely many times for each x.

Definition 38

A sequence of measurable functions f_n **converges in probability** to f on a set E if for all $\varepsilon > 0$, there is some N such that $\mu\{x \in E : |f_n(x) - f(x)| > \varepsilon\} < \varepsilon$ for all $n > N$.

Theorem 39

If f_n converges in probability to f on a set E, then there is a subsequence n_k such that we have pointwise convergence $f_{n_k} \to g$ almost everywhere on E.

The point is that in an example like Example 37, we should take a subsequence that only visits every point finitely many times. That prompts the following approach:

Proof. For any j, we can find some N_j such that for all $n \ge N_j$ we have

$$\mu\left(x \in E : |g(x) - f_n(x)| \ge \frac{1}{2^j}\right) < \frac{1}{2^j}.$$

We can then take our subsequence to be $\phi_j(x) = f_{N_j}(x)$. Then defining the sets $E_j(x) = \left\{x \in E : |g(x) - f_{N_j}(x)| \ge \frac{1}{2^j}\right\}$, we get a decreasing sequence of sets $D_k = \bigcup_{j=k}^{\infty} E_j$. If $x \notin D_k$, then we have $|g(x) - f_{N_j}(x)| < \frac{1}{2^j}$ for all $j \ge k$, which means that $f_{N_j}(x)$ converges to $g(x)$ as long as $x \notin \bigcap_{k=1}^{\infty} D_k$.

But the measure of $\bigcap_{k=1}^{\infty} D_k$ is at most the measure of D_k, and the measure of each D_k is at most $\sum_{j=k}^{\infty} \frac{1}{2^j} \le \frac{2}{2^k}$ by the definition of N_j. Thus the measure of this intersection must be zero, and our function converges almost everywhere on this subsequence. $\square$

We'll now turn our attention to the Lebesgue integral (following the Royden treatment) – the idea is that instead of computing with the Riemann integral and dividing up our x-axis into small intervals, calculating areas of rectangles, we'll split the y-axis into small intervals and look at the measure horizontally instead. (And this is better because we don't need to worry about issues with discontinuities like in the Riemann integral.)

To define the integral, we'll build it up step by step:

Definition 40

For any nonnegative simple function $f(x) = \sum_j \alpha_j 1\{x \in A_j\}$, we define

$$\int f(x)d\mu = \sum_j \alpha_j \mu(A_j).$$

(We may also just write this as $\int f$.) More generally for a simple function $f = f^+ - f^-$ with both $\int f^+$ and $\int f^-$ finite, we define $\int f d\mu = \int f^+ d\mu - \int f d\mu$.

Proposition 41

Let f be a bounded function defined on a measurable set E with $\mu(E) < \infty$. Then the "upper" and "lower" approximations $\int_E^* f\,d\mu = \inf_{f\le\psi}\int_E \psi d\mu$ and $\int_{*,E} f\,d\mu = \sup_{f\ge\phi}\int_E \phi d\mu$ are equal (where we take the inf and sup over all simple functions satisfying the corresponding inequalities) if and only if f is measurable.

Proof. First suppose that f is measurable. Fix some positive integer n; because $|f| \le M$, we can define

$$E_k = \left\{x : \frac{k}{n} \le f(x) < \frac{k+1}{n}\right\}.$$

There are finitely many nonempty E_ks (for a fixed n), the E_ks are disjoint and measurable, and we can consider the "upper approximation" $\psi_n(x) = \sum_k \frac{k+1}{n} 1\{x \in E_k\}$ and "lower approximation" $\phi_n(x) = \sum_k \frac{k}{n} 1\{x \in E_k\}$. By definition of the set E_k we know that $\phi_n(x) \le f(x) \le \psi_n(x)$, and

$$\int_E \psi_n d\mu - \int_E \phi_n d\mu = \sum_{k=-\infty}^{\infty} \frac{1}{n}\mu(E_k) < \frac{1}{n}\mu(E).$$

We then find that

$$\boxed{\int_E^* f\,d\mu} \le \int_E \psi_n d\mu \le \frac{1}{n}\mu(E) + \int_E \phi_n d\mu \le \boxed{\frac{1}{n}\mu(E) + \int_{*,E} f\,d\mu}.$$

Since $\mu(E)$ is finite, taking $n \to \infty$ shows that the two approximations are equal.

On the other hand, suppose that $\int_{*,E} f\,d\mu = \int_E^* f\,d\mu$. Then we can find $\phi_n \le f$ and $\psi_n \ge f$ simple functions so that $|\int_E(\psi_n - \phi_n)d\mu| < \frac{1}{n}$. Let $\psi^*(x) = \liminf_n \psi_n(x)$ and $\phi_*(x) = \limsup_n \phi_n(x)$. Since $\phi_n(x) \le f(x) \le \psi_n(x)$ for all n, taking limits tells us that $\phi_*(x) \le f(x) \le \psi^*(x)$ as well.

But for any j, we can define $A_j = \left\{\psi^*(x) > \phi_*(x) + \frac{1}{2^j}\right\}$. Then (by definition of liminf and limsup) for $n > N_j$ large enough, we have $\psi_n(x) \ge \psi^*(x) - \frac{1}{10\cdot 2^j} > \phi_*(x) + \frac{1}{10\cdot 2^j} > \phi_n(x)$ on all of A_j, because ψ^* and ϕ_* differ by at least $\frac{1}{2^j}$ and bringing them closer by $\frac{1}{10\cdot 2^j}$ on each side still preserves $\psi^* > \phi_*$. But this means that (because we're integrating nonnegative functions)

$$\int_E (\psi_n - \phi_n)d\mu \ge \int_{A_j} (\psi_n - \phi_n)d\mu \ge \frac{1}{100\cdot 2^j}\mu(A_j)$$

for all $n > N_j$. (Here the $\frac{1}{100}$ can really just be $\frac{4}{5}$, but we don't want to worry about constants.) So we can only have $\lim\int\psi_n = \lim\int\phi_n$ if $\mu(A_j) = 0$ for all j, which means that $\psi^*(x) = \phi_*(x)$ almost everywhere. Since f is sandwiched between them, we find that $f(x) = \psi^*(x) = \phi_*(x)$, and in particular it is measurable because it is a liminf of simple functions (which are measurable), as desired. □

In words, this result tells us that measurability is the same as being able to be approximated from above and below by simple functions. So we've found another way to work with the abstract Caratheodory notion that is "more physical."

Definition 42

For any bounded measurable function f defined on a measurable set E of finite measure, we define

$$\int_E f\,d\mu = \sup_{\psi \le f, \psi \text{ simple}} \int_E \psi\,d\mu.$$

Similarly, for **any nonnegative** measurable function f defined on **any** measurable set E, we have

$$\int_E f\,d\mu = \sup_{\substack{\psi \le f, \psi \text{ simple and vanishes} \\ \text{everywhere except a set of finite measure}}} \int_E \psi\,d\mu.$$

Finally, for any measurable function $f = f^+ - f^-$, we say that f is **integrable** if $\int_E f^+$ and $\int_E f^-$ are both defined, and we set $\int_E f\,d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$.

In other words, we approximate from below by simple functions for which we know how to calculate the integral. (And we approximate from below to avoid worrying about infinite integrals, though we have seen that for measurable functions approximating from below and above are equivalent.) Here, we should note that there are different conventions for whether "integrability" requires that the integrals are finite, so we should be careful.

We'll now mention two useful results that are often used in probability:

Theorem 43 (Markov's inequality)

For any nonnegative measurable function $f \ge 0$ and any $\lambda > 0$, we have $\mu\{x : f(x) > \lambda\} \le \frac{1}{\lambda} \int f(x) d\mu$.

In probabilistic terms, this tells us that the probability a random variable is large is bounded in terms of the expectation of the random variable (since $\int f(x) d\mu$ is the average value of f if μ is a probability measure with total measure 1).

Proof. Notice that $f(x) > \lambda \cdot 1_{E_\lambda}$, so

$$\int f(x) d\mu \ge \lambda \int 1_{E_\lambda} d\mu = \lambda\mu(x : f(x) > \lambda).$$

Dividing by λ yields the result. $\square$

Not all random variables are always positive, so there is a related result that can also get us tail bounds more generally:

Theorem 44 (Chebyshev's inequality)

Suppose f is a measurable function. Define $\overline{f} = \int f\,d\mu$ and $m_2 = \int (f - \overline{f})^2 d\mu$ (these are the "mean" and "variance," respectively). Then for any $\lambda > 0$,

$$\mu\left(x : |f(x) - \overline{f}| > \lambda\sqrt{m_2}\right) \le \frac{1}{\lambda^2}.$$

In other words, the probability we're more than λ standard deviations away from the mean is at most $\frac{1}{\lambda^2}$. This result does not require that we have a probability measure μ (meaning that $\mu(E) = 1$), but notice that this inequality is strongest when we minimize m_2 (as a function of $\overline{f}$), and this in fact occurs where $\frac{1}{\mu(E)} \int f\,d\mu$ is plugged in for $\overline{f}$.

Proof. Let $g(x) = (f(x) - \overline{f})^2$. Then by Markov's inequality applied to g,

$$\mu(x : (f - \overline{f})^2 > \lambda^2 m_2) \leq \frac{1}{\lambda^2 m_2} \int (f(x) - \overline{f})^2 d\mu,$$

and this right-hand side is $\frac{1}{\lambda^2 m_2} \cdot m_2 = \frac{1}{\lambda^2}$. $\square$

6 October 18, 2022

Our topic for today is the **Lebesgue dominated convergence theorem**, which we'll prove in baby steps. Much like we defined the Lebesgue integral in a sequence of steps, we'll need to slowly build up towards our general result. The question we're basically asking is whether convergence $f_n \to f$ almost everywhere on a set E implies $\int_E f_n d\mu \to \int_E f d\mu$. There's a few obvious counterexamples which show that this isn't true in general:

Example 45

If $f_n = 1_{[n,n+1]}$ for all n, then $\int f_n = 1$ for all n, but $f_n \to 0$ for all $x \in \mathbb{R}$ so $\int f = 0$.

Example 46

If f_n is a triangle of height n and base $\frac{1}{n}$ (that is, we define $f_n(x) = 2nx$ for $x \in [0, \frac{1}{2}]$, then again $\int f_n = 1$ for all n but $f_n \to 0$ for all $x \in \mathbb{R}$ so $\int f = 0$ again.

In the first example, we should think of the issue as "the function escaping to infinity horizontally," and in the second example, we can think of the issue as "frequency going to infinity" or "the function escaping to infinity vertically." So whatever convergence theorems we have must rule these issues out.

Theorem 47 (Bounded convergence theorem)

Let E be a set of finite measure. Assume there is some $M > 0$ such that $|f_n(x)| \leq M$ for all $x \in E$ and all positive integers n. Then if $f_n(x) \to f$ almost everywhere on E, then $\int_E f_n d\mu \to \int_E f d\mu$.

Proof. By Egorov's theorem, for any $\varepsilon > 0$ we can find a subset $A_\varepsilon \subseteq E$ such that $f_n \to f$ uniformly on A_ε and $\mu(E \setminus A_\varepsilon) < \varepsilon$. Then we can break up the integral and use the triangle inequality to find

$$\left| \int_E f_n d\mu - \int_E f d\mu \right| \leq \int_{A_\varepsilon} |f_n - f| d\mu + \int_{E \setminus A_\varepsilon} |f_n - f| d\mu.$$

We may choose N large enough so that $|f_n - f| < \frac{\varepsilon}{100\mu(E)}$ on all of A_ε for all $n \geq N$ and pick a set A_ε so that $\mu(E - A_\varepsilon) < \frac{\varepsilon}{100M}$. Then the first integral can be bounded by $\frac{\varepsilon\mu(A_\varepsilon)}{100\mu(E)} \leq \frac{\varepsilon}{100}$, while the second integral can be bounded by $2M\mu(E \setminus A_\varepsilon) < \frac{2\varepsilon}{100}$. Thus the overall expression is less than ε, and thus we must indeed have $\int_E f_n d\mu \to \int_E f d\mu$. $\square$

Theorem 48 (Fatou's lemma)

Let $f_n \geq 0$ be measurable, and suppose $f_n \to f$ converges μ-almost-everywhere on a measurable set E. Then

$$\int_E f d\mu \leq \liminf_{n \to \infty} \int_E f_n d\mu.$$

In other words, if we have positive functions, then we cannot "gain mass" by taking the limit (though we can lose mass, as we can see from the examples above). So this is a very useful tool for obtaining estimates on a limiting object.

Proof. Let $h \le f$ be any simple function that vanishes outside a set of finite measure. Define $h_n(x) = \min(h(x), f_n(x))$. We know that $h_n(x) \to h(x)$ μ-almost-everywhere (because if $h(x) = f(x)$ then we converge to $f(x)$, and otherwise eventually $h_n(x) = f_n(x)$), but h is bounded and vanishes outside a set of finite measure E', so by the bounded convergence theorem we have

$$\int_E h d\mu = \int_{E'} h d\mu = \lim_{n\to\infty} \int_{E'} h_n d\mu \le \liminf_{n\to\infty} \int_{E'} f_n d\mu,$$

since $h_n \le f_n$ for each n. But now increasing the set E' to E only increases the integral because the f_n are positive, so $\int_E h d\mu \le \liminf_{n\to\infty} \int_E f_n d\mu$ for any simple function of the form of h. Taking the sup over all such h, we find that $\int f d\mu \le \liminf_{n\to\infty} \int_E f_n d\mu$, as desired. $\square$

Theorem 49 (Monotone convergence theorem)
Let $f_n(x)$ be an increasing sequence of nonnegative-valued functions on E (where the f_n and E need to be measurable, but we won't keep writing this down). Then

$$\int_E f d\mu = \liminf_{n\to\infty} \int_E f_n d\mu.$$

Proof. This follows directly from Fatou's lemma, since one direction already holds and then $f \ge f_n$ for all n yields the other inequality. $\square$

Since partial sums of nonnegative functions form an increasing sequence, we thus find that if f_n are a set of nonnegative functions, then

$$\int_E \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu.$$

Theorem 50 (Lebesgue dominated convergence theorem)
Let $f_n \to f$ almost everywhere on a set E, and assume there is some function $g(x)$ such that (1) $|f_n(x)| \le g(x)$ for μ-almost-every $x \in E$ and (2) $\int_E |g(x)| d\mu < \infty$. Then $\int_E f_n d\mu \to \int_E f d\mu$.

In other words, it's okay for g to not have compact support and also not be bounded, but it must be defined in a way that keeps g integrable. Then if the f_ns are stuck in the finite volume contained between $-g$ and g, then we get convergence of the integral. (Notice that the bounded convergence theorem is a special case of this where E is a set of finite measure and g is a constant.)

Proof. The functions $g - f_n$ are all nonnegative (because $|f_n| \le g$ for all n), so by Fatou's lemma we have

$$\int_E (g - f) d\mu \le \liminf \int_E (g - f_n) d\mu.$$

But because g is integrable and $|f_n|$ and $|f|$ are all bounded by g so they are all integrable, we can separate the integrals and find that

$$\int_E g d\mu - \int_E f d\mu \le \int_E g d\mu - \limsup_{n\to\infty} \int f_n d\mu \implies \int_E f d\mu \ge \limsup_{n\to\infty} \int_E f_n d\mu.$$

On the other hand, $g + f_n$ are all also nonnegative functions, so Fatou's lemma tells us that

$$\int_E (g+f)d\mu \le \liminf \int_E (g+f_n)d\mu \implies \int_E f d\mu \le \liminf \int_E f_n d\mu.$$

Thus the liminf is at least the limsup of $\int_E f_n d\mu$, and $\int_E f d\mu$ is between them, so the limit must actually exist and be equal to $\int_E f d\mu$. $\square$

Notice that we did need to build up this result slowly (we used Fatou's lemma, which was obtained from the bounded convergence theorem), and the "real content" of the result is not in this proof but rather in Fatou's lemma and perhaps the definition of the Lebesgue integral as an approximation by simple functions.

We'll next discuss **absolute continuity** of the integral, which in particular tells us that the Dirac delta function is not an integrable function:

Theorem 51

Let $f \ge 0$ be integrable. Then for any $\varepsilon > 0$, there is some $\delta > 0$ such that if we have any set A with $\mu(A) < \delta$, then $\int_A f d\mu < \varepsilon$.

Proof. Suppose otherwise. Then there is some $\varepsilon_0 > 0$ such that there are sets A_n of measure $\mu(A_n) < \frac{1}{2^n}$ where $\int_{A_n} f d\mu \ge \varepsilon_0$ for all n (since we can find sets of arbitrarily small measure that achieve integral ε). Defining $g_n(x) = f(x)1_{A_n}(x)$, we see that $g_n(x) \to 0$ except if x is in infinitely many of the A_ns, meaning that the "bad set" is

$$A = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j$$

(since being in infinitely many A_js is equivalent to being in every $\bigcup_{j=n}^{\infty} A_j$). But then by countable subadditivity we have

$$\mu(A) \le \sum_{j=n}^{\infty} \mu(A_j) \le \frac{2}{2^n},$$

so we must have $\mu(A) = 0$ and thus $g_n(x) \to 0$ μ-almost-everywhere. But applying Fatou's lemma to $f - g_n$, we find that because $f(x) - g_n(x)$ converges to $f(x)$ for almost all x, we have

$$\int f d\mu \le \liminf_{n\to\infty} \int (f - g_n) = \int f d\mu - \limsup_{n\to\infty} \int g_n d\mu$$

(because f and $g_n \le f$ are all integrable). But each $\int g_n d\mu$ is at least ε_0, so the limsup is also at least ε_0, which is a contradiction. $\square$

With that, we're ready to turn to calculus, specifically looking at the **Newton-Leibniz formula** (that is, the fundamental theorem of calculus)

$$f(b) - f(a) \overset{?}{=} \int_a^b f'(x)dx, \quad f(x) \overset{?}{=} \frac{d}{dx}\int_a^x f(z)dz.$$

We want to ask whether these identities hold – for example, we want to ask if f' is integrable when it exists and whether it's equal to $f(b) - f(a)$ (as well as whether we can extend the result to a larger class of functions by density), and we want to ask whether $\int_a^x f(z)dz$ is differentiable and equal to $f(x)$ almost everywhere. But first we need to define the derivative:

Definition 52

Define the functions $D^+f(x) = \limsup_{h\to 0^+} \frac{f(x+h)-f(x)}{h}$, $D_+f(x) = \liminf_{h\to 0^+} \frac{f(x+h)-f(x)}{h}$ (upper and lower derivatives on the right) and $D^-f(x) = \limsup_{h\to 0^-} \frac{f(x)-f(x-h)}{h}$, $D_-f(x) = \liminf_{h\to 0^+} \frac{f(x)-f(x-h)}{h}$ (upper and lower derivatives on the left). If all four functions coincide, then we say that f is differentiable and call its derivative $Df(x)$.

Theorem 53

Suppose $f \in L^1([a,b])$ (meaning that f is integrable on $[a,b]$), and suppose $F(t) = \int_a^t f(x)dx$. Then $F'(t) = f(t)$ almost everywhere.

Definition 54

A function $g(x)$ is **absolutely continuous** on an interval $[a,b]$ if for any $\varepsilon > 0$, there is some $\delta > 0$ such that whenever we have a finite collection of nonnegative h_ns with $\sum_{n=1}^N h_n < \delta$ and x_n arbitrary so that (x_n, x_n+h_n) do not overlap,

$$\sum_{n=1}^N |f(x_n+h_n) - f(x_n)| < \varepsilon.$$

This is similar to uniform continuity but even stronger, since we can "split up the total length" into arbitrarily small intervals. In particular, notice that if $f \in L^1$, then $F(x) = \int_a^x f(t)dt$ is absolutely continuous by absolute continuity of the integral, since $F(x+h_n) - F(x)$ is the integral of f over the interval $[x, x+h_n]$.

Theorem 55

A function $F(t)$ can be written as $\int_a^t f(x)dx$ with some $f \in L^1$ if and only if F is absolutely continuous.

We'll prove these results about absolute continuity in future lectures, but we'll first state a benign-looking result here:

Theorem 56

Any increasing function f on $[a,b]$ is differentiable almost everywhere on $[a,b]$, and f' is a measurable function.

This result might seem similar to the fact that any increasing function can only have a countable set of discontinuities (because each one is a jump discontinuity and the sum of uncountably many positive numbers is infinite), but it turns out to be more complicated. We'll need Vitali's covering lemma for this:

Definition 57

A **cover** of a set E by nontrivial closed intervals is a **fine cover** of E if for any $x \in E$ and any $\varepsilon > 0$, there is some interval I of length less than ε containing x.

Theorem 58 (Vitali's lemma)

Let E be a set with $m^*(E) < \infty$. Any fine cover of E by closed intervals has the following property: for any $\varepsilon > 0$, there is a finite subcollection $I_1, \cdots, I_{N_\varepsilon}$ such that the I_ks are disjoint and $m^*(E \setminus \bigcup_{k=1}^{N_\varepsilon} I_\varepsilon) < \varepsilon$.

In other words, we get a result similar to Heine-Borel with a compromise, and the advantage is that we can get disjoint intervals I_k without losing too much measure.

Proof. Since $m^*(E) < \infty$ and our cover is fine, there exists some open set $O \supseteq E$ such that $m(O) \subset \infty$ and we can assume that $I \subseteq O$ for all intervals I (because we can just include the small-enough intervals containing each x that don't go outside O and we'll still cover E). Pick an arbitrary closed interval I_1, and then repeat the following iterative process: if $I_1, I_2, \cdots, I_n$ have been chosen already, then let k_n be the supremum of the lengths of all intervals not intersecting $I_1 \cup \cdots \cup I_n$. Then $O \setminus (I_1 \cup \cdots \cup I_n)$ is open, so if $k_n = 0$ then we have covered everything (because if there were a point uncovered, there would be some finite interval around it contained in $O \setminus (I_1 \cup \cdots \cup I_n)$ and thus $k_n > 0$). Otherwise, choose I_n so that $|I_{n+1}| > \frac{k_n}{2}$ and I_{n+1} does not intersect $I_1 \cup \cdots \cup I_k$. If this process ever stops, we are done; otherwise we get some countable collection of intervals $\{I_n\}_{n \geq 1}$. Then $\sum_{k=1}^{\infty} |I_k| < m(O)$ because the intervals are pairwise disjoint, so we can choose N and ε so that $\sum_{k=N_{\varepsilon}+1}^{\infty} |I_k| < \frac{\varepsilon}{100}$.

Now if we set $A_\varepsilon = \bigcup_{k=1}^{N_\varepsilon} I_k$ and $B_\varepsilon = E \setminus A_\varepsilon$, we claim that $m(B_\varepsilon) < \varepsilon$. Indeed, for any $x \in B_\varepsilon$, we can find some interval I_x in the cover containing x that doesn't intersect any of $I_1, \cdots, I_{N_\varepsilon}$ (since the union of those intervals is closed). But it was not chosen as a candidate interval at any point, meaning that there is some smallest $m \geq N_\varepsilon + 1$ so that $I_x \cap I_m \neq \varnothing$. But then $|I_x| \leq 2|I_m|$ (because both I_x and I_m were candidates for being added to A_ε, but I_m was chosen), so I_x is contained in the interval $10I_m$ (where we scale each interval I_m from its center). Since this argument applies to all $x \in B_\varepsilon$, we find that B_ε is contained in the union $\bigcup_{k=M_\varepsilon+1}^{\infty} 10I_k$, meaning that

$$m(B_\varepsilon) \leq 10 \cdot \sum_{k=M_\varepsilon+1}^{\infty} |I_m| < 10 \cdot \frac{\varepsilon}{100} < \varepsilon.$$

Thus our finite subcollection A_ε of disjoint intervals covers everything in E except a set of measure less than ε. $\square$

The same result also holds for closed boxes in $\mathbb{R}^n$ by a very similar argument.

Corollary 59

Let $O \subseteq \mathbb{R}^n$ be an open set of finite measure, and let $\delta > 0$. Then there is a countable collection of disjoint closed balls B_n contained in O, such that the diameter of each ball is at most δ and $m(O \setminus \bigcup B_n) = 0$.

Proof. By Vitali's lemma, we can find balls $B_{1,1}, \cdots, B_{1,n_1}$ such that all balls have diameter at most δ and $m(O \setminus \bigcup_{k=1}^{n_1} B_{1,k}) < \frac{m(\theta)}{10}$. Next, we can find $B_{2,1}, \cdots, B_{2,N_2}$ such that $m(O \setminus \bigcup_{k=1}^{n_1} B_{1,k} \setminus \bigcup_{i=1}^{n_2} B_{2,k}) < \frac{m(\theta)}{100}$. Repeating this process gives us a countable collection with the "missing set" of measure 0. $\square$

The main idea here is that in the Lebesgue measure, blowing up a set by a factor of c multiplies the measure by a corresponding factor – this is a property of "doubling measures." So Vitali's lemma is really only relevant for a special class of measures, and we may see a stronger covering lemma later on in the course.

7 October 25, 2022

We're currently discussing differentiation and the Newton-Leibniz formula – last time, we defined the derivative functions D^+, D^-, D_+, D_- (which are upper and lower derivatives on the right and left), and we say that f is differentiable at x if these functions all coincide. In particular, then we have

$$Df(x) = \lim_{n \to \infty} n \left(f\left(x + \frac{1}{n}\right) - f(x) \right),$$

so if f is measurable and the derivative $Df(x)$ exists, then $Df(x)$ is measurable. Alternatively writing $Df(x) = f'(x)$, we may then ask whether the Newton-Leibniz formula ("fundamental theorem of calculus") $f(b) - f(a) = \int_a^b f'(x)dx$ holds as long as f' exists. What we started to show last time is that the derivative f' exists (Lebesgue-)almost everywhere for a monotonic function f. Specifically we showed Vitali's lemma last time, which shows that we can always take a finite subcollection of a fine cover of some set E of finite measure such that the uncovered part has arbitrarily small measure. We'll now show how to use this to prove our claim:

Proof sketch of Theorem 56. Consider the set of x where the limsup from the right is greater than the limsup from the left, so we define $E = \{x : D^+f(x) > D^-f(x)\}$. We wish to show that $m(E) = 0$. Notice that whenever this occurs, we can find two rationals between those values, so

$$E = \bigcup_{r,s\in\mathbb{Q}} D_{rs}, \quad D_{rs} = \{x : D^+f(x) > r > s > D^-f(x)\},$$

and by countable additivity it suffices to show that $m(D_{rs}) = 0$ for all $r, s \in \mathbb{Q}$. Let $\ell = m^*(D_{rs})$, which is finite because it's a subset of $[a, b]$. We may cover D_{rs} by an open set $\mathcal{O}$ such that $m^*(\mathcal{O}) < m^*(D_{rs})+\varepsilon$; because $D^-(f(x))$ is bounded by s from above, for every $x \in D_{rs}$, there must be arbitrarily small $h_n \downarrow 0$ such that $f(x)-f(x-h_n) < sh_n$, and the intervals $[x - h_n, x]$ cover D_{rs}. So by Vitali's lemma, we may pick a finite disjoint subcollection $\{I_1, \cdots, I_N\}$ so that $A = \bigcup_{k=1}^N I_k \cap D_{rs}$ has measure in $[\ell - \varepsilon, \ell + \varepsilon)$. But for any $y \in A$, we have $D^+f(y) > r$, so there are some $\delta \to 0$ with $f(y + \delta) - f(y) > r\delta$.

The point now is that the total increase across one endpoint of an interval I_i is at most $s\ell(I_i)$, but now we can choose a finite subcollection of intervals covering A, where on each of those intervals we go up by r times the length of any of those intervals. Since f is monotonic (so between the intervals we cannot go down), we can do some careful bounding to show that the total jump cannot be both this small and this large unless the total length ℓ was zero. Thus D^+ and D^- must agree almost everywhere. A similar argument works for showing that the other derivative functions also agree. □

Fact 60

As we saw on our homework, there is a continuous monotonic function $f(x)$ where $f(0) = 0$, $f(1) = 1$, and $f'(x) = 0$ almost everywhere (the Cantor function). So it is possible that $f(1) - f(0) = 1$ but $\int f'(x)dx = 0$.

Theorem 61

Let f be monotonically increasing on $[a, b]$. Then $\int_a^b f'(x)dx \le f(b) - f(a)$.

Proof. Extend the function f by defining $f(x) = f(b)$ for all $x > b$, and define

$$g_n(x) = n\left(f\left(x + \frac{1}{n}\right) - f(x)\right).$$

Then $f'(x) = \lim_{n\to\infty} g_n(x)$ almost everywhere, and $g_n(x) \ge 0$ because f is increasing. Thus by Fatou's lemma,

$$\int_a^b f'(x)dx \le \liminf_{n\to\infty} \int_a^b g_n(x)dx \liminf_{n\to\infty} n\left(\int_b^{b+1/n} f(x)dx - \int_a^{a+1/n} f(x)dx\right).$$

But by monotonicity we know that $f\left(a + \frac{1}{n}\right) \ge f(a)$, so we can bound this by (noting that the first term is just the integral of a constant)

$$\le \liminf_{n\to\infty}\left(n \cdot \frac{1}{n}f(b) - n \cdot \frac{1}{n}f(a)\right) = f(b) - f(a),$$

as desired. $\square$

So even though this result looks elementary, it takes quite a bit of work to arrive at in the Lebesgue formalism. And now we want to think about the case where equality holds:

Definition 62

Let $f(x)$ be a function on $[a, b]$, and consider a partition $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. Define the sum of the positive and negative jumps to be

$$p = \sum_{k=0}^{m-1} [f(x_{k+1} - f(x_k)]_+, \quad n = \sum_{k=0}^{m-1} [f(x_{k+1} - f(x_k)]_-,$$

where a_+ and a_- denote the "positive part" and "negative part" of a. We know that $p - n = f(b) - f(a)$; we let $t = p + n$ be the **total variation** $\sum_{k=0}^{m-1} |f(x_{k+1}) - f(x_k)|$ with respect to the partition. Then define $P_a^x[f], N_a^x[f], T_a^x[f]$ to be the supremum of p, n, t over all partitions on $[a, x]$, and say that f has **bounded variation (BV)** if $T_a^b[f] < \infty$. Let the set of functions on $[a, b]$ of bounded variation be denoted $\text{BV}([a, b])$.

Theorem 63

A function f has bounded variation on $[a, b]$ if and only if it is the difference of two monotonic functions.

Proof. First, write $f = g_1 - g_2$ for increasing functions g_1, g_2. Then $T_a^b[f] \le T_a^b[g_1] + T_a^b[g_2]$ (by the triangle inequality on the expressions of p and n), but the total variation of a monotonic function is just the overall difference and thus the right-hand side is $g_1(b) - g_1(a) + g_2(b) - g_2(a)$.

On the other hand, suppose f is of bounded variation on $[a, b]$. Then it is also of bounded variation on $[a, x]$ for all $a \le x \le b$ because using a smaller partition can only give us smaller positive and negative parts. So now $P_a^x[f]$ and $N_a^x[f]$ are increasing functions (this is where we use bounded variation), and we want to write f in terms of them. Now for any partition of $[a, x]$ we have

$$p - n = f(x) - f(a) \implies p = n + f(x) - f(a) \le N_a^x[f] + f(x) - f(a),$$

and the right-hand side does not depend on our partition so we can take the supremum over all partitions and get $P_a^x \le N_a^x + f(x) - f(a)$. Similarly, because $n = p + f(a) - f(x) \le P_a^x + f(a) - f(x)$, taking sup gives us $N_a^x \le P_a^x + f(a) - f(x)$. Combining these equations tells us that $P_a^x - N_a^x = f(x) - f(a)$, so $f(x) = f(a) + P_a^x - N_a^x$, and $f(a) + P_a^x$ and N_a^x are both monotonic functions. $\square$

Corollary 64

Every function of bounded variation is differentiable almost everywhere (because it is the difference of monotonic functions, which are differentiable almost everywhere).

This result can be applied to something like optimization in sharpening images (trying to have clear boundaries while not changing the image too much); what we want to improve in those cases turns out to be the total variation norm.

Theorem 65

Let $f \in L^1([a, b])$ (meaning that f is (absolutely) integrable) and set $F(x) = \int_a^x f(t)dt$. Then $F'(x) = f(x)$ almost everywhere.

Proof. We first show that F is actually differentiable everywhere. By absolute continuity of the integral, F is continuous (because for any $\varepsilon > 0$ we can find some δ so that $\int_{x_1}^{x_1+\delta} f(t)dt$ is small enough). Additionally, F has bounded total variation

$$t = \sum_{k=1}^{N-1} \left| \int_{x_k}^{x_{k+1}} f(t)dt \right| \le \sum_{k=1}^{N-1} \int_{x_k}^{x_{k+1}} |f(t)dt| = \int_a^b |f(t)|dt$$

by assumption on f, so the total variation is finite and our previous result shows that F is differentiable. We must now prove a lemma:

Lemma 66

Let $f(t)$ be integrable on $[a, b]$, and assume that $\int_a^x f(x)dx = 0$ for almost every x. Then $f(x) = 0$ almost everywhere.

Proof of lemma. Suppose that $f(x) > 0$ on a set E with positive measure $m(E) > 0$. Then there is some compact set $K \subset E$ such that $m(K) > \frac{m(E)}{2}$, and we can write

$$0 = \int_a^b f(t)dt = \int_K f(t)dt + \int_{K^c} f(t)dt.$$

The first term is always positive (because on E the sets where $f(t) > \frac{1}{n}$ approach the whole set, so there is some n where this overlaps K with positive measure). But then K^c is an at most countable union of open intervals, so there is an open interval I such that $\int_I f(t)dt \neq 0$, a contradiction. A similar argument shows that $f(x) < 0$ only on a set of measure zero. $\square$

Returning now to the proof, if we first assume that $|f(x)| \le K$ on $[a, b]$, then the function $f_n(x) = \frac{F\left(x+\frac{1}{n}\right)-F(x)}{1/n} = n\int_x^{x+1/n} f(t)dt$ is also bounded by that same K. But since $f_n(x) \to F'(x)$ almost everywhere by definition, the bounded convergence theorem tells us that

$$\int_a^x F'(t)dt = \lim_{n\to\infty} \int_a^x f_n(t)dt = \lim_{n\to\infty} \left(n\int_x^{x+1/n} F'(t)dt - n\int_a^{a+1/n} F'(t)dt \right),$$

which converges to $F(x) - F(a)$ because F is continuous. But because $F(a) = 0$, this means that

$$F(x) = \int_a^x F'(t)dt = \int_a^x f(t)dt \implies \int_a^x (F'(t) - f(t))dt = 0$$

and by our lemma this can only happen if $F'(t) = f(t)$ almost everywhere.

More generally, if f is not bounded, we can still assume without loss of generality that $f \ge 0$ (since we can do the problem separately for the positive and negative parts of f), and now we can define $g_n(x) = \min(f(x), n)$ and $\boxed{G_n(x) = \int_a^x (f(t) - g_n(t))dt}$. Each $G_n(x)$ is increasing (because the integrand is nonnegative), so it has a derivative $g_n(x)$ which is nonnegative almost everywhere; by the previous case (since g_n is bounded) we know that

$$\frac{d}{dx}\int_a^x g_n(t)dt = g_n(x).$$

But now we have almost everywhere that (restating the boxed equality)

$$G_n(x) = F(x) - \int_a^x g_n(x)dx \implies F(x) = G_n(x) + \int_a^x g_n(x)dx,$$

and $G_n(x)$ and $\int_a^x g_n(x)$ are both differentiable so we must have $F'(x) = G_n'(x) + g_n(x)$ almost everywhere, meaning $F'(x) \ge g_n(x)$ almost everywhere. Taking the limit as $n \to \infty$ (because $g_n \to f$ pointwise), we find that $F'(x) \ge f(x)$

almost everywhere. Thus

$$\int_a^x F'(t)dt \ge \int_a^x f(t)dt = F(x) - F(a) = F(x).$$

But $f \ge 0$ almost everywhere, so F is increasing. Thus by our previous result we also know that $\int_a^x F'(t)dt \le F(x) - F(a) = F(x)$, so we can put these together to find

$$F(x) \ge \int_a^x F'(t)dt \ge \int_a^x f(t)dt = F(x),$$

meaning that we must have equality. ☐

We'll study the class of functions for which this kind of equality holds in more detail next time!

8 October 27, 2022

Last time, we showed that derivatives of monotone functions (on $[a, b]$) exist almost everywhere, and in fact $\int_a^b f'(x)dx \le f(b) - f(a)$ by a Fatou argument. We also showed that if f is absolutely integrable, then $F(x) = \int_a^x f(t)dt$ is differentiable almost everywhere with $F'(x) = f(x)$. So now we want to go backwards, figuring out whether a function F gives rise to a function f. To do so, we need to make an additional definition:

Definition 67

A function f is **absolutely continuous** (sometimes denoted **AC**) if for any $\varepsilon > 0$, there is some δ such that for any finite collection of intervals $[x_i^(\ell), x_i^{(r)}]$ with $\sum_{i=1}^N |x_i^{(r)} - x_i^{(\ell)}| < \delta$, we have $\sum_{i=1}^N |f(x_i^{(r)}) - f(x_i^{(\ell)})| < \varepsilon$.

This is a stronger notion than uniform continuity, since that only allowed us to use a single interval instead of arbitrary collections.

Example 68

By absolute continuity of the integral, if $f \in L^1([a, b])$ (is absolutely integrable) and $F(x) = \int_a^x f(t)dt$, then $F(x)$ is absolutely continuous on $[a, b]$.

Theorem 69

Let F be absolutely continuous on $[a, b]$. Then F is differentiable almost everywhere on $[a, b]$ with $F(x) - F(a) = \int_a^x F'(t)dt$.

Proof. If F is absolutely continuous, then f has bounded total variation (useful exercise to check from the definition), meaning that we can write $F(x) = F_1(x) - F_2(x)$ with F_1, F_2 increasing. But then $F'(x) = F_1'(x) - F_2'(x)$, and F_1', F_2' are integrable (bounded from above by $F_1(b) - F_1(a)$ and $F_2(b) - F_2(a)$, respectively), so $G(x) = \int_a^x F'(x)dx$ is a well-defined function. Define

$$R(x) = F(x) - G(x) = F(x) - \int_a^x F'(x)dx.$$

Since F and G are differentiable almost everywhere (because F is absolutely continuous and G is the integral of an integrable function), $R'(x) = F'(x) - G'(x) = F'(x) - F'(x) = 0$ almost everywhere. This alone does not prove that R is a constant, but now $R(x)$ is absolutely continuous because F and G are, and we claim this implies $R(x) = R(a)$ for all $x \in [a, b]$.

Indeed, fix some $\varepsilon > 0$. For almost every $x \in [a, b]$, we can find $h_n \to 0$ such that $|R(x + h_n) - R(x)| < \varepsilon h_n$. Then the set $E \subseteq [a, b]$ on which $R'(x) = 0$ is covered by these intervals $[x, x + h_n]$, so by Vitali's covering lemma (we need $h_n \to 0$ to have a fine cover) we can take a finite subcollection of these intervals so that $m^*(E \setminus \cup_{k=1}^N I_k) < \frac{\delta}{10}$. We know that E has full measure (equal to the measure of $[a, b]$), so the sum of the complement (which is some collection of intervals) is of length less than $\frac{\delta}{5}$. But if we choose δ to be small enough to satisfy absolute continuity for R', then notice that the variation over the good intervals I_k is at most $\varepsilon(b - a)$, and on the complement we can bound the increase by at most ε. Thus $|R(b) - R(a)| \le \varepsilon + \varepsilon(b - a)$ for all ε, so we must have $R(b) = R(a)$. This argument works for any point $x \in [a, b]$ instead of just b, so $F(x) - G(x)$ must indeed be constant and that constant must be $F(a)$. $\square$

(The main idea of this proof is to split up the contributions to R on a "good set," which we obtain here using Vitali's lemma, and a "bad set," which we control with absolute continuity.)

We'll now turn to product measures and Fubini's theorem – the main idea is motivated by calculus, in which we can find a function $f(x, y)$ such that $\int_0^1 \int_0^1 f(x, y)dydx$ and $\int_0^1 \int_0^1 f(x, y)dxdy$ both exist but are not equal. (We just pick a function f which blows up in some way.) To make sense of this with our integration theory, we must first define a product measure:

Definition 70

Let μ be a measure on X and ν a measure on Y. We define the product measure $\mu \times \nu$ by setting, for any $S \subseteq X \times Y$,

$$(\mu \times \nu)(S)^* = \inf \left(\sum_{j=1}^{\infty} \mu(A_j)\nu(B_j) \right),$$

where the infimum is taken over all measurable sets $A_j \subset X, B_j \subset Y$ with $S \subseteq \bigcup_{j=1}^{\infty} A_j \times B_j$.

Our goal is to show that we can actually write this measure as an integral

$$(\mu \times \nu)(S) = \int_Y \left(\int_X 1_S(x, y)d\mu(x) \right) d\nu(y) = \int_X \left(\int_Y 1_S(x, y)d\nu(y) \right) d\mu(x).$$

(We'll write $d\mu(x)$ or $d\mu_x$ interchangeably.) To do this, we'll let $\mathcal{F}$ be the collection of sets $S \subseteq X \times Y$ with $1_S(x, y)$ μ-measurable for ν-almost-every-y such that $s(y) = \int_X 1_S(x, y)d\mu_x$ is ν-integrable. (So we're looking at which sets we can actually calculate in this way.) For any such set S, we then define $\rho(S) = \int_Y s(y)d\nu(y) = \int_Y \left(\int_X 1_S(x, y)d\mu_x \right) d\nu_y$. (The principle here is that if we have a two-dimensional body, we calculate the length along a cross-section and then integrate that along the other axis – this is Cavalieri's principle. And we're trying to show that computing cross-sections in either direction gives us the same answer as covering by rectangles, but we need to set up this notation to figure out when such a process is actually possible.)

We'll prove that $\mathcal{F}$ contains all sets which are measurable with respect to $\mu \times \nu$. First, notice that if $U, V \in \mathcal{F}$ and $U \subseteq V$, then $\rho(U) \le \rho(V)$ (because $1_U(x, y) \le 1_V(x, y)$) – this will be useful later. Additionally, if A is μ-measurable and B is ν-measurable, then $S = A \times B \in \mathcal{F}$ (because $\int 1_{A \times B} d\mu_X = 1_B(y)\mu(A)$, which is measurable because B is measurable, and then integrating it over Y gives us $\mu(A)\nu(B)$). So now let $\mathcal{P}_1$ be the set of countable unions of such sets:

$$\mathcal{P}_1 = \left\{ \bigcup_{j=1}^{\infty} (A_j \times B_j) : A_j \ \mu\text{-measurable and } B_j \ \nu\text{-measurable} \right\}.$$

For any set $S \in \mathcal{P}_1$, we can do set theory manipulation and subdivide our rectangles to write $S = \bigcup_{i=1}^{\infty} A_i' \times B_i'$, so that the intersection of $A'_i \times B'_i$ and $A'_j \times B'_j$ is empty if $i \neq j$. Then

$$\int_X 1_S(x, y) d\mu = \int \sum_{j=1}^{\infty} 1_{A_j \times B_j}(x, y) d\mu_X,$$

and now we have a nonnegative sum in the integral, so we can evaluate it termwise to get $\sum_{j=1}^{\infty} \mu(A_j) 1_{B_i}(y)$. In particular, all B_is are measurable, so this means anything in $\mathcal{P}_1$ is indeed in $\mathcal{F}$. Integrating this over Y then gives us $\rho(S) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j)$ for any such set S written as a disjoint union of rectangles.

Now for any set U, we can write its outer measure by approximating from above:

$$(\mu \times \nu)^*(U) = \inf \{p(s) : U \subseteq S, S \in \mathcal{P}_1\}.$$

(Indeed, the measure was originally defined by taking a not-necessarily disjoint union, but we can always write any covering of U using a disjoint union.) We can now relate the notion of "being in $\mathcal{F}$" to "being measurable:"

Proposition 71

If A is μ-measurable, and B is ν-measurable, then $A \times B$ is $\mu \times \nu$-measurable.

Proof. We know that $(\mu \times \nu)(A \times B) = \rho(A \times B) = \mu(A)\nu(B)$ (because $A \times B$ covers itself, and for any other subset covering it we use the fact that $U \subseteq V$ implies $\rho(U) \leq \rho(V)$). To show measurability, take any arbitrary subset $T \subseteq X \times Y$, and let R be any set in $\mathcal{P}_1$ with $R \supseteq T$. Then $R \cap (A \times B)^c$ and $R \cap (A \times B)$ are in P_1 and are disjoint, so

$$(\mu \times \nu)^*(T \cap (A \times B)^c) + (\mu \times \nu)^*(T \cap A \times B) \leq \rho(R \cap (A \times B)^c) + \rho(R \cap (A \times B)) = \rho(R)$$

because ρ is additive. But if we take the infimum over all R, we find that

$$(\mu * \nu)^*(T \cap (A \times B)^c) + (\mu * \nu)^*(T \cap (A \times B)) \leq (\mu \times \nu)^*(T),$$

so $A \times B$ is measurable. $\square$

We're now ready to calculate areas of more than just rectangles by defining

$$\mathcal{P}_2 = \left\{ \bigcap_{i=1}^{\infty} S_j : S_j \in \mathcal{P}_1 \right\}.$$

Proposition 72

For each $S \subseteq X \times Y$, there is some set $R \in \mathcal{P}_2 \cap \mathcal{F}$ with $S \subseteq R$ and $\rho(R) = (\mu \times \nu)^*(S)$.

Proof. We proved in the last argument that the outer measure of S is the infimum of all $\rho(R)$ where R covers S and is in $\mathcal{P}_1$, so we may pick a sequence of such sets that approximates the measure from above. Choose $R_j \in \mathcal{P}_1$ so that $\rho(R_j) < (\mu \times \nu)^*(S) + \frac{1}{j}$ for all j, and define the intersection and partial intersections

$$R = \bigcap_{j=1}^{\infty} R_j, \quad Q_k = \bigcap_{i=1}^{k} R_k.$$

We know that Q_k decreases to R and $\rho(Q_1) = \rho(R_1) < \infty$, so we must have $1_R(x, y) = \lim_{k \to \infty} 1_{Q_k}(x, y)$. In particular, since each D_{R_i} is measurable, so is D_{Q_i}. That means there is a set of full measure for each k such that R_k

is measurable, so there is some $S_0 \subseteq Y$ of full ν-measure so that for all $y \in S_0$, $1_R(x,y)$ is μ-measurable. Thus by the bounded convergence theorem, for all $y \in S_0$ we have

$$\rho_R(y) = \int_X 1_R(x,y) d\mu = \lim_{k\to\infty} \rho_k(y),$$

where $\rho_k(y) = \int_X 1_{Q_k}(x,y) d\mu_x$, so $\rho_R(y)$ is ν-measurable almost everywhere (being a limit of measurable functions almost everywhere). Thus

$$\rho(R) = \int_Y \rho_R(y) = \int_Y \lim_{k\to\infty} \rho_k(y) dy = \lim_{k\to\infty} \rho_k(y) dy = \lim_{k\to\infty} \rho(Q_k) \le (\mu \times \nu)^*(S),$$

but each Q_k covers S and is in $\mathcal{P}_1$, so $(\mu \times \nu)^*(S) \le \rho(Q_k)$ for all k, meaning $(\mu \times \nu)^*(S) \le \rho(R)$. Thus combining the inequalities we see that $\rho(R) = (\mu \times \nu)^*(S)$. □

Theorem 73

Let $S \subseteq X \times Y$ be ***σ*-finite** with respect to $\mu \times \nu$ (meaning that $S = \bigcup_{j=1}^\infty S_j$ for measurable sets S_j with $\mu \times \nu(S_j)$ finite for all j). Then the cross-section $S_y = \{x : (x,y) \in S\}$ is μ-measurable for ν-almost-everywhere y and vice versa, with

$$(\mu \times \nu)(S) = \int_Y \mu(S_y) d\nu_y = \int_X \nu(S_x) d\mu_x.$$

Proof. First of all, if $(\mu \times \nu)(S) = 0$, then there is some $R \in \mathcal{P}_2$ with $\rho(R) = 0$ and $S \subseteq R$. But we have $S_y \subseteq R_y$ for ν-almost-every y, so $\mu(S_y) \le \mu(R_y) \le 0$ for ν-almost-every y because $\rho(R) = 0$.

On the other hand, if $(\mu \times \nu)(S)$ is nonzero and finite, then there is some $R \in P_2$ such that $S \subseteq R$ and $\rho(R) = (\mu \times \nu)(S)$. Thus $(\mu \times \nu)(R \setminus S) = 0$, so $\mu(S_y) = \mu(R_y)$ for ν-almost-every y. Thus

$$(\mu \times \nu)(S) = \rho(R) = \int_Y \left(\int_R (x,y) d\mu(x) \right) d\nu_y = \int_Y \left(\int_X 1_S(x,y) d\mu_x \right) d\nu_y,$$

as desired. Finally, for σ-finite measures as opposed to finite ones, we just make approximations of this result using the S_js. □

We thus get an analogous result for functions:

Theorem 74 (Fubini)

Suppose f is $(\mu \times \nu)$-measurable and integrable and $X \times Y$ is σ-finite. Then

$$\int f d(\mu \times \nu) = \int_X \left(\int_Y f(x,y) d\nu_y \right) d\nu_x = \int_Y \left(\int_X f(x,y) d\mu_x \right) d\nu_y.$$

And similarly, we also have **Tonelli's theorem**, which tells us the same result if f is nonnegative instead of integrable.

9 November 1, 2022

We'll discuss the **Besicovitch covering theorem** today, which doesn't need the scaling argument that we did in the Vitali covering lemma (where we said that blowing up the balls by a factor of 5 only multiplies the measure by 5^d).

Definition 75

A measure is **doubling** if there is some constant $C > 0$ such that for any $x \in \mathbb{R}^n$ and $R > 0$, we have

$$\frac{1}{C}\mu(B(x, 2R)) \leq \mu(x, R) \leq \mu(x, 2R).$$

Vitali's lemma then holds for any doubling measure, but for non-doubling measures we have to compensate in another way.

Theorem 76 (Besicovitch)

There is a number $N(n)$ (depending only on the dimension n) such that the following property holds: let $\mathcal{F}$ be a nondegenerate collection of closed balls in $\mathbb{R}^n$, such that $D = \sup_{\overline{B} \in \mathcal{F}}[\text{diam}(\overline{B})]$ is finite. Let A be the set of centers of balls in $\mathcal{F}$. Then there are subcollections $\mathcal{J}_1, \cdots, \mathcal{J}_{N(n)}$ of these balls such that

$$A \subseteq \bigcup_{k=1}^{N(n)} \bigcup_{B \in \mathcal{J}_k} B,$$

such that each $\mathcal{J}_k$ contains only pairwise disjoint balls. (In particular, this implies that $\mathcal{J}_k$ must be countable because each ball contains some point in $\mathbb{Q}^n$.)

Remember that in Vitali's lemma, we only chose one subcollection and covered "most" of the set, and in this case we have finitely many collections of these balls but they completely cover our set A. And the number of balls we need depends only on the dimension n, not on the particular set A.

Corollary 77

Let μ be a Borel regular measure on $\mathbb{R}^n$, let $\mathcal{F}$ be any collection of nondegenerate closed balls, and let A be the set of centers of balls in $\mathcal{F}$. Assume that $\mu^*(A)$ is finite, and for each $a \in A$ there are arbitrarily small balls centered at a (that is, $\inf\{r : \overline{B}(a, r) \in \mathcal{F}\} = 0$). Then for each open set $U \subseteq \mathbb{R}^n$, there is an at most countable collection $\mathcal{J}$ of pairwise disjoint balls in $\mathcal{F}$ (just one collection) such that

$$\mu^*\left((A \cap U) \setminus \bigcup_{B \in \mathcal{J}} \overline{B}\right) = 0.$$

There is no requirement written here that the balls are contained in U, but we can just choose our collection initially so that it only contains balls in U. And in particular we can take $U = \mathbb{R}^n$ as a typical application, and we see that we in fact **do not lose anything when not requiring a doubling measure**.

Proof of corollary. First, only keep the balls $\overline{B}(a, r)$ of radius $r \leq 1$ in our collection. By the Besicovitch theorem, $A \cap U$ can be covered by $N(n)$ subcollections, so by the pigeonhole principle we can choose one of the subcollections $\mathcal{J}_k$ so that we cover more than $\frac{1}{N(n)}$ of the total:

$$\mu^*\left((A \cap U) \cap \bigcup_{B \in \mathcal{J}_k} \overline{B}\right) \geq \frac{1}{N(n)}\mu^*(A \cap U).$$

But now we can take some finite subcollection of the balls $\overline{B}_{1,k}, \cdots, \overline{B}_{M_1,k}$ in $\mathcal{J}_k$ so that

$$\mu^*\left((A \cap U) \cap \bigcup_{p=1}^{M_1} \overline{B}_{p,k}\right) \geq \frac{1}{2N(n)}\mu^*(A \cap U),$$

because $\mathcal{J}_k$ is countable and looking at the measure of the partial unions must approve But now we have covered a positive fraction of $A \cap U$ with finitely many balls in $\mathcal{F}$, and now we can repeat the same argument to $(A \cap U) \setminus \bigcup_{p=1}^{M_1} B_{p,k}$ since any point in that set is a positive distance away from the balls we remove (and thus the property still holds). Notice that

$$m^* \left((A \cap U) \setminus \bigcup_{p=1}^{M_1} B_{p,k} \right) \leq \left(1 - \frac{1}{2N(n)} \right) \mu^*(A \cap U),$$

so repeatedly again getting another finite collection of balls to gain another factor of $\left(1 - \frac{1}{2N(n)}\right)$ shows that we can get

$$m^* \left((A \cap U) \setminus \bigcup_{p=1}^{M_a} B_{p,k} \right) \leq \left(1 - \frac{1}{2N(n)} \right)^a \mu^*(A \cap U)$$

for any positive integer a; repeating this gives us the countable collection that we desire since the right-hand side goes to 0 as $n \to \infty$. $\square$

We'll now turn to the proof of the Besicovitch theorem, which will not require any particularly advanced techniques but will be a long proof:

Proof of Theorem 76. Here's the strategy: we'll choose a sequence of balls $B_1, B_2, \cdots$ that are "large enough" in some sense, covering A but not caring about intersections between them, and we'll prove that we can do so in a way so that each B_k intersects at most $N(n) - 1$ balls in $B_1, \cdots, B_{k-1}$. But then we can distribute the B_ks into $N(n)$ buckets inductively, always putting B_k into a bucket where it does not intersect with any of the previous balls.

We can assume without loss of generality that A is bounded. Indeed, consider concentric annuli of radius 10Δ, and alternate coloring them red and blue. Then the balls within different red annuli cannot intersect each other, and similarly the balls within different blue annuli cannot intersect each other. So if we can do the job separately with $N(n)$ buckets for each annulus (which is bounded), we can do the job with $2N(n)$ buckets ($N(n)$ of them for red and $N(n0$ of them for blue).

To actually choose the balls B_i, first choose any $\overline{B}_1(a, R_1)$. Now inductively, after choosing $\overline{B}_1, \cdots, \overline{B}_k$, choose $\overline{B}_{k+1}(a_{k+1}, R_{k+1})$ so that $a_{k+1} \notin \bigcup_{j=1}^{k} \overline{B_j}$ and with radius "near-maximal:"

$$R_{k+1} \geq \frac{3}{4} \sup \left[R : B(a, r), a \notin \bigcup_{j=1}^{k} \overline{B}_j \right].$$

In particular, we don't need the whole ball $B(a, R)$ to be disjoint from the previous balls, but we do need a to be, and here is where we use that D is finite. We'll now study this process in more detail:

- The balls $\overline{B}\left(a_i, \frac{R_i}{3}\right)$ are pairwise disjoint. Indeed, for any $j > i$, we have $a_j \notin B(a_j, R_j)$, so $|a_j - a_i| > r_i$. But $r_j \leq \frac{4}{3} r_i$ (because B_j was chosen later than B_i, so B_j was a valid candidate in place of B_i so couldn't have been much bigger). This means that

$$|a_j - a_i| > R_i = \frac{R_i}{3} + \frac{2R_i}{3} \geq \frac{R_i}{3} + \frac{R_j}{2} \geq \frac{R_i + R_j}{3},$$

  so indeed the shrunk balls do not intersect.

- The process can terminate (if we've covered all of the centers), but if it does not terminate, then we must have $\lim_{j \to \infty} R_j = 0$. Indeed, we know that $\bigcup_{j=1}^{\infty} B\left(a_j, \frac{R_j}{3}\right)$ must be a bounded set because A is assumed to be bounded (without loss of generality) and the diameters are also bounded, and the balls are all disjoint so we must have $\sum R_j^n < \infty$.

- Thus, this process does indeed cover A (we have $A \subseteq \bigcup_{j=1}^{\infty} \overline{B}(a_j, R_j)$). This is because any point $a \in A$ contains some ball $B(a, R)$ centered around it, and we must have either chosen that ball or covered a with another ball. Indeed, if we have only finitely many balls we cover everything, and otherwise $\lim_{j \to \infty} R_j \to 0$ so eventually $R_j < \frac{R}{100}$ meaning that $B(a, R)$ was a candidate and stopped being one.

So our balls cover A, and we now claim that $\overline{B}(a_j, R_j)$ intersects at most $N(n)$ previous balls $\overline{B}(a_i, R_i)$ for $i < j$ (this is the crux of the argument). Consider the "bad indices"

$$I_m = \{j : 1 \le j \le m, \overline{B}_j \cap \overline{B}_m \ne \varnothing\},$$

the "small bad balls"

$$K_m = I_m \cap \{j : R_j \le 3R_m\},$$

and the "large bad balls"

$$P_m = I_m \cap \{j : R_j > 3R_m\}.$$

- We claim that $|K_m| \le 20^n$ – for any $j \in K_m$, we can show that the shrunk ball $\overline{B}(a_j, \frac{R_j}{3})$ is contained in the stretched ball $\overline{B}(a_m, 5R_m)$. Indeed, for any $x \in B(a_j, \frac{R_j}{3})$, we know that

  $$|x - a_m| \le |x - a_j| + |a_j - a_m| \le \frac{R_j}{3} + (R_j + R_m)$$

  (because the original balls around a_j and a_m intersected by definition of K_m), and we can bound this as

  $$|x - a_m| \le R_m + (3R_m + R_m) = 5R_m \implies B\left(a_j, \frac{R_j}{3}\right) \subset \overline{B}(a_m, 5R_m).$$

  This means that

  $$5^n |R_m|^n \ge \sum_{j \in K_m} \frac{1}{3^n} R_j^n$$

  because all of the shrunk balls in K_m are disjoint and their union is contained within the stretched ball $\overline{B}(a_m, 5R_m)$. But furthermore $R_j \le \frac{4}{3} R_i$ whenever $j > i$, so we find that

  $$5^n |R_m|^n \ge |K_m| \frac{1}{3^n} \left(\frac{3}{4}\right)^n R_m^n \implies 20^n \ge |K_m|.$$

  So there can be at most 20^n small balls intersecting the mth ball.

- For the large balls P_m, here is the key claim: if $i, j \in P_m$ with $|a_i| < |a_j|$, and θ is the angle between the lines connecting a_i and a_j to a_m, then $\theta \ge \cos^{-1} \frac{61}{64} > 0$. This proves the theorem because the number of points on the unit sphere where any two are at an angle at least $\cos^{-1} \frac{61}{64}$ apart is finite and depends only on the dimension n. To prove this claim, we prove that **(1)** if $\cos(\theta) > \frac{5}{6}$ (meaning the balls are somewhat collinear), then we have $a_i \in \overline{B}(a_j, R_j)$ (so in particular that implies $j > i$ or else we couldn't pick the balls in this order), and **(2)** if $a_i \in \overline{B}_j$, then $\cos\theta \le \frac{61}{64}$ (they cannot be too collinear).

  For sake of notation, pick $a_m = 0$ to be the origin in $\mathbb{R}^n$. For **(1)**, a_m was chosen after a_i and a_j, so $a_m \notin \overline{B}_i \cup \overline{B}_j$, meaning that $R_i < |a_i|$ and $R_j < |a_j|$. Additionally, $|a_i - a_m| = |a_i| < R_m + R_i$ (because B_m and B_i intersect) and similarly $|a_j| < R_m + R_j$; also by definition of large balls we have $R_i > 3R_m, R_j > 3R_m$. So chaining these facts together we see that

  $$3R_m < R_i < |a_i| < R_i + R_m, \quad 3R_m < R_j < |a_j| < R_j + R_m.$$

We claim that $|a_i - a_j| \le |a_j|$ if $\cos\theta > \frac{5}{6}$. Indeed, if $|a_i - a_j| > a_j$, we have

$$\cos\theta = \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} \le \frac{a_i|^2}{2|a_i||a_j|} \le \frac{1}{2} < \frac{5}{6},$$

a contradiction. Thus $|a_i - a_j| \le |a_j|$, meaning that $a_i \in \overline{B}(a_j, a_j)$. But now if $a_i \notin \overline{B}(a_j, R_j)$ (so $|a_i - a_j| > R_j$), we must have

$$\begin{aligned}\cos\theta = \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} &= \frac{|a_i|}{|a_j|} + \frac{(|a_j| - |a_i - a_j|)(|a_j| + |a_i - a_j|)}{2|a_i||a_j|} \\ &\le \frac{1}{2} + \frac{|a_j| - |a_i - a_j|}{|a_i|} \le \frac{1}{2} + \frac{|a_j| - R_j}{|a_i|} \\ &\le \frac{1}{2} + \frac{R_j + R_m - R_j}{R_i} \\ &\le \frac{1}{2} + \frac{1}{3} = \frac{5}{6},\end{aligned}$$

a contradiction and thus we do have a_i in the ball $\overline{B}(a_j, R_j)$.

Finally, to prove **(2)**, similar arguments show that $|a_i - a_j| + |a_i| - |a_j| \ge \frac{|a_j|}{8}$, but also $|a_i - a_j| + |a_i| - |a_j| \le \frac{8}{3}|a_j|(1 - \cos\theta)$; putting these equalities shows the result.

Thus we've successfully obtained our constant $N(n)$ and proven the theorem with our "buckets" argument. $\square$

In particular, we claim that $N(1) = 2$ and $N(2) = 19$ (though this may be somewhat painful to check).

10 November 3, 2022

We'll see an application of (the corollary of) Besicovitch's theorem today to the Radon–Nikodym theorem, which asks us to recover one measure from another given a relation between them:

Definition 78

Let μ, ν be two Borel regular measures on $\mathbb{R}^n$. The upper and lower Radon–Nikodym derivative are defined as

$$\overline{D}_\mu \nu(x) = \limsup_{r \to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))}, \quad \underline{D}_\mu \nu(x) = \liminf_{r \to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))}.$$

If $\overline{D}_\mu \nu = \underline{D}_\mu \nu$, then we call that function the **Radon-Nikodym derivative** $D_\mu \nu(x)$ of ν with respect to μ.

(All balls B here will be **closed** and **nontrivial**.) We wish to reconstruct the Riemann integral now in the following way: for any set A, take small balls in μ-measure around points in A. Then it is reasonable to assume that the ν-measure is approximately the μ-measure multiplied by the Radon-Nikodym derivative at each point, so

$$\nu(A) = \int_A D_\mu \nu(x) d\mu.$$

For example, suppose $\nu(A) = \int_A f(x)dx$ for some continuous function $f(x) \in L^1(\mathbb{R}^n)$. Then the Radon-Nikodym derivative of ν with respect to the Lebesgue measure is just $D_\mu \nu(x) = f(x)$ by continuity at x. But if we just have $f \in L^1(\mathbb{R}^n)$, we now want to ask whether we still have $D_\mu \nu(x) = f(x)$ (now we must say μ-almost-everywhere, since f is only defined up to a set of measure zero). Then the values of f in a small ball look like they may not have to do with f in the center.

Theorem 79

Let μ and ν be Radon measures on $\mathbb{R}^n$. Then $D_\mu \nu(x)$ exists μ-almost-everywhere and is μ-measurable.

Proof. The definition of $\overline{D}_\mu \nu(x)$ and $\underline{D}_\mu \nu(x)$ are local at x, so we can assume without loss of generality that $\mu(\mathbb{R}^n)$ and $\nu(\mathbb{R}^n)$ are finite.

Lemma 80

Let μ, ν be two finite Radon measures on $\mathbb{R}^n$, and fix $0 < s < \infty$. Then we have the following for any (not necessarily measurable) set A:

- If $A \subseteq \{x \in \mathbb{R}^n : \underline{D}_\mu \nu \le s\}$, then $\nu^*(A) \le s\mu^*(A)$.
- If $A \subseteq \{x \in \mathbb{R}^n : \overline{D}_\mu \nu \ge s\}$, then $\nu^*(A) \ge s\mu^*(A)$.

Proof of lemma. We'll just do the first argument (the second one is basically identical). For any $x \in A$, we can find a sequence $r_n \to 0$ so that $\nu(B(x, r_n)) \le (s + \varepsilon)\mu(B(x, r_n))$. By Besicovitch, there is a countable collection of disjoint balls B such that $\mu^*\left(A \setminus \bigcup_j B_j\right) = 0$. Thus by countable subadditivity and because the balls are measurable, $\mu^*(A) \le \mu\left(\bigcup_j B_j\right) = \sum_j \nu(B_j)$. But we chose our balls B_j so that the ν-measure is no more than $(s + \varepsilon)$ times the μ-measure, so

$$\nu^*(A) \le (s + \varepsilon) \sum_{j=1}^{\infty} \mu(B_j).$$

So now we see that we can take any open set $U \supseteq A$, and we can choose our r_ns such that $B(x, r_n) \subseteq U$ (since we can just start with arbitrarily small radii). But all of these balls are contained in U and are pairwise disjoint, so we actually have

$$\nu^*(A) \le (s + \varepsilon) \sum_{j=1}^{\infty} \mu(B_j) \le (s + \varepsilon)\mu(U);$$

taking the infimum over all $U \supseteq A$ yields $\nu^*(A) \le (s + \varepsilon)\mu^*(A)$. Taking $\varepsilon \to 0$ yields the desired result. $\square$

Now returning to our theorem, we can fix any rational $a < b$ and consider

$$R_{a,b} = \left\{x : \underline{D}_\mu \nu(x) \le a, \overline{D}_\mu \nu(x) \ge b\right\}.$$

By our lemma, we know that

$$\nu^*(R_{a,b}) \le a\mu^*(R_{a,b}), \quad \nu^*(R_{a,b}) \ge b\mu^*(R_{a,b}) \implies b\mu^*(R_{a,b}) \le a\mu^*(R_{a,b})$$

which can only occur if $\mu^*(R_{a,b}) = 0$. Since this argument works for all $a, b \ge 0$, we see that

$$\mu^*\left(x : \underline{D}_\mu \nu(x) < \overline{D}_\mu \nu(x)\right\} = 0$$

because this set is in the countable union of the $R_{a,b}$s, which each have measure zero. Since $\mu^*\left(x : \overline{D}_\mu \nu(x) < \underline{D}_\mu \nu(x)\right\} = 0$ by definition, we see that the lower and upper Radon–Nikodym derivatives agree μ-almost-everywhere, as desired.

It remains to show that $D_\mu \nu(x)$ is actually μ-measurable, and we'll do so in the following way:

Lemma 81

For all $x \in \mathbb{R}^n$ and $r > 0$, we have (here remember B is a closed ball)

$$\limsup_{y \to x} \mu(B(y,r)) \le \mu(B(x,r)), \quad \limsup_{y \to x} \nu(B(y,r)) \le \nu(B(x,r)).$$

In other words, if we look at balls of the same radius near x, we can gain measure when we shift the balls, but we cannot lose measure.

Proof. Let $y_k \to x$ be a sequence of points and define $f_k(z) = 1\{z \in B(y_k, r)\}$. We claim that $\limsup_{k \to \infty} f_k(z) \le 1\{z \in B(x,r)\}$. This equality is clear if $z \in B(x,r)$ (because the right-hand side is 1 and the left-hand side is always at most 1). And if $z \notin B(x,r)$, then $|z - y_k| > r$, so $f_k(z) = 0$ for large enough k. But now by Fatou's lemma,

$$\liminf_{k \to \infty}(1 - f_k(z)) \ge 1 - 1\{z \in B(x,r))\} \implies \int_{B(x,2r)} 1 - 1\{z \in B(x,z)\} d\mu \le \liminf_{k \to \infty} \int_{B(x,2r)} (1 - f_k(z)) d\mu,$$

so evaluating the integral we find that

$$\mu(B(x,2r)) - \mu(B(x,r) \le \liminf_{k \to \infty} \left[\mu(B(x,2r)) - \mu(B(y_k, r))\right];$$

because our measures are finite (because μ is Radon) we can cancel the $\mu(B(x,2r))$s and get the desired result. ☐

The point is that this shows that $f_r(x) = \mu(B(x,r))$ is an upper semicontinuous function (which are measurable with respect to any Radon measure – this was on our homework), and so is $g_r(x) = \nu(B(x,r))$, so their ratio is measurable. But then this implies $D_\mu \nu(x)$ is a limit of μ-measurable functions and is thus measurable. (It's true that $D_\nu \mu(x)$ is measurable as well, but that Radon–Nikodym derivative can be infinite unlike $D_\mu \nu(x)$.) ☐

So we now want to go back to our original question, asking if we can obtain ν by integrating $D_\mu \nu(x)$ with respect to μ. This cannot always be true – for example, if $\mu(A) = 0$ that would mean $\nu(A) = 0$ (because integral over a set of measure zero is always zero). So we'll take that as an assumption instead:

Definition 82

A measure ν is **absolutely continuous** with respect to μ if whenever $\mu(A) = 0$, we have $\nu(A) = 0$. (We will write this as $\nu \ll \mu$.)

Theorem 83 (Radon–Nikodym)

Let μ and ν be two Radon measures on $\mathbb{R}^n$, and suppose $\nu \ll \mu$. Then for any μ-measurable set A, we have

$$\nu(A) = \int_A D_\mu \nu(x) d\mu.$$

(This result also holds over more general measure spaces, but we won't talk about that here.)

Proof. Notice that if a set A is μ-measurable and $\nu \ll \mu$, then A will be ν-measurable as well. (This is not true for measures in general – for example, a set in $\text{supp}(\mu)$ which is not μ-measurable but is outside $\text{supp}(\nu)$ is ν-measurable of measure zero.) To prove this, take a Borel set $B \supseteq A$ with $\mu(B \setminus A) = 0$. Then $\nu(B \setminus A) = 0$ as well by absolute continuity, and $B \setminus A$ is ν-measurable (because it is of measure zero) so $A = B \setminus (B \setminus A)$ is also ν-measurable (since B is Borel and thus measurable). So we indeed don't need to use outer measure in the statement of this theorem.

Now consider the sets

$$Z = \{x : D_\mu \nu(x) = 0\}, \quad I = \{x : D_\mu \nu(x) = \infty\}.$$

Then we know that $\mu(I) = 0$ (because we've just previously shown that $D_\mu \nu(x)$ exists – that is, is finite – μ-almost-everywhere), so $\nu(I) = 0$ by absolute continuity. On the other hand, $\nu(Z) \le \varepsilon \mu(Z)$ for any $\varepsilon > 0$, so $\nu(Z \cap B(0,R)) \le \varepsilon \mu(Z \cap B(0,R))$ for any ball $B(0,R)$. But because μ is Radon, $\mu(Z \cap B(0,R))$ is finite for any R, so $\nu(Z \cap B(0,R)) = 0$. That implies that $\nu(Z) = 0$ by continuity from below. This means we don't have to worry about the sets where the Radon–Nikodym derivative is zero or infinite.

We can now fix some $t > 1$ and write our set A as a **disjoint** union

$$A = \bigcup_{m=-\infty}^{\infty} A_m \cup (A \cap Z) \cup (A \cap I) \cup \{D_\mu \nu \text{ does not exist}\}, \quad A_m = \{x : t^m \le D_\mu \nu(x) < t^{m+1}\}.$$

But the sets in $(A \cap Z) \cup (A \cap I) \cup \{D_\mu \nu \text{ does not exist}\}$ all have ν-measure-zero, so

$$\nu(A) = \sum_{m=-\infty}^{\infty} \nu(A_m).$$

But now for each m, we know that $\nu(A_m) \le t^{m+1}\mu(A_m)$ and $\nu(A_m) \ge t^m \mu(A_m)$, so

$$\nu(A) \le \sum_m t^{m+1}\mu(A_m) \le t \sum_m t^m \mu(A^m).$$

Now on A^m we know the Radon-Nikodym derivative is at least t^m, so

$$\boxed{\nu(A)} \le t \sum_m \int_{A_m} (D_\mu \nu(x)) d\mu = \boxed{t \int_A D_\mu \nu(x) d\mu}.$$

(In the last part we notice that $A \cap Z$ doesn't necessarily have μ-measure zero, but the Radon–Nikodym derivative integrates to zero on that set anyway.) Similarly, we can flip things around and write

$$\boxed{\nu(A)} = \sum_m \nu(A_m) \ge \sum_m t^m \mu(A_m) \ge \frac{1}{t} \sum_m t^{m+1} \mu(A_m) \ge \frac{1}{t} \sum \int_{A_m} D_\mu \nu(x) d\mu = \boxed{\frac{1}{t} \int_A D_\mu \nu(x) d\mu}.$$

Taking $t \to 1$ and combining the two inequalities yields the result. □

The next question to ask is what happens when the absolute continuity condition $\nu \ll \mu$ is violated. One naive example is where the supports of μ and ν are in disjoint intervals, but things can be much more complicated. For example, consider the usual ternary Cantor function f (a monotonic function) and define the measure of any interval $[x, y]$ to be the difference of $f(y)$ and $f(x)$. Then its relation to the Lebesgue measure is somewhat complicated. The idea is that given any μ, ν, we want to split up a measure ν into a part that is absolutely continuous to μ and the "remaining part:"

This is indeed something we can do:

Definition 84

We say that two measures μ, ν are **mutually singular**, denoted $\mu \perp \nu$, if there is a set $B \subseteq \mathbb{R}^n$ such that $\mu(B) = \nu(\mathbb{R}^n \setminus B) = 0$.

Theorem 85 (Lebesgue decomposition theorem)

Let μ, ν be two Radon measures. Then we can decompose ν as $\nu_{\text{ac}} + \nu_s$, where $\nu_{\text{ac}} \ll \mu$ and $\nu_s \perp \mu$.

The idea of the proof is to choose a set B such that $\nu_{\text{ac}} = \nu|_B$ and $\nu_s = \nu|_{B^c}$ – specifically, we choose the largest such set B that makes our restriction still absolutely continuous, and we do so by considering all candidates

$$\mathcal{F} = \{A \subseteq \mathbb{R}^n : \mu(\mathbb{R}^n \setminus A) = 0\}.$$

Specifically, we want the ν-measure of A to be as small as possible, or else we can find some positive-ν-measure set outside of A with μ-measure zero, breaking absolute continuity when we restrict to A. What's more complicated in all of this discussion is the structure of the singular measure, but that's more delicate and we won't get much into it.

11 November 10, 2022

Last time, we discussed the Radon–Nikodym theorem: we say that $\nu \ll \mu$ if whenever $\mu(E) = 0$, we also have $\nu(E) = 0$, and Radon–Nikodym says that whenever this **absolutely continuous** condition holds for two Radon measures, we have

$$\nu(E) = \int_E D_\mu \nu d\mu.$$

We said that two measures were **mutually singular** if there is some Borel set B such that $\mu(\mathbb{R}^n \setminus B) = \nu(B) = 0$, and we stated the Lebesgue decomposition theorem, which says that we can decompose any measure ν into $\nu_{\text{ac}} + \nu_s$ (for $\nu_{\text{ac}} \ll \mu$ and $\nu_s \perp \mu$); in other words, for any E we have

$$\nu(E) + \int_E D_\mu \nu_{\text{ac}}(x) d\mu + \nu_s(E).$$

We'll write the proof out more explicitly now:

Proof of Theorem 85. Wihout loss of generality we can assume that $\mu(\mathbb{R}^n)$ and $\nu(\mathbb{R}^n)$ are both finite; otherwise, we can consider the ball $B(0, R)$ and take $R \to \infty$ (exercise). Notice that if we take any set B such that $\mu(B^c) = 0$, then the restriction of ν to B^c is singular to μ. But there's no reason to expect $\nu|_B$ to be absolutely continuous with respect to μ unless B is small enough – we don't want a place inside B where μ has no measure but ν has positive measure. So our goal is to minimize the ν-measure inside B to avoid those problem spots. Define

$$\mathcal{F} = \{B \subseteq \mathbb{R}^n \text{ Borel} : \mu(B^c) = 0\}.$$

(We're restricting to Borel sets because we then know for sure that everything is both μ- and ν-measurable.) Now for each k, define $B_k \in \mathcal{F}$ such that

$$\nu(B_k) \le \inf_{A \in \mathcal{F}} \nu(A) + \frac{1}{k}$$

and define $B = \bigcap_{k=1}^\infty B_k$. Then we know that

$$\mu(B^c) \le \sum_{k=1}^\infty \mu(B_k^c) = \sum 0 = 0,$$

so $B \in \mathcal{F}$ as well and in fact B achieves the infimum $\boxed{\nu(B) = \inf_{A \in \mathcal{F}} \nu(A)}$. We can now define $\nu_{\text{ac}} = \nu|_B$ and $\nu_s = \nu|_{B^c}$; we know that $\nu_s \perp \mu$ because $B \in \mathcal{F}$ and we want to show that ν_{ac} is actually absolutely continuous with respect to μ. Suppose otherwise; then there would be some Borel set E such that $\nu_{\text{ac}}(E) > 0$ but $\mu(E) = 0$. That means E contains some subset of B; define $\tilde{B} = B \setminus E$. Then $\tilde{B}^c = B \cup (B \cap E)$, so in particular $\mu(\tilde{B}^c) \le \mu(B^c) + \mu(E) = 0 + 0 = 0$ (by definition of B and E). Thus $\tilde{B} \in \mathcal{F}$ as well, but $\nu(\tilde{B}^c) = \nu(B) - \nu(B \cap E)$ is smaller than $\nu(B)$, since $\nu(B \cap E) = \nu_{\text{ac}}(E) > 0$ by assumption. This contradicts B achieving the infimum boxed above, so we do indeed have absolute continuity.

And now we can also notice that $D_\mu \nu_s = 0$. Indeed, if we look at the set

$$C_z = \{D_\mu \nu_s > z\} = (C_z \cap B) \cup (C_z \cap B^c),$$

we see that $\mu(C_z \cap B^c) = 0$ because $\mu(B^c) = 0$, and $\nu_s(C_z \cap B) \ge z\mu(C_z \cap B)$ by Lemma 80 but the left-hand side is zero because ν_s is supported only on B^c. Thus both terms on the right-hand side have measure zero and thus $\mu(C_z) = 0$ for all z; this implies that the Radon–Nikodym derivative is zero μ-almost-everywhere, and by the Radon–Nikodym theorem $D_\mu \nu = D_\mu \nu_{\text{ac}}$. That gives us the desired equation $\nu(E) + \int_E D_\mu \nu_{\text{ac}}(x) d\mu + \nu_s(E)$. $\square$

(What we've done is applicable more generally to measure spaces – nothing here is specific to Radon measures or $\mathbb{R}^n$ – and there's only a few places where we've really used any regularity.) And with that, we're now ready to turn to the question that we started last week, asking what happens when we average a function over a ball. Define

$$⨍_E f \, d\mu = \frac{1}{\mu(E)} \int_E f \, d\mu;$$

what we're curious about is whether averages around small balls get us the value of the function at the center.

Theorem 86 (Lebesgue differentiation theorem)
Let μ be a Radon measure, and suppose $f \in L^1_{\text{loc}}(d\mu)$ (meaning we're integrable around a small ball). Then

$$f(x) = \lim_{R \downarrow 0} ⨍_{B(x,R)} f(y) d\mu_y$$

for μ-almost-every $x \in \mathbb{R}^n$. (Here μ_y indicates integrating in y but using the measure μ.)

This result is not surprising for continuous functions, but it is surprising for L^1 functions that integrability and measurability is enough to get the value at the center to be equal to the limit of the averages. The proof is "abstract nonsense:"

Proof. Let $f = f_+ - f_-$ – we'll assume $f \in L^1$ to avoid some pesky details; in general, we can just restrict to a small ball and do everything within that. Define

$$\nu_+(E) = \int_E f_+ d\mu, \quad \nu_-(E) = \int_E f_- d\mu$$

for all Borel sets E. Now for general sets A define

$$\nu^*(A) = \inf \{\nu(B) : B \supseteq A, B \text{ Borel}\}$$

to get a measure ν. Since $f \in L^1(\mathbb{R}^n)$ (and thus f^+ and f^- are), $\nu_\pm$ are absolutely continuous with respect to μ, meaning that

$$\nu_+(A) = \int_A D_\mu \nu_+ d\mu, \quad \nu_-(A) = \int_A D_\mu \nu_- d\mu.$$

But at the same time, we also know that

$$\nu_+(A) = \int_A f_+ d\mu, \quad \nu_-(A) = \int_A f_- d\mu.$$

So if we look at a set of the form $S_q = \{D_\mu \nu_+ > f_+ + q\}$ for any $q > 0$, then

$$\nu_+(S_q) = \int_{S_q} f_+ d\mu = \int_{S_q} D_\mu \nu_+ d\mu > \int_{S_q} (f_+ + q) d\mu,$$

which can only occur if $\mu(S_q) = 0$ for all positive q. Flipping the roles of f_+ and $D_\mu\nu_+$, this means $D_\mu\nu_+ = f_+$ μ-almost-everywhere, and similarly $D_\mu\nu_- = f_-$ μ-almost everywhere. So $f = D_\mu\nu$ μ-almost-everywhere, which is actually exactly what we wanted to prove by the definition of the function $D_\mu\nu$. □

Definition 87

Let $f \in L^p_{\text{loc}}(d\mu)$ for some $1 \le p < \infty$. A point $x \in \mathbb{R}^n$ is a **Lebesgue point** if

$$\lim_{R\downarrow 0} \fint_{B(x,R)} |f(y) - f(x)|^p d\mu_y = 0.$$

Corollary 88

For any Radon measure μ and $f \in L^p_{\text{loc}}(d\mu)$, we have that μ-almost-every $x \in \mathbb{R}^n$ is a Lebesgue point of f.

Bad proof. We cannot directly apply the previous result because $|f(y) - f(x)|^p$ depends on x, and if we tried to fix x and apply the Lebesgue differentiation theorem with $g(z) = |f(z) - f(x)|^p$ (which is in L^1_{loc}), then the previous result tells us that

$$\lim_{R\downarrow 0} \fint_{B(z,R)} |f(y) - f(x)|^p d\mu = |f(z) - f(x)|^p$$

for μ-almost-every $z \in \mathbb{R}^n$. But there's no reason for this to be true at x specifically, because x could be in the "bad set." □

Proof. Instead, we have to be a bit more careful: start with a dense set of values ξ_j in $\mathbb{R}$ and set

$$g_j(z) = |f(z) - \xi_j|^p.$$

Then for each j, there is some set $G_j \subseteq \mathbb{R}^n$ such that $\mu(G_j^c) = 0$ and

$$\lim_{R\downarrow 0} \fint_{B(z,R)} |f(y) - \xi_j|^p d\mu_y = |f(z) - \xi_j|^p$$

for all $z \in G_j$. We know that $G = \bigcap G_j$ is a set of full measure, since $\mu(G^c) \le \sum \mu(G_j^c) = 0$, and now we know that for any $x \in G$, by the triangle inequality

$$\lim_{R\downarrow 0} \fint_{B(x,R)} |f(y) - f(x)|^p d\mu_y \le C \lim_{R\downarrow 0} \fint_{B(x,R)} |f(y) - \xi_j|^p d\mu_y + C|f(x) - \xi_j|^p$$

for some constant C depending on p – here there's no integral on the second term because $|f(x) - \xi_j|^p$ doesn't depend on y. Because ξ_j is a dense set, for any ε we can find some ξ_j such that $|f(x) - \xi_j| < C|f(x) - \xi_j|^p + \varepsilon$, and the integral on the right-hand side goes to zero as $R \downarrow 0$ because we're in the good set G. Thus $\lim_{R\downarrow 0} \fint_{B(x,R)} |f(y) - f(x)|^p d\mu_y \le C\varepsilon^p$ for all ε and thus must be zero for any $x \in G$. □

What's important is that we do not have uniformity in this statement – the rate of convergence at the different ξ_js are different, so we can't find a uniformly small R so that the averages around various points x are uniformly close to their respective $f(x)$s. But it's still a powerful result – for example, applying this result to the characteristic function of $y \in B(x, R)$ yields the following:

Corollary 89

Let E be a Lebesgue measurable set in $\mathbb{R}^n$. Then for almost every x,

$$\lim \frac{|B(x,R)\cap E|}{|B(x,R)|} = \begin{cases} 1 & x \in E, \\ 0 & x \notin E. \end{cases}$$

In other words, except for a set of measure zero, the points around any point in E fill everything, and the points around any point not in E are almost nothing.

We'll now switch gears and talk about **bounded linear functionals on function spaces** – for example, fix some $\mu \in \mathbb{R}^n$ and a function g and define

$$L_g(f) = \int_{\mathbb{R}^n} fg d\mu.$$

We're then curious to see the set of functions f for which this is defined. By Hodler's inequality, we know that if $f \in L^p$ and $g \in L^q$, where $\frac{1}{p} + \frac{1}{q} = 1$, then $|L_g(f)| \le ||f||_{L^p}||g||_{L^q}$. So for any $g \in L^q$, L_g is a **bounded linear functional** on L^p (where $\frac{1}{p} + \frac{1}{q} = 1$). So we want to ask if the converse is true as well – in other words, we want to know if any bounded linear functional $F : L^p \to \mathbb{R}$ is of the form $F(f) = L_g(f)$ for some $g \in L^q$. It turns out the answer is yes for any $1 \le p < \infty$, but the case $p = \infty$ is more complicated. There's no particularly good answer, and we'll instead rephrase this case by thinking not about $L^\infty(\mathbb{R}^n)$ but about $C_c(\mathbb{R}^n)$, the space of continuous functions with compact support. Then the linear functionals on $C_c(\mathbb{R}^n)$ do have a nice description – it turns out they are all integrals against a signed measure, meaning that we always have

$$F(f) = \int f d\nu \;\forall f \in C_c(\mathbb{R}^n).$$

To get to this result, we'll first talk about signed measures, then discuss the Riesz representation theorem for L^p ($1 \le p < \infty$), and finally discuss the Riesz representation theorem for $C_c(\mathbb{R}^n)$.

Remark 90. *There were two brother mathematicians both named Riesz, who sometimes worked together but had different proof strategies. This one comes from F. Riesz, who had a more fundamental approach, but we'll see M. Riesz's trickier proofs later on in the course.*

Definition 91

A **signed measure** is a function defined on a σ-algebra $\mathcal{B}$ of sets, such that (1) ν takes at most one of the values ∞ and $-\infty$, (2) $\nu(\varnothing) = 0$, and (3) for any countable collection of pairwise disjoint sets in $\mathcal{B}$, $\sum_{i=1}^\infty \nu(E_j)$ converges absolutely and is equal to $\nu\left(\bigcup_{j=1}^\infty E_j\right)$.

The point of assumption (1) is to avoid issues with computing measures if we have $\infty - \infty$, and notice that absolute convergence is part of the assumption in (3).

Definition 92

A set A is **positive** with respect to a signed measure ν if $\nu(E) \ge 0$ for all $E \subseteq A$ with $E \in \mathcal{B}$.

In words, a set is positive if none of its subsets have negative measure.

Proposition 93

Let $A \in \mathcal{B}$ be a set with $\nu(A) > 0$. Then there is some $E \subseteq A$ (with $E \in \mathcal{B}$) such that E is positive and $\nu(E) > 0$

We should think of $\nu(E) = \int_E f d\mu$ for some $f \in L^1(d\mu)$, where f does not necessarily need to be a positive function. So what we're saying is basically that there is a set on which f is positive.

Proof. The point is to throw away sets of negative measure. Take the smallest n_1 such that A contains a set B_1 with $\nu(B_1) < -\frac{1}{n_1}$, and remove B_1 from A. Then take the largest n_2 such that $A \setminus B_1$ contains a set B_2 with $\nu(B_2) < -\frac{1}{n_2}$; throw out B_2. Repeat this process and define $\tilde{A} = A \setminus (\bigcup B_k)$. We know that

$$\nu(\tilde{A}) + \sum \nu(B_k) = \nu(A) \implies \sum |\nu(B_k)| = \nu(\tilde{A}) - \nu(A) < \nu(\tilde{A}),$$

and we know that $\nu(\tilde{A})$ is finite because the series converges absolutely. So $\sum_{k=1}^{\infty} |\nu(B_k)|$ is finite and converges to some finite value, meaning $\nu(B_k) \to 0$ as $k \to \infty$. We now claim that $\tilde{A}$ cannot contain a set of negative measure. Suppose it did contain such a set S; then there is some k at which S could have been chosen but wasn't (because it beats out the set of measure $-\frac{1}{k}$ by being more negative). Thus no S can exist and $\tilde{A}$ is the set we want.

(In the case where $\nu(\tilde{A}) = \infty$, not much changes – the measure cannot take the value $-\infty$, so $\sum(B_k)$ should still converge absolutely.) $\square$

We'll use this next time to prove the Hahn decomposition theorem, which lets us split up a space into a "positive part" and a "negative part" with respect to ν (specifically, any signed measure is the difference of two measures with disjoint support).

12 November 15, 2022

Our first topic today is the Riesz representation theorem on L^p. Recall that a bounded linear functional on $L^p(\mathbb{R}^n, \mu)$ is a linear functional $L^p(\mathbb{R}^n, \mu) \to \mathbb{R}$ such that $|L(f)| \le C||f||_{L^p}$ for all f and some constant C; we then call the norm of f the infimum of all such constants C. More precisely, this means

$$||F|| = \sup_{f \neq 0, f \in L^p} \frac{|F(f)|}{||f||}_{L^p}.$$

One simple example we discussed last time is to fix a function g and define $F_g(f) = \int fg d\mu$. In the cases where $g \in L^q(\mathbb{R}^n, d\mu)$ where $\frac{1}{p} + \frac{1}{q} = 1$, we know that F_g is a bounded linear functional with norm at most $||g||_q$ by Holder's inequality, since $F_g(f) \le (\int |g|^q)^{1/q} (\int |f|^p)^{1/p} = ||q||_{L^q} ||f||_{L^p}$. For another point of view, we can define a new measure ν via $d\nu = g d\mu$, so that $F_g(f) = \int f d\nu$. Then ν is a **signed** measure with a density (living in L^q) with respect to μ – on $C_c(\mathbb{R}^n)$ we'll see that this view is useful.

Now suppose we're given a bounded linear functional $F : L^p \to \mathbb{R}$ which is of the form F_g. Then we can always reconstruct g, because we can take f to be some indicator function 1_E to find that

$$F(1_E) = \int_E g d\mu,$$

so that we can define a signed measure $\nu(E) = F(1_E)$, and then g must be be the Radon–Nikodym derivative of ν with respect to μ (if it exists). Noticing now that if we know $F(1_E)$ for all E, we know the values of F on all simple functions (meaning $F(f) = \int fg$ for all simple f), and thus we know the values of F everywhere as long as simple functions are dense in our space. So this whole thing works as long as $1 \le p < \infty$, but it doesn't when $p = \infty$, and that explains why the situation with linear functionals is different for L^∞.

All of this has been motivation for what we actually need to prove:

Theorem 94

Let μ be a Radon measure, let $1 \le p < \infty$, and let $F : L^p(\mathbb{R}^n, d\mu) \to \mathbb{R}$ be a bounded linear functional. Then there exists $g \in L^q(\mathbb{R}^n, d\mu)$ with $\frac{1}{p} + \frac{1}{q} = 1$, such that $F(f) = \int fgd\mu$ for all $f \in L^p(\mathbb{R}^n, d\mu)$ and $||F|| = ||g||_{L^q}$.

Proof. First assume that $\mu(\mathbb{R}^n) < \infty$ so that $f = 1$ is in all L^p spaces, and $1_E \in L^p(\mathbb{R}^n, d\mu)$ for all Borel sets as well. We will also assume that $p > 1$ so that $q \ne \infty$. Thus we can define $\nu(E) = F(1_E)$ for all Borel sets E. ν is a signed measure by linearity of F, and by definition of the norm of our bounded linear operator F,

$$|\nu(E)| \le ||F|| \cdot ||\chi_E||_{L^p} = ||F||(\mu(E))^{1/p}.$$

Thus $\nu \ll \mu$, because if $\mu(E) = 0$ then $\nu(E) = 0$ as well (and notice that this part of the argument does not work if $p = \infty$). But that's a sloppy thing to say when we have signed measures involved, so instead by Hahn decomposition we can write $\nu = \nu^+ - \nu^-$ and split up $\mathbb{R}^n = A \cup B$, where A is a positive set and B is a negative set. We then have

$$\nu^+(E) = \nu(A \cap E) \le ||F||(\mu(A \cap E))^{1/p} \le ||F||(\mu(E))^{1/p}$$

so by the same argument $\nu^+ \ll \mu$ and similarly $\nu^- \ll \mu$. Therefore by Radon–Nikodym we have $\nu^+ = \int g^+ d\mu$ and $\nu^- = \int g^- d\mu$, with g^+ and g^- both in $L^1(d\mu)$. Thus $g \in L^1(d\mu)$ as well, and we've constructed a function satisfying the property

$$F(\chi_E) = \int_E g d\mu$$

for all Borel sets E. We claim that g is actually in $L^q(d\mu)$ (which is a stricter condition than being in L^1 if μ is finite). We take $\psi_n(x)$ to be a nondecreasing sequence of simple functions, taking finitely many values, so that $\psi_n^{1/q} \to g$ (from below). For example, we can take

$$\psi_n(x) = \frac{j}{2^n} \text{ if } |x| \le n, \frac{j}{2^n} \le |g(x)|^q < \frac{j+1}{2^n}$$

and define $\psi_n(x) = 0$ otherwise (if $|x| \ge n$ or $|g(x)|^q > 2^n$); notice that this does make ψ_n only take on finitely many values. If we then define $\phi_n = (\psi_n)^{1/p}\text{sgn}(g)$, then $||\phi_n||_{L^p} = (\int \psi_n d\mu)^{1/p}$, and

$$\int \psi_n d\mu = \int \psi_n^{1/p} \psi_n^{1/q} d\mu = \int |\psi_n|^{1/q} |\phi_n| d\mu \le \int |g||\phi_n| d\mu = \int g\phi_n d\mu$$

because ϕ_n and g always have the same sign. But ϕ_n is a simple function taking on finitely many values, so in fact this last expression is $F(\phi_n)$ by our definition on simple functions. Since ϕ_n is in L^p, we can thus bound this by $||F|| \cdot ||\phi_n||_{L^p}$. Putting this together,

$$\int \psi_n d\mu \le ||F|| \cdot ||\phi_n||_{L^p} = ||F|| \left(\int \psi_n d\mu \right)^{1/p}.$$

Dividing by $\left(\int \psi_n d\mu\right)^{1/p}$ on both sides yields $\left(\int \psi_n d\mu\right)^{1/q} \le ||F||$, so we see that

$$\left(\int \left(|\psi_n|^{1/q} \right)^q d\mu \right)^{1/q} \le ||F|| \implies || |\psi_n|^{1/q} ||_{L^q} \le ||F||.$$

Since the blue function converges to g, we must also have $\boxed{||g||_{L^q} \le ||F||}$ by Fatou's lemma, and thus $g \in L^q$.

This means that $F(f) = \int fgd\mu$ for all simple functions for some $g \in L^q(\mathbb{R}^n, d\mu)$, and by density this means $F(f) = \int fgd\mu$ for all $f \in L^p(\mathbb{R}^n, d\mu)$ (a bounded linear functional defined on a dense set can be extended to the whole space, and $||F(f)|| \le ||g||_{L^q}||f||_{L^p}$ by Holder so $\boxed{||F|| \le ||g||_{L^q}}$. Putting the boxed statements together shows that the two norms are equal.

Now we remove the additional assumptions that we placed. If the total measure $\mu(\mathbb{R}^n)$ is infinite, we now define $1_R(x) = 1_{B(0,R)}(x)$ for any R. we claim that $||f \cdot 1_R - f||_p \to 0$ as $R \to \infty$. Indeed, if we define $F_R(f) = F(f \cdot 1_R)$ and define μ_R to be the restriction of μ to $B(0, R)$, then we can check that F_R is a bounded linear functional on $L^p(d\mu_R)$. Since the original measure was Radon, μ_R is finite, and thus $F_R(f) = \int f \cdot 1_R g_R d\mu_R = \int f \cdot 1_R g_R d\mu$. But if $R_1 > R_2$, then $g_{R_1} = g_{R_2}$ on $B(0, R_2)$, so if we define $g_R = g1_R$ we have

$$\int |g_r|^q d\mu \le ||F_R|| \le ||F|| \implies \int |g|^q d\mu \le ||F||,$$

and the rest of this just follows from the above case.

The more interesting question is to ask about the case $p = 1$. Then $F(f) = \int fgd\mu$ for all simple functions for f taking on finitely many values, and we can consider the sets $\{x : |g(x)| \le ||F|| + \varepsilon\}$. We then know that

$$|\int_E gd\mu| \le ||F|| \cdot ||1_E||_{L^1} = ||F||\mu(E),$$

but taking E to be the set $\{x : |g(x)| \le ||F|| + \varepsilon\}$ we have $(||F|| + \varepsilon)\mu(E) \le ||F||\mu(E)$, which can only happen if $\mu(E) = 0$. So $|g(x)| \le ||F||$ almost everywhere, meaning $g \in L^\infty$ and $||g||_{L^\infty} \le ||F||$. The rest of the proof follows as in the above case, and to show that the norm of F is exactly $||g||_\infty$, we notice that $F(f) = \int fgd\mu \le \int f||g||_\infty d\mu = ||g||_\infty ||f||_{L^1}$, so $||F|| \le ||g||_{L^\infty}$ and the matching inequality is proven. $\square$

Notice that none of this proof really relies on our space being $\mathbb{R}^n$, as long as we have the Radon–Nikodym theorem.

Now we're ready to look at the more challenging case of $C_c(\mathbb{R}^n)$. We know that we can integrate against L^1, but just having L^1 is in some sense too restrictive because any function of compact support can integrate against 1 or in general bounded functions. So we need a **different sense** of "bounded linear functional" as we do for functional analysis:

Definition 95

A linear function $L : C_c(\mathbb{R}^n, \mathbb{R}^m)$ (using $\mathbb{R}^m$ instead of $\mathbb{R}$ doesn't complicate things too much more) is **bounded** if for every compact set K, the quantity

$$L_K = \sup\{L(f) \in C_c(\mathbb{R}{\cdot}\mathbb{R}^m) : \text{supp}(f) \subseteq K, |f(x)| \le 1 \ \forall x \in \mathbb{R}^n\}$$

is finite.

(This is the same as being continuous in the compact open topology.)

Theorem 96

For any bounded linear function $L : C_c(\mathbb{R}^n, \mathbb{R}^m)$, there is some Radon measure μ and a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$, such that the following properties hold:

1. $|\sigma(x)| = 1$ for μ-almost-every $x \in \mathbb{R}^n$,
2. $L(f) = \int_{\mathbb{R}^n}(f \cdot \sigma)d\mu$ for all $f \in C_c(\mathbb{R}^n; \mathbb{R}^m)$.

In particular, if $m = 1$, then $d\nu = \sigma d\mu$ is a signed measure (because σ takes values in $\mathbb{R}$), and $L(f) = \int f d\nu$ for all $f \in C_c(\mathbb{R}^n)$.

Beginning of proof. We'll just do the case $m = 1$. The key construction here is finding the Radon measure μ. Like in the last proof, we want to say that $L(f) = \int f d\nu$ and define $L(1_E) = \nu(E)$ for all E. But 1_E is not in $C_c(\mathbb{R}^n)$, so this is not a valid step. Instead, we define the **variation measure** for any open set V

$$\mu^*(V) = \sup\{L(f) : f \in C_c(\mathbb{R}^n, \mathbb{R}^m), |f| \le 1, \operatorname{supp}(f) \subseteq V\}$$

(this is basically an approximation to what we wanted to do in the previous proof), and then for an arbitrary set A we define

$$\mu^*(A) = \inf\{\mu^*(V) : V \text{ open}, V \supseteq A\}.$$

We must show that μ is a Radon measure. We do this by generalization: notice that the condition $|f| \le 1, \operatorname{supp}(f) \subseteq V$ is the same as saying that $|f| \le 1_V$, so we should think of $\mu^*(V)$ as (in some sense) the "integral of 1_V.") Motivated by this, define for any positive function $f \in C_c^+(\mathbb{R}^n)$ (not just the indicator functions)

$$\lambda(f) = \{\sup L(g) : |g| \le f, g \in C_c(\mathbb{R}^n)\}.$$

We claim that λ is a linear functional and that $\lambda(f) = \int f d\mu$. And now to get our function σ, we define for any scalar f and vector $e \in \mathbb{R}^n$

$$\lambda_e(f) = L(fe) = \int f\sigma_e d\mu.$$

Then defining $\sigma = \sum \sigma_{e_j}(x)e_j(x)$, we see that

$$f = \sum (f \cdot e_j)e_j \implies L(f) + \sum L((f \cdot e_j)e_j) = \sum \lambda_{e_j}(f \cdot e_j)$$

$$= \sum \int (f \cdot e_j)\sigma_{e_i} d\mu = \int (f \cdot \sigma)d\mu.$$

From there, the next step is to show that $|\sigma| = 1$. □

We'll continue the proof next time, but the point is to first set up an approximation μ^* and then realize that we need more, leading us to define λ. But the point is to represent λ as an integral against some measure, and that is the key step.

13 November 17, 2022

We'll continue proving Theorem 96 today, showing that for any linear functional L on $C_c(\mathbb{R}^n, \mathbb{R}^m)$ such that $\sup\{L(f) : f \in C_c(\mathbb{R}^n; \mathbb{R}^m), |f| \le 1, \operatorname{supp}(f) \subseteq K\}$ is finite for any compact set K, we have a Radon measure μ and a function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ so that $|\sigma(x)| = 1$ μ-almost-everywhere and $L(f) = \int (f \cdot \sigma)d\mu$ for all $f \in C_c(\mathbb{R}^n)$. In particular, we get a signed measure $\nu = \sigma\mu$ in dimension $m = 1$, meaning that such linear functional can be represented as integration against a signed measure ν.

Proof continued. As we mentioned last time, the difficulty in constructing this μ is that we want to insert 1_E instead of f, but 1_E is not continuous so we can't apply L to it. So we defined

$$\mu^*(V) = \sup\{L(g) : g \in C_c(\mathbb{R}^n; \mathbb{R}^m), |g| \le 1, \operatorname{supp}(g) \subseteq V\}$$

for every open set V and let $\mu^*(A)$ be the infimum of $\mu^*(V)$ over all open sets $V \supseteq A$ for a general A. The point is that g is supposed to play the role of plugging in an arbitrary 1_E, so now we can define

$$\lambda(f) = \sup\{L(g) : g \in C_c(\mathbb{R}^n, \mathbb{R}^m), |g| \le f\}.$$

But before we go on, we need to prove a few properties. First we verify that as defined on V and A above, μ is a measure. We must check subadditivity first on open sets: if $V_j \subseteq \mathbb{R}^n$ and $V = \bigcup_{i=1}^{\infty} V_j$, then we can choose g supported in V such that $|g| \le 1$. Then $K_g = \text{supp}(g)$ is a compact set with $K_g \subseteq V$, so by compactness $K_g \subseteq \bigcup_{i=1}^{k} V_j$ for some finite subset. Taking ζ_j to be continuous functions such that $\text{supp}(\zeta_j) \subseteq V_j$ and $\sum_{j=1}^{k} \zeta_j = 1$ on all of K_g (this is a **partition of unity**), we see that

$$g = \sum_{j=1}^{k} g\zeta_j.$$

But $|g\zeta_j| \le 1$ because $|g| \le 1$ and $|\zeta| \le 1$, and the support of $g\zeta_j$ is contained in V_j, so $|L(g(\zeta_j))| \le \mu^*(V_j)$ by the definition of our measure μ. So $|L(g)| \le \sum_{j=1}^{k} \mu^*(V_j) \le \sum_{j=1}^{\infty} \mu^*(V_j)$. Finally taking a supremum over g, we see that $\mu^*(V) \le \sum_{j=1}^{\infty} \mu^*(V_j)$.

Now for showing subadditivity on general sets A_j, pick $V_j \supseteq A_j$ so that $\mu^*(A_j) \ge \mu^*(V_j) - \frac{\varepsilon}{2^j}$. Then

$$A = \bigcup_{i=1}^{\infty} A_j \subseteq \bigcup_{j=1}^{\infty} \implies \mu^*(A) \le \mu^* \left(\bigcup_{j=1}^{\infty} \right) V_j \le \sum \mu^*(V_j) \le \sum \mu^*(A_j) + \varepsilon,$$

so taking $\varepsilon \to 0$ yields the desired subadditivity in general.

Now we check that we have a Borel measure – for that, we use a nice criterion by Caratheodory:

Lemma 97

Let μ be a measure such that $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ whenever $\text{dist}(A, B) > 0$. Then μ is a Borel measure.

(This is basically an exercise in the measure theoretic properties of being a closed set.) So if V_1, V_2 are open sets that are a positive distance away from each other, then any function supported on the union of them can be broken up into functions on each part, so $\mu^*(V) = \mu^*(V_1) + \mu^*(V_2)$. Then if general A_1, A_2 are such that they are a positive distance away from each other, we can find $U_1 \supseteq A_1$ and $U_2 \supseteq A_2$ so that $\text{dist}(U_1, U_2) > 0$ (for example take a small ball of radius $\frac{\text{dist}(U_1, U_2)}{10}$ around every point and take the union. Then for any $V \supseteq A_1 \cup A_2$, we have

$$V = (V \cap U_1) \cup (V \cap U_2) \supseteq A_1 \cup A_2$$

a union of two open sets, where $A_1 \subseteq V \cap U_1$ and $A_2 \subseteq V \cap U_2$ so

$$\mu^*(V) = \mu^*(V_1) + \mu^*(V_2) \implies \mu^*(A_1 \cup A_2) \ge \mu^*(A_1) + \mu^*(A_2)$$

(since we can always choose V so that $\mu^*(A_1 \cup A_2) \ge \mu^*(V) - \varepsilon$). Combining this with subadditivity, Caratheodory's criterion is satisfied and we have a Borel measure; Radon-ness follows by intersecting open sets in the definition of μ^*.

Next we check that λ as defined is a linear functional. Recall that

$$\lambda(f) = \sup\{L(g) : g \in C_c(\mathbb{R}^n, \mathbb{R}^m), |g| \le f, f \in C_c^+(\mathbb{R}^n)\}$$

where f is a positive linear function on $C_c(\mathbb{R}^n)$. We claim that $\lambda(f_1 + f_2) = \lambda(f_1) + \lambda(f_2)$; indeed, take g_1 with $|g_1| \le f_1$ and g_2 with $|g_2| \le |f_2|$. We will not actually take $g = g_1 + g_2$; instead we will take the "larger"

$$g = g_1 \text{sgn}(L(g_1)) + g_2 \text{sgn}(L(g_2)).$$

Call the two terms g_1' and g_2'. Then $|g| \le g_1' + g_2'| \le f_1 + f_2$, and

$$L(g) = |L(g_1)| + |L(g_2)| = L(g_1') + L(g_2'),$$

and by linearity of L this is equal to $L(g_1' + g_2') \le \lambda(f_1 + f_2)$, so $\boxed{\lambda(f_1) + \lambda(f_2) \le \lambda(f_1 + f_2)}$. For the other direction, if we take some g with $|g| \leq f_1+f_2$, we can define $g_k = \frac{f_k}{f_1+f_2} g$ if $f_1+f_2 > 0$ and 0 otherwise for $k = 1, 2$. Then $g = g_1+g_2$ because the two sides are both zero if $f_1 + f_2 = 0$, and $|g_1| \leq f_1$ and $|g_2| \leq f_2$, so $|L(g)| \leq |L(g_1)| + |L(g_2)| =\leq \lambda(f_1) + \lambda(f_2)$. So taking the supremum over all g yields $\boxed{\lambda(f_1 + f_2) \leq \lambda(f_1) + \lambda(f_2)}$. Putting those together shows the result.

So λ is linear, and the next step is to show that λ is integration against this μ that we've defined. The construction should somehow tell us what μ's measure is on indicator functions, but as we mentioned previously we have to dance around the issues of continuity. Choose a partition (fine mesh) $0 < t_0 < t_1 < \cdots < t_N = 2||f||_{L^\infty}$, such that $\mu\left(f^{-1}(t_k)\right) = 0$ for all k. (We do this so we don't hit values where the preimage has a positive measure – this can only happen on a countable set so there's no issues.) Then the sets $U_i = f^{-1}(t_{i-1}, t_i)$ are open and bounded because f is compactly supported, so $\mu(U_j) < \infty$ for all j.

We want to construct (scalar functions) h_j such that $0 \leq h_j \leq 1$, $\text{supp}(h_j) \subseteq U_j$, and $\mu(U_j) - \frac{\varepsilon}{N} \leq \lambda(h_j) \leq \mu(U_j)$. (The idea is that we are making a good approximation to the characteristic function of h_j.) Taking K_j compact so that $K_j \subseteq U_j$ for all j and such that $\mu(U_j \setminus K_j) < \frac{\varepsilon}{N}$, and finding (vector functions) $g_j \in C_c(\mathbb{R}^n; \mathbb{R}^m)$ so that $\text{supp}(g_j) \subseteq U_j$ and and $|L(g_j)| \geq \mu(U_j) - \frac{\varepsilon}{N}$, we can indeed choose h_j of the form to be 1 on $K_j \cup \text{supp}(g_j)$ (we want h_j to be 1 on K_j so it's a good approximation of the indicator of U_j, and we want it to be 1 on $\text{supp}(g_j)$ so that $h_j g_j = g_j$). Then $h_j \geq |g_j|$ means that $\lambda(h_j) \geq |L(g_j)| \geq \mu(U_j) - \frac{\varepsilon}{N}$, and we get the other bound $\lambda(h_j) \leq \mu(U_j)$ for free because we're defining on a more restrictive set in the supremum.

If we now look at the set where we don't have $h_j = 1$, namely

$$A = \{f(x) > 0, \quad 0 \leq h_j < 1 \ \forall 1 \leq j \leq N\}$$

(remember that the h_js have disjoint support so we just have to keep track of one at a time), then

$$\mu(A) = \mu\left(\bigcup_{j=1}^{N} (U_j \setminus \{h_j = 1\}\right),$$

since f needs to be positive so it's in one of the U_js and h_j cannot be 1. But this can be bounded by

$$\mu^*\left(\bigcup_{i=1}^{N} U_j \setminus K_j\right) \leq \sum_{i=1}^{N} \frac{\varepsilon}{N} = \varepsilon,$$

so the set A is small. And now if we plug into λ, by definition we have

$$\lambda\left(f - f\sum_{j=1}^{N} h_j\right) = \sup\left\{|L(g)| : g \in C_c(\mathbb{R}^n; \mathbb{R}^m), |g| \leq f - f\sum h_j\right\}$$

and $f - f\sum h_j$ is nonzero only on A, so $\lambda\left(f - f\sum_{j=1}^{N} h_j\right) \leq \mu(A)||f||_{L^\infty} \leq \varepsilon||f||_{L^\infty}$. That means

$$\lambda(f) \leq \sum_{i=1}^{N} \lambda(fh_j) + \varepsilon||f||_{L^\infty} \leq \sum_{i=1}^{N} t_j\mu(U_j) + \varepsilon||f||_{L^\infty}$$

by definition of the sets U_j. But also by the upper bound on the sets U_j we know that

$$\lambda(f) \geq \sum_{i=1}^{N} \lambda(fh_i) - \varepsilon||f||_{L^\infty} \geq \sum_{j=1}^{N} t_{j-1}\mu(U_j) - \varepsilon||f||_{L^\infty}.$$

And since we basically did a Riemann sum approximation $\sum t_j\mu(U_j)$ and found upper and lower bounds, we now see why we needed the preimage of each t_k to have measure zero – we needed $\int f d\mu = \sum_{j=1}^{N} \int_{U_j} f d\mu$. This means that

$$\sum t_{j-1}\mu(U_j) \int f d\mu \le \sum t_j \mu(U_j)$$

with upper and lower bound differing by at most $\varepsilon(\text{supp}(f))$, so

$$\left| \lambda(f) - \int f d\mu \right| \le C\varepsilon$$

for some constant C independent of ε. Taking $\varepsilon \to 0$ proves $\lambda(f) = \int f d\mu$.

Finally, we must construct the function σ. For any $\tilde{f} \in C_c(\mathbb{R}^n)$ (scalar functions are with tildes) and $e \in S^{m-1}$, we can define

$$\lambda_e(\tilde{f}) = L(\tilde{f}e).$$

Since $|\tilde{f}e| \le |\tilde{f}|$, we know that $|\lambda_e(\tilde{f})| \le \lambda(\tilde{f}) = \int |\tilde{f}| d\mu$ (by definition of λ) and thus λ_e is a bounded linear functional on $L^1(d\mu)$ (it's linear because L is linear). So by the previous Riesz representation theorem on L^p spaces, there is some $\sigma_e \in L^\infty(\mathbb{R}^n)$ such that $L(\tilde{f}e) = \int \tilde{f}\sigma_e d\mu$. But now $f = (f \cdot e_1)e_1 + (f \cdot e_2)e_2 + \cdots + (f \cdot e_m)e_m$, so we do indeed have condition (2) of the theorem

$$L(f) = \sum_{j=1}^{n} \int (f \cdot e_k)\sigma_{e_k} d\mu = \int (f \cdot \sigma) d\mu,$$

as long as we define $\boxed{\sigma = \sum \sigma_{e_k} e_k}$. So we now need to show condition (1), namely that $|\sigma(x)| = 1$ μ-almost-everywhere. Taking any open set U, we can consider

$$\sigma'(x) = \begin{cases} \sigma(x)/|\sigma(x)| & \sigma(x) \neq 0, \\ 0 & \sigma(x) = 0. \end{cases}$$

By Lusin's theorem, we can find $K_j \subseteq U$ such that $\sigma'(x)$ is continuous on K_j with $\mu(U \setminus K_j) < \frac{1}{j}$. Furthermore, we can extend $\sigma'(x)$ to a continuous function f_j on all of $\mathbb{R}^n$ such that $|f_j(x)| \le 1$ (the norm is not increased). Now take some "cutoff function" h_j which is 1 on K_j and where the support of h_j is contained in U (this we can always do), such that $0 \le h_j \le 1$ everywhere. Then the sequence of functions $g_j = f_j h_j$ has $|g_j| \le 1$, $\text{supp}(g_j) \subseteq U$, and $g_j \cdot \sigma$ only differs from $|\sigma|$ is on the set $U \setminus K_j$, which has vanishing measure. Thus $g_j \cdot \sigma \to |\sigma|$ in probability (the set where the difference is large goes to zero), so there is a subsequence g_{jk} where we have convergence almost everywhere.

Now $\int |\sigma| d\mu \le \lim_{k\to\infty}(g_{jk} \cdot \sigma) d\mu$, but because g_{jk} is continuous this is just $\lim_{k\to\infty} L(g_{jk})$. This is at most $\mu(U)$ because $g_{jk} \le 1$ and is supported inside U, which means $\int |\sigma| d\mu \le \mu(U)$. But $\mu(U)$ is a supremum of $\int f \cdot \sigma d\mu$ over all f satisfying our conditions, and because $|f| \le 1$ that is always at most $|\sigma| d\mu$. Putting these inequalities together, we must actually have $\mu(U) = \int_U |\sigma| d\mu$ for all open sets U, which means $|\sigma| = 1$ μ-almost-everywhere as desired. □

14 November 29, 2022

We'll discuss the Fourier transform today, first analyzing everything on the circle S^1. The idea is the following: if $f(x)$ is a periodic function (so that it can be defined on S^1), we wish to write $f(x) = \sum_{k\in\mathbb{Z}} f_k e^{2\pi i k x}$. We know that the inner product $\langle e^{2\pi i k x}, e^{2\pi i m x}\rangle = \delta_{km}$ (we have an orthonormal set), so if the coefficients f_k exist, then we must have $\boxed{f_k = \langle f, e^{2\pi i k x}\rangle = \int_0^1 f(x) e^{-2\pi i k x} dx}$. Then there are two separate questions here – we want to ask whether $\sum_{k\in\mathbb{Z}} f_k e^{2\pi ikx}$ actually converges to $f(x)$, and we also want to know if a function can be approximated by trigonometric polynomials (for example, approximating sounds in music with pure sounds). We'll start with the first of these questions:

Lemma 98 (Riemann-Lebesgue)
Let $f \in L^1(S^1)$. Then $f_k \to 0$ as $k \to \infty$.

Proof. We know that $f_k = \int_0^1 f(x)e^{-2\pi ikx}dx$. An "engineer's proof" would use integration by parts, saying that $\int_0^1 f(x)\left(-\frac{1}{2\pi ik}\frac{d}{dx}(e^{-2\pi ikx}\right) dx = \frac{1}{2\pi ikx}\int_0^1 f'(x)e^{-2\pi ikx}dx$ and thus $|f_k| \le \frac{1}{2\pi|k|}||f'||_{L^1}$. But of course this doesn't work because we don't know that f' is in L^1.

Instead, we can write the coefficient as

$$f_k = \int f\left(x + \frac{1}{2k}\right)e^{-2\pi ik(x+\frac{1}{2k})}dx = -\int f\left(x+\frac{1}{2k}\right)e^{-2\pi ikx}dx,$$

where we've used that $e^{\pi i} = -1$. Thus $f_k = \frac{1}{2}\int\left(f(x) - f(x+\frac{1}{2k})\right)e^{-2\pi ikx}dx$ (by averaging the two expressions we got), but now that means by triangle inequality that

$$|f_k| \le \frac{1}{2}\int_0^1 \left|f(x) - f\left(x+\frac{1}{2k}\right)\right| dx$$

goes to 0 as $k \to \infty$ for any $f \in L^1(S^1)$ as we've proved more generally before. ☐

So the decay rate depends on how quickly $\int |f(x) - f(x+\varepsilon)|dx$ goes to 0, which is a question of how quickly f oscillates. So Fourier coefficients decay faster or slower **depending on scales of variation of f**.

Corollary 99
As $m \to \infty$, we have $\int f(x)\sin(2\pi mx)dx \to 0$ and $\int f(x)\cos(2\pi mx)dx \to 0$.

Towards showing convergence, we can now define the partial sums

$$S_N f(x) = \sum_{k=-N}^{N} f_k e^{2\pi ikx}$$

for each N, which we can write out explicitly as

$$\int_0^1 f(t)\sum_{k=-N}^{N} e^{-2\pi ik(x-t)}dt.$$

Changing variables, this is the same as

$$= \int_0^1 f(x-t)\sum_{k=-N}^{N} e^{2i\pi kt}dt = \int_0^1 f(x-t)D_N(t)dt.$$

So we are basically doing a convolution of f with the **Dini kernel** $D_N(t) = \sum_{k=-N}^{N} e^{2\pi ikt}$, so in order for this to converge to f we should have basically an approximation of identity (so D_N should approach the delta function, looking like $\frac{1}{\varepsilon}g\left(\frac{x}{\varepsilon}\right)$ for some function g with $\int g dx = 1$). Since the Dini kernel is a geometric series, we can explicitly calculate and find that

$$D_N(t) = \frac{\sin((2N+1)\pi t)}{\sin(\pi t)}.$$

Unfortunately, such a function oscillates between positive and negative values near $t = 0$, but it is true that $\int_{-1/2}^{1/2} D_N(t)dt = 1$ from the original formula for the Dini kernel. On the other hand, the L^1 norm of D_N is $\int_{-1/2}^{1/2} |D_N(t)|dt$ can be calculated by noticing that

$$\frac{\sin((2N+1)\pi t)}{\pi t} + \left(\frac{1}{\sin(\pi t)} - \frac{1}{\pi t}\right)\sin((2N+1)\pi t),$$

where the second term here is bounded, so we roughly have $\int_{-1/2}^{1/2} |D_N(t)|dt = O(\log N)$. So that's not good, because we're convolving with something not of L^1 norm 1. Instead, having any hope of convergence $S_N f(x) \to f(x)$ would rely on having lots of cancellations between the different $e^{2\pi ikt}$s. And if $f(x-t)$ follows the oscillations of the Dini kernel, then $\int_0^1 f(x-t)D_N(t)dt$ could be huge. So we need a way to prohibit that, ensuring that f is "almost a constant" near x:

Theorem 100 (Dini criterion)

Let $f \in L^1(S^1)$. Suppose that at some given x and for some $\delta > 0$, we have $\int_{|t|<\delta} \frac{|f(x+t)-f(x)|}{|t|} dt < \infty$. (For example, this holds if f is Lipschitz or even Holder with any exponent.) Then $\lim_{N\to\infty} S_N f(x) = f(x)$.

Proof. Since D_N integrates to 1, we have

$$S_N f(x) - f(x) = \int_{-1/2}^{1/2} D_N(t)(f(x-t) - f(x))dx.$$

By the explicit form of the Dini kernel, this can be written as

$$= \int_{-1/2}^{1/2} (f(x-t) - f(x))\frac{\sin((2N+1)\pi t)}{\sin(\pi t)}dt.$$

Splitting this integral into the part A where $|t| < \delta$ and the part B where $|t| > \delta$, the first integral can be written as

$$A = \int \frac{f(x-t) - f(x)}{\sin(\pi t)} 1\{|t| < \delta\} \sin(\pi(2N+1)t)dt.$$

Calling the blue part $g_x(t)$, by the Dini criterion we know that $g_x(t) \in L^1$ (because $\sin(\pi t)$ and t are basically constantly related near 0), so by Riemann-Lebesgue (specifically Corollary 99) this goes to zero as $t \to 0$. And for the second integral, we have

$$B = \int_{|t|>\delta} \frac{f(x-t) - f(x)}{\sin(\pi t)} \sin(\pi(2N+1)t)dt,$$

and this also goes to zero by Riemann-Lebesgue because for $|t| > \delta$, $\frac{f(x-t)-f(x)}{\sin(\pi t)}$ is not even singular (it's bounded away from 0) and thus is integrable. $\square$

Corollary 101

For any $\delta \in (0, \frac{1}{2})$, we have $\int_{-\delta}^{\delta} D_N(t)dt \to 1$ as $N \to \infty$.

Proof. The function $f = 1\{|t| \le \delta\}$ satisfies the Dini criterion at $x = 0$, and $f(0) = 1$. $\square$

In particular, this shows that all of the oscillations cancel between the terms in the Dini kernel except at $t = 0$.

Theorem 102 (Jordan's criterion)

Let f be a function of bounded variation. Then $S_N f(x)$ converges to $\frac{1}{2}(f(x^+) + f(x^-))$ for all x.

(This statement is actually a bit misleading – if we have a jump in our function f at some point x, then representing the jump as a Fourier series will actually have spikes in the **opposite directions** from the jump which get closer and closer to the point of discontinuity but do not decrease in height. This is called the **Gibbs phenomenon**, and it's why Fourier series are not used in approximations for things like photos.)

Proof. Without loss of generality, we may set $x = 0$ and say that $f(x)$ is increasing near $x = 0$. We can then write

$$S_N f(0) = \int_0^{1/2} [(f(t) + f(-t)] D_N(t) dt = \int_0^{1/2} f(t) D_N(t) + \int_0^{1/2} f(-t) D_N(t)$$

(here we use that the Dini kernel is even so $D(x) = D(-x)$). Denote these two terms by A and B. Since f is increasing, it has left and right limits, so we can say that $f(0^+) \le f(t) \le f(0^+) + \varepsilon$ for all $0 \le t \le \delta_\varepsilon$. We then see that

$$A = \int_0^{1/2} f(t) D_N(t) dt = \int_0^{\delta_\varepsilon} f(t) D_N(t) + \int_{\delta_\varepsilon}^{1/2} f(t) D_N(t).$$

For any fixed ε, the second term goes to zero because we've cut off the singularity and then can use Riemann-Lebesgue after moving the $\sin(\pi t)$ to $f(t)$ like before. We can now check (exercise) that if h is increasing and ϕ is continuous, then $\int_a^b h(x)\phi(x)dx = h(a^+) \int_a^c \phi(x)dx + h(b^-) \int_c^b \phi(x)dx$ for some intermediate c. So the first term of A is

$$f(0^+) \int_0^{c_\varepsilon} D_N(t)dt + f(\delta_\varepsilon^-) \int_{c_\varepsilon}^{\delta_\varepsilon} D_N(t)dt = f(0^+) \int_0^{\delta_\varepsilon} D_N(t)dt + ((f(\delta_\varepsilon^-) - f(0^+)) \int_{c_\varepsilon}^{d_\varepsilon} D_N(t)dt.$$

So as $N \to \infty$ and ε is fixed, the first term goes to $\frac{1}{2} f(0^+)$ (because we know the integral from $-\delta_\varepsilon$ to δ_ε of $D_N(t)$ goes to 1, and we're taking hal fof that). And now $|\int_a^b D_N(t)dt|$ is bounded by a constant C for all N (exercise again), so the second term goes to 0 by making ε sufficiently small. (So taking ε small to make the second term small, then taking N large to make the other terms large, gives us the result.) Similarly the $f(0^-)$ term comes from $\tilde{T}_N$. ☐

Theorem 103 (Localization principle)
Suppose $f \in L^1$ with $f(x) = 0$ for all $x \in (-\delta, \delta)$. Then $S_N f(0) \to 0$.

Proof. Since $S_N f(0) = \int f(-t) D_N(t) dt$ and f vanishes near 0, we can rewrite as $\int \frac{f(-t)1\{|t| \ge \delta\}}{\sin(\pi t)} \sin((2N+1)\pi t)dt$, which goes to zero by Riemann-Lebesgue. ☐

This is an interesting result – it tells us that changing f away from 0 does not change the value of the Fourier series, so the Fourier series is global in its definition but local in the sense that the Fourier series only depends on an interval around x.

On the other hand, not all Fourier series will converge pointwise:

Theorem 104 (Du Bois-Raymond)
There is a continuous function $f \in C(S^1)$ whose Fourier series diverges at $x = 0$.

There are various examples – the idea is to build a function that is continuous but with lots of oscillations "separated in frequency." And we'll use a completely different result to prove it:

Theorem 105 (Banach-Steinhaus)
Let X be a Banach space, and let T_α be a collection of bounded linear operators on $X \to Y$. Then either $\sup_\alpha ||T_\alpha|| < \infty$, or there is some $x_0 \in X$ such that $\sup_\alpha ||T_\alpha(x_0)|| = \infty$.

In other words, we can notice that the norms of the operators can blow up by looking at some point x_0.

Proof. Fix α, and define $\phi_\alpha = ||T_\alpha x||$. Each ϕ_α is continuous as a function $X \to \mathbb{R}$, and we can define $\phi(x) = \sup_\alpha \phi_\alpha(x)$. Consider the set $V_n = \{\phi(x) > n\}$, which is the union $\bigcup_\alpha \{\phi_\alpha(x) > n\}$. Each $\{\phi_\alpha(x) > n\}$ is open, so V_n is open. So there are two possibilities: in the first case, there is some N such that V_N is not dense, meaning there is a ball with $B(x_0, r) \cap V_N = \varnothing$. But then $||T_\alpha(x_0 + z)|| \le N$ for all $||z|| \le r$, so $||T_\alpha(z)|| \le N + ||T_\alpha(x_0)|| \le 2N$, meaning that each $||T_\alpha||$ has norm at most $\frac{2N}{r}$ and thus the operator norms are bounded. Otherwise, all V_N are dense, meaning that by Baire category theorem $\bigcap V_N$ is nonempty. Thus there is some x_0 with $\sup \phi_\alpha(x_0) = \infty$. ☐

Proof of Theorem 104. The operator $S_N f(0)$ maps continuous functions to $\mathbb{R}$ via $f \mapsto \int f(x-t) D_N(t) dt$. Then $|S_N f(0)| \le ||f||_{C(S^1)} ||D_N||_{L^1}$, and in fact the norms $||S_N||_{C[0,1] \to \mathbb{R}}$ and $||D_N||_{L^1}$ are equal (because we can take f to basically be an approximation of $\text{sgn}(D_N(t))$ since N is finite, and then the supremum norm of $S_N f$ is 1 and the L^1 norm of D_N is basically the integral of D_N, which is 1). But then $||D_N||_{L^1}$ blows up as $O(\log N)$, so the first case of Banach-Steinhaus does not hold. Thus there is some f where $\sup |S_N f(0)| = \infty$, meaning the Fourier series does not converge. ☐

The point is that the Dini kernel has some problems because of its oscillations, and there is a well-established way to get rid of those oscillations. This is related to the idea of summing the divergent series $1 - 1 + 1 - 1 + \cdots$; we can either taking the limit of $1 + r + r^2 + \cdots = \frac{1}{1-r}$ as $r \to -1$ or we can look at the **Cesaro** sums $C_N = \frac{S_1 + \cdots + S_N}{N}$ (that is, averaging the partial sums $1, 0, 1, 0, \cdots$, which also converges to $\frac{1}{2}$). This latter approach turns out to work well if we try taking Cesaro averages of the Fourier series

$$\frac{S_1 f(x) + \cdots + S_N f(x)}{N} = \int_{-1/2}^{1/2} f(x-t) \frac{1}{N} \sum_{k=-N}^{N} D_k(t) dt.$$

By working with some trigonometric identities, it turns out that the **Fejer kernel**

$$F_N(t) = \frac{1}{N+1} \sum_{k=0}^{N} D_k(t) = \frac{1}{N+1} \frac{\sin^2(\pi(N+1)t)}{\sin^2(\pi t)}$$

is much more well-behaved – $\int_{-1/2}^{1/2} f(x-t) F_N(t) dt$ is now convolution of f against a **nonnegative** function, and it's also a trigonometric polynomial because D_k is. And we'll see that it does, and while it's not the Fourier series, it's still a good approximation of f by trigonometric polynomials.

15 December 1, 2022

We'll start today by going into more detail about Cesaro summmation and the Fejer kernel. Recall that the Fourier series on a torus are defined via

$$f_k = \int_0^1 f(x) e^{-2\pi i k x} dx,$$

and we found that the partial sums $S_N f = \sum_{k=-N}^{N} f_k e^{2\pi i k x}$ can be written as convolution of f with the Dini kernel D_N. This kernel has the nice property that $\int_{-1/2}^{1/2} D_N(t) dt = 1$, but unfortunately its oscillation means that its L^1 norm grows as $\log N$, and thus there is a continuous function whose Fourier series diverges at a point. So instead, we now consider

$$\sigma_N f(x) = \frac{1}{N+1} \sum_{k=0}^{n} S_k f(x),$$

which we can write again as convolution over the whole circle $\int_{-1/2}^{1/2} F_N(t)f(x-t)dt$, where this time we're convolving with the Fejer kernel

$$F_N(t) = \frac{1}{N+1}\frac{\sin^2(\pi(N+1)t)}{\sin^2(\pi t)}.$$

This is a better kernel because we still have $\int_{-1/2}^{1/2} F_N(t)dt = 1$ (it's an average of $D_0, D_1, \cdots, D_N$), but F_N is nonnegative and goes to zero as $N \to \infty$ uniformly on any interval $|\delta| < t < 1/2$ for $\delta > 0$ (that is, any interval not containing 1). Therefore $\int_{\delta<|t|<1/2} F_N(t)dt \to 0$ as $N \to \infty$, giving us the following result:

Theorem 106

If $f \in L^p(S^1)$ for some $1 \le p < \infty$, then $\lim_{N\to\infty} ||\sigma_N f - f||_p = 0$. Additionally, if $f \in C(S^1)$, then $\lim_{N\to\infty} ||\sigma_N f - f||_{C(S^1)} = 0$.

(Here, the $C(S^1)$ norm is the supremum norm.)

Proof. We may write (using that $\int F_N(t)dt = 1$)

$$\sigma_N f(x) - f(x) = \int_{-1/2}^{1/2} F_N(t)[f(x-t) - f(x)]dt.$$

By the Minkowski inequality, we then have

$$||\sigma_N f - f||_p \le \int_{-1/2}^{1/2} F_N(t)||f(\cdot - t), f(\cdot)||_{L^p} dt.$$

We can now split up this integral into the parts where $|t| < \delta$ and $\delta < |t| < 1/2$. The latter of those terms is bounded by $2||f||_p \int_{\delta<|t|<1/2} F_N(t)dt$, which we've already shown goes to 0 as $N \to \infty$ by uniform convergence. And for the first term, we use the fact that $||f(\cdot - t) - f(t)||_{L^p} = 0$ for any function $f \in L^p$ (as we've proven in various situations). Thus the first term is at most $\int_{|t|<\delta_\varepsilon} F_N(t)dt \cdot \varepsilon < \varepsilon$ for all N (and here notice that it's important we're using F_N instead of D_N, because the latter would have required an absolute value in the integrand and we'd grow with N). So basically we choose δ small enough to make the first term small, then choose N to make the second term small. The continuous norm case is basically the same argument verbatim. $\square$

Corollary 107

Trigonometric polynomials are dense in $L^p(S^1)$ and $C(S^1)$ for any $1 \le p < \infty$.

Corollary 108 (Parseval's identity)

The Fourier transform is an isometry from L^2 to ℓ^2.

Proof. If we define $e_k(x) = e^{2\pi i k x}$ for each integer k, then the span of all e_k are dense in $L^2([0,1])$ (we don't need to impose periodicity at $x = 0$ because we're defined up to a set of measure zero anyway), so we get a basis for L^2. Then for any f, we have $f = \sum_k \langle f, e_k\rangle e_k$, meaning that

$$||f||_{L^2} = \sum |\langle f, e_k\rangle|^2 = \sum |f_k|^2$$

as desired. $\square$

Corollary 109

For all $f \in L^2([0,1])$, we have $S_N f \to f$ in $L^2[0,1])$.

In other words, the Fourier series always converges in L^2.

Proof. We have
$$\langle S_N f - f, S_N f - f\rangle = \langle S_N f, S_N f\rangle - 2\langle f, S_N f\rangle + ||f||^2$$
and now by orthogonality of the basis $\langle f, S_N f\rangle = \langle S_N f, S_N f\rangle$ (the subsequent terms are all zero after the Nth partial sum) and thus this is
$$= ||f||^2 - 2\langle S_N f, S_N f\rangle + ||f||^2 = 2(||f||^2 - ||S_N f||^2) = 2\sum_{k>n} ||f_k||^2,$$
which goes to zero because we're taking a tail sum. □

Corollary 110

If $f \in L^1([0,1])$, and $f_k = 0$ for all k, then $f = 0$.

Proof. If all Fourier series are zero, then all Cesaro sums are zero, and thus $\sigma_N f = 0$ for all f meaning $||f||_1 = 0$. □

We can apply these results to **ergodicity of irrational rotations**: for any irrational α, consider the map $T_\alpha(x) = (x+\alpha) \bmod 1$ on $[0,1]$ (or equivalentl S^1). We may ask if there are any invariant sets for T_α – suppose that R is such a set. Then if we define $f(x) = 1\{x \in R\}$, then $f(x+\alpha) = f(x)$ by definition. But then the Fourier coefficient f_k is
$$f_k = \int f(x+\alpha)e^{-2\pi i k x}dx = \int f(x)e^{-2\pi i k(x-\alpha)} = e^{2\pi i k\alpha} f_k,$$
and α is irrational so $e^{2\pi i k\alpha}$ is not 1 whenever $k \neq 0$, so $f_k = 0$ for all nonzero k. By the above corollary, that means f_k is constant, so $\mu(R)$ is either 1 (if $f = 1$ almost everywhere) or 0 (if $f = 0$ almost everywhere).

Remark 111. *However, there are invariant sets that are not just the empty set or the whole set – for example, the orbit of any single point will be countable, thus of measure zero, and it will be invariant under T_α.*

With that, we're now ready to talk about Fourier **transforms**:

Definition 112

For any $f \in L^1(\mathbb{R}^n)$, define the **Fourier transform**
$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}dx.$$

For any ξ, we have $|\hat{f}(\xi)| \le ||f||_{L^1}$ by the triangle inequality for integrals, so $||\hat{f}||_{L^\infty} \le ||f||_{L^1}$. Also,
$$|\hat{f}(\xi) - \hat{f}(\xi')| \le \int_{\mathbb{R}^n} |f(x)||e^{-2\pi i x\xi} - e^{-2\pi i x\xi'}|dx,$$
and the integrand is bounded by $2|f(x)|$ and converges pointwise, so it goes to zero as $\xi' \to \xi$ by the dominated convergence theorem. Thus $\hat{f}$ is continuous (in other words, the Fourier transform is a bounded map into continuous functions).

One of the most useful properties of the Fourier transform is algebraic, and to understand that we must define a class of functions called the **Schwartz class**. For some motivation, notice that if $f \in C_c^\infty(\mathbb{R}^n)$, then we can write the Fourier transform of its derivative

$$\widehat{\frac{\partial f}{\partial x_k}}(\xi) = \int \frac{\partial f}{\partial x_k}(x) e^{-2\pi i \xi \cdot x} dx = 2\pi i \xi_k \hat{f}(\xi)$$

by integration by parts (the negative sign from integration by parts and the exponent cancel out). On the other hand, the integration by parts in the reverse direction yields

$$\widehat{x_k f}(\xi) = \int x_k f(x) e^{-2\pi i \xi x} dx = \int f(x) \left(-\frac{1}{2\pi i}\right) \frac{\partial}{\partial \xi_k} e^{-2\pi i \xi \cdot x}$$

and because f behaves nicely we can pull the derivative out of the integral to find that

$$\widehat{x_k f}(\xi) = -\frac{1}{2\pi i} \frac{\partial}{\partial \xi_k} \hat{f}(\xi).$$

So there is a duality between multiplication in one space (regular function space and Fourier space) and differentiation in the other, and we want to make sure that we can repeatedly apply these operations without any issues.

Definition 113

A function f is in the **Schwartz class** $\mathcal{S}(\mathbb{R}^n)$ if for any multi-indices α, β , we have $|x|^\alpha \left|\frac{\partial^\beta f}{\partial x^\beta}\right|$ bounded by some constant $p_{\alpha\beta}(f)$ (more precise definition below).

In other words, all derivatives of f must decay faster than any polynomial in the Schwartz class – more properly, we actually define

$$p_{\alpha\beta}(f) = \sup_x \langle x \rangle^\alpha \left|\frac{\partial^\beta f}{\partial x^\beta}\right|, \quad \langle x \rangle^\alpha = (1 + x_1^2)^{\alpha_1/2} \cdots (1 + x_n^2)^{\alpha_n/2}$$

to avoid having issues with singularity at $x = 0$. (And this is not a normed space, but we can define a topology on it.) We say that $f_n \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ if for all α, β fixed, $p_{\alpha\beta}(f_n) \to 0$ as $n \to \infty$. And it turns out that f_n converge to zero in the Schwartz class if and only if the Fourier transforms $\hat{f}_n$ converge to zero, so the Fourier transform $\mathcal{F}$ is actually a continuous linear operator $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$.

Theorem 114

For all $f, g \in \mathcal{S}(\mathbb{R}^n)$, we have $\int_{\mathbb{R}^n} f(x)\hat{g}(x)dx = \int_{\mathbb{R}^n} \hat{f}(x)g(x)dx$ and $f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi)e^{2\pi i \xi \cdot x}d\xi$.

The first statement here is another statement of Parseval, and the second is the Fourier inversion theorem.

Proof. By Fubini's theorem,

$$\int f(x)\hat{g}(x)dx = \int \int f(x)g(\xi)e^{2\pi i \xi \cdot x} dx d\xi$$

and now because all of these functions decay we are integrable and we can change the order of integration to

$$= \int \int f(x)g(\xi)e^{2\pi i \xi \cdot x} d\xi dx = \int \hat{f}(x)g(x)dx.$$

The second result is more interesting – we start by proving it for just a single function:

Lemma 115

If $f(x) = e^{-\pi |x|^2}$, then $\hat{f}(\xi) = e^{-\pi |\xi|^2}$.

Proof of lemma. Since f is the product of the functions $e^{-\pi x_i^2}$ for $1 \le i \le n$, it suffices to prove this for $n = 1$. We know that $\hat{f}(0) = \int_{\mathbb{R}} f(x)dx = 1$ by multivariable calculus arguments, and $\frac{\partial f}{\partial x} = -2\pi x f$, meaning that we have the differential equation

$$f'(x) = 2\pi x f = 0, \quad f(0) = 1.$$

Taking the Fourier transform of this equation, we see that

$$2\pi i \xi f \hat{f}(\xi) + (2\pi)\left(-\frac{1}{2\pi i}\right)\frac{d\hat{f}(\xi)}{\partial \xi} = 0 \implies 2\pi i \xi \hat{f}(\xi) + \frac{\partial \hat{f}}{\partial \xi} = 0,$$

and we also know that $\hat{f}(0) = 1$. Thus the first-order ODEs for f and $\hat{f}$ are identical, including the initial condition, and thus f and $\hat{f}$ are the same function. $\square$

(The lemma can also be proved with direct computation, but that does require some calculation with contour integrals.) Returning to the proof, we essentially want to choose $\hat{g}$ to be a delta function, but that requires some work with distributions and so on, and the idea is to scale g so that it looks more and more peaked. We know that

$$\int f(x)\hat{g}(\lambda x)dx = \int\int f(x)g(\xi)e^{-2\pi i \xi \lambda x}d\xi dx$$

again by the same trick, and now we can integrate in x first to get

$$= \int \hat{f}(\lambda \xi)g(\xi)d\xi = \frac{1}{\lambda^n}\int \hat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)d\xi.$$

Multiplying by λ^n and making λx our new variable, we see that

$$\int f\left(\frac{x}{\lambda}\right)\hat{g}(x)dx = \int \hat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)d\xi.$$

Taking $\lambda \to \infty$, because f is continuous and $\hat{g}$ is integrable the left-hand side becomes $f(0)\int \hat{g}(x)dx$, and similarly the right-hand side becomes $\hat{g}(0)\int \hat{f}(\xi)d\xi$. Moving all the terms involving f to one side and those involving g to the other, this means that $\frac{\int \hat{f}(\xi)}{f(0)}$ is a constant independent of f, and if we plug in $f = e^{-\pi|x|^2}$, we know that $f(0) = 1$ and $\int \hat{f}(x) = \int f(x)dx = 1$, so $f(0) = \hat{f}(\xi)d\xi$ for any function f. Finally, if we define $f_y(x) = f(x+y)$, then $f_y(0) = f(y)$ and $\hat{f}_y(\xi) = e^{-2\pi i \xi y}f(\xi)$, so replacing f with f_y above gives us the desired inversion formula. $\square$

Since the characteristic function of a random variable is the Fourier transform of its associated probability measure, we can prove the central limit theorem and law of large numbers using Fourier transforms. But that'll be left to a probability theory class, and here our next step is to try to extend the Fourier transform beyond the definition on L^1 or L^2. The idea is to try to define actions on the "intermediate" spaces in between two L^p spaces if we have an operator defined on the extremes, and this is the **Riesz-Thorin interpolation theorem**.

Example 116

Suppose $f \in L^{p_0}(\mathbb{R}^n)$ and $f \in L^{p-1}(\mathbb{R}^n)$, and consider some $p_0 \le p \le p_1$. Then we can write $p = (1-t)p_0 + tp_1$ for some $t \in [0, 1]$, meaning that

$$\int |f|^p d\mu = \int |f|^{(1-t)p_0 + tp_1} \le \left(\int |f|^{(1-t)\alpha p_0}\right)^{1/\alpha}\left(\int |f|^{tp_1\beta}\right)^{1/\beta}$$

by Holder's inequality whenever $\frac{1}{\alpha} + \frac{1}{\beta} = 1$.

The point is that we want to apply this when $\frac{1}{\alpha} = 1 - t$, which automatically makes $\frac{1}{\beta} = t$, because we then have an inequality of the form

$$\int |f|^p d\mu \le \left(\int |f|^{p_0}\right)^{1-t}\left(\int |f|^{p_1}\right)^t \implies \boxed{||f||_p^p \le ||f||_{p_0}^{(1-t)p_0}||f||_{p_1}^{tp_1}}$$

So we can bound a function's L^p norm by its p_0 and p_1 norms. So now we may ask a more general question: if $A : L^{p_0}(M,\mu) \to L^{q_0}(N,\nu)$ and $A : L^{p_1}(M,\mu) \to L^{q_1}(N,\nu)$, and we have some p with $p_0 \le p \le p_1$, we may want to ask if A maps $L^p(M,\mu)$ to some $L^q(N,\nu)$ with $q_0 \le q \le q_1$. Specifically, we wish to get a bound of the form

$$||A||_{L^p\to L^q} \stackrel{?}{\to} \le ||A||^s_{L^{p_0}\to L^{q_0}}||A||^t_{L^{p_1}\to L^{q_1}}$$

for some powers s,t depending on the ps and qs. First of all, by scaling of A we know that we must have $s+t=1$ (or else we could multiply A by some large or small λ), but there is also a scaling invariance associated with the "dimension" of the measures μ and ν. If we assume f has no dimension, which we denote $[f]=1$, then the dimension of the L^p norm of our function is

$$\left[||f||_{L^p(M,\mu)}\right] = \left[\int |f(x)|^p d\mu\right]^{1/p} = [\mu]^{1/p},$$

and thus the dimension of A has units of $\frac{||A||_{L^q}}{||A||_{L^p}} = \frac{[\nu]^{1/q}}{[\mu]^{1/p}}$. Thus we must actually have

$$\frac{[\nu]^{1/q}}{[\mu]^{1/p}} = \left(\frac{[\nu]^{1/q_0}}{[\mu]^{1/p_0}}\right)^s\left(\frac{[\nu]^{1/q_1}}{[\mu]^{1/p_0}}\right)^t,$$

Thus we must actually have

$$\frac{1}{q} = \frac{s}{q_0} + \frac{t}{q_1}, \quad \frac{1}{p} = \frac{s}{p_0} + \frac{t}{p_1}, \quad s+t=1.$$

In other words, we can only hope to have an interpolation result if $\frac{1}{p} = \frac{1-t}{p_0} + \frac{t}{p_1}$ and $\frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}$ – for any value of p we can determine t and there is only one possible value of q for which we can get a bound on the operator. And it turns out that such a bound does always hold, and in fact there is no additional constant required:

Theorem 117 (Riesz-Thorin)
Let A be a bounded operator $L^{p_0}(M,\mu) \to L^{q_0}(N,\nu)$ of norm k_0 and also a bounded operator $L^{p_1}(M,\mu) \to L^{q_1}(N,\nu)$ of norm k_1. Let p be such that $\frac{1}{p} = \frac{1-t}{p_0} + \frac{t}{p_1}$ and q such that $\frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}$ for some $00 \le t \le 1$. Then A is a bounded operator with

$$||A||_{L^p\to L^q} \le k_0^{1-t}k_1^t.$$

(And indeed, when $t=0$ this bound is k_0, and when $t=1$ this bound is k_1, so this is the best we can hope the inequality to be.) We'll discuss some corollaries and the proof next time – even though the statement is functional analysis, we'll prove it using complex analysis.

16 December 6, 2022

We'll continue discussing Riesz-Thorin interpolation today – recall the motivation that any function $f \in L^{p_1} \cap L^{p_2}$ is also in L^p for all $p_1 \le p \le p_2$ by Holder's inequality. We then want to see whether an operator $A : L^{p_0}(M,\mu) \to L^{q_0}(N,\nu)$ which also maps $L^{p_1}(M,\mu) \to L^{q_1}(N,\nu)$ also maps $L^p \to L^q$ for $p \in [p_0,p_1]$ and $q \in [q_0,q_1]$. By some dimensional analysis arguments, we showed last time that this could only hold if $\frac{1}{q_t} = \frac{1-t}{q_0} + \frac{t}{q_1}$ and $\frac{1}{p_t} = \frac{1-t}{p_0} = \frac{t}{p_1}$ for some $0 \le t \le 1$, and Theorem 117 asserts that in fact this does happen. Specifically, if $A : L^{p_0}(M,\mu) \to L^{q_0}(N,\nu)$ has norm k_0 and $A : L^{p_1}(M,\mu) \to L^{q_1}(N,\nu)$ has norm k_1, then $A : L^{p_t}(M,\mu) \to L^{q_t}(N,\nu)$ will have norm $k_0^{1-t}k_1^t$.

Corollary 118

The Fourier transform maps $L^p(\mathbb{R}^n) \to L^{p'}(\mathbb{R}^n)$, where $\frac{1}{p} + \frac{1}{p'} = 1$, for all $1 \le p \le 2$.

Proof. We know $\mathcal{F}$ maps $L^1 \to L^\infty$, since $||\hat{f}||_{L^\infty} \le \int |f(x)|dx = ||f||_{L^1}$ by definition. Also, $\mathcal{F}$ maps $L^2 \to L^2$ – in fact it is an isometry on L^2 by Parseval's identity. So $p_0 = 1, p_1 = 2, q_0 = \infty, q_1 = 2$, meaning that

$$\frac{1}{p_t} = \frac{1-t}{1} + \frac{t}{2}, \quad \frac{1}{q_t} = \frac{1-t}{\infty} + \frac{t}{2}$$

and we see that these indeed always add to 1 and p_t ranges from 1 to 2, meaning Riesz-Thorin applies automatically. $\square$

Corollary 119

Suppose $f \in L^p(\mathbb{R}^n)$, and $g \in L^q(\mathbb{R}^n)$. Then $f * g \in L^r(\mathbb{R}^n)$, where $\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}$.

Proof. Notice that

$$||f * g||_{L^1} \le \int |f(x)||g(x-y)|dydx = \|f||_{L^1}||g||_{L^1},$$

and on the other hand

$$||f * g||_{L^\infty} \le \sup_x \left| \int f(y)g(x-y)dy \right| \le ||f||_{L^\infty}||g||_{L^1}.$$

So for any $g \in L^1$, the map $f \mapsto f * g$ is bounded in norm by $||g||_1$ as a map $L^1 \to L^1$ and also as a map $L^\infty \to L^\infty$. Thus by Riesz-Thorin, this means that (notice $p_t = q_t$ for all t here) $\boxed{||f * g||_{L^p} \le ||g||_{L^1}||f||_{L^p}}$, which proves the case $q = 1$.

For the general case, notice that

$$\boxed{||f * g||_{L^\infty} \le ||f||_{L^p}||g||_{L^{p'}}}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$, by applying Holder's inequality on $\int f(y)g(x-y)dy$ for some fixed x. So $g \mapsto f * g$ is bounded as a map $L^{p'} \to L^\infty$ and as a map $L^1 \to L^p$, with both norms given by $||f||_p$. Thus by Riesz-Thorin again, for

$$\frac{1}{p_t} = \frac{1-t}{p'} + \frac{t}{1}, \quad \frac{1}{q_t} = \frac{1-t}{\infty} + \frac{t}{p},$$

we have $||f * g||_{L^{q_t}} \le ||f||_{L^p}||g||_{L^{p_t}}$. Setting $p_t = r$ and $q_t = s$, we find $\frac{1}{r} = \frac{1-t}{p'} + t = (1-t) - \frac{1-t}{p} + t = 1 - \frac{1-t}{p} = 1 - \frac{1}{p} + \frac{1}{s}$ (by definition of p'), so we indeed have $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. $\square$

The point is that in practice, estimating L^1, L^2, L^∞ norms are the easiest to do, and the rest are a lot trickier. But this gives us a tool to analyze the spaces in between.

For proving Riesz-Thorin, we'll use a complex analytic fact:

Theorem 120 (Three lines theorem)

Let $F(z)$ be a bounded analytic function in the infinite strip $S = \{0 \le \text{Re}(z) \le 1\}$, and suppose that $|F(iy)| \le m_0$ and $|F(1+iy)| \le m_1$ for all $y \in \mathbb{R}$ (with $m_0, m_1 > 0$). Then $|F(x+iy)| \le m_0^{1-x}m_1^x$ for all $x + iy \in S$.

Basically, if our analytic function is bounded on the left and right line, then it's bounded in the middle in an interpolated way.

Proof. Define $F(z) = m_0^{1-z}m_1^zG(z)$, so that (letting $z+x+iy$) $|G(z)| = \frac{|F(z)|}{m_0^{1-x}m_1^x}$, meaning $|G(iy)| \le 1$ and $|G(1+iy)| \le 1$ for all y. Then we just need to show that $|G(z)| \le 1$ for all $z \in S$. This looks a lot like the maximum modulus principle, but that result only applies to bounded domains.

So instead, first assume that $G(z) \to 0$ as $y \to \infty$, uniformly in x. If this happens, then we can find some large enough R such that we can restrict to the rectangle where $|\text{Im}(z)| \le R$, and then the maximum modulus principle tells us G is less than 1 inside the rectangle, and above or below this rectangle uniform convergence to zero already tells us the result. Otherwise, define $G_n(z) = G(z)e^{(z^2-1)/n}$. Then

$$|G_n(iy)| \le e^{(-y^2-1)/n} \le 1, \quad |G_n(1+iy)| \le e^{-(1-y^2-1)/n} \le 1,$$

and $|G_n(x+iy)|$ is bounded by a constant because F was bounded by a constant C and m_0^{1-x}, m_1^x are also bounded by constants from below. More specifically

$$|G_n(x+iy)| \le Ce^{(x^2-1-y^2)/n} \le Ce^{-y^2/n},$$

which goes to zero as $|y| \to \infty$, and thus by the previous case $|G_n(z)| \le 1$ for all n. Since $G_n(z) \to G(z)$ for each z, this means $|G(z)| \le 1$ as well, as desired. $\square$

Proof of Theorem 117. We first want to show that A maps L^{p_t} into L^{q_t}. For any $f \in L^{p_t}(M, \mu)$, we can write

$$f = f_0 + f_1 = f1\{|f(x)| \ge 1\} + f1\{|f(x)| < 1\}.$$

We know that $|f_0|^{p_0} \le |f_0|^{p_1}$ (because both sides are nonzero only when $f \ge 1$), so $\int |f_0|^{p_0} \le \int |f_0|^{p_1} \le \int |f|^{p_1}$, which is finite. Similarly we see that $\int |f_1|^{p_1} \le \int |f_1|^{p_0} \le \int |f|^{p_0} < \infty$. This means that any function $f \in L^{p_t}$ can be written as a sum of a function in L^{p_0} and a function on L^{p_1}, and we can define $Af = Af_0 + Af_1$. This makes A defined on $L^{p_0} \cap L^{p_1}$, which is dense in L^{p_t}.

Now recall by the Riesz representation theorem (this is the other Riesz from the one in Riesz-Thorin) that the operator L_f defined by $L_f(g) = \int fg d\mu$ is an operator $L^{q'} \to \mathbb{R}$ with norm $||f||_{L^q}$. The definition of the operator norm of A can thus be thought of as

$$||A||_{L^p(\mu) \to L^q(N)} = \sup_{||f||_{L^p}=1} ||Af||_{L^q} = \sup_{\substack{||f||_{L^p(M)}=1 \\ ||g||_{L^{q'}(N)}=1}} \int (Af) g d\nu.$$

Since simple functions are dense in $L^p(M)$ and $L^{q'}(N)$ (notice that we only care about the "intermediate" spaces, so we can assume p, q' are not ∞), we can just take the supremum over simple functions f and g of appropriate norm 1. Letting $a_j, \alpha_j, b_j, \beta_j$ be arbitrary real numbers, we can write

$$f(x) = \sum_{j=1}^n a_j e^{i\alpha_j} 1\{x \in A_j\}, \quad g(y) = \sum_{j=1}^m b_j e^{i\beta_j} 1\{y \in B_j\}.$$

If we then define $\frac{1}{p(\zeta)} = \frac{1-\zeta}{p_0} + \frac{\zeta}{p_1}$ and $\frac{1}{q(\zeta)} = \frac{1-\zeta}{q_0} + \frac{\zeta}{q_1}$, then consider the function

$$u(x, \zeta) = \sum a_j^{p_t/p(\zeta)} e^{i\alpha_j} 1\{x \in A_j\}, \quad v(y, \zeta) = \sum b_j^{q_t/q(\zeta)} e^{i\beta_j} 1\{x \in B_j\}.$$

Then $u(x, t) = f(x)$ and $v(y, t) = g(x)$ (because $p(\zeta) = p_t$ and $q(\zeta) = q_t$ by definition), and both u and v are a sum of things that are bounded and supported in finite measure, so they're in all L^p spaces. Thus we can define

$$F(\zeta) = \int Au(y, \zeta) v(y, \zeta) d\nu(y).$$

We are really only interested in this function when $\zeta = t$, because then we are in fact computing $\int (Af) g d\nu$. But we can look at the whole strip $\text{Re}(\zeta) \in [0,1]$ and notice that

$$F(\zeta) = \sum_{j=1}^{n} \sum_{k=1}^{m} a_j^{p_t/p(\zeta)} b_j^{q_t/q(\zeta)} \int_N (A\psi_j)(y) e^{i\beta_k(y)} 1\{y \in B_k\} dy,$$

where we define $\psi_j(x) = e^{i\alpha_j} 1\{x \in A_j\}$. And the point is that

$$|a_j^{p_t/(p(\zeta)}| = \left| a_j^{p_t\xi/p_1 + p_t(1-\xi)/p_0} \right|,$$

which is uniformly bounded over the range of ζs (but depending on a_j) by some constant C. Then

$$||u(\cdot, i\xi)||_{L^{p_0}} = \left(\int_M \sum_{j=1}^{m} \left| a_j^{p_t(i\xi)/p_1 + p_t(1-i\xi)/p_0} \right|^{p_0} 1\{x \in A_j\} d\mu \right)^{1/p_0},$$

which is just

$$= \left(\sum_{j=1}^{n} \int_M |a_j|^{p_t} 1\{x \in A_j\} d\mu \right)^{1/p_0} = ||f||_{L^{p_t}}^{p_t/p_0} = 1,$$

since $||f||_{L^{p_t}} = 1$ (from the definition of the simple function to begin with). Similarly $||v(\cdot, i\xi)||_{L^{q_0'}} = 1$, and that tells us that $|F(i\xi)| \le k_0$. (This was the key computation!) Similarly we can calculate and find that $||u(\cdot, 1+i\xi)||_{L^{p_1}} = ||v(\cdot, 1+i\xi)||_{L^{q_1'}} = 1$ so $|F(1+i\xi)| \le k_1$. So we have an analytic function in ζ and by the three-lines lemma we have $|F(x+iy)| \le k_0^{1-x} k_1^x$, and thus $|F(t)| \le k_0^{1-t} k_1^t$. Plugging this back in, we get the same bound for A itself. $\square$

There's another interpolation theorem due to Marcinkiewicz, where we can weaken the assumptions to **weak** L^1 bounds on the boundary $p_0 \to q_0$ and $p_1 \to q_1$ and still get **strong** bounds in the middle, but we lose some constants. (We might discuss this next lecture if we have time.)

We'll finish this class by discussing the **Hilbert transform**. There are two ways to define it:

Definition 121

For any $f \in \mathcal{S}(\mathbb{R}^n)$, we can define

$$u(x,t) = \int e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi$$

where $x \in \mathbb{R}^n$ and $t \ge 0$.

Notice that $u(x,0) = f(x)$. Furthermore, if we look at the Laplacian $\Delta_{x,t} u$, we can differentiate under the integral because the integral converges in every sense we care about, and we get $\Delta_{x,t} u = 0$ (because we get a $|\xi|^2$ from the t-derivative and a $-|\xi|^2$ from the x-derivative). In other words, any function f can be **extended to a harmonic function in the upper half-plane**.

Specializing to the case $n = 1$, a function in the upper half-plane has a corresponding conjugate harmonic function $\eta(x+it) = u(x+it) + iv(x+it)$ which is analytic and thus $\Delta_{x,t} v = 0$. If we then restrict v to $\{t = 0\}$, we get a function of x again, and that's what we call the **Hilbert transform**

$$Hf(x) = v(x,0).$$

In other words, extend from the real line to get a harmonic function, find the dual, and restrict again – we'll write down the formula for it next time. It sounds plausible that this is fine if f is Schwartz, and the question is how nice this operator actually is – it turns out to appear in fluid mechanics, and it's the most basic example of a singular integral operator, satisfying

$$\hat{Hf}(\xi) = \pm i\text{sgn}(\xi)\hat{f}(\xi).$$

So because of the discontinuity at zero, the Fourier transform is never going to give something in the Schwartz class because it won't decay fast enough. But $||Hf||_{L^2} = ||f||_{L^2}$ (by Parseval, since other than the sign the Fourier transforms are the same), so this is still a bounded operator. And by interpolation, even though we only have the single point $p = 2$, we can actually show that H is a bounded linear operator on L^p for all $1 < p < \infty$. This bound in fact extends to all singular operators, and we'll need the Marcinkiewicz theorem to do it more generally.

17 December 8, 2022

We'll continue our discussion of the Hilbert transform which we defined last time. Basically, we start with a Schwartz function $f \in \mathcal{S}(\mathbb{R})$ and define

$$u(x, t) = \int_{\mathbb{R}} e^{-2\pi t|\xi|} e^{2\pi i x\xi} \hat{f}(\xi) d\xi$$

for all $t > 0$. Then differentiating under the integral, we find that $\Delta_{x,t} u = 0$, so we have a solution to the Laplace equation in the upper half-plane with the prescribed boundary condition $u(x, 0) = f(x)$. But we know that $\hat{u}(\xi, t) = e^{-2\pi t|\xi|}\hat{f}(\xi)$, so on the Fourier side we are just multiplying by the function $e^{-2\pi t|\xi|}$ and thus we can write $u(x, t)$ as a convolution $P_t * f(x)$, where $\hat{P}_t(\xi) = e^{-2\pi t|\xi|}$ (since convolution is multiplication on the Fourier side). Specifically, by the inverse Fourier transform we have

$$P_t(x) = \int_{\mathbb{R}} e^{-2\pi t|\xi|+2\pi i\xi x} d\xi = \frac{1}{\pi}\frac{t}{x^2 + t^2} = \frac{1}{\pi t}\frac{1}{1 + \frac{x^2}{t^2}}.$$

Since $\frac{1}{\pi}\int \frac{dx}{1+x^2} = 1$, we see that $P_t(x)$ is in fact an approximation of identity as $t \to 0$, known as the **Poisson kernel**. And we can look at all of this in a slightly different way: write $z = x + it$ and $u(z) = \int e^{-2\pi t|\xi|+2\pi i x\xi}\hat{f}(\xi)d\xi$, so that (splitting up the integral into the two regions for ξ)

$$\begin{aligned}
u(z) &= \int_0^\infty e^{-2\pi t\xi+2\pi i x\xi}\hat{f}(\xi)d\xi + \int_0^\infty e^{2\pi t\xi+2\pi i x\xi}\hat{f}(\xi)d\xi \\
&= \int_0^\infty \hat{f}(\xi)e^{2\pi i\xi(x+it)} + \int_{-\infty}^0 \hat{f}(\xi)e^{2\pi i\xi(x-it)} \\
&= \int_0^\infty \hat{f}(\xi)e^{2\pi iz\xi} + \int_{-\infty}^0 \hat{f}(\xi)e^{2\pi i\overline{z}\xi}d\xi.
\end{aligned}$$

If we then define the function $v(z)$ via

$$iv(z) = \int_0^\infty \hat{f}(\xi)e^{2\pi iz\xi} - \int_{-\infty}^0 \hat{f}(\xi)e^{2\pi i\overline{z}\xi}d\xi$$

(just switching the sign for the second term), then $G(z) = u(z) + iv(z)$ is analytic in z (in the region $\text{Im}(z) > 0$) because the $\overline{z}$-dependence has gone away, as long as f is in the Schwartz class. And we see that $v(z)$ is real-valued, with

$$v(z) = \int (-i\text{sgn}(\xi))e^{-2\pi t|\xi|}e^{2\pi ix\xi}\hat{f}(\xi)d\xi.$$

We then have

$$\hat{v}(\xi, t) = (-i\text{sgn}(\xi))e^{-2\pi t|\xi|}\hat{f}(\xi),$$

so just like before we can define Q_t so that $\hat{Q}_t(\xi) = (-i\text{sgn}(\xi))e^{-2\pi t|\xi|}$ and say that $v(x, t) = Q_t * f(x)$. By Fourier inversion, we find that

$$Q_t(x) = \frac{1}{\pi}\frac{x}{t^2+x^2}.$$

This turns out to be totally different from the Poisson kernel – this is not even in L^1 as a function of x, corresponding to the discontinuous Fourier transform. So we're curious what happens to $v(x,t)$ as $t \to 0$. We can indeed integrate the expression for $v(z)$ and everything will be integrable, but $Q_t(x)$ converges to $\frac{1}{\pi x}$, so this in fact has something to do with principal values:

Definition 122

The principal value of $\frac{1}{x}$, denoted p.v.$\frac{1}{x}$, is a distribution defined by setting

$$\text{p.v.}\frac{1}{x}(\phi) = \lim_{\varepsilon\to 0}\int_{|x|>\varepsilon}\frac{\phi(x)}{x}dx.$$

for any Schwartz function ϕ.

To check that this is well-defined, we can write

$$\text{p.v.}\frac{1}{x}(\phi) = \lim_{\varepsilon\to 0}\int_{\varepsilon<|x|<1}\frac{\phi(x)-\phi(0)}{x}dx + \int_{|x|>1}\frac{\phi(x)dx}{x}$$

(since the integral of $\frac{\phi(0)}{x}$ is zero on the region $\varepsilon < |x| < 1$) and now in fact there is no problem having the integral include the origin in the first term as well:

$$\text{p.v.}\frac{1}{x}(\phi) = \int_{|x|<1}\frac{\phi(x)-\phi(0)}{x}dx + \int_{|x|>1}\frac{\phi(x)dx}{x}.$$

Proposition 123

We have

$$\frac{1}{\pi}\text{p.v.}\frac{1}{x}(\phi) = \lim_{t\to 0} Q_t * \phi(0) = \lim_{t\to 0}\int Q_t(x)\phi(x)dx$$

Proof. We know that

$$\frac{1}{\pi}\text{p.v.}\frac{1}{x}(\phi) = \lim_{t\to 0}\psi_t(x)\phi(x)dx,$$

where $\psi_t(x) = \frac{1}{x}1\{|x| \ge t\}$. But then we can write

$$\begin{aligned}
\frac{1}{\pi}\text{p.v.}\frac{1}{x}(\phi) - Q_t*\phi(0) &= \lim_{t\to 0}\int(\psi_t(x) - Q_t(x))\phi(x)dx \\
&= \lim_{t\to 0}\frac{1}{\pi}\int_{|x|>t}\frac{\phi(x)}{x}dx - \frac{1}{\pi}\int\frac{x}{t^2+x^2}\phi(x)dx \\
&= \lim_{t\to 0}\frac{1}{\pi}\int_{|x|<t}\frac{x\phi(x)}{t^2+x^2}dx + \lim_{t\to 0}\frac{1}{\pi}\int_{|x|>t}\left(\frac{1}{x} - \frac{x}{x^2+t^2}\right)\phi(x)dx.
\end{aligned}$$

For the first term, we can do a change of variables $x = ty$ to get

$$\int_{|x|<t}\frac{x\phi(x)}{t^2+x^2} = \int_{|y|<1}\frac{ty\phi(ty)}{t^2+t^2y^2}tdy = \int_{|y|<1}\frac{y\phi(ty)dy}{1+y^2},$$

which converges to $\int_{|y|<1}\frac{y\phi(0)dy}{1+y^2} = 0$ by the dominated convergence theorem. On the other hand,

$$\int_{|x|>t}\left(\frac{1}{x} - \frac{x}{x^2+t^2}\right)\phi(x)dx = \int_{|x|>t}\frac{t^2}{x(x^2+t^2)}\phi(x)dx$$

and now doing the same change-of-variables yields

$$= \int_{|y|>1} \frac{t^2\phi(ty)tdy}{t^3(1+y^2)} = \int_{|y|>1} \frac{\phi(ty)}{y(1+y^2)}dy$$

which is also zero by the dominated convergence theorem. So we indeed have the desired equality, and what we've basically used is that we have odd functions so their integrals over symmetric intervals is zero. $\square$

We can thus define the Hilbert transform this way:

Definition 124

For any Schwartz function f, we define the Hilbert transform

$$Hf(x) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{|y|>\varepsilon} \frac{f(x-y)}{y}dy = \frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}dy.$$

There are two reasons for studying this operator – one is that

$$Hf(x) = \lim_{t\to 0} Q_t * f(x),$$

so (as discussed last lecture) we take a function f, get a harmonic function u, compute the analytic function $G(z) = u(z) + iv(z)$ that is analytic, and then restrict $v(x,t)$ as $t \to 0$. But on the other hand, we now also have a formula for it coming from the principal value distribution, and it's a **singular integral operator**. Integral operators look like $Af(x) = \int g(x,y)f(y)dy$ for some kernel $g(x,y)$, and we usually require g to be well-behaved. But here our kernel is not integrable, and it's "barely" singular (on the order of $\frac{1}{y}$) and it is odd. So there is a lot of cancellation in the integral defining $Hf(x)$, giving us some hope to work with it and deduce some regularity. The Fourier transform of Hf is then

$$\widehat{Hf}(\xi) = (-i\text{sgn}(\xi))\hat{f}(\xi),$$

so by Parseval's identity we see that $||Hf||_{L^2} = ||f||_{L^2}$. Additionally, because $(-i\text{sgn}(\xi))^2 = -1$ we have $H^2 f = -f$, and we get the antisymmetry $(Hf, f) = -(f, Hf)$. And the big result is that we can get boundedness of the Hilbert transform not just on L^2:

Theorem 125

For any $1 < p < \infty$, we have $||Hf||_p \le C_p||f|_p$ for some constant p.

Proof. We wish to use Riesz-Thorin, but manifestly there is only one point for which we already have a bound. So the strategy is to consider the class of Schwartz functions vanishing around the origin:

$$\mathcal{S}_0 = \{f \in \mathcal{S}(\mathbb{R}^n) : \exists \varepsilon_0 > 0 \text{ such that } \hat{f}(\xi) = 0 \text{ for } |\xi| \le \varepsilon_0\}.$$

We see that H maps $\mathcal{S}_0 \to \mathcal{S}_0$ (because of the formula for $\widehat{Hf}$), and we claim $\mathcal{S}_0$ is dense in L^p for all $2 \le p < \infty$. Indeed, we know that $\mathcal{S}$ is dense in each L^p, and now if we define the smooth cutoff function $\chi_n(\xi)$ which is 1 for $|\xi| > \frac{2}{n}$ and 0 for $|\xi| < \frac{1}{n}$ and define $\hat{f}_n(\xi) = \hat{f}(\xi)\chi_n(\xi)$, we see that each $\hat{f}_n$ is in $\mathcal{S}_0$. By Parseval, we have

$$||f_n - f||_{L^2} \le \int_{|\xi|\le\frac{2}{n}} |\hat{f}(\xi)|^2\xi,$$

which goes to zero as $n \to \infty$. Also, because $f_n(x) = \int \hat{f}_n(\xi)e^{2\pi i\xi x}d\xi$ and $f(x) = \hat{f}(\xi)e^{2\pi i\xi x}d\xi$,

$$||f_n - f||_{L^\infty} \le \int_{|\xi|\le\frac{2}{n}} |\hat{f}(\xi)|d\xi$$

which also goes to zero because we have a Schwartz function and we can use absolute continuity of the integral. This means that $||f_n - f||_{L^p} \to 0$ as $n \to \infty$, so we've approximated $f \in \mathcal{S}$ by elements of $\mathcal{S}_0$ and shown that $\mathcal{S}_0$ is dense in L^p as desired. So it suffices to prove the result for functions $f \in \mathcal{S}_0$.

Now we can define $G(z) = u(z) + iv(z)$ as before, but the advantage of having $f \in \mathcal{S}_0$ is that $Hf \in \mathcal{S}_0$ (in particular, $Hf \in \mathcal{S}$). If we then consider $\int_C G^4(z)dz$ over the semicircular contour of radius R in the upper half-plane, we have $G(z) = 2\int_0^\infty 2\pi e^{i\xi z}\hat{f}(\xi)d\xi$ and $\hat{f}(\xi) = 0$ for all sufficiently small $|\xi| \le \varepsilon_0$. Thus $|G(x+iy)| \le Ce^{-2\pi\varepsilon_0 y}$ for some constant C (because the iy in the exponent gives us exponential decay), and thus we get a similar bound for G^4. But $\int_C G^4(z)dz = 0$ because G^4 is analytic, and the standard contour integral calculations tell us that the integral over the semicircular arc part goes to zero as $R \to \infty$. Thus looking at the real part and noting that $u = f$ and $v = Hf$ on the real line

$$\lim_{R\to\infty} \int_{-R}^{R} (u^4(x) + v^4(x) - 6u^2(x)v^2(x))dx = 0$$

means that (we can drop the $\int v^4(x)dx$ term)

$$\int_{-\infty}^{\infty} (Hf)^4(x) \le 6\int f^2(x)(Hf)^2(x)dx \le \frac{6}{1000}\int (Hf)^4(x)dx + 6\cdot 1000\int f^4(x)dx,$$

and now we can move the first term on the right-hand side to the left-hand side and find that $\int (Hf)^4 dx \le C\int f^4(x)dx$, meaning that $||Hf||_{L^4} \le C_4||f||_{L^4}$. Since we also have a bound in L^2, we see that H is a bounded operator in L^p for all $2 \le p \le 4$. But we can repeat this trick with any even power in place of 4 by doing the same type of expansion of $\int_C G^p(z)dz = 0$ (for example next we can do $p = 8$, then $p = 16$, and so on) – this shows the result for all $2 \le p < \infty$.

Finally, we use duality: notice that $H^* = -H$ (because the adjoint operator turns the i into a $-i$), and $H^* = -H$ is a bounded operator from $L^p \to L^p$ for $2 \le p < \infty$. But this means H will be bounded from $L^{p'} \to L^{p'}$ for all $2 \le p \le \infty$, meaning it is bounded for all $1 < p \le 2$. $\square$

We may ask about the endpoints $p = 1, \infty$ – it is easy to see that we do not have bounded operators in those cases, because we can compute the Hilbert transform of $f(x) = \chi_{[0,1]}(x)$ and we get a logarithmic singularity:

$$Hf(x) = \frac{1}{\pi}\log\frac{x}{1-x} \text{ for } 0 < x < 1.$$

So H does not map L^∞ to L^∞ – we in fact map into a space of "bounded mean oscillation" – and in general this argument extends to operators of the form (these are what we call **singular integral operators** in $\mathbb{R}^n$)

$$\int \frac{K(x-y)}{|x-y|^n} f(y)dy$$

for any K odd on the unit sphere. Unfortunately the trick in the previous proof (by M. Riesz) only works for the Hilbert transform, and the argument in general requires more machinery with harmonic analysis. And singular integral operators appear in a variety of applied problems (like studying compressable fluids in 3-dimensional fluid mechanics, studying vorticity and other physical phenomena).