A course in
Real Analysis

Taught by Prof. P. Kronheimer

Fall 2011

Contents

1. August 31

What is the volume of any set $E \subset \mathbb{R}^3$? We want some properties to hold:

- If $E = E_1 \cup E_2$ and $E_1 \cap E_2$, then we want $vol(D) = vol(E_1) + vol(E_2)$.
- If E' and E are related to each other by rotations and translations, then their volumes should be equal.
- We want some normalization by specifying the volume of the unit ball to be $4\pi/3$ (or, equivalently, saying that the unit cube has volume 1).

Sadly, there is no such way to define volume: the Banach-Tarski paradox says that we can take the unit ball in 3-space, cut it into finitely many pieces, and reassemble them to form two copies of the unit ball, disjoint. (This only works if we accept the axiom of choice.)

So we need a definition for subspaces of $\mathbb{R}^3$ that is restricted enough to rule out the Banach-Tarski paradox, but general enough to be useful.

Measure should be thought of as the n-dimensional analogue of volume (for subsets of $\mathbb{R}^3$) and area (for subsets of $\mathbb{R}^2$). The idea is that we can define a measure with the above properties, if we stick to *measurable* subsets.

Let's first define a *rectangle* to be a Cartesian product of closed intervals. (So the rotation of a rectangle is probably not be a rectangle.) A d-dimensional rectangle is the product of at most d intervals (or exactly d intervals, if some of them are allowed to be $[a, a]$). An open rectangle R^0 is the product of open intervals, and can be $\emptyset$. It's easy to define the volume of a rectangle:

$$|R| = |[a_1, b_1] \times \cdots \times [a_d, b_d]| = \prod_{i=1}^{d}(b_i - a_i)$$

If we have a random set $E \subset \mathbb{R}^d$, let's cover it with at most countably many rectangles: $E \subset R_1 \cup R_2 \cup \cdots$. There are many ways to do this, and some of them have less overlap than others.

DEFINITION 1.1. The *exterior measure* $m_*(E)$ is defined as

$$\inf_{\{R_i\} covering\ E} \sum_{1}^{\infty} |R_n|$$

What does this mean? If $m_*(E) = X$, then for any $\varepsilon > 0$ we can cover E by rectangles with total volume $\leq X + \varepsilon$. (But it doesn't work for any $\varepsilon < 0$.) It is clear from the definition that $m_*(E) \geq 0$ and it has the monotonic property that

$$E' \supset E \implies m_*(E') \geq m_*(E)$$

DEFINITION 1.2. E is *null* if $m_*(E) = 0$. (So your set can be covered by rectangles of arbitrarily small total size.)

The key example of an uncountable null set is the Cantor set. To define this, define $C_0 = [0,1]$, $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, and so on, where to get C_n you delete the middle third of all the disjoint intervals in C_{n-1}. The Cantor set is the intersection of all of these. In base 3, these numbers have a ternary expansion that looks like $x = 0.02200200222\cdots$: that is, there are no 1's in this expansion. To show that it is null, note that $m_*(C) \leq m_*(C_r) \leq 2^r \times (\frac{1}{3})^r$ because it is covered by 2^r rectangles of length $\frac{1}{3^r}$. (This is completely by definition.)

Now, we really hope that $m_*(R) = |R|$. We know that $m_*(R) \leq |R|$, but what if this is $<$? Annoyingly, this is not obvious; it's not even obvious why it is nonzero at all. BUT, it's true!

First, let's forget the part about being able to use infinitely many rectangles. Can we find a set $\{R_n\}$ such that $\sum_1^N |R_n| \leq |R|$. It is clear, that it is silly to have the rectangles stick out over the boundaries of R: so $R_1 \cup \cdots R_n = R$. Now, draw enough lines/ hyperplanes so that each cube belongs to *all* of some set collection of rectangles. From here, there is some awkwardness, but we can get a contradiction.

Now the problem is in allowing infinite covering sets. Suppose $R \subset \cup_1^\infty R_n$, such that $\mathbb{R}| > \sum_1^\infty |R_n|$. Let $\varepsilon = |R| = \sum_1^\infty |R_n|$. Choose some $\widetilde{R}_n \supset R_n$ so that $R_n \subset \widetilde{R}_n^0$ and $|R_n| = |\widetilde{R}_n| - \frac{\varepsilon}{2}n$. So $R \subset \cup_1^\infty \widetilde{R}_n^0$ and $|R| = \sum_1^\infty |\widetilde{R}_n|$. Now we appeal to the Heine-Borel theorem: because R is closed and bounded, it is compact, and is covered by finitely many of those open $\widetilde{R}_n^0$. So we're back to the first situation, with finitely many rectangles.

2. September 2

RECALL we were in $\mathbb{R}^d$, and we had an arbitrary subset E on which we defined the exterior measure $m_*(E)$. Also recall that we can't expect both properties

- $m_*(E) = m_*(E) + m_*(E)$ for $E = E_1 \cup E_2$ disjoint
- $m_*(E)$ is invariant under rigid motions.

PROPOSITION 2.1 (Countable additivity of exterior measure). *If $E = \cup_{j=1}^\infty E_j$ then $m_*(E) \leq \sum m_*(E_j)$*

Recall that the sum could diverge, and then the proposition doesn't say very much.

PROOF. Given $\varepsilon > 0$, for each j, we can find rectangles $R_{j,n}$ such that $E_j \subset \cup_n R_{j,n}$ such that $m_*(E_j) \geq \sum_n |R_{j,n}| - \frac{\varepsilon}{2^j}$. So we have that E is covered by this doubly-indexed collection $E \subset \cup_n \cup_j R_{j,n}$, and

$$\begin{aligned} m_*(E) &\le \sum_j \sum_n |R_{j,n}| \\ &\le \sum_j m_*(E_j) + \frac{\varepsilon}{2^j} \end{aligned}$$

So $m_*(E) \le \sum_j m_*(E_j) + \varepsilon$. Since this works for all ε, we have the desired inequality. ⊟

2.1. Measurable sets. There are many ways to define measurable sets; here we will give Carathéodory's.

DEFINITION 2.2. $E \subset \mathbb{R}^d$ is *measurable* if for every other subset A of $\mathbb{R}^d$ we have:

$$m_*(A) = m_*(A \cap E) + m_*(A \cap E^c)$$

(where E^c is the complement of E in $\mathbb{R}^d$).

So, to check that E is measurable, it is enough to check

$$m_*(A) \ge m_*(A \cap E) + m_*(A \cap E^c)$$

because the other inequality just comes from countable (finite!) additivity above. This means that it is also enough to check for all A of finite exterior measure.

LEMMA 2.3. *A half space (a space of the form $E = \{x_i \ge a\}$) is measurable.*

PROOF. Given $\varepsilon > 0$ we can find rectangles R_n with $\sum |R_n| \le m_*(A) + \varepsilon$. Divide each rectangle R_n along the line $x_i = a$, so you have two pieces $R_n' = R_n \cap \{x_i \ge a\}$ and $R_n'' = R_n \cap \{x_i \le a\}$. Since $A \cap E$ is covered by rectangles, we have

$$\begin{aligned} m_*(A \cap E) + m_*(A \cap E^c) &\le \sum_n |R_n'| + \sum_n |R_n''| \\ &= \sum_n |R_n| \le m_*(A) + \varepsilon \end{aligned}$$

Since this works for all ε, the inequality holds in general, showing that E is measurable. ⊟

LEMMA 2.4. *If E is measurable, so its its complement E^c.*

PROOF. The definition is symmetric in terms of E and E^c. ⊟

LEMMA 2.5. *If E_1 and E_2 are measurable, so are $E_1 \cup E_2$ and $E_1 \cap E_2$.*

PROOF. Since $E_1 \cup E_2 = (E_1^c \cap E_2^c)^c$, we will do the intersection case only (as per the previous lemma). So take a space A of finite measure. E_1 is measurable, so $m_*(A) = m_*(A \cap E_1) + m_*(A \cap E_1^c)$. Now decompose this further, taking $A \cap E$ and $A \cap E^c$ as the arbitrary set:

$$\begin{aligned} m_*(A) &= m_*((A \cap E_1) \cap E_2) + m_*((A \cap E_1) \cap E_2^c) + m_*(A \cap E_1^c \cap E_2) + m_*(A \cap E_1^c \cap E_2^c) \\ &\ge m_*(A \cap (E_1 \cap E_2)) + m_*(A \cap (E_1 \cap E_2)^c) \end{aligned}$$

where the last inequality is by sub-additivity. Done! $\boxminus$

Summary: Measurable sets form an *algebra of sets* – they are closed under complements, finite intersection, and finite union.

We have done this for half-spaces; taking intersections gives us "rectangles". However, we do not have the unit disk yet. The key is to replace finite intersections with countable ones.

PROPOSITION 2.6. *If $E = \cup_{n=1}^{\infty} E_n$ and if each E_n is measurable, then so is E. Furthermore, if the E_n are disjoint, then $m_*(E) = \sum m_*(E_n)$.*

COROLLARY 2.7. *Every open set in $\mathbb{R}^d$ is measurable.*

This is because every open set is a countable union of rectangles (remember the definition of a topology!).

PROOF OF PROPOSITION. We may as well assume from the beginning that the E_n are disjoint: replace the n^{th} set by $E_n \setminus \cup_{m<n} E_m$. We have not messed up measurability, because we just showed we can take finite unions and complements. To prove measurability, we want to fix some set A of finite measure. (Recall that A need not be measurable itself.) We have

$$m_*(A \cap (E_1 \cup E_2)) = m_*(A \cap E_1) + m_*(A \cap E_2)$$

This is a consequence of the measurability of E_1 (or E_2) and the disjoint-ness of E_1 and E_2, with the "test set" being $A \cap (E_1 \cup E_2)$. This shows additivity of two terms, and hence additivity of finitely many terms. That is, for every $N \in \mathbb{N}$ we have

$$m_*(A \cap \cup_{n=1}^{N} E_n) = \sum_{n=1}^{N} m_*(A \cap E_n)$$

$$\begin{aligned} m_*(A \cap E) &\geq m_*(A \cap \cup_1^N E_2) \\ &= \sum_1^N m_*(A \cap E_n) \end{aligned}$$

where the first inequality is by containment. Taking the limit as $N \to \infty$, we have

$$m_*(A \cap E) \geq \sum_1^{\infty} m_*(A \cap E_n)$$

Countable sub-additivity gives the other inequality, so we get

$$m_*(A \cap E) = \sum_1^{\infty} m_*(A \cap E_n)$$

We haven't yet proved measurability, however. We have to play a similar game with $A \cap E^c$:

$$\begin{aligned} m_*(A \cap E^c) &\leq m_*(A \cap (\cup_1^N E_n)^c) \\ &= m_*(A \cap \cup_1^N E_n) \\ &= m_*(A) - \sum m_*(A \cap E_n) \end{aligned}$$

where the first inequality is by inclusion. Then we applied the definition of measurability to $\cup_1^N E_n$. Note that we are assuming that A has finite exterior measure. (In using the simple trick $x = y + z \implies x - y = z$, we have to watch out if two of the things are infinite.)

Now take the limit with $N \to \infty$, which gives

$$\begin{aligned} m_*(A \cap E^c) &\leq m_*(A) - \sum_1^\infty m_*(A \cap E_n) \\ &= m_*(A) - m_*(A \cap E) \end{aligned}$$

by the additivity proven earlier. (Again, we have proven only the nontrivial inequality.) So the countable union E is measurable.

Note that *in just the first part* (countable additivity), we didn't use the fact that A had finite measure; so we can take $A = \mathbb{R}^n$, which gives the countable additivity

$$m_*(E) = \sum m_*(E_n)$$

$\boxminus$

If E is measurable, we define the measure $m(E) = m_*(E)$.

3. September 7

3.1. More about measurable sets. RECALL we had defined outer measure for all $E \subset \mathbb{R}^d$. When E was measurable, we defined the measure to be $m(E) = m_*(E)$. We showed that measurability is preserved by complements and countable unions and intersections. We also had the key formula

$$m(E) = \sum_1^\infty m(E_n)$$

if $E = \cup E_n$ and the E_n were disjoint. As examples of measurable sets, we had rectangles and open sets. The definition of open-ness can be reformulated to say that all points x in the open set $\mathcal{O}$ can be surrounded by a rectangle $R \subset \mathcal{O}$; moreover, this rectangle can be taken to have rational coordinates for the vertices. There are only countably many such R; so every open set can be written as a countable union of rectangles.

Null sets are measurable: these are precisely the sets such that $m_*(E) = 0$. To check measurability, we require

$$m_*(A) \geq m_*(A \cap E) + m_*(A \cap E^c)$$

If E is null, then $m_*(A \cap E) = 0$, so this inequality holds. So null sets are measurable.

We say that $E_n \nearrow E$ as $n \to \infty$ if $E_1 \subset E_2 \subset \cdots$ and $E = \cup E_n$. Similarly, we say that $E_n \searrow E$ if $E_1 \supset E_2 \supset \cdots$ and $E = \cap E_n$.

Lemma 3.1.

- *If E_n is measurable and $E_n \nearrow E$ then E is measurable and*
$$m(E) = \lim_{n\to\infty} m(E_n).$$
(We are using measure in the extended sense: it is allowed to be infinite.)
- *If $E_n \searrow E$ and $m(E_1)$ is finite, then*
$$m(E) = \lim_{n\to\infty} m(E_n).$$

Note that the condition $m(E_1) < \infty$ is necessary: take $E_n = [n, \infty)$, and note that $m(E_n) = \infty$ but $m(\cap E_n) = 0$.

Proof.

- If $E_n \nearrow E$, set $F_n = E_n \backslash E_{n-1}$, where $F_1 = E_1$. By countable additivity,
$$m(E) = \sum m(F_n) = \lim_{n\to\infty} \sum_{i=1}^{n} m(F_i).$$
But the first n F_i's are disjoint, and their union is E_n.
- If $E_n \searrow E$ and $m(E_1) < \infty$, write $H_1 = E_1 \backslash E_2$, $H_2 = E_2 \backslash E_3$, etc. We have that $H = \cup_1^\infty H_2$.
$$E_1 = E \cup H$$
$$m(E) = m(E_1) - m(H)$$
We can't rearrange this formula unless $m(E_1)$ is finite. So
$$\begin{aligned} m(E) &= m(E_1) - \sum_1^\infty m(H_n) \\ &= \lim(m(E_1) - \sum_{i=1}^{n} m(H_i)) \\ &= \lim m(E_{n+1}) \end{aligned}$$
$$E_{n+1} = E_1 \backslash \cup_{i=1}^n H_i$$

$\boxminus$

Stein and Shakarchi have a "different" definition of measurability:

Definition 3.2. $E \subset \mathbb{R}^d$ is SS-measurable if: for all $\varepsilon > 0$, there is an open $\mathcal{O} \supset E$ with $m_*(\mathcal{O} \backslash E) \leq \varepsilon$.

LEMMA 3.3. *A set is SS-measurable iff it is measurable (in the previous sense).*

PROOF. If E is SS-measurable, then for every n choose $\mathcal{O}_n \supset E$ open with $m_*(\mathcal{O}_n \backslash E) \leq \frac{1}{n}$. Write

$$G = \bigcap_{n=1}^{\infty} \mathcal{O}_n$$

We have that $G \supset E$ and $m_*(G \backslash E) \leq m_*(\mathcal{O}_n \backslash E) \leq \frac{1}{n}$ for all n. So $m_*(G\backslash E) = 0$. Null sets are measurable (in the old sense), and G is measurable too, being the intersection of measurable sets. We can write

$$E = G \cap (G \backslash E)^c$$

which shows that E is measurable in the old sense.

Conversely, suppose that E is measurable in our sense. First suppose that E has finite measure. Given $\varepsilon > 0$, cover E with rectangles R_n with

$$\sum_1^{\infty} m(R_n) \leq m(E) + \frac{\varepsilon}{2}$$

Now we want to find larger rectangles $\widetilde{R}_n$ such that $R_n \subset \widetilde{R}_n^o$. We can also require that $\sum m(\widetilde{R}_n^o) \leq m(E) + \varepsilon$. Set $\mathcal{O} = \cup_{n=1}^{\infty} \widetilde{R}_n^o$. So $\mathcal{O} \supset E$ and $m(\mathcal{O}) \leq m(E) + \varepsilon$. Since E is measurable, we have

$$m(\mathcal{O}) = m(E) + m(\mathcal{O} \backslash E)$$

So we have that

$$m(\mathcal{O} \backslash E) \leq \varepsilon.$$

Note that it's important that $m(E)$ is finite in this argument; otherwise we keep running into trouble by trying to "subtract infinity." $\boxminus$

DEFINITION 3.4. A subset of $\mathbb{R}^d$ is called a G_δ set if it is a countable intersection of open sets. A subset of $\mathbb{R}^d$ is called an F_σ set if it is a countable union of closed sets.

Notice that in the previous proof we had a line

$$E = G \backslash (G \backslash E)$$

where E was measurable, and G was a G_δ set: $G_\delta = \cup \mathcal{O}_n$. Also remember that $G \backslash E$ was null. This gives a nice decomposition of measurable sets.

COROLLARY 3.5. *If E is measurable, then $E \subset G$ where G is a G_δ set and $G \backslash E$ is null. (And conversely.)*

3.2. Integration. Integrals are about *signed* area. We can imagine breaking up the area under a graph into the positive and negative parts:

$$U_+ = \{(x, y) \,:\, 0 \leq y \leq f(x), y \in \mathbb{R}, x \in \mathbb{R}^d\}$$
$$U_- = \{(x, y) \,:\, 0 > y \geq f(x), y \in \mathbb{R}, x \in \mathbb{R}^d\}$$

But instead of talking about area, let's talk about measure! We *could* define (but we won't) f integrable to mean that U_+ and U_- are measurable subsets of $\mathbb{R}^{d+1}$ of finite measure. If f is integrable, we *could* write

$$\int_{\mathbb{R}^d} f = m(U_+) - m(U_-)$$

[Note that the dx is omitted, and sometimes we will even omit $\mathbb{R}^d$ when that is understood.] *But we're not going to do this!* This definition has the problem of not being able to handle infinite upper and lower areas very well; we will give another definition.

DEFINITION 3.6. A function $f : \mathbb{R}^d \to \mathbb{R}$ is *measurable* if for every $a \in \mathbb{R}$,

$$E_a = \{x \in \mathbb{R}^d \,:\, f(x) \leq a\}$$

is measurable in $\mathbb{R}^d$.

(So the constant function 1 is measurable, but not integrable.) Observe

$$\{x \,:\, f(x) < a\} \leq \bigcup_1^\infty E_{a-\frac{1}{n}}$$

(That is, saying that $y \leq a - \frac{1}{n}$ for all n is equivalent to $y < a$.) So if f is measurable, this is a measurable set. Similarly,

$$\{x \,:\, f(x) \geq a\}$$
$$\{x \,:\, f(x) \in [a, b]\}$$

are also measurable if f is measurable. So we could have written our definition a bit differently. If $A \subset \mathbb{R}$ is open, then it is a countable union of closed intervals $[a, b]$. So $\{x \,:\, f(x) \in A\}$ is also measurable. The other direction is also true.

LEMMA 3.7. *f is measurable iff $f^{-1}(A) \subset \mathbb{R}^d$ is measurable for all open sets $A \subset \mathbb{R}$.*

PROOF. Boring. ⊟

4. September 9

4.1. Measurable functions.

DEFINITION 4.1. We say that $f : \mathbb{R} \to \mathbb{R}$ is measurable if $f^{-1}(A)$ is a measurable set for all open A. (Equivalently, for all closed intervals, or all half-intervals $[a, \infty)$.)

We can also consider functions $f : \mathbb{R}^d \to \mathbb{R}^k$. Such a function $f = (f_1 \cdots f_k)$ is measurable if $f^{-1}(A)$ is measurable for all open $A \subset \mathbb{R}^k$, or all closed rectangles, or all half-spaces. Thus f is measurable iff $f_1 \cdots f_k$ are measurable for all its components f_i.

REMARK 4.2. If $C \subset \mathbb{R}$ is measurable and f is a measurable function, then $f^{-1}(C)$ need not be measurable. (The test sets need to be simple things like opens, etc.)

EXAMPLE 4.3. Examples of $f : \mathbb{R}^d \to \mathbb{R}$ measurable:

- The characteristic function χ_E for a measurable set $E \subset \mathbb{R}^d$. This is
$$\chi_E = \begin{cases} 1 & \text{on } E \\ 0 & \text{else.} \end{cases}$$
- The linear combination of measurable functions is measurable. Why?

PROPOSITION 4.4. *If $\mathbb{R}^d \xrightarrow{f} \mathbb{R}^k$ is measurable, and $\mathbb{R}^k \xrightarrow{g} \mathbb{R}^\ell$ is continuous, then the composite $\mathbb{R}^d \xrightarrow{g\circ f} \mathbb{R}^\ell$ is measurable.*

PROOF. To test whether the composition is measurable, we need a test open set in $\mathbb{R}^\ell$, which is pulled back to an open set in $\mathbb{R}^k$, by continuity. Now use the fact that f is measurable. ⊟

Now use the continuous function $(f, g) : \mathbb{R}^d \to \mathbb{R}^2$ where $(f, g)(x) = (f(x), g(x))$, and then compose with the continuous function $(x, y) \mapsto ax + by$.
- It is *not* true that the composite of measurable maps is measurable.

DEFINITION 4.5. When we say that something holds for *almost all x* (a.a. x), we mean that it holds for all x outside a null set in $\mathbb{R}^d$. Equivalently, we say that a property holds *almost everywhere* (a.e.).

PROPOSITION 4.6. *Suppose that f_n are measurable functions $\mathbb{R}^d \to \mathbb{R}$ for $n \in \mathbb{N}$, and $f_n \to f$ a.e. (that is, $f_n(x) \to f(x)$ for almost all x). Then f is measurable too.*

Note first: if $g = \widetilde{g}$ a.e.then one of them is measurable iff the other one is. (The inverse image of an open set will be the same under both, $\pm$ a null set.) We can change f_n and f so that they are zero on the null "bad set," so we can assume that $f_n \to f$ everywhere. (Note that our convergence is pointwise.)

PROOF. Using the definition of convergence, $f(x) \geq a$ means: for all k there exists some n_0 such that for all $f_n(x) \leq a + \frac{1}{k}$ for $n \geq n_0$. Let's write the quantifiers in terms of intersections:
$$\{x \,:\, f(x) \leq a\} = \bigcap_{k\geq 1} \bigcup_{n_0 \geq 1} \bigcap_{n \geq n_0} E_{n,k}$$
where $E_{n,k} = \{f_n(x) \leq a + \frac{1}{k}\}$. Since countable unions and intersections of measurable things are measurable, . ⊟

4.2. Integrals of simple functions. First we will define the integral of simple functions, then non-negative functions, and finally treat the general case.

DEFINITION 4.7. A *simple function* is a finite linear combination $\mathbb{R}^d \to \mathbb{R}$
$$g = \sum_1^k a_i \chi_{E_i}$$
where E_i is measurable and of finite measure.

This takes finitely many values. The set where this takes a given value is measurable. There are many ways to write a simple function in this form, but there is always a most efficient expression: take a_i to be the distinct non-zero values that g takes, and let $E_i = g^{-1}(a_i)$. (This produces a decomposition in which the E_i are disjoint.) We call this the *standard form* of a simple function. Of course, you can always reduce something to standard form: if $E = E' \cup E''$ is a disjoint decomposition, then $a\chi_E \to a\chi_{E'} + a\chi_{E''}$.

DEFINITION 4.8. If $g = \sum a_i \chi_{E_i}$ (not necessarily in standard form) define the integral

$$\int g = \sum a_i m(E_i).$$

We need to check that this is well-defined among other representations of something in standard form. Reducing a representation by the operation described above does not change this integral. Note that this only makes sense if the E_i have *finite* measure: otherwise you could be adding $\infty + (-\infty)$.

Linearity of the integral is obvious by the definition of simple functions:

PROPOSITION 4.9. (1) $\int (ag + bh) = a\int g + b\int h$
(2) $\int g \geq 0$ *if* $g \geq 0$

LEMMA 4.10. *If (g_n) is a sequence of simple functions and $g_n \searrow 0$ then $\int g_n \to 0$ as $n \to \infty$.*

For all x outside a null set, $g_n(x)$ is decreasing and converges to zero as $n \to \infty$.

If I take a simple function, and modify it on a null set, then it is still a simple function and the integral is unchanged. If we make all $g_n(x) = 0$ for all x in the bad null set, you can replace all claims of "almost everywhere" by "everywhere."

PROOF. Let M be the measure of the set where $g_1 \neq 0$, and let S be the set where it happens. We call this the support of g_1. Let $C = \max_{x \in \mathbb{R}^d} g_1(x)$. By the assumption of decreasing-ness, all other g's are contained in the set of measure M where g_1 is zero. If $Q \geq g_n$, then $\int Q \geq \int g_n$. Choose $\delta > 0$, and overestimate g_n by filling in the box $S \times [0, \delta]$. That is,

$$\int g_n \leq \delta M + C m\{x : g_n(x) \geq \delta\} \tag{1}$$

This holds for all $\delta > 0$. Look at

$$E_n = \{g_n(x) \geq \delta\}$$

We have $E_n \searrow \emptyset$ for $\delta > 0$. So $m(E_n) \to 0$ as $n \to \infty$. This is the continuity of measure as we described in lecture 2. So the second term on the right of (1) approaches zero as $n \to \infty$. Eventually, $\int g_n < 2\delta M$; this is true for any δ, so

$$\lim_{n\to\infty} \int g_n = 0.$$

$\boxminus$

COROLLARY 4.11. *(Monotone convergence theorem for simple functions) If $\{g_n\}$ is a sequence of simple functions where $g_n \nearrow g$ a.e. with g also simple, then $\lim \int g_n = \int g$. Put $h_n = g - g_n$. This is simple, nonnegative, and $h_n \searrow 0$ a.e. So by the last lemma, $\int h_n \to 0$ so $\int g - \int g_n \to 0$.*

5. September 12

5.1. Non-negative functions $f : \mathbb{R}^d \to \mathbb{R}$.

We are going to approximate our general non-negative f by simple functions $g_n \leq f$.

DEFINITION 5.1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be non-negative and measurable. We say that f is *integrable* if

$$\sup_g \int g < \infty$$

where the supremum is over all simple $g \geq 0, g \leq f$. When this quantity is finite, we will define

$$\int f = \sup_g \int g$$

A simple example of a function that isn't integrable is

$$f(x) = \begin{cases} \lfloor x \rfloor & \text{if } x - \lfloor x \rfloor < \frac{1}{\lfloor x \rfloor} \\ x & \text{elsewhere} \end{cases}$$

When this does work, we want to be able to approximate this by functions that are zero outside a finite box. Inside the box defined by $|x_i| \leq n$ for all i, define

$$g_n(x) = \max\{\frac{k}{n} \, : \, 0 \leq k \leq n^2, \frac{k}{n \leq f(x)}\}$$

where we are thinking of the graph living in $\mathbb{R}^n \times \mathbb{R} \ni (x_1, \cdots, x_d, y)$. (Notice that the range of g_n is indeed $\leq n$.)

LEMMA 5.2. *Suppose $\{f_n\}$ are integrable, nonnegative, and non-decreasing: $f_1 \leq f_2 \leq \cdots$. Let's look at the integrals. Suppose $\lim_{n\to\infty} \int f_n < \infty$. Then, for almost all x,*

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

exists. Define

$$f(x) = \lim f_n(x).$$

Then f is integrable, and

$$\int f = \lim_{n\to\infty} \int f_n.$$

Notice that we do not even assume that all the functions converge, only that the integrals do.

PROOF. Define $f(x) = \lim_{n\to\infty} f_n(x)$. We will continue this in the extended sense, where the limit may be ∞ for some x. Let h be any simple function with $h \leq f$. We'll show

$$\int h \leq \lim_{n\to\infty} \int f_n. \tag{2}$$

If we can prove (2), we're pretty much done:

- f can't be infinite on a set E of positive measure, or else we could set $h_C = C\cdot\chi_E$ with C very large, and see that $\int h_C$ has no upper bound (whereas $\lim_{n\to\infty} f_n$ does by hypothesis).
- Also, $\int f = \sup_h \int h \leq \lim_n \int f_n = K$; i.e., f is integrable and
$$\int f \leq \lim_n \int f_n.$$
- Also, $f_n \leq f$, so
$$\lim \int f_n \leq \int f.$$
So equality holds where we want.

So now let's try to prove (2). For each f_n there are simple functions $g_{n,m}$ with $g_{n,m} \to f_n$ as $m \to \infty$. Look at the diagonal sequence. $g_{n,n}$ are not necessarily increasing, but we can define

$$\widetilde{g}_{n,n} = \max_{m,m'\leq n} g_{m,m'}.$$

We claim that $\widetilde{g}_{n,n} \nearrow f$ a.e. By construction, $\widetilde{g}_{n,n}$ are simple and non-decreasing. Given simple $h \leq f$, consider

$$F = \min\{\widetilde{g}_{n,n}, h\}$$

which is a simple function. As $n \to \infty$, $F(x) \nearrow h(x)$, since the $\widetilde{g}_{n,n}$ will eventually lie above h. Now we can apply the monotone convergence theorem for simple functions:

$$\begin{aligned}\int h &= \lim \int \min\{\widetilde{g}_{n,n}, h\}\\ &\leq \lim_n \int \widetilde{g}_{n,n}\\ &\leq \lim_n \int f_n = K\end{aligned}$$

which is what we wanted at (2). (In the end we used the fact that $\widetilde{g}_{n,n} \leq f_n$ by construction: $g_{m,m'} \leq f_m$ and $f_m \leq f_n$.)

$\boxminus$

APPLICATION 5.3. If f, g are ≥ 0 are integrable and $a, b \geq 0$, and

$$F = af + bg$$

then F is integrable and

$$\int F = a\int f + b\int g.$$

PROOF. We have simple functions $f_n \nearrow f$ and $g_n \nearrow g$ so we can set $F_n = af_n + bg_n$ so $F_n \nearrow F$. By linearity of the integral for simple functions, we have

$$\int F_n = a\int f_n + b\int g_n \nearrow a\int f + b\int g < \infty.$$

In the form just proven, the Monotone Convergence Theorem says that $\lim F_n$ exists a.e. But we already knew that: the F_n converge to F. But it also tells us that

$$\int F = \lim \int F_n.$$

$\boxminus$

6. September 14

We know that $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. Define $I_n \subset [0,1]$ to have length $\frac{1}{n}$. Consider the function

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{n}\chi_{I_n}(x)$$

Let the f_n's be the partial sums. Then we have an increasing sequence $f_n \nearrow f$. But it looks like f might not be well-defined somewhere! But remember the set $\cap_n I_n$ has zero measure, and everywhere else this limit is perfectly well-defined. The monotone convergence theorem says that f is finite for a.a. $x \in [0,1]$ (no matter how we chose I_n). Furthermore, it says that

$$\begin{aligned}\int f_n &= \sum_1^n \int \frac{1}{m}\chi_{I_m} \\ &= \sum_1^n \frac{1}{m^2} = \frac{n}{m^2}\end{aligned}$$

Each $\int f_n$ converges, and so does f.

DEFINITION 6.1 (General case of the Lebesgue integral). Let $f : \mathbb{R}^d \to \mathbb{R}$ be measurable. We can write

$$f = f_+ - f_-$$

where

$$f_+(x) = \max\{f(x), 0\},$$
$$f_0(x) = \max\{-f(x), 0\}.$$

We say that f as above is integrable if f_+ and f_- are. In this case, we define

$$\int f = \int f_+ - \int f_-.$$

NOTATION 6.2. $|f|$ means that the function $|f|(x) := |f(x)|$.

REMARK 6.3. *f integrable implies that $|f|$ is also integrable*, just because $|f| = f_+ + f_-$ and each of those are integrable.

If $|f|$ is integrable and f is measurable, then f is also integrable, because $f_+ \leq |f|$ and f_+ is measurable. (Any measurable function (here, f_+) that lies below an integrable function is also integrable.)

If $f = h_+ - h_-$ with h_+ and h_- both integrable, then f is too. f is measurable, and $|f|$ is integrable because $|f| \leq |h_+| + |h_-|$, and the RHS of this is integrable. From here we can say that f_+ and f_- are integrable, which shows that f is integrable.

PROPOSITION 6.4 (Properties of the integral).

(1) $f \geq 0 \implies \int \geq 0$
(2) *If $f \leq g$ then $\int f \leq \int g$.*
(3) *$\int(af + bg) = a\int f + b\int g$ for integrable f and g.*
(4) *f integrable $\iff$ it is measurable and $|f|$ is integrable*

PROOF. Just the third part. The annoyance here is that $(af+bg)_+ \neq af_+ + bg_+$. But we can write

$$af + bg = h_+ - h_-$$

and then $h_+ = af_+ + bg_+$. So

$$\begin{aligned}\int (af + bg) &= \int h_+ - \int h_- \\ &= a\left(\int f_+ - \int f_-\right)_b \left(\int g_+ - \int g_-\right)\end{aligned}$$

$\boxminus$

We sometimes write $\int f = +\infty$ to mean that f is measurable and $f = f_+ - f_-$ and either f_+ or f_- is not integrable, and the other one is finite. (Think about $f = 1$, modified to have a little squiggle below the x-axis in the middle.)

THEOREM 6.5 (Monotone Convergence Theorem a.k.a. Bepo Levi's Theorem). *Let f_n be a non-decreasing sequence of integrable functions (a.e., of course), and suppose that $\int f_n$ (an increasing sequence) does not diverge to $+\infty$ as $n \to \infty$; i.e. $\lim_{n\to\infty} \int f_n$ exists. Then for a.a. $x \in \mathbb{R}^d$, $\lim_{n\to\infty} f_n(x)$ exists (it is finite). If f is defined a.e. by $f(x) = \lim_n f_n(x)$ then*

(1) *f is integrable*
(2) *$\int f = \lim \int f_n$.*

This is just a slightly fuller version of the monotone convergence theorem we had earlier; here the functions can be negative. Again, the interesting thing is that we don't assume that the functions converge, only that their integrals do.

PROOF. Set $\widetilde{f_n} = f_n - f_1$. Then the $\widetilde{f_n} \geq 0$ a.e. and are increasing a.e. By the monotone convergence theorem for non-negative functions,

$$\widetilde{f}(x) = (\lim f_n(x)) - f_1(x)$$

exists a.a. x and

$$\int \widetilde{f} = \lim \int (f_n - f_1) = \left(\lim \int f_n\right) - \int f_1.$$

$\boxminus$

THEOREM 6.6 (Dominated convergence theorem). *Suppose $\{f_n\}$ is a sequence of integrable functions $\mathbb{R}^d \to \mathbb{R}$. Suppose that there is some function f such that*

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

Suppose there exists some function g which is integrable (and non-negative) such that

$$|f_n| \leq g \ \forall n.$$

Then f is integrable and $\int f = \int \lim f_n = \lim \int f_n$.

Without the existence of the dominating function g, this is really wrong. For example, take $f_n = \chi_{[n,n+1]}$. Then $f_n \to 0$ everywhere, but

$$\lim \int f_n \neq \int \lim f_n,$$

because the LHS is 1 and the RHS is 0.

PROOF. Write

$$f_n = f_{n,+} - f_{n,-}$$
$$f = f_+ - f_-$$

In this way we can reduce to the case that all of the f_n and f are nonnegative functions. f is measurable because it is a limit of f_n (a.e.). By the dominated hypothesis, $|f| = f \leq g$ and is thus integrable. So $\int f$ exists. We may as well assume that $f_n \to f$ everywhere, by changing things on a null set. Set

$$\widehat{f_n}(x) = \sup_{m \geq n} f_m(x)$$

$$\breve{f}_n(x) = \inf_{m \geq n} f_m(x)$$

So $f_n(x) \to f(x)$ implies that $\widehat{f_n}(x) \searrow f(x)$ and $\breve{f}_n(x) \nearrow f(x)$. The monotone convergence theorem says that

$$\lim_{n\to\infty} \int \widehat{f_n} = \int f = \lim_{n\to\infty} \int \breve{f}_n$$

But also, $\breve{f}_n \leq f_n \leq \widehat{f_n}$ everywhere, and so

$$\int \breve{f}_n \leq \int f_n \leq \int \widehat{f_n}$$

Now we use the squeeze theorem to say that

$$\int f_n = \int f$$

$\boxminus$

APPLICATION 6.7. Consider a series $\sum_1^\infty g_n$ where g_n is integrable for all n and $\sum_n \int |g_n| < \infty$. Then $\sum_1^\infty g_n(x)$ exists a.a. $x \in \mathbb{R}^d$ and

$$\int \left(\sum g_n\right) = \sum \left(\int g_n\right).$$

PROOF. Consider

$$G_n = \sum_1^n |g_m|.$$

Then $G_n \nearrow$ something for all x and

$$\int G_n = \sum_1^n \int |g_m| \nearrow K < \infty$$

By the MCT, there is some G such that $G_n(x) \nearrow G(x)$ for a.a. x. So $\sum_1^\infty |g_n(x)|$ is convergent a.a. $x \in \mathbb{R}^d$ and that sum is $G(x)$. Therefore, $\sum_1^\infty g_n(x)$ converges a.a. x also.

To see why you can change the limit with the integral use the DCT. Set

$$f_n = \sum_{m=1}^n g_m$$

to be the partial sums. We know that $f_n \to f = \sum_1^\infty g_n$ a.e., and $|f_n| \le \sum_1^\infty |g_n| = G$ a.e. G is integrable (it came from the MCT) so the conditions of DCT are satisfied with G as the dominating function. $\boxminus$

7. September 16

7.1. More on convergence theorems.

PROPOSITION 7.1. *If g_n are integrable for all n and $\sum_n \int |g_n| < \infty$. Then $\sum_1^\infty g_n$ is absolutely convergent a.a. x. Writing*

$$f = \sum g_n$$

then f is integrable and

$$\int f = \sum \int g_n.$$

Last time we had $G_n(x) = \sum_{k=1}^n |g_k(x)|$. Apply the MCT to $G_n \nearrow$. See that $G(x) = \sum_1^\infty |g_n(x)|$ exists a.e. and is an integrable function. So $f(x)$ exists a.a. x. Set $f_n(x) =$

$\sum_1^n g_k(x)$. Apply the DCT to f_n, $f_n \to f$ a.e. Use $|f_n(x)| \leq \sum_1^n |g_k(x)| \leq G(x)$. That is, f_n is dominated by G for all n. Conclude f is integrable and $\int f = \lim \int f_n$.

LEMMA 7.2. *(Fatou's lemma) Let f_n be a sequence of non-negative functions, integrable for all n. Suppose* $\liminf_{n\to\infty} \int f_n$ *is finite. Then*

$$f(x) := \liminf_{n\to\infty} f_n(x)$$

is integrable, and

$$\int f \leq \liminf_{n\to\infty} \int f_n$$

PROOF. Lt $\breve{f}_n(x) = \inf_{k\geq n} f_k(x)$. So $\breve{f}_n$ are increasing: $\breve{f}_n(x) \leq \breve{f}_{n+1}$ for all x. By definition of $\liminf$, $\breve{f}_n(x) \nearrow f(x)$ for all x (for things possibly being infinite). Since $\breve{f}_n$ is the infimum of functions, the first of which is f_n, we have $\breve{f}_n \leq f_n$ and hence $\int \breve{f}_n \leq \int f_n$. The $\breve{f}_n$ are increasing. So

$$\begin{aligned}\lim_{n\to\infty} \int \breve{f}_n &= \liminf_{n\to\infty} \int \breve{f}_n \\ &\leq \liminf_{n\to\infty} \int f_n < \infty\end{aligned}$$

By the MCT, f is integrable and

$$\int f = \lim_{n\to\infty} \int \breve{f}_n \leq \liminf \int f_n$$

$\boxminus$

7.2. The Lebesgue space L^1. Write $\mathcal{L}^1(\mathbb{R}^d, \mathbb{R})$ or just $\mathcal{L}^1(\mathbb{R}^d)$ or just $\mathcal{L}^1$ for the integrable functions $f : \mathbb{R}^d \to \mathbb{R}$. Write

$$\mathcal{N} \subset \mathcal{L}^1$$

for the space of functions f which are 0 a.e.

$\mathcal{L}^1$ and $\mathcal{N}$ are both vector spaces. We can define

$$L^1 = \mathcal{L}^1/\mathcal{N}$$

to be the set of equivalence classes of integrable functions f. That is, $f \sim g$ iff $f = g$ a.e. We will write $f \in L^1$ when we mean, "the equivalence class of f is in L^1."

PROPOSITION 7.3. *If f is integrable, and $\int |f| = 0$ then $f = 0$ a.e.; i.e. $f \in \mathcal{N}$ which means "$f = 0$ in L^1."*

PROOF. Use the MCT. Set $f_n = n|f|$. So $f_n(x)$ is increasing for all x. Then $\int f_n = 0$ for all n $(= n \int |f| = 0)$. So $\lim_{n\to\infty} n|f(x)|$ exists a.e. by the MCT. So $|f(x)| = 0$ a.e. $\boxminus$

DEFINITION 7.4. For $f \in L^1$ define the norm $||f||$ (or $||f||_{L^1}$) as

$$||f|| = \int |f|.$$

(We have to be careful with equivalence classes. But if $f = g$ a.e. then $\int |f| = \int |g|$, so the norm is well-defined on L^1.) We can restate what we just proved. If $||f|| = 0$ then $f = 0$ in L^1.

PROPOSITION 7.5. *$|| \cdot ||$ is a norm on the vector space L^1:*

(1) $||\lambda f|| = |\lambda| \cdot ||f||$ *for* $\lambda \in \mathbb{R}$
(2) *(Triangle inequality:)* $||f + g|| \leq ||f|| + ||g||$
(3) $||f|| = 0 \iff f = 0$

PROOF. (1) is immediately clear, and (3) we just proved. (2) is straightforward:

$$\begin{aligned} ||f+g|| &= \int |f+g| \\ &\leq \int (|f| + |g| \\ &= \int |f| + \int |g|) \\ &= ||f|| + ||g|| \end{aligned}$$

∎

THEOREM 7.6. *Suppose $g_n \in L^1$ for all n. Suppose*

$$\sum ||g_n|| < \infty. \tag{3}$$

Then there is some $f \in L^1$ such that the partial sums $f_n = \sum_1^n g_k$ converge in norm to f. That is, $f_n \in L^1$, $f \in L^1$ and $||f_n - f|| \to 0$ as $n \to \infty$. (This is metric space convergence.)

PROOF. This is mainly a direct translation of what we just said. (3) means that $\sum \int |g_n| < \infty$. We saw earlier that, for all $f \in L^1$ with $\sum_1^\infty g_n(x) = f(x)$ a.e. (This gives $f_n \in L^1$ and $f \in L^1$.) To show that this converges in norm, note that if $f_n(x) := \sum_1^n g_k(x)$ then $f_n(x) \to f(x)$ a.e. We want $\int |f_n - f| \to 0$. This is a consequence of the DCT. Recall from earlier that there is some dominating function G with $|f_n| \leq G$ a.e. and $\int G < \infty$. So $|f| \leq G$ a.e. and $|f_n - f| \leq 2G$ a.e. So apply the DCT to the functions $f_n - f|$. These are all dominated by $2G$ and $|f_n - f|(x) \to 0$ a.a. x by what we proved today. The conclusion of the DCT is exactly what we want:

$$\lim \int |f_n - f| = \int 0 = 0$$

∎

This is related to the completeness of L^1. There are some other notions of convergence:

- Convergence almost everywhere: $f_n(x) \to f(x)$ a.a. x
- Convergence in measure: for a sequence of measurable functions f_n, we say that $f_n \to f$ in measure if for all $\varepsilon > 0$,
$$m\{x : |f_n(x) - f(x)| \geq \varepsilon\} \to 0 \text{ as } n \to \infty$$
- Convergence in norm (as above)

On the HW you will see that convergence in norm implies convergence in measure. Beyond that, in full generality there are no other implications among these. You can improve this a bit by considering only convergent subsequences.

PROPOSITION 7.7. *If $f_n \to f$ in measure then there is a subsequence $f_{n'}$ for $n' \in N \subset \mathcal{N}$ which converges a.e.*

EXAMPLE 7.8. First consider the finite sequence S_k where $f_1 = \chi_{[0,\frac{1}{k}]}, f_2 = \chi_{[\frac{1}{k},\frac{2}{k}]} \cdots f_{k-1} = [\frac{k-1}{k}, 1]$. Concatenate all the S_k to get a larger sequence.

8. September 19

PROPOSITION 8.1. *If $f_n \to f$ in measure on $\mathbb{R}^d$ then there is a subsequence $\{n'\} \subset \mathbb{N}$ such that $f_n' \to f$ a.e.*

PROOF. Without loss of generality $f_n \geq 0$ for all n and $f_n \to 0$ in measure. (Just take the difference.) Fix $\delta > 0$. Convergence to zero in measure means that for all k there is some $n_k \geq 1$ such that
$$m(\{x : f_{n_k}(x) \geq \delta\}) \leq 2^{-k}$$
The point of choosing 2^{-k} is that these numbers are summable. The n_k should be increasing with k. So for all k_0
$$m \bigcup_{k \geq k_0} \{x : f_{n_k}(x) \geq \delta\} \leq \sum_{k \geq k_0} 2^k = 2^{-k_0+1}$$
This union is decreasing as k_0 increases. So
$$m(\bigcap_{k=1}^{\infty} \bigcup_{k \geq k_0} \{x : f_{n_k}(x) \geq \delta\}) = \lim_{k_0 \to \infty} 2^{-k_0+1} = 0$$
That is,
$$m(\{x : \limsup_{k \to \infty} f_{n_k}(x) \geq \delta\}) = 0$$
which implies
$$\limsup_{k \to \infty} f_{n_k}(x) \leq \delta$$
But δ could be anything: we can repeat this process with a smaller δ. We get a nested series of subsequences:
$$\{n\} \supset \{n_k\} \supset \{n_k^{(2)}\} \supset \{n_k^{(3)}\} \supset \cdots$$

such that for all r
$$\limsup_{k\to\infty} f_{n_k^{(r)}}(x) \le 2^{-r}\delta$$
a.a.x. (The almost all thing works across the limits, because a countable union of null sets is null.) Now look at a diagonal sequence:

$$n_1^{(1)}, n_2^{(2)}, n_3^{(3)}, \cdots$$
We have, for a.a.x,
$$\limsup_{k\to\infty} f_{n_k^{(k)}}(x) = 0$$
i.e. (I think because they converge)
$$\lim_{k\to\infty} f_{n_k^{(k)}}(x) = 0$$

Now do the converse. Write $E_n \to E$ as $n \to \infty$ if $\lim \chi_{E_n} = \chi_E$ a.e. on $\mathbb{R}^d$. More verbosely, for all x there is some n_0 ("eventually") such that
$$\forall n \ge n_0, x \in E_n \iff x \in E$$
If there is some set F of finite measure such that $E, E_n \subset F$ for all n and $E_n \to E$ then $m(E_n) \to m(E)$. (This is because of the DCT: $\int \chi_{E_n} \to \int \chi_E$ where the dominating function is χ_E.)

Claim: *If $f_n \to f$ a.e. and Supp$(f_n) \subset F$ for all n where $m(F) < \infty$ then $f_n \to f$ in measure.* Fix $\varepsilon > 0$. Look at
$$E_n = \{x : |f(x) - f_n(x)| \ge \varepsilon\}$$
By hypothesis, $E_n \subset F$ and $E_n \to \emptyset$ as $n \to \infty$. So $m(E_n) \to 0$ as $n \to \infty$, i.e. $f_n \to f$ in measure.

$\boxminus$

8.1. Iterated integrals. When considering integration in $\mathbb{R}^2$, we have a problem of measurability: if $S \subset [0,1]$ is a non-measurable set in $\mathbb{R}$, but it is measurable after being embedded in $\mathbb{R}^2$ because it is null: it is contained in the null set $[0,1] \times \{0\}$. So if $f : \mathbb{R}^2 \to \mathbb{R}$ then it is possible for $x \mapsto f(x,y)$ not to be measurable when $y = 0$ but measurable for $y \neq 0$.

THEOREM 8.2. *(Fubini's theorem) Let $f \in \mathcal{L}^1(\mathbb{R}^2)$. Then*

(1) *for a.a. y the function $x \mapsto f(x,y)$ is in $\mathcal{L}^1(\mathbb{R})$; so the function*
$$F(y) = \int_{\mathbb{R}} f(x,y)\,dx$$
is defined a.a.y.

(2) *Defining $F(y)$ as above, $F \in \mathcal{L}^1(\mathbb{R})$ and $\int_{\mathbb{R}} F = \int_{\mathbb{R}^2} f$. That is,*
$$\int_{\mathbb{R}} F(y)\, dy = \int \int f(x,y)\, dx\, dy$$

The statement that this is integrable a.a.y means that it is measurable for a.a.y.

PROOF. We will say that f is "good" if the conclusions of Fubini's theorem hold for f.

(1) *If $R \subset \mathbb{R}^2$ is a rectangle, then χ_R is good.* This is straightforward: $\int R =$ *width* $\cdot$ *height*.
(2) *If f is a step function (a finite linear combination of characteristic functions of rectangles), then f is good.* This is just because an integral of sums is the sum of the integrals, etc. (Basically, linear combinations preserve goodness.)
(3) *Monotone limits of good functions are good.* Suppose $\{f_n\}$ is a sequence of good functions for $n \in \mathbb{N}$ and $f_n \nearrow f$ or $f_n \searrow f$, and $f \in \mathcal{L}^1(\mathbb{R}^2)$ then we claim that f is good. We will just do the case $f_n \nearrow f$. For all n, a.a.y, $x \mapsto f_n(x,y)$ is integrable. Let A be the set of $y$such that $x \mapsto f_n(x,y)$. (This is the union of countably many bad sets, one for each n; and A has null complement.) Define
$$F_n(y) = \begin{cases} \int f_n(x,y)\, \mathrm{d}x & y \in A \\ 0 & \text{else} \end{cases}$$
We have $f_1 \le f_2 \le \cdots$ by hypothesis. So for all n, we have
$$F_1(y) \le F_2(y) \le \cdots$$
Since f_n are good, then $F_n \in \mathcal{L}^1$. Also $\int_{\mathbb{R}} F_n = \int_{\mathbb{R}^2} f_n$. So the integrals $\int_{\mathbb{R}} F_n$ are increasing and bounded by the "constant" $\int_{\mathbb{R}^2} f$. By the MCT, a.a.y, $\lim_{n\to\infty} F_n(y)$ exists. That is, $F_n \nearrow G \in \mathcal{L}^1(\mathbb{R})$.

For $y \in B$, apply MCT to the functions
$$x \mapsto f_n(x,y)$$
(because $\int f_n(x,y)\, \mathrm{d}x$ is bounded above by $G(y)$). So for all $y \in B$ these functions converge a.a.x as $n \mapsto \infty$ to an integrable function (MCT). We know that
$$f_n(x,y) \mapsto f(x,y)$$
So $x \mapsto f(x,y)$ is an integrable function of x for all $y \in B$. So (1) holds for f. To do the second part:
$$\begin{aligned} \int_{\mathbb{R}} G &= \lim_{n\to\infty} \int_{\mathbb{R}} F_n \\ &= \lim \int_{\mathbb{R}^2} f_n \\ &= \int_{\mathbb{R}^2} f \end{aligned}$$

where MCT was invoked in the first and third lines. Recall $G(y) = \lim \int f_n(x, y)\,\mathrm{d}x = \int f(x, y)\,\mathrm{d}x = F(y)$. So

$$\int_{\mathbb{R}} F = \int G = \int_{\mathbb{R}^2 f}$$

which satisfies the second condition.

$\boxminus$

9. September 21

9.1. Finishing Fubini.

Recall, f is "good" if $f \in \mathcal{L}^1$ and

- a.a.y $\in \mathbb{R}$ the function $x \mapsto f(x, y)$ is integrable
- $F(y) := \int f(x, y)\,\mathrm{d}x$ is integrable and $\int \int f(x, y)\,\mathrm{d}x\,\mathrm{d}y := \int_{\mathbb{R}} F = \int_{\mathbb{R}^2} f$

We showed that χ_R, step functions, and (a.e.) monotone limits of step are good.

COROLLARY 9.1. *χ_E is good if E is*

- *a rectangle,*
- *a finite union of rectangles (this is a linear combination of characteristic functions, with the intersections subtracted),*
- *an open set $\mathcal{O}$ of finite measure(a countable union of rectangles; there are some K_n, each a finite union of rectangles, such that $\chi_{K_n} \nearrow \chi_{\mathcal{O}}$)*
- *a G_δ set G of finite measure (find open sets $\mathcal{O}_n$ with $\chi_{\mathcal{O}_n} \searrow \chi_G$)*
- *We would like to say that we can get to any measurable set of finite measure... Let E be a measurable set of finite measure. We know that $E \subset G$ with G a G_δ set such that $G \backslash E$ is null. $\chi_G = \chi_E$ a.e. and χ_G is good. We want to show that χ_E is good. But this is not obvious. But once this is proven, we have:*
- *simple*
- *If $f \in \mathcal{L}^1$ and $g \geq 0$ then there are simple functions g_n with $g_n \nearrow f$. So f is good.*
- *All measurable functions: if $f \in \mathcal{L}^1$ write $f = f_+ - f_-$.*

PROPOSITION 9.2. *If f is good and $f' = f$ a.e. then f' is also good.*

PROOF. To see that f' is good, we need to know: for a.a.y, the two functions

$$x \mapsto f(x, y)$$
$$x \mapsto f'(x, y)$$

are equal a.e. as functions of x. $f \neq f'$ on some $N \subset \mathbb{R}^2$ null. To prove the previous statement, we need to look at the intersection of N with the line $y = b$; these cross sections need to be null a.a. y. This will be done in the following lemma. $\boxminus$

Lemma 9.3. *If $N \subset \mathbb{R}^2$ is null then a.a. $b \in \mathbb{R}$ the set*

$$N^b = \{x : (x,b) \in N\} \subset \mathbb{R}$$

is null.

Proof. We will replace the null set by a slightly fatter null set. We have $N \subset \widehat{N}$ where $\widehat{N}$ is a G_δ set where $\widehat{N} \backslash N$ is null; i.e. $\widehat{N}$ is null. Apply Fubini to the "good" function $\chi_{\widehat{N}}$. So $\int_{\mathbb{R}^2} \chi_{\widehat{N}} = 0 = \int \int \chi_{\widehat{N}} \, \mathrm{d}x \, \mathrm{d}y$. So $\int \chi_{\widehat{N}}(x,y) \, \mathrm{d}x$ exists a.a. y i.e. $\widehat{N}^y$ is measurable a.a.y. And $\int m(\widehat{N}^y) \, \mathrm{d}y = 0$. So $m(\widehat{N}^y) = 0$ a.a.y. Since $N^b \subset \widehat{N}^b$ for all b and a subset of a null set is null, the lemma follows. ⊟

Remark 9.4. Fubini's theorem is not a criterion for integrability of f. Here is such a criterion, that is often combined with Fubini's theorem.

Theorem 9.5 (Tonelli's theorem). *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be measurable, and suppose the repeated integral $\int_{\mathbb{R}^2} |f(x,y)| \, dx \, dy$ exists (i.e. a.a.y $x \mapsto |f(x,y)|$ is in $\mathcal{L}^1(\mathbb{R})$ so $G(y) = \int |f(x,y)| \, dx$ exists a.e. and $\int G < \infty$).*

Then $f \in \mathcal{L}^1(\mathbb{R}^2)$

Proof. WLOG $f \geq 0$. To show something is integrable, look at the functions that lie below it, and show that their integrals are bounded above. That is, we must show that there is an upper bound C such that $\int g \leq C$ for all simple g which have $0 \leq g \leq f$ everywhere. By Fubini for g

$$\int g = \int \int g(x,y) \, \mathrm{d}x \, \mathrm{d}y$$

in the sense that $\int g(x,y) \, \mathrm{d}x$ exists a.a.y. We have

$$\int \int g(x,y) \, \mathrm{d}x \, \mathrm{d}y \leq \int \int |f(x,y)| \, \mathrm{d}x \, \mathrm{d}y = C$$

⊟

Remark 9.6. This is false without the absolute values: the values might cancel.

9.2. Littlewood's three principles.

Proposition 9.7. *(First principle)*

Any measurable set E of finite measure is "almost a finite union of rectangles"; i.e., for all $\varepsilon > 0$ there is some set K that is a finite union of rectangles, such that $m((E \backslash K) \cup (K \backslash E)) \leq \varepsilon$. (Call this the symmetric difference, and notate it $E \Theta K$.)

Proof. First approximate from the outside: there is some $\mathcal{O} \supset E$ open with $m(\mathcal{O} \backslash E) \leq \frac{\varepsilon}{2}$. Now approximate from the inside: there are some K_n, each a finite union of rectangles, with $K_n \nearrow \mathcal{O}$. Then $\mathcal{O} \backslash K_n \searrow \emptyset$. So $m(\mathcal{O} \backslash K_n) \to 0$ since $\mathcal{O}$ has finite measure. So

$m(\mathcal{O}\backslash K) \leq \frac{\varepsilon}{2}$ for some K. So

$$E\Theta K \subset (\mathcal{O}\backslash E) \cup (\mathcal{O}\backslash K) \to m(E\Theta K) \leq \varepsilon$$

$\boxminus$

PROPOSITION 9.8. *(Third principle) If $f_n \to f$ a.e. (f_n measurable) and if Supp(f_n), Supp(f) $\subset F$, where F is a set of finite measure then f_n converges to f almost uniformly: that is, for all $\varepsilon > 0$ there is some bad set $B \subset \mathbb{R}^d$ with $m(B) \leq \varepsilon$, such that*

$$f_n|_{\mathbb{R}^d\backslash B} \to f|_{\mathbb{R}^d\backslash B}$$

uniformly.

PROOF. $f_n \to f$ in measure (proven last time that a.e. convergence $\implies$ convergence in measure if everything is in a finite-measurable set). So for all k if

$$E_{n,k} = \{x \,:\, \exists n' \geq n \,:\, |f_{n'}(x) - f(x)| \geq \frac{1}{k}\}$$

then $\lim_{n\to\infty} m(E_{n,k}) \to 0$. ($x$ is eventually in none of these sets, a.a.x.) So there are some n_k depending on k such that $m(E_{n,k}) \leq 2^{-k}\varepsilon$ for $n \geq n_k$. So on $\mathbb{R}^d\backslash E_{n_k}$ we have

$$|f_n(x) - f(x)| \leq \frac{1}{k}$$

for all $n \geq n_k$. Set $B = \bigcup_1^\infty E_{n_k,k}$. Then $m(B) \leq \varepsilon$. For all k there is some n_k such that for all $x \in \mathbb{R}^d\backslash B$ we have $|f_n(x) - f(x)| \leq \frac{1}{k}$ for $n \geq n_k$.

$\boxminus$

10. September 23

THEOREM 10.1 (Littlewood's third principle/ Egorov's theorem). *If $f_n \to f$ a.e. and f_n, f measurable and Supp(f_n), Supp(f) $\subset F$ where F is some set of finite measure, then for all $\varepsilon > 0$, there is some B with $m(B) \leq \varepsilon$ such that*

$$f_n|_{\mathbb{R}^d\backslash B} \to f|_{\mathbb{R}^d\backslash B}$$

uniformly as $n \to \infty$.

THEOREM 10.2 (Littlewood's second principle / Lusin's theorem). *If $f : \mathbb{R}^d \to \mathbb{R}$ is measurable, and $\varepsilon > 0$, then there is some $B \subset \mathbb{R}^d$ with $m(B) < \varepsilon$ such that $f|_{\mathbb{R}^d\backslash B}$ is continuous from $\mathbb{R}^d\backslash B \to \mathbb{R}^d$. (i.e. for all $x \in \mathbb{R}^d\backslash B$ and $x_n \in \mathbb{R}^d\backslash B$ with $x_n \to x$, $n \to \infty$ we have $f(x_n) \to f(x)$.)*

REMARK 10.3. Note that this does NOT say that $f : \mathbb{R}^d \to \mathbb{R}$ is continuous at point of $\mathbb{R}^d\backslash B$. Take

$$f = \begin{cases} 1 & \text{on } \mathbb{Q} \subset \mathbb{R} \\ 0 & \text{elsewhere} \end{cases}$$

PROOF. It's true for $f = \chi_R$; for all ε you can use the rectangle minus its boundary. It's true for $f = \chi_K$ where K is a finite union of rectangles. It's true if f is simple. It's true for $f = \chi_E$ where E is measurable of finite measure. Why? For all ε, there is some finite union of rectangles K with

$$m(E\Theta K) < \varepsilon$$

Then we have $\chi_E = \chi_K$ except on a set $E\Theta K = B$ of measure $< \varepsilon$.

CLAIM: If $f_n \to f$ a.e., all supported inside of a set F of finite measure, and it then holds for all f_n, then it holds for f. Use Egorov. For all n, there is some B_n with $m(B_n) \leq 2^{-n-1}\varepsilon$ such that $f_n|_{\mathbb{R}^d \setminus B}$ continuous. $\forall f_n|_{\mathbb{R}^n \setminus C}$ continuous on $\mathbb{R}^n \setminus C$ where $C = \cup B_n$ and $m(C) \leq \frac{\varepsilon}{2}$. By Egorov, there is some D with $m(D) \leq \frac{\varepsilon}{2}$ such that $f_n \to f$ uniformly on $\mathbb{R}^d \setminus D$. $B - C \cup D$. On $\mathbb{R}^d \setminus B$ we have continuous f_n converging uniformly to f. So $f|_{\mathbb{R}^d \setminus B}$ is continuous too. So it holds for all f where $Supp(f)$ has finite measure.

General case: exercise. (Use $\frac{\varepsilon}{2^n}$'s.)

⊟

Suppose we want to find a measurable subset $E \subset [0,1] \times [0,1]$ with $m(E) = \frac{1}{2}$, such that for all rectangles $R \subset [0,1]^2$, $m(E \cap R) = \frac{1}{2}m(R)$. (So E is very uniformly distributed.) Sadly, there's no such set. The first principle tells you that measurable sets are not so far from a finite union of rectangles. Suppose E exists. Then for any finite union of rectangles $J \subset [0,1]^2$, $m(E \cap J) = \frac{1}{2}m(J)$. Littlewood's first principle says that there is some finite union of rectangles K such that $m(E\Theta K) \leq 0.01$. So $m(E \cap K)$ and $m(K)$ are within 0.02 of each other. That is,

$$m(E \cap K) \geq m(K) - 0.01$$

and $m(K) \geq 0.49$. There's no way that $m(E \cap K) = \frac{1}{2}m(K)$. (This result holds for all dimensions, not just 2.)

There's a version of this that looks like the fundamental theorem of calculus. First some notation.

NOTATION 10.4. Note that $\int_E f$ for $E \subset \mathbb{R}^d$ measurable means $\int_{\mathbb{R}^d} f \cdot \chi_E$. When $d = 1$ we use the familiar notation

$$\int_0^x f(y)\,\mathrm{d}y = \int f \cdot \chi_{[0,x]}$$

NOTATION 10.5. $\mathcal{L}^1_{loc}$ is the set of locally integrable functions: that is, the set of f such that f is integrable on any bounded measurable set.

THEOREM 10.6 (Lebesgue). *Let f be locally integrable. Define*

$$F(x) = \int_0^x f(y)\,dy$$

(if $x < 0$ this is $-\int_x^0$). Then F is differentiable a.e.; so $F'(x)$ is defined a.a.x and $F'(x) = f(x)$ a.a.x.

Let's use this to get another contradiction to our hypothetical evenly-distributed set E. If $f = \chi_E$ then define $F(x) = \int_0^x \chi_E$. $F(x) = \frac{1}{2}x$ and $F'(x) = \frac{1}{2}$ on $[0,1]$. $f = 0$ or 1 for all x, so this contradicts the theorem.

Can you swap the integral and the derivative, in the theorem? If $F : \mathbb{R}^d \to \mathbb{R}$ is continuous and differentiable a.e., and if the function $f = F'$ is locally integrable, then **it does not follow that** $\int_0^x f(y)\,\mathrm{d}y = F(x) - F(0)$.

Counterexample: Let F be the limit of functions $F_n : [0,1] \to [0,1]$ where $F_0(x) = x$, $F_1(x)$ is $\frac{3}{2}x$ on $[0, \frac{1}{3}]$, is $\frac{1}{2}$ on $[\frac{1}{3}, \frac{2}{3}]$, and then increases linearly to $(1,1)$ again on $[\frac{2}{3}, 1]$. Now iterate this construction: replace every strictly increasing interval with a piecewise function with three equal pieces, such that the function increases linearly on the first piece, stays constant on the second piece, and then increases linearly to the old high point on the last piece. We call this the *Cantor-Lebesgue function.* F is constant on each interval contained in $[0,1]\backslash C$, so $F' = 0$ on the interiors of all these intervals a.e. So the indefinite integral of the derivative $= \int 0 = 0$, which is not the original function.

Now back to the theorem. It says that a.a.$x \in \mathbb{R}$,

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = f(x)$$

You can rephrase this:

$$\lim \frac{F(b_i) - F(a_i)}{b_i - a_i} = f(x)$$

for any sequence of intervals $[a_i, b_i] \ni x$, with $|b_i - a_i| \to 0$. (This is equivalent to it being differentiable at that point.) Rewrite this with the aim to generalize to more dimensions:

$$\frac{\int_{[a_i,b_i]} f}{m([a_i,b_i])} \to f(x)$$

for any $[a_i, b_i] \ni x$ with $|b_i - a_i| \to 0$. This suggests the following d-dimensional version:

THEOREM 10.7 (Lebesgue differentiation theorem on $\mathbb{R}^d$). *Let $f \in \mathcal{L}^1_{loc}(\mathbb{R}^d)$. Then a.a.$x \in \mathbb{R}^d$, the following holds:*

$$\lim_{m(B) \to 0} \frac{1}{m(B)} \int_B f = f(x)$$

where the limit is over balls $B = \{y : |y - x_0| < r\}$ containing x. (i.e. for any sequence of balls $B_i \ni x$ with radius$(B_i) \to 0$, we have avg$(f; B_i) \to f(x)$, where avg$(f; B) = \frac{\int_B f}{\int_B 1}$.)

In one dimension, this implies the previous theorem (with the caveat of dealing with open vs. closed balls). In d dimensions, it's hard to interpret it as a statement about derivatives per se, but that's the spirit.

Apply this to $f = \chi_E$ where E is a measurable set. For a.a.x, we find

$$\lim_{\substack{m(B) \to 0 \\ B \ni x}} = \frac{m(E \cap B)}{m(B)} = \chi_E(x) \subset \{0,1\}$$

Again, we get a contradiction to our uniformly distributed set E: if you zoom in on a particular x, and ask how filled up your balls are, almost everywhere the answer is either 0 or 1.

11. September 26

In the homework: you can integrate $f : \mathbb{R}^d \to \mathbb{C}$ by treating $\mathbb{C}$ as $\mathbb{R}^2$: $a + bi = (a, b) \in \mathbb{R}^2$. Then the DCT says: for a sequence of integrable functions $f_n : \mathbb{R}^d \to \mathbb{C}$ with $f_n \to f$ a.e. and dominated by a (real) integrable function f ($|f_n| \leq g\ \forall n$) we have $\int f_n \to \int f$. This is exactly what you'd get if you applied the DCT to each part of $\mathbb{C}$ separately.

11.1. Hardy-Littlewood maximal function.

Guy plays one baseball game per week, and records the number of hits per week. This results in a discrete function, which we can think of as a simple function. What is your average "around" week$=n$? Choose a finite interval around n, not necessarily centered around n, and compute. What interval does he choose to maximize this "average?" We can define a new function: $h(x)$ is the number of hits in week x, and $h^*(x)$ is the "best" average over intervals containing x. We will be able to bound the best averages given a small total integral.

Let $B = \{x : |x - x_0| < r\}$; as we discussed last time, we can define the average

$$avg(f; B) = \frac{\int_B f}{m(B)}$$

for $f \in \mathcal{L}^1_{loc}$.

DEFINITION 11.1. If $h : \mathbb{R}^d \to \mathbb{R}$ with $h \geq 0$ locally integrable,

$$h^*(x) = \sup_{B \ni x} avg(h; B)$$

(This could be either nonnegative, or infinite, as when your function is $y = x^2$.)

THEOREM 11.2. *If $h : \mathbb{R}^d \to \mathbb{R}$ is nonnegative and integrable, then for any $\alpha > 0$,*

$$m(\{x : h^*(X) > \alpha\}) \leq \frac{3^d}{\alpha} \|h\|_L$$

(3 might not be the best constant. Also note that this really only works for balls*, not some weird ellipse things.)*

(So this bounds the number of weeks when his "best average" is $> \alpha$.)

COROLLARY 11.3. *If $h_n \to 0$ as $n \to \infty$ in L^1 norm, then $h_n^* \to 0$ in measure.*

To prove the theorem we need

LEMMA 11.4. *Suppose $B_1 \cdots B_N$ are balls in $\mathbb{R}^d$ and $m(\bigcup_1^N B_i) = \mu$. (That is, the balls may overlap.) Then there is a subcollection of these balls*

$$B_{i_1} \cdots B_{i_\ell}$$

which are disjoint and

$$m(\bigcup_{k=1}^{\ell} B_{i_k}) \geq \frac{\mu}{3^d}$$

PROOF. Be greedy. Let B_{i_1} be the ball of largest measure. Then remove from the collection this ball, and all those which intersect it. Let B_{i_2} be the largest remaining, and so forth, continuing until there are no balls. The claim is that at least $\frac{1}{9}$ of the total measure is contained in these balls. At step k, the balls meeting any B_{i_k} are all contained in $\widetilde{B}_i$, which is the ball with the same center and three times the radius. (Dilating a ball by a factor of 3 multiplies its measure by a factor of 3^d.) All of the original balls are contained in $\bigcup_{k=1}^{\ell} \widetilde{B}_{i_k}$.

$$\mu = m(\bigcup_1^N B_i) \leq m(\bigcup_1^{\ell} \widetilde{B}_{i_\ell}) \leq \sum m(\widetilde{B}_{i_k}) = \sum_{k=1}^{\ell} 3^d m(B_{i_k}) = 3^d m(\bigcup_{k=1}^{\ell} B_{i_k})$$

The greedy algorithm isn't necessarily the best selection to make; you can possibly reduce the constant 3. ⊟

PROOF. Let h be nonnegative and integrable. Fix $\alpha > 0$ and let

$$E = \{x : h^*(x) > \alpha\}$$

$h^*(x)$ is a supremum, so $h^*(x) > \alpha$ iff there is some $B \ni x$ with $avg(h; B) > \alpha$. So

$$E = \bigcup \{B : avg(h; B) > \alpha\}$$

This is an open set. Let $K \subset E$ be any compact set. Use the open cover of B as above. Then $K \subset B_1 \cup \cdots B_N$, where $avg(h; B_i) > \alpha$ for all i. The lemma says that we can find a subcollection $B_{i_1} \cdots B_{i_\ell}$ which are disjoint, so

$$\sum_1^{\ell} m(B_{i_k}) = m(\cup_1^{\ell} B_{i_k}) \geq \frac{1}{3^d} m(\bigcup_1^N B_i)$$

So

$$m(K) \leq m(\cup_1^N B_i) \leq 3^d \sum_{k=1}^{\ell} m(B_{i_k})$$

Now using the definition of avg, and the fact that it is $< \alpha$, we have

$$\cdots \leq 3^d \sum_{k=1}^{\ell} \frac{\int_{B_{i_k}} h}{\alpha} = \frac{3^d}{\alpha} \int_{\cup B_{i_k}} h \leq \frac{3^d}{\alpha} \|h\|_{L^1}$$

To summarize, we have

$$m(K) \leq \frac{3^d}{\alpha} \|h\|_{L^1}$$

Since E is open of bounded measure, we can take an increasing set of compact sets K that eventually approach E, so So

$$m(E) \leq \frac{3^d}{\alpha} \|h\|_{L^1}$$

$\boxminus$

THEOREM 11.5 (Lebesgue differentiation theorem). *Let f be locally integrable on $\mathbb{R}^d$. Then for a.a. $x \in \mathbb{R}^d$,*

$$\lim_{\substack{B \ni x \\ m(B) \to 0}} avg(f; B) = f(x)$$

(The idea is that you can recover the value at x by looking at the average in small balls around x.)

PROOF. This holds for all x if f is continuous. Given x, you can pick δ such that on $B_\delta \ni x$ the value of $f(x)$ does not deviate from $f(x)$ by more than ε. Furthermore, you can approximate any integrable f on $\mathbb{R}^d$ in the L^1 norm by continuous functions. That is,

CLAIM 11.6. *There are continuous $g_n \in \mathcal{L}^1$ such that*

$$\int |f - g_n| \to 0$$

PROOF. You can approximate f by simple functions, $f_n \to f$ a.e., where f_n are dominated by $|f|$. This ensures that $\int |f - f_n| \to 0$. Simple functions are $\sum a_i \chi_{E_i}$, so we will be done if we can approximate χ_{E_i} in the L^1 norm by continuous functions. We don't know how to do this, but we can approximate them by χ_K where K is a finite union of rectangles. So, it's enough to do this for a finite union of rectangles, or indeed just a single rectangle.

Given ε, find some f_ε such that $\int |f_\varepsilon - \chi_R| \to 0$ as $\varepsilon \to 0$. For example, choose f to be 0 outside the rectangle, a steep linear function on an ε-wide border of R, and 1 inside the rest of R.

$\boxminus$

Rest of proof: *to be continued*!

12. September 28

THEOREM 12.1 (Lebesgue Differentiation). *Let f be locally integrable on $\mathbb{R}^d$. Then a.a.x in $\mathbb{R}^d$,*

$$\lim_{\substack{B \ni x \\ m(B) \to 0}} avg(f; B) = f(x)$$

PROOF. The statement is "local," so we may replace f with a function that is zero outside some bounded set. The proof is by contradiction, so assume the conclusion fails for a set of nonzero measure. This could be either because the limit is something else, or because the limit doesn't exist. That is, either

- $\limsup_{\substack{B \ni x \\ m(B)\to 0}} avg(f;B) > f(x)$
- $\liminf_{\substack{B \ni x \\ m(B)\to 0}} avg(f;B) < f(x)$

(If both are false, then the statement is true.) Without loss of generality assume the first possibility holds on a set E of positive measure. That is, if E_k is the set where $\limsup avg(f;B) > f(x) + \frac{1}{k}$ then $E = \cup E_k$, so one of these E_k has positive measure as well. On some set E' of positive measure, $\limsup avg(f;B) > f(x) + \alpha$, where α is some fixed quantity > 0.

We can find continuous functions $g_n \to f$, where convergence is both in the L^1 norm, and a.e. (Convergence in norm implies convergence in measure; convergence in measure implies that a subsequence converges a.e.) The theorem holds for each g_n.

So $\limsup avg(f - g_n; B) > f(x) - g_n(x) + \alpha$. Define $h_n = |f - g_n|$.

$$\begin{aligned}
\limsup avg(h_n; B) &\geq \limsup |avg(f - g_n; B)| \\
&\geq \limsup avg(f - g_n; B) \\
&> f(x) - g_n(x) + \alpha \\
&\geq \alpha - |f(x) - g_n(x)| \\
&= \alpha - h_n(x)
\end{aligned}$$

Recall $h_n^*(x) = \sup_{B \ni x} avg(h_n; B) \geq \limsup_{\substack{B \ni x \\ m(B)\to 0}} avg(h_n; B) \geq \alpha - h_n(x)$. So for $x \in E'$ (the set of positive measure),

$$h_n^*(x) + h_n(x) \geq \alpha$$

So $h_n \to 0$ in the L^1 norm, which implies that $h_n \to 0$ in measure. Using the corollary to the Hardy-Littlewood thing, we know that this implies $h_n^*(n) \to 0$ in measure. This contradicts the previous statement that $h_n^*(x) + h_n(x) \geq \alpha$. ⊟

Why did we do this in the first place? The definition of the Hardy-Littlewood maximal function used open balls. But you can use closed balls instead. The only difference is that there are some open balls that don't contain x, but the closure does. But by continuity of measure, the average on the closed ball can be approximated by the average on open balls that are slightly larger than it. The average of $f : \mathbb{R} \to \mathbb{R}$ on $[x, x+h]$ is

$$\frac{F(x+h) - F(x)}{h}$$

COROLLARY 12.2. *If $f : \mathbb{R} \to \mathbb{R}$ is in $\mathcal{L}^1_{loc}$ and F is its indefinite integral, then $F'(x)$ exists and equals $f(x)$ a.a.$x \in \mathbb{R}$.*

We could also apply this to $f = \chi_E$ for $E \subset \mathbb{R}^d$.

COROLLARY 12.3 (Lebesgue density theorem). *For a.a.$x \in E$,*

$$\lim_{\substack{B \ni x \\ m(B) \to 0}} \frac{m(E \cap B)}{m(B)} \to 1$$

That is, the "density" of E at x is 1, and this fails on the "boundary."

THEOREM 12.4. *If $f \in \mathcal{L}^1_{loc}(\mathbb{R}^d)$, then a.a.$x_0 \in \mathbb{R}^d$,*

$$\lim_{\substack{B \ni x_0 \\ m(B) \to 0}} avg(|f - f(x_0)|, B) = 0$$

(If you took away the absolute values, this would be the same as before.) To clarify, this says:

$$\lim_{B \ni x_0} \frac{\int_B |f(x) - f(x_0)|\, dx}{m(B)} = 0$$

For any $r \in \mathbb{Q}$ look at $h(r) = |f - r|$ and apply the result of the previous theorem: for all $x \in G_r$, where $(\mathbb{R}^d \backslash G_r)$ is null,

$$\lim avg(|f - r|, B) = |f(x_0) - r| \tag{4}$$

Let $G = \bigcap_{r \in \mathbb{Q}} G_r$, so $\mathbb{R}^d \backslash G$ is still null (the union of countably many null sets is null). (4) holds for all $r \in \mathbb{Q}$ if $x_0 \in G$. Given $\varepsilon > 0$ and $x_0 \in G$ we can find $r \in \mathbb{Q}$ with

$$|f(x_0) - r| \le \frac{\varepsilon}{2}.$$

Given a sequence of balls B_k with $m(B_k) \to 0$ $x_0 \ni B_k$ we have $avg(f - r; B_k)$ is eventually $< \frac{\varepsilon}{2}$ for $k \ge k_0$. So $avg(|f - f(x_0)|; B) \le \varepsilon$ for $k \ge k_0$.

DEFINITION 12.5. $x_0 \in \mathbb{R}^d$ is a Lebesgue point of the function f if

$$\lim_{\substack{m(B) \to 0 \\ B \ni x_0}} avg(|f(x) - f(x_0)|; x \in B) = 0$$

So the theorem says that almost every point is a Lebesgue point. (Sometimes this is defined without the absolute value signs.)

Week from Friday: Kronheimer will not be in class.

13. September 30

13.1. Convolution.

$$(f * g)(x) = \int f(x - y) g(y)\, \mathrm{d}y$$

For $f, g \in \mathcal{L}^1(\mathbb{R})$ (or $\mathcal{L}^1(\mathbb{R}^d)$),

$$y \mapsto f(x - y) g(y)$$

is an $\mathcal{L}^1$ function at y for a.a.x. (This was in the HW, and it comes out of Fubini's theorem.)

- $f * g \in L^1$
- $\|f * g\|_{L^1} \leq \|f\|_{L^1}\|g\|_{L^1}$
- If $|f| \leq C$ everywhere, then $|f * g| \leq C\|g\|_{L^1}$ for all x. (This is because the integral is bounded by $|g|$.)

Define $Z_1 = \frac{1}{2}\chi_{[-1,1]}$. Similarly, define $Z_\varepsilon = \frac{1}{2\varepsilon}\chi_{[-\varepsilon,\varepsilon]}$; we can write $Z_\varepsilon = \frac{1}{\varepsilon}Z_1(\frac{x}{\varepsilon})$. Look at $f * Z_\varepsilon$ for $f \in \mathcal{L}^1$. Then

$$(f * Z_\varepsilon)(x) = \int f(x-y)Z_\varepsilon(y)\,\mathrm{d}y$$

Looking at where things are nonzero,

$$\begin{aligned}\int f(x+y)Z_\varepsilon(-y)\,\mathrm{d}y &= \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} f(x+y)\,\mathrm{d}y \\ &= avg(f;[x-\varepsilon,x+\varepsilon]) \to f(x) \text{ as } \varepsilon \to 0\end{aligned}$$

if x is a Lebesgue point. This proves

PROPOSITION 13.1. *$f * Z_\varepsilon \to f$ a.e. for all $\varepsilon > 0$*

REMARK 13.2. There is no $g \in \mathcal{L}^1$ such that $f * g = f$ a.e. for all f.

It is also true that $f * Z_\varepsilon \to f$ in the L^1 norm as $\varepsilon \to 0$.

PROOF. It's true if $f = \chi_{[a,b]}$. The convolution looks like the original f, with the interval $[a-\varepsilon, a+\varepsilon]$ turned into an increasing line, and there is a downward line from $[b-\varepsilon, b+\varepsilon]$. So you turn a box into a trapezoid; think about Z_ε moving, and getting an average on that ε-interval as you go along. It's also true for step functions.

Given $f \in \mathcal{L}^1$ and $\eta > 0$, we can find f and a step function with

$$\|f - g\|_{L^1} \leq \frac{\eta}{3}$$

Find ε_0 so that for all $\delta \leq \delta_0$

$$\begin{aligned}\|g - g * Z_\varepsilon\|_{L^1} &\leq \frac{\eta}{3} \\ \|f * Z_\varepsilon - g * Z_\varepsilon\|_{L^1} &= \|(f-g) * Z_\varepsilon\| \\ &\leq \|f-g\|\|Z_\varepsilon\| \\ &= \|f-g\| \leq \frac{\eta}{3}\end{aligned}$$

$\boxminus$

Note that for any $f \in \mathcal{L}^1$, $f * Z_\varepsilon$ is continuous. $F(x) = \int_0^x f(y)\,\mathrm{d}y$ is continuous in x.

13.2. Integration by parts.

TERMINOLOGY 13.3. We say that F is *differentiable in the L^1 sense* (with derivative f) if $f \in \mathcal{L}^1_{loc}$ and

$$F(b) - F(a) = \int_a^b f$$

for all a, b.

This implies that F is differentiable a.e. and has a.e. derivative f. (But the converse of this implication is FALSE!)

THEOREM 13.4. *Suppose F, G are differentiable in the L^1 sense, with derivative $f, g \in \mathcal{L}^1_{loc}$. Then*

$$\int_a^b fG = -\int Fg + [FG]_a^b$$

PROOF. Use Fubini. (Problem set.) ⊟

Recall, from the first lecture, that we want "measure" for $E \subset \mathbb{R}^d$ to be invariant under rigid motions. More generally, integrals are invariant under rigid motions.

THEOREM 13.5. *For $A : \mathbb{R}^d \to \mathbb{R}^d$ linear, invertible, and $b \in \mathbb{R}^d$ and $f \in \mathcal{L}^1(\mathbb{R}^d)$*

$$\int f(Ax + b)\, dx = \frac{1}{|det(A)|} \int f(x)\, dx$$

(A rigid motion is a transformation of determinant 1.)

PROOF. Section/ course notes. ⊟

13.3. Fourier transforms. For now, $\mathcal{L}^1(\mathbb{R}^d)$ denotes the integrable functions $f : \mathbb{R}^d \to \mathbb{C}$. Ditto, $L^1(\mathbb{R}^d)$ denotes equivalence classes of such functions. Given $f \in \mathcal{L}^1(\mathbb{R})$, the Fourier transform is a new function

$$\widehat{f}(\xi) = \int e^{-2\pi i \xi x} f(x)\, \mathrm{d}x$$

We have $|f| = |e^{-2\pi i \xi x} f(x)|$, and the integrand, as a function of x, is in $\mathcal{L}^1$.

EXAMPLE 13.6. Take $f = \chi_{[-1,1]}$. Then

$$\widehat{f}(\xi) = \int_{-1}^1 e^{-2\pi i \xi x}\, \mathrm{d}x = [\frac{e^{-2\pi i \xi x}}{-2\pi i \xi}]_{-1}^1$$

$$= \frac{\sin(2\pi \xi)}{\pi \xi}$$

Take $Z_1 = \frac{1}{2}\chi_{[-1,1]}$. Then $\widehat{X_1} = \frac{\sin(2\pi\xi)}{2\pi\xi}$ (which is 1 at $\xi = 0$, by L'Hospital's rule).

$$\widehat{f}(0) = \int_{\mathbb{R}} f$$

In $\dim d$ $f \in \mathcal{L}^1(\mathbb{R}^d)$ and $\xi \in \mathbb{R}^d$

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} e^{-2\pi x\xi} f(x)\,\mathrm{d}x$$

Note that the Fourier transform is not necessarily integrable; for example, $\frac{\sin(2\pi\xi)}{\pi\xi}$ seen previously. Basically, it doesn't decay fast enough; the decay is of the order $\frac{1}{|\xi|}$. Take the absolute value, and try to integrate; it diverges.

FACTS 13.7.

- For a constant $c \in \mathbb{R}$, $f(x+c)$ has Fourier transform $e^{2\pi ic\xi}\widehat{f}(\xi)$. (Basically, make a linear substitution $x \mapsto x - c$ in $\int e^{-2\pi ix\xi} f(x+c)\,\mathrm{d}x$.)
- The Fourier transform of $f(Ax)$ is $\frac{1}{|A|}\widehat{f}(A^{-1}\xi)$ ($f \in \mathcal{L}^1(\mathbb{R})$). Or, in more dimensions, you could interpret A as a matrix. You figure it out.
- For all ξ, $|\widehat{f}(\xi)| \leq \|f\|_{L^1}$, so $\widehat{f}$ is a bounded function of ξ.
- $\widehat{f}$ is continuous. (This is an application of DCT: if $\xi_n \to \xi$ as $n \to \infty$, then $\int e^{-2\pi ix\xi} f(x)\,\mathrm{d}x \to \int e^{-2\pi ix\xi} f(x)\,\mathrm{d}x$, because it converges pointwise.)

LEMMA 13.8 (Riemann-Lebesgue lemma). *For $f \in \mathcal{L}^1$, $\widehat{f}(\xi) \to 0$ as $|\xi| \to \infty$.*

PROOF. First, it's true for $f = \chi_{[a,b]}$: use the earlier calculation for $\chi_{[-1,1]}$ and translate/dilate. So, it's true for step functions. Given $f \in \mathcal{L}^1$ and $\eta > 0$ first approximate f by a step function g, so that $\|f - g\|_{L^1} \leq \frac{\eta}{3}$. $\widehat{g}(\xi) \to 0$ as $|\xi| \to \infty$ (R-L lemma for step functions), so there exists some R so that $|\widehat{g}(\xi)| \leq \frac{\eta}{3}$ for $|\xi| \geq R$. For $|\xi| > R$,

$$|\widehat{f}(\xi) - \widehat{g}(\xi)| \leq \|(f - g)\|_{L^1} \leq \frac{\eta}{3}$$

So $|\widehat{f}(\xi)| \leq \frac{2\eta}{3}$ for $|\xi| > R$. ⊟

No class next Friday. Midterm 2 weeks from Monday.

14. October 3

PROPOSITION 14.1.

(1) *Suppose $F \in L^1(\mathbb{R})$ and is differentiable in the L^1 sense with derivative (a.e.) $f(x) = \frac{d}{dx}F(x) \in \mathcal{L}^1$. Then the Fourier transform of f (i.e. $\widehat{\frac{d}{dx}F}$) is $2\pi i\xi\widehat{F}(\xi)$.*

(2) *Suppose $f \in \mathcal{L}^1$ and $xf(x) \in \mathcal{L}^1$. Then $\widehat{f}(\xi)$ is differentiable and $\frac{d}{d\xi}\widehat{f}(\xi) = -\widehat{2\pi ixf}(x)$.*

PROOF. $\lim_{R\to\infty} F(R)$ exists, because it's $\lim_{R\to\infty}(F(R)-F(0)) = \lim_{R\to\infty}\int_0^R f = \int_0^\infty f$. Ditto for $\lim F(-R) : F \in \mathcal{L}^1 \implies \lim_{R\to\pm\infty} F(R) = 0$.

$$\begin{aligned}
\widehat{f}(\xi) &= \int e^{-2\pi i x\xi} f(x)dx \\
&= \lim_{R\to\infty}\int_{-R}^{R} e^{-2\pi i x\xi} f(x)dx \\
&\overset{parts}{=} \lim_{R\to\infty} -\int_{-R}^{R}(-2\pi i\xi)e^{-2\pi i x\xi}F(x)dx + [e^{-2\pi i x\xi}F(x)]_{-R}^{R} \\
&- \lim_{R\to\infty} 2\pi i\xi\int_{-R}^{R} e^{-2\pi i x\xi}F(x)dx = 2\pi i\xi\widehat{F}(x\xi)
\end{aligned}$$

Second one:

$$\begin{aligned}
\frac{d}{d\xi}\widehat{f}(\xi) &= \lim_{h\to 0}\frac{1}{h}\left(\int (e^{-2\pi i(\xi+h)x} - e^{-2\pi i\xi x})f(x)\right)dx \\
&= \lim_{h\to 0}\int e^{-2\pi i x\xi}(\frac{e^{-2\pi i x h}}{xh} - 1)f(x) \\
&= \int e^{-2\pi i x\xi}(-2\pi i)xf(x)dx
\end{aligned}$$

Think about $\theta = xh$ and $\frac{e^{-2\pi i\theta}-1}{\theta} \to -2\pi i$ as $\theta \to 0$. (Take the ratio of the arc length to the straight line between points on a circle; this approaches 1.) Use the DCT with $2\pi|xf(x)|$ as the dominating function.

SO all of this is the Fourier transform of $-2\pi i x f(x)$ at ξ. $\boxminus$

We're relating the integrability of $xf(x)$ with the summability of $\sum_{-\infty}^{\infty} n|a_n|$ where $a_n = \int_n^{n+1}|f|$.

14.1. Schwartz Space.

DEFINITION 14.2. f is *rapidly decreasing* if for all $k \geq 0$, $|x^k f(x)| \to 0$ as $x \mapsto \pm\infty$ (i.e. it beats polynomial decay in general).

EXAMPLE 14.3. The Gaussian: $G(x) = e^{-\pi x^2}$

Define $\mathcal{S}(\mathbb{R})$ to be the Schwartz space: the space of infinitely differentiable functions $f : \mathbb{R} \to \mathbb{C}$ such that f and $(\frac{d}{dx})^n f(x)$ are rapidly decreasing.

PROPOSITION 14.4. *The following are equivalent:*

1. $f \in \mathcal{S}(\mathbb{R})$
2. $x^k f^{(n)}(x) \to 0$ *as* $x \to \pm\infty$, *for all* $k, n \geq 0$
3. $(\frac{d}{dx})^n(x^k f(x)) \to 0$ *(the equivalence of this and the previous just comes from how you differentiate things)*

(4) *f is C^∞ and $x^k f^{(n)}(x) \in L^1$ for all n, k*
(5) *f is C^∞ and $(\frac{d}{dx})^n (x^k f(x)) \in L^1$ for all n, k*

PROOF. (2) $\iff$ (3) by Leibniz rule.

(4) $\iff$ (5) also by Leibniz.

(2) $\iff$ (4). As an example, if $f, f' \in L^1$ then $f(x) \to 0$ as $x \to \infty$. The indefinite integral of an L^1 function approaches a constant as $x \to \infty$. Now do this a bunch of times; it gets (2) $\to$ (4). For example, if $|x^2 f(x)| \to 0$ as $x \to \pm\infty$ and f is continuous, then $f \in L^1$, because $\frac{1}{x^2}$ is integrable. $\boxminus$

The equivalence of (4) and (5), plus the first of the two propositions, gives $f \in \mathcal{S}(\mathbb{R})$ implies $\widehat{f} \in \mathcal{S}(\mathbb{R})$. The key example is the Gaussian. The normalization $G(x) = e^{-\pi x^2}$ is set up so that the integral is 1.

PROOF.

$$\begin{aligned}
(\int G)^2 &= \int G(x)G(y)dxdy \\
&= \int e^{-\pi(x^2+y^2)} dxdy \\
&= \int e^{-\pi r^2} r dr d\theta
\end{aligned}$$

Now we can do this by substituting $s = r^2$ and $rdr = \frac{1}{2} ds$. $\boxminus$

To calculate the FT of G note that $\frac{d}{dx} G(x) = -2\pi x G(x)$. Now taking the FT we have $2\pi i \xi \widehat{G}(\xi) = \frac{1}{i} \frac{d}{d\xi} \widehat{G}(\xi)$.

$$\frac{d}{d\xi} \widehat{G}(\xi) = -2\pi i \xi \widehat{G}(\xi)$$

is not hard to solve:

$$\widehat{G}(\xi) = ce^{-\pi \xi^2}$$

for some c. But we do know that $\widehat{G}(0) = \int G = 1$, so the constant is 1. So

$$\widehat{G} = G$$

From last time, $\chi_{[-\frac{1}{2}, \frac{1}{2}]}$ has FT $\frac{\sin(\pi\xi)}{\pi\xi}$.

Now consider the tent function $x + 1$ on $[-1, 0]$ and $1 - x$ on $[0, 1]$; alternatively,

$$H(x) = \begin{cases} 1 - |x| & \text{if } |x| < 1 \\ 0 & \text{else} \end{cases}$$

We could differentiate, etc. Or regard it as $\chi_{[-\frac{1}{2},\frac{1}{2}]} * \chi_{[-\frac{1}{2},\frac{1}{2}]}$ (just think of this as averages on small intervals).

The FT takes convolutions to products. So

$$\widehat{H}(\xi) = \left(\frac{\sin \pi\xi}{\pi\xi}\right)^2$$

This function is called the Féjer kernel. Let $F_1 = (\frac{\sin \pi x}{\pi x})^2$, and in general, $F_R(x) = R(\frac{\sin \pi R x}{\pi R x})^2$. (The scaling is set up so that $\int F_R = \int F_1$); so it takes the value R at 0 and vanishes at $\frac{n}{R}$.

14.2. Inverting the Fourier transform.

DEFINITION 14.5. For $F \in L^1(\mathbb{R})$ define the inverse Fourier transform of F to be

$$\check{f}(\xi) = \int e^{2\pi i x\xi} f(x)\,\mathrm{d}x$$

Is this actually an inverse? We've already seen that if $f \in L^1$, that does not imply that $\widehat{f} \in L^1$. So we can't always compute/ define $(\widehat{f})^\vee$. Also $\check{f}$ is continuous if $f \in L^1$. So we can't expect to recover all f directly. Eventually, we will prove:

THEOREM 14.6. *If $f \in \mathcal{L}^1$ and continuous* and $\widehat{f} \in L^1$, *then* $\widehat{f}^\vee = f$.

Note: if you have a function that looks like $\int K(x,y)f(x)\,\mathrm{d}x$, then $K(x,y)$ is referred to as the *kernel*.

15. October 5

15.1. Pointwise convergence in the inversion of the Fourier transformation.

Motivation: Suppose we have $\sum_{-\infty}^{\infty} a_n$ with $|a_n| \to 0$ as $|n| \to \infty$. Absolute convergence is if $\sum_{-\infty}^{\infty} |a_n| < \infty$. Convergence as a series is if $s_N = \sum_{-N}^{N} a_n \to$ a limit as $N \to \infty$.

DEFINITION 15.1. A series is Cesaro summable if when we define

$$\sigma_n = \frac{1}{N}\sum_{1}^{N} s_n$$

then $\sigma_N \to$ a limit as $N \to \infty$. This is basically taking the average value over a large range.

This is even more susceptible to rearrangements of terms. Absolute convergence implies convergence as a series implies Cesaro summability.

EXAMPLE 15.2. Let $a_n = (-1)^n$. Then s_n is 1 or -1 (n even or odd). Then this is Cesaro summable, because $\sum \sigma_n = \sum (-1)^n \frac{1}{n}$ and the alternating harmonic function converges.

These relate to integrability. Absolute convergence is like ordinary integrability: $\int_{\mathbb{R}} |f| < \infty$. Or, we could take $s_R = \int_{-R}^{R} f$ and ask whether $s_R \to$ a limit as $R \to \infty$. Or, define

$$\sigma_R = \int_0^R s_r \, \mathrm{d}r$$

and ask if σ_R converges to a limit as R is large. For example, the sawtooth function is integrable in this last sense. Write

$$s_R = \int F \cdot \chi_{[-R,R]}$$
$$\sigma_R = \frac{1}{R} \int F \cdot H^R$$

where

$$H^R(x) = \frac{1}{r} \int_0^R \chi_{[-r,r]}(x) \, \mathrm{d}r = \begin{cases} 1 - \frac{|x|}{R} & \text{if } |x| < R \\ 0 & \text{else} \end{cases}$$

is the tent function.

The same implications hold as in the sum case: integrability implies s_r converge implies σ_R converge.

If $F \in \mathcal{L}^1(\mathbb{R})$ then $\sigma^R \to \int F$ as $R \to \infty$, by the DCT.

15.2. Fourier inversion. Start with $f \in \mathcal{L}^1(\mathbb{R})$. Construct the Fourier transform

$$\widehat{f}(\xi) = \int e^{-2\pi i x \xi} f(x) dx$$

But this might not be in $\mathcal{L}^1$. We want

$$f(x) = \int e^{2\pi i x \xi} \widehat{f}(\xi) \, \mathrm{d}x$$

but this might not exist. Apply the same strategies as before. Define

$$s_R(f)(x) = \int_{-R}^{R} e^{2\pi i x \xi} \widehat{f}(\xi) \, \mathrm{d}x$$

Question: Does $s_R(f) \to f$ as $R \to \infty$ maybe a.e.? NO. You can find f such that this happens *nowhere*. But we can define the Cesaro mean:

$$\sigma_R(f)(x) = \int_{\mathbb{R}} e^{2\pi i x \xi} \widehat{f}(x) H^R(x) \, \mathrm{d}x$$

This is the average of $s_r(x)$ for $0 \leq r \leq R$.

THEOREM 15.3. *For $f \in \mathcal{L}^1(\mathbb{R})$ and any point x_0 in the Lebesgue set of f (i.e. for a.a.x_0) these Cesaro means converge to the correct value:*

$$\sigma_R(f)(x_0) \to f(x_0) \quad \textit{as } R \to \infty$$

COROLLARY 15.4.

(1) *f can be recovered from its Fourier transform; i.e. if $\widehat{f} = 0$ then $f(x) = 0$ a.e.*
(2) *If $f \in \mathcal{L}^1(\mathbb{R})$ and $\widehat{f} \in \mathcal{L}^1(\mathbb{R})$ then we can define the inverse Fourier transform as last time, and it's equal to $f(x)$.*

REMARK 15.5. $\widehat{f}^\vee$ is automatically a continuous function. So in order for this to work, the starting function must be a.e. equal to a continuous function. So we can restate this: if $f \in L^1$ and is continuous, and $\widehat{f} \in \mathcal{L}^1$, then $\widehat{f}^\vee = f$ everywhere.

Recall that H^1 has Fourier transform

$$F_1(x) = \left(\frac{\sin(\pi x)}{\pi x}\right)^2$$

and H^R has Fourier transform

$$F_R(x) = R\left(\frac{\sin(\pi R x)}{\pi R x}\right)^2$$

$$\sigma_R(f)(x_0) = \int_R \int e^{2\pi i x_0 \xi} e^{-2\pi i y \xi} f(y) H^R(\xi)\,\mathrm{d}y\,\mathrm{d}\xi$$

But $|integrand| = |H^R(\xi)| \cdot |f(y)| \in \mathcal{L}^1(\mathbb{R}^2)$ and it's absolutely integrable. So we can apply Fubini:

$$\begin{aligned}
\cdots &= \int \int e^{2\pi i x_0 \xi} e^{-2\pi i y \xi} f(y) H^R(\xi)\,\mathrm{d}\xi\,\mathrm{d}y \\
&= \int \left(\int e^{2\pi i \xi (x-y)} H^R(\xi)\,\mathrm{d}\xi\right) f(y)\,\mathrm{d}y \\
&= \int \widehat{H^R} f(y)\,\mathrm{d}y \\
&= \int F_R(y - x_0) f(y)\,\mathrm{d}y \\
&= F_R(x_0 - y) f(y)\,\mathrm{d}y = (F_R * f)(x_0)
\end{aligned}$$

So the theorem says

COROLLARY 15.6. *$(F_R * f) \to f$ at all Lebesgue points of f.*

This should remind us of $Z_\varepsilon * f \to f$ at Lebesgue points of f as $\varepsilon \to 0$, where ε is playing the role of $\frac{1}{R}$.

We want general conditions on the family of functions $K^\delta(x)$ to ensure that

$$K_\delta * f \to f$$

as $f \to 0$ at all Lebesgue points of f. Unneeded assumption: assume that K_δ is obtained by scaling.

$$K_\delta(x) = \frac{1}{\delta} K_1\left(\frac{x}{\delta}\right)$$

We also want $\int K_1 = 1$ (and $K_1 \in \mathcal{L}^1$). The scaling is set up so that $\int K_\delta = 1$.

REMARK 15.7. $\int F_1 = 1$, which is not quite obvious. (Later.)

Look at

$$\begin{aligned}|(K_\delta * f)(x_0) - f(x_0)| &= |\int K_\delta(y)(f(x_0 - y) - f(x_0))\,\mathrm{d}y| \\ &\leq \int |K_\delta(y)| \cdot |f(x_0 - y) - f(x_0)|\,\mathrm{d}y\end{aligned}$$

We seek conditions on K_1 such that

$$\int |K_\delta(y)| \cdot |f(x_0 - y) - f(x_0)|\,\mathrm{d}y \to 0 \tag{5}$$

as $\delta \to 0$.

LEMMA 15.8. (5) *holds if x_0 is a Lebesgue point of f provided K_1 satisfies*

$$|K_1| \leq \sum_{k=1}^{\infty} a_k Z_{r_k}$$

where $a_k \geq 0$ and $\sum |a_k| < \infty$ and r_k are increasing as k increases.

Recall $Z_1 = \frac{1}{2}\chi_{[-1,1]}$ and $Z_{r_1} = \frac{1}{2r_1}\chi_{[-1,1]}$ so $\frac{1}{2}\sum_1^\infty \frac{a_k}{r_k}$ is a pyramidal step function where the step gaps are $\frac{1}{2}\frac{a_1}{r_1}, \frac{1}{2}\frac{a_2}{r_2}$. Since $\sum a_k < \infty$ we have $\sum a_k Z_{r_k} \in \mathcal{L}^1$ by MCT. In (5)

$$K^1 \leq \sum_1^\infty a_k Z_{r_k}(y)$$

so

$$\begin{aligned}LHS &\leq \int |f(x_0 - y) - f(x_0)| \cdot (\sum_1^\infty a_k Z_{\delta r_k})\,\mathrm{d}y \\ &= \sum_{k=1}^\infty a_k \int |f(x_0 - y) - f(x_0)| \cdot Z_{\delta r_k}(y)\,\mathrm{d}y \\ &= \sum_1^\infty a_k \cdot \mathcal{A}(\delta r_k)\end{aligned}$$

where $\mathcal{A}(r)$ is the average value of $|f(x_0 - y) - f(x_0)|$ for $y \in [-r, r]$. (An integrable function times something bounded is an integrable function. So by the MCT we can interchange the two things.)

Next time we will finish the proof by showing that

$$\sum a_k \mathcal{A}(\delta r_k) \to 0$$

as $\delta \to 0$.

16. October 12

We had an $\mathcal{L}^1$ function $f : \mathbb{R} \to \mathbb{C}$, and x_0 was a Lebesgue point. We had defined the averages:

$$\mathcal{A}(r) = \frac{1}{2r} \int_{x_0-r}^{x_0+r} |f(x) - f(x_0)| \, \mathrm{d}x$$

We were looking at expressions of the form

$$\alpha(\delta) = \sum_{k=1}^{\infty} a_k \mathcal{A}(\delta r_k)$$

where $\sum |a_k| < \infty$ and r_k increasing.

Lemma 16.1. *$\alpha(\delta) \to 0$ as $\delta \to 0$*

Proof. $\mathcal{A}(r)$ is continuous on $r > 0$ (indefinite integral is a continuous function of the endpoints). Also $\mathcal{A}(r) \to 0$ as $r \to 0$ (this is a corollary of being a Lebesgue point). As $r \to \infty$, $\mathcal{A}(r) \to |f(x_0)|$ (you're taking the average value on a shrinking set, so the error is decreasing). So $\mathcal{A}(r)$ extends to a continuous function on $[0, \infty)$, with $\mathcal{A}(0) = 0$ and $\mathcal{A}(r) \leq M$ (it approaches $|f(x_0)|$, and you should think of $f \to 0$ as $r \to \infty$ because it is integrable).

So the sum

$$\sum a_k \mathcal{A}(\delta r_k)$$

as a function of δ, is uniformly convergent, and the k^{th} term is $\leq a_k M$ for all δ. So it defines a continuous function on $[0, \infty)$ at δ nd is zero at $\delta = 0$. That is,

$$\lim_{\delta \to 0} \sum a_k \mathcal{A}(\delta r_k) = 0$$

Recall where this came from. If K_1 is a function which satisfies $K_1 \leq \sum a_k Z_{r_k}$ (where $\sum |a_k| < \infty$ and the r_k are increasing), and

$$K_\delta(x) = \frac{1}{\delta} K_1(\frac{x}{\delta})$$

then

$$\int |f(x_0 - u) - f(x_0)| |K_\delta(y)| \, \mathrm{d}y \to 0 \text{ as } \delta \to 0$$

K_1 lies under a summable simple function, and the further K_δ get increasingly spiky, and go to zero on most of the real line. ⊟

Proposition 16.2. *If $K_\delta = \frac{1}{\delta} K_1(\frac{x}{\delta})$ and K_1 satisfies $|K_1| \leq J$ where J is bounded, integrable, and a decreasing function of $|x|$ then the same conclusion holds:*

$$\int |f(x_0 - u) - f(x_0)| |K_\delta(y)| \, dy \to 0 \text{ as } \delta \to 0$$

The idea is that any such J is $\leq \sum a_k Z_{r_k}$ for suitable summable coefficients a_k and increasing r_k. Why? If you make the original step functions by dividing up the real line into $0 < r_1 < r_2 < \cdots$, use the same divisions r_i but make the boxes just as large as necessary for J to fit inside. (The difference between this step function and the original ones will go to zero.) Since $\int J < \infty$ then $\sum |a_k| < \infty$; this is equivalent to saying that the box function has finite integral, equivalently the a_k are summable.

COROLLARY 16.3. *If K_1 satisfies these hypotheses, and $\int K_1 = 1$, then $(f * K_\delta)(x_0) \to f(x_0)$ for all Lebesgue points x_0. (The $\int K_1 = 1$ hypothesis is to ensure that you're taking properly normalized averages.)*

COROLLARY 16.4. *For the Fejer kernel F_R, assume $\int F_1 = 1$; then we have*

$$f * F_R \to f$$

as $R \to \infty$ at Lebesgue points.

Recall

$$F_1 = \left(\frac{\sin \pi x}{\pi x}\right)^2 \leq \begin{cases} \frac{1}{(\pi x)^2} & \text{if } x \geq 1 \\ \mathit{constant} & \text{if } x \leq 1 \end{cases}$$

Hence $\sigma_R(f) \to f$ a.e. (at Lebesgue points)

$$\sigma_R = \int e^{2\pi i x \xi} \widehat{f}(\xi) H^R \, \mathrm{d}\xi$$

where H^R was the tent function with support $[-R, R]$ and apex 1. Convergence happens, in particular, at all points of continuity of f:

$$\sigma_k(f)(x_0) \to f(x_0)$$

What's so great about H^R? We could look instead at

$$\gamma_R(f)(x) \int e^{2\pi i x \xi} \widehat{f}(\xi) \underbrace{e^{-\pi (x/R)^2}}_{G(x/R)} \, \mathrm{d}\xi$$

We could have done the same thing for these: instead of F_R (the Fourier transform of the tent), use G_R, the transform of the Gaussian. Also, we want to scale the Gaussian:

$$G_R(x) = R \cdot G_1(Rx)$$

$$\gamma_R(f) = f * G_R$$

(To see that this expression converges, recall that $\widehat{f}(\xi)$ is bounded, so you can integrate on $[-R, R]$.) Our general criterion about $\leq J$ tells us

$$f * G_R \to f$$

at Lebesgue points. What's important is the shape of the tent-function or its replacement: it has to be integrable, etc.

Earlier, we compared

$$\int f(x_0 - y) F_R(y) \, \mathrm{d}y \to f(x_0)$$

$$\int (f(x_0 - y) - f(x_0)) F_R(y) \, \mathrm{d}y$$

so why can you move $f(x_0)$ inside? F_R has some integral…Suppose the first thing converged to $mf(x_0)$ instead. If $m \neq 1$ we can prove

$$\sigma_R(f) \to mf$$

But then you could take the inverse Fourier transform, and this m would crop up again! $(\widehat{f})^\vee = mf$. We know that $\widehat{G} = G = G^\vee$, so there's no magic constant m.

Remark 16.5. We've only talked about the one-dimensional case for Fourier inversion. But, replacing multiplication by dot product, we can do this in d dimensions:

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} e^{-2\pi i \xi \cdot x} f(x)\,\mathrm{d}x$$

Define

$$H^1(\xi) = \begin{cases} 1 - |\xi| & \text{if } |\xi| \leq 1 \\ 0 & \text{else} \end{cases}$$

Look at Cesaro means:

$$\sigma_R(f) = \int e^{2\pi i x \cdot \xi} \widehat{f}(\xi) H^R(\xi)\,\mathrm{d}\xi$$

Does $\sigma_R(f) \to f$ a.e.? This is not guaranteed if $d \geq 3$. Look at $\widehat{H}$: this is the Fejer kernel in $d = 1$: what was important is that it was integrable (there was an $\frac{1}{x^2}$ in the denominator). It's also in L^1 for $d = 2$, but for $d = 3$ it doesn't decay fast enough: $\widehat{H} \notin L^1$. So we don't have an integrable J dominating the Fejer kernel. In fact, the theorem is false for $d \geq 3$.

If you have $f = \chi_E$, and ask whether $\sigma_R(f) \to f$ on E (i.e. you're asking whether it $\to 1$). You could have a problem at points when the normals all focus inwards. In higher dimensions, these oscillate wildly as R increases. So the tent function doesn't work…but the Gaussian does.

$$\gamma_R(f)(x) \to f(x)$$

does converge at Lebesgue points.

$$\gamma_R(f) = \int e^{2\pi i x \cdot \xi} \widehat{f}(\xi) \cdot e^{-\pi |\xi/R|^2}\,\mathrm{d}\xi = f * G_R$$

where $G_R = RG(Rx)$. This works because G_R is integrable.

17. October 17

Theorem 17.1. *Let $f : \mathbb{R} \to \mathbb{C}$ be in $\mathcal{L}^1$ and $x^- \in \mathbb{R}$ a Lebesgue point, and K_δ a family of functions $\delta \in (0, 1]$ on $\delta = \delta_1 \cdots \delta_2, \cdots$ with $\delta_k \to 0$ then*

$$(f * K_\delta)(x_0) \to f(x_0)$$

provided $\int K_\delta = 1$ for all δ, there exists J such that

$$|K_\delta(x)| \leq \frac{1}{\delta} J(\frac{x}{\delta})$$

for all x, δ where J is integrable, bounded (a consequence of the other conditions since the maximum is at zero), and a decreasing function of $|x|$.

We can also do this in d dimensions: $f : \mathbb{R}^d \to \mathbb{C}$*, and just change the condition on* J *to*

$$|K_\delta(x)| \leq \frac{1}{\delta^d} J(\frac{x}{\delta})$$

For example, take $J = \min(A, |x|^{-d-0.001})$ (i.e. it's cut off so as to avoid the asymptote). Instead of $K_\delta = \frac{1}{\delta} K_1(\frac{x}{\delta})$ we have

$$K_\delta(x) \leq A/\delta^d$$

or in the example

$$K_\delta(x) \leq \frac{C}{\delta^d} |\frac{x}{\delta}|^{-(d+0.001)}$$

(Last time we were assuming that the boundedness was in K_δ, but it just needs to be bounded above by some J that satisfies the conditions, as above.)

17.1. Fourier series. Unless otherwise stated, "periodic" means that f satisfies

$$f(x) = f(x+1)$$

for all x. Equivalently, you could think of $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ or $\mathbb{R}$. Write $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Write $\mathcal{L}^1(\mathbb{T})$ for the set of measurable functions f which are periodic with

$$\int_0^1 |f| < \infty$$

Similarly to the real case, we have an L^1-norm

$$\|f\|_{L^1(\mathbb{T})} = \int_0^1 |f|$$

Now do Fourier analysis. $e^{2\pi i n x}$ is periodic.

Definition 17.2. Define the Fourier coefficients of $f \in \mathcal{L}^1(\mathbb{T})$ to be

$$\widehat{f}(n) = \int_0^1 f(x) e^{-2\pi i n x} \, \mathrm{d}x$$

for $n \in \mathbb{Z}$.

The idea is to think of $f(x)$ as being recovered from $\sum_{-\infty}^{\infty} \widehat{f}(n) e^{2\pi i n x}$. If we can write $f(x) = \sum_{-\infty}^{\infty} a_n e^{2\pi i n x}$ where the sum is absolutely convergent, then the coefficients a_n have to be given by that formula:

$$\widehat{f}(m) = \int_0^n (\sum a_n e^{2\pi i n x}) e^{-2\pi i m x} \, \mathrm{d}x$$

By the DCT we can interchange the sum with the integral:

$$\begin{aligned} \widehat{f}(m) &= \sum_{n=-\infty}^{\infty} \int_0^1 a_n (e^{2\pi i (n-m)}) \, \mathrm{d}x \\ &= \sum_n a_n \delta_{n=m} = a_m \end{aligned}$$

where $\delta_{n=m} = \begin{cases} 1 & n = m \\ 0 & \text{else} \end{cases}$.

PROPOSITION 17.3 (Properties of $\widehat{f}(n)$).

(1) $|\widehat{f}(n)| \leq \|f\|_{L^1(\mathbb{T})}$
(2) $\widehat{f}(n) \to 0$ *as* $n \to \pm\infty$ *(This is the Riemann-Lebesgue lemma for the Fourier series world; you obtain it by direct computation in the step function case, and then take limits.)*
(3) *Suppose F is periodic and differentiable in the L^1 sense with derivative f (i.e. F is an indefinite integral $F(x) = \int_0^x f + C$ where $f \in L^1(\mathbb{T})$ and $\int_0^1 f = 0$). Then* $\widehat{f}(n) = 2\pi in\widehat{F}(n)$. $\widehat{f}(0) = \int_0^1 f = 0$.

$$\begin{aligned} \widehat{F}(n) = \int_0^1 F(x)e^{-2\pi inx}\,dx \\ &= -\int_0^1 f(x)\frac{e^{-2\pi inx}}{-2\pi in}\,dx + \left[F(x)e^{-2\pi inx}\right]_0^1 \\ &= \frac{1}{2\pi in}\widehat{f}(n) \end{aligned}$$

So if a function has a derivative in the L^1 sense, then this means it decays faster.

$$\widehat{F}(n) = \frac{1}{2\pi in}\widehat{f}(n)$$

If F is k times continuously differentiable, then

$$n^k|\widehat{F}(n)| \to 0$$

as $n \to \infty$. In particular, this ensures that the Fourier series is summable.

We can also think of $L^1(\mathbb{T})$ as containing functions $f \in L^1$ that are supported in $[0, 1]$.

17.2. Square wave.

$$f = \begin{cases} 1 & [-\frac{1}{4}, \frac{1}{4}] \\ -1 & [\frac{1}{4}, \frac{3}{4}] \end{cases}$$

$$\widehat{f}(n) = \begin{cases} 0 & n \text{ even} \\ \frac{2(-1)^{\frac{n-1}{2}}}{n\pi} & n \text{ odd} \end{cases}$$

Taking the partial sums, you get a bumpy thing where the straight $y = 1$ region should be, etc. As you add more terms, the height of the bumps stays the same, but they get more frequent.

17.3. Triangle wave. This is the indefinite integral of the square wave. In the Fourier approximation, the corners are slightly rounded. So the Fourier coefficients are the indefinite integral of the previous ones, and so they decay as $\frac{1}{cn^2}$.

17.4. Cesaro sums. Define

$$(s_n f) = \sum_{-m=-n}^{n} \widehat{f}(m)e^{2\pi imx}$$

and the Cesaro means:

$$(\sigma_n f) = \frac{1}{n}(s_0 f + \cdots s_{n-1} f)$$

Theorem 17.4. *If $f \in \mathcal{L}^1(\mathbb{T})$ and x_0 is a Lebesgue point of f then*

$$(\sigma_n f)(x_0) \to f(x_0)$$

as $n \to \infty$.

This is easier with the hypothesis that f is continuous. Very hard:

Theorem 17.5. *If f is continuous and periodic, then $s_n f \to f$ a.e. as $n \to \infty$.*

Corollary 17.6. *If x_0 is a Lebesgue point of f, and $(s_n f)(x_0) \to c$ as $n \to \infty$ (i.e. if the Fourier series does converge) then $c = f(x_0)$ (i.e. then the Fourier series converges to the right thing).*

We know that $s_n f(x_0) \to c$ implies $\sigma_n f(x_0) \to f(x_0)$ (i.e. convergence of partial sums implies the convergence of the Cesaro means to the same limit).

$$\begin{aligned}
(s_n f)(x) &= \sum_{m=-n}^{n} \left(\int_0^1 f(y)e^{-2\pi imy}\,\mathrm{d}y \right) e^{2\pi imx} \\
&= \int_0^1 \underbrace{\left(\sum_{-n}^{n} e^{2\pi im(x-y)} \right)}_{D_n(x-y)} f(y)\,\mathrm{d}y \\
&= (D_n *_{\mathbb{T}} f)(x)
\end{aligned}$$

where $(g *_{\mathbb{T}} f)(x)) = \int_0^1 g(x-y)f(y)\,\mathrm{d}y$. By Fubini's theorem, this might not exist for all x, but it exists for a.a.x.

$$D_n(y) = \sum_{-n}^{n} e^{2\pi imy} = \frac{\sin(\pi(2n+1)y)}{\sin(\pi y)}$$

This is the *Dirichlet kernel.*

$$\begin{aligned}
\sigma_n f &= \frac{1}{n}(s_0 f + \cdots + s_{m-1} f) \\
&= \frac{1}{n}(D_0 + \cdots + D_{n-1}) * f \\
&= F_n * f
\end{aligned}$$

where F_n is the Fejer kernel

$$F_n = \frac{1}{n}\sum_0^{n-1} \frac{\sin(\pi(2m+1)y)}{\sin(\pi y)} = \frac{1}{n}\left(\frac{\sin(\pi n y)}{\sin(\pi y)}\right)^2$$

Proving the theorem comes down to a statement about convolutions:

$$(F_n * f)(x_0) \to f(x_0)$$

if x_0 is Lebesgue point. The Fejer kernel is periodic. $\int_0^1 F_n = 1$ because it is the average of D_n's, and those integrate to 1 (look at definition). For example take $x_0 = 0$. We claim that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} F_n(y) f(y)\, \mathrm{d}y \to f(0)$$

as $n \to \infty$. So we can forget about everything outside $[-\frac{1}{2}, \frac{1}{2}]$ and we can instead consider

$$\widetilde{F}_n = \chi_{[-\frac{1}{2},\frac{1}{2}]} F_n$$
$$\widetilde{f}_n = \chi_{[-\frac{1}{2},\frac{1}{2}]} f_n$$

So now the previous claim is

$$\int_{\mathbb{R}} \widetilde{F}_n(y) \widetilde{f}(y)\, \mathrm{d}y \to f(0)$$

This follows from today's first theorem. The parameter that was previously $\delta \to 0$ is now $n \to \infty$; set $\delta = \frac{1}{n}$ and we have an inequality

$$|\widetilde{F}_n(x)| \leq \frac{1}{\delta} J(\frac{x}{\delta})$$

We could define

$$J(x) = \begin{cases} A & |x| \leq 1 \\ A|x|^{-2} & |x| \geq 1 \end{cases}$$

so $\int_{\mathbb{R}} J < \infty$. The difference between this Fejer kernel and the old one is having $\sin \pi y$ in the denominator, instead of πy.

18. October 19

We have seen for $f \in \mathcal{L}^1(\mathbb{T})$ and $(s_n f) = \sum_{-n}^n \widehat{f}(m) e^{2\pi i m x}$, $\sigma_n f = \frac{1}{n}\sum_0^{n-1} s_n f$ that $\sigma_n \to f$ a.e.

THEOREM 18.1 (Dini's Criterion). *Let $f \in \mathcal{L}^1(\mathbb{T})$ and $x_0 \in \mathbb{R}$. Suppose that*

$$\frac{f(x) - f(x_0)}{x - x_0}$$

is an integrable function of x on $[x_0 - \frac{1}{2}, x_0 + \frac{1}{2}]$. Then

$$(s_n f)(x_0) \to f(x_0)$$

as $n \to \infty$.

REMARK 18.2.

- It's enough to test integrability on $[x_0-\delta, x_0+\delta]$
- The hypothesis holds if f is differentiable at x_0 (the quotient is bounded in a neighborhood of x_0 because the quotient is convergin to $f'(x_0)$. Or , if f is continuous at x_0 and the left and right derivatives exist.)
- If f is Hölder continuous at x_0 with exponent $\alpha \in (0,1)$. What?

  DEFINITION 18.3. A function is Hölder continuous at x_0 with exponent α if
  $$|f(x)-f(x_0)| \leq C \cdot |x-x_0|^\alpha$$
  for $x \in [x_0-\delta, x_0+\delta]$.

  (For example, $f(x) = \sqrt{|x|}$ with any exponent $\alpha \leq \frac{1}{2}$.) Dini's criterion is:
  $$\frac{C\cdot|x-x_0|^\alpha}{|x-x_0|} = C\cdot|x-x_0|^{-1+\alpha}$$
  so it's OK where $\alpha < -1$ (?).
- Instead of the normal Dini criterion, look at
  $$\frac{f(x)-f(x_0)}{\sin(\pi(x-x_0))}$$
  because we can bound the derivatives in the interval $[-\frac{1}{2}, \frac{1}{2}]$: $2y \leq \sin(\pi y) \leq \pi y$.

PROOF. We will use the last criterion.
$$\begin{aligned}
(s_n f)(x_0) &= (f *_{\mathbb{T}} D_n)(x_0) \\
&= \int_{\frac{-1}{2}}^{\frac{1}{2}} f(x_0-y)D_n(y)\,\mathrm{d}y \\
&\overset{y\mapsto -y}{=} \int_{\frac{-1}{2}}^{\frac{1}{2}} f(y+x_0)\frac{\sin((2n+1)\pi y)}{\sin(\pi y)}\,\mathrm{d}y \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} (f(x_0-y)-f(x_0))D_n(y)\,\mathrm{d}y \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} (\frac{f(y+x_0)-f(x_0)}{\sin(\pi y)})\sin((2n+1)\pi y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}} h(y)\sin((2n+1)\pi y)\,\mathrm{d}y
\end{aligned}$$
where $h(y) = (\frac{f(y+x_0)-f(x_0)}{\sin(\pi y)}) \cdot \chi_{[-\frac{1}{2},\frac{1}{2}]}(y)$. The criterion just asks for h to be integrable. This approaches zero as $n \to \infty$ by the Riemann Lebesgue lemma, because $\widehat{\xi} \to 0$ as $|\xi| \to \infty$. $\boxminus$

Note that the converse is not true, and continuity is not enough of a hypothesis here.

18.1. Poisson summation formula. The idea is that

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \widehat{f}(n)$$

but the left hand side doesn't even make sense, because you can change finitely many values of f and get the same Fourier transform.

Given $f \in \mathcal{L}^1(\mathbb{R})$ we can try to form a periodic function F by

$$F(x) = \sum_{n=-\infty}^{\infty} f(x+n)$$

Does the RHS converge? Because the functions are periodic, look at $[0,1]$. Define

$$f_n(x) = \begin{cases} f(x+n) & x \in [0,1] \\ 0 & \text{else} \end{cases}$$

$\int_0^1 |f_n| = \int_n^{n+1} |f|$ so $\sum_n \int_0^1 |f_n| = \sum_n \int_n^{n+1} |f|$ and the RHS converges to $\int_{\mathbb{R}} |f|$. This means that $\sum \|f_n\|_{L^1} < \infty$. Recall that this implies the partial sums of $\sum f_n$ converge a.e. and in L^1 norm, to some $F \in L^1([0,1])$. So $F(x)$ makes sense (the RHS is defined for a.a.x.)

Let $F \in L^1(\mathbb{T})$. Suppose that $\sum_{-\infty}^{\infty} f(x+n)$ converges at $x=0$; call this $F(0)$. The LHS in the Poisson summation formula is $F(0)$. On the RHS,

$$\begin{aligned} \widehat{f}(n) &= \int_{-\infty}^{\infty} f(x) e^{-2\pi i x n} \, \mathrm{d}x \\ &= \sum_{m=-\infty}^{\infty} \int_0^1 f_m(x) e^{-2\pi i n x} \, \mathrm{d}x \\ &= \int_0^1 (\sum_m f_m(x)) e^{-2\pi i n x} \, \mathrm{d}x \end{aligned}$$

because the partial sums $\sum_{m=-M}^{M} f_m$ converge in the L^1 norm to $\sum f_m = F$. SO this is

$$\cdots = \int_0^1 F(x) e^{-2\pi i n x} \, \mathrm{d}x = \widehat{F}(n)$$

The RHS of the PSF (without further hypotheses) is just $\sum \widehat{F}(n)$, the Fourier coefficients of the periodic function $F \in L^1(\mathbb{T})$. So the PSF is trying to say

$$F(0) = \sum_{-\infty}^{\infty} \widehat{F}(n)$$

This holds for any $F \in \mathcal{L}^1(\mathbb{T})$ satisfying Dini's criterion at $x_0 = 0$. To recap:

THEOREM 18.4 (Poisson summation formula). *If $\sum_{-\infty}^{\infty} f(n)$ converges, and f satisfies Dini's criterion at $x=0$, then*

$$\sum_{-\infty}^{\infty} f(n) = \sum_{-\infty}^{\infty} \widehat{f}(n)$$

EXAMPLE 18.5. Let $f(x) = e^{-\pi(x/R)^2}$. Recall $\widehat{\xi} = Re^{-\pi(Rx)^2}$. We get that
$$\sum_n e^{-\pi(n/R)^2} = R\sum_n e^{-\pi(Rn)^2}$$
for $R > 0$. Change coordinates so that $R = \frac{1}{\sqrt{t}}$. So this turns into
$$\sum_n e^{-\pi n^2 t} = \frac{1}{\sqrt{t}}\sum_{-\infty}^{\infty} e^{-\pi n^2/t}$$
Writing $\theta(t)$ for the LHS, this can be written $\theta(t) = \frac{1}{\sqrt{t}}\theta(\frac{1}{t})$. This is a version of Jacobi's identity for θ.

We can apply the PSF to $e^{-\pi(Rx)^2}e^{2\pi iax}$. The Fourier transform just gets translated by a.
$$\sum e_{-\pi n^2 t}e^{2\pi ina} = \frac{1}{Nt}\sum e^{-\frac{\pi(n-a)^2}{t}}$$
Now the LHS is a function of a and t, and the RHS is a different expression for the same thing. This is the fundamental solution of the heat equation.

19. October 21

19.1. Norms and Banach spaces. Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$ today. X will be a (possibly infinite-dimensional) vector space over $\mathbb{F}$.

DEFINITION 19.1. A *norm* on X is a function $X \to \mathbb{R}$, $x \mapsto \|x\|$, such that

(1) $\|x\| = 0 \iff x = 0$
(2) $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbb{F}$, $x \in X$
(3) (Triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$

It follows from (2) and (3) that the norm is always nonnegative. (We will use $|f|$ to mean absolute value, not norm.)

DEFINITION 19.2. A normed vector space is a pair $(X, \|\cdot\|)$ of a vector space with a norm on it.

DEFINITION 19.3. The distance between x, y is $d(x, y) = \|x - y\|$. This makes X into a metric space. So we have notions of open and closed that come from X being a metric space: closed balls are
$$B(x;\delta) = \{y : \|x - y\| \leq \delta\}$$
S is open iff for all $x \in S$, there is some $\delta > 0$ such that $B(x;\delta) \subset S$.

It is possible for a space to have two different norms $\|\|$ and $\|\|'$ on X; we say these are equivalent if there are constants $C_1, C_2 > 0$ such that $C_1\|x\| \leq \|x\|' \leq C_2\|x\|$. Equivalent norms define the same topology.

Example 19.4. C_0 is the set of sequences $\{(a_n)_{n\in\mathbb{N}} : a_n \in \mathbb{F}\}$ which converge to zero as $n \to \infty$. This is a vector space. The norm on this space is

$$\|(a)\|_{C_0} = \sup_n |a_n|$$

$$\ell^\infty = \{(a_n) : a_n \in \mathbb{F}, n \in \mathbb{N}, \sup_n |a_n| < \infty\}$$

where the norm is $\|a\|_\infty = \sup_n |a_n|$.

$$\ell^p = \{(a_n) : \sum_n |a_n|^p < \infty\}$$

where the norm is $\|(a)\|_{\ell^p} = (\sum |a_n|^p)^{\frac{1}{p}}$. Why is this a vector space? We need $\sum |a_n + b_n|^p < \infty$. We also need the triangle inequality:

$$\left(\sum |a_n|^p\right)^{\frac{1}{p}} + \left(\sum |b_n|^p\right)^{\frac{1}{p}} \geq \left(\sum |a_n + b_n|^p\right)^{\frac{1}{p}}$$

This is the Minkowski inequality.

It doesn't matter if the sequences are indexed by $\mathbb{N}$ or $\mathbb{Z}$, or any countable set. To avoid confusion, you can write $\ell^p(\mathbb{N})$ to mean the $\mathbb{N}$-indexed ones, etc.

Example 19.5. Let Ω be a compact metric space. Then define $C(\Omega)$ to be the set of continuous functions $f : \Omega \to \mathbb{F}$. We use the supremum norm: $\|f\| = \sup_{\omega\in\Omega} |f(\omega)|$.

Let $C_0(\mathbb{R}^d)$ be the set of continuous functions $f : \mathbb{R}^d \to \mathbb{F}$ such that $f(x) \to 0$ as $|x| \to \infty$. We again take the supremum norm: $\|f\| = \sup_{\mathbb{R}^d} |f|$.

Example 19.6. We have already seen L^1. For $p \geq 1$ define $L^p(\mathbb{R}^d)$ to be the space of equivalence classes of measurable functions $f : \mathbb{R}^d \to \mathbb{F}$ such that $\int |f|^p < \infty$. Use the norm $\|f\|_p = \left(\int |f|^p\right)^{\frac{1}{p}}$. Later, we will check Minkowski's inequality, which will give the triangle inequality. (As before use $\mathcal{L}^p$ for the functions themselves, and L^p for the equivalence classes.)

We can also define $L^\infty(\mathbb{R}^d)$ to be the set of equivalence classes of functions $f : \mathbb{R}^d \to \mathbb{F}$ which are measurable and *essentially bounded*, i.e. there is some null set N for which $f|_{\mathbb{R}^d\setminus N}$ is bounded. Equivalently, f is equivalent to a bounded function. For norm use the *essential supremum*

$$\|f\|_\infty = \sup\{M : |f| \leq M a.e.\}$$

(Even when the function is actually bounded, we still use the essential supremum norm.)

Example 19.7. $X = \mathbb{F}$ is a one-dimensional vector space, where $\|x\| = |x|$ (absolute value or complex modulus, depending on $\mathbb{F}$).

$\mathbb{F}^n$ can be a normed space, with norms $(\sum |x_i|^p)^{\frac{1}{p}}$ (if $p = 2$ this is the usual Euclidean norm) or $\max |x_i|$, which is like the ℓ^∞ norm. You can also take an ellipse, and consider the norm that makes it "a ball of radius 1."

All the norms given have the additional property that they are complete.

DEFINITION 19.8. A sequence x_n, for $n \in \mathbb{N}$ and $x_n \in X_n$, is Cauchy if:

$$\forall \varepsilon > 0,\ \exists n_0 \text{ such that } \forall n, n' \geq n_0, \quad \|x_n - x_{n'}\| \leq \varepsilon$$

A sequence is convergent if there is some $x \in X$ such that $\|x_n - x\| \to 0$ as $n \to \infty$.

A metric space is *complete* if every Cauchy sequence in X is convergent.

A *Banach space* is a complete, normed space.

All of our examples above are complete. But this is not entirely obvious: it's easy to see this for the ones based on the sup norm, but for the others you have to do something; we will do this when we do the Minkowski inequality.

19.2. A criterion for completeness.

DEFINITION 19.9. A series $\sum_0^\infty t_n$, for $t_n \in X$ (a normed space), is called *absolutely summable*, if $\sum_0^\infty \|t_n\| < \infty$. A series is *convergent* if the partial sums $s_n = \sum_0^n t_m$ are a convergent sequence.

PROPOSITION 19.10. *A normed space X is complete iff every absolutely summable series is convergent.*

Having partial sums converge is NOT the same as being Cauchy.

PROOF. (Just the "if" direction.) Let x_n be a Cauchy sequence in X, $x_n \in \mathbb{N}$. For each such j there exists some n_j such that $n, n' \geq n_j$ then $\|x_n - x_{n'}\| \leq 2^{-j}$. Arrange $n_j \leq n_{j+1}$. Consider $x_{n_1}, x_{n_2}, \cdots$. Write $t_j = x_{n_j} - x_{n_{j-1}}$, where $t_1 = x_{n_1}$. So the sequence of partials sums of (t_j) is the sequence (x_{n_j}). We have $\|t_j\| \leq 2^{-(j-1)}$ for $j > 1$ so $\sum \|t_j\| < \infty$. Now use the hypothesis: $\sum_{j=1}^J t_j \to x \in X$ as $J \to \infty$. That is, $x_{n_J} \to x \in X$ as $J \to \infty$. So we have shown that there is a convergent subsequence.

The next claim is that a Cauchy sequence with a convergent subsequence is itself convergent, to the same limit.

$\boxminus$

20. October 24

20.1. Young's inequality. This is a version of the "arithmetic - geometric mean."

THEOREM 20.1 (Young's inequality). *For $0 \leq \lambda \leq 1$ and $a, b > 0$:*

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1-\lambda) b$$

PROOF. Take logs of both sides:

$$\lambda \log a + (1-\lambda)\log b \le \log(\lambda a + (1-\lambda)b)$$

This is a statement that the graph of the logarithm restricted to $[a, b]$ lies above the line joining the two points. Analytically, this is because the second derivative is negative. ⊟

THEOREM 20.2 (Hölder's inequality). *If $p > 1$ and $q > 1$ are* dual exponents *(i.e. $\frac{1}{p} + \frac{1}{q} = 1$) and if $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$ then $fg \in L^1(\mathbb{R}^d)$ and*

$$\left|\int fg\right| \le \|f\|_p \cdot \|g\|_q$$

If $p = q = 2$ this is called Cauchy-Schwartz.

PROOF. Without loss of generality we may take $\|f\|_p = 1$ and $\|g\|_q = 1$, and $f, g \ge 0$. Write $a = f^p$ and $b = g^q$.

$$\begin{aligned}
\int fg &= \int a^{\frac{1}{p}} b^{\frac{1}{q}} \\
&\le \int \left(\frac{1}{p}a + \frac{1}{q}b\right) && Young, \lambda = \frac{1}{p} \\
&= \frac{1}{p}\int a + \frac{1}{q}\int b \\
&= \frac{1}{p}\int f^p + \frac{1}{q}\int g^q \\
&= \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p\|g\|_q
\end{aligned}$$

⊟

THEOREM 20.3 (Triangle inequality for L^p). *If $1 \le p \le \infty$ and $f, g \in L^p$ then $f + g \in L^p$ and*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p$$

This doesn't work for $p < 1$. Assume $p < \infty$.

PROOF. Let q be the dual exponent. Suppose $f, g \ge 0$ (worst case scenario).

$$\begin{aligned}
\|f+g\|_p^p = \int (f+g)^p &= \int f(f+g)^{p-1} + \int g(f+g)^{p-1} \\
&\le \left(\int f^p\right)^{\frac{1}{p}} \left(\int (f+g)^{(p-1)q}\right)^{\frac{1}{q}} + \left(\int g^p\right)^{\frac{1}{p}} \left(\int (f+g)^{(p-1)q}\right)^{\frac{1}{q}} \\
&= (\|f\|_p + \|g\|_p)\left(\int (f+g)^p\right)^{\frac{p-1}{p}} && pq = p + q \\
&= (\|f\|_p + \|g\|_p)\|f+g\|_p^{p-1}
\end{aligned}$$

So we get

$$\|f+g\|_p \le \|f\|_p + \|g\|_p$$

$\boxminus$

Thus $L^p(\mathbb{R}^d)$ is a normed space.

THEOREM 20.4. *$L^p(\mathbb{R}^d)$ is a complete normed (i.e. Banach) space.*

This is kind of the whole point of defining the Lebesgue integral (as opposed to the Riemann integral.)

PROOF. By last time, it's enough to show that all absolutely summable series are convergent. So let $f_n \in L^p$ for all $n \in \mathbb{N}$ and consider $\sum_n f_n$ and suppose that

$$\sum_n \|f_n\|_p = C < \infty$$

Let $s_n = \sum_{m=1}^n f_m$. We must show that there is some $s \in L^p$ such that $s_n \to s$ in the L^p norm. First apply the DCT. Let $t_n = \sum_{m=1}^n |f_m|$. Then t_n are in L^p (since $f \in L^p$ and it's a sum of L^p functions). Minkowski's inequality says that $\|t_n\|_p \leq \sum_{m=1}^n \|f_m\|_p$. In particular,

$$\left(\int t_n^p\right)^{\frac{1}{p}} \leq C$$

so $\int t_n^p \leq C^p$. Now apply the MCT to t_n^p: these are integrable for each n, and they are an increasing sequence, pointwise, by definition. So there is some limit t^p, where $t_n^p \nearrow t^p$ and $\int t^p \leq C^p$. That is, for a.a.x,

$$\sum_{m=1}^{\infty} |f_m(x)| \leq t(x) < \infty$$

This says that $\sum_1^\infty f_n(x)$ exists for a.a. x: that is, there is some s such that $s_n(x) \to s(x)$ a.a.x. and $|s| \leq \|t\|$ which implies $s \in L^p$.

This gives convergence a.e., not convergence in norm. Just apply the DCT to see that the convergence is also in norm. Look at $|s_n - s|^p$: this converges to zero a.e. and $|s_n - s|^p \leq (2t)^p$, which is integrable. So by the DCT,

$$\int |s_n - s|^p \to 0$$

This is the same as $\|s_n - s\|_p^p \to 0$ as $n \to \infty$. (This proof is identical to the one for L^1.)

$\boxminus$

20.2. Separability. Take a metric space X, e.g. a normed space. Recall that if $S \subset X$ is a subset, then the closure $\overline{S}$ is defined as:

$$x \in \overline{S} \iff \exists x_n \in S \ \ \forall n s.t. x_n \to x \text{ as } n \to \infty$$

DEFINITION 20.5. X is *separable* if it has a countable dense subset S. (A dense subset is one such that $\overline{S} = X$.)

$\mathbb{R}^2$ is closed in $\mathbb{R}^3$. But in general, in a normed space, linear subspaces need not be closed: it may be a proper subset that is dense.

Observation: If there is a countable set $\{x_n : x_n \in X, n \in \mathbb{N}\}$ whose linear span is dense, then X is separable. (A *span* is all finite linear combinations $a_1 x_{n_1} + \cdots + a_k x_{n_k}$.) The *countable* dense subset is just all linear combinations $a_1 x_{n_1} + \cdots + a_k x_{n_k}$ with the restriction that the a_i's in some countable dense subset like $\mathbb{Q} \subset \mathbb{R}$ or $\mathbb{Q} + \sqrt{-1}\mathbb{Q} \subset \mathbb{C}$.

EXAMPLE 20.6. In L^p, $1 \le p < \infty$, the following subspaces are dense:

- The functions f of bounded support (i.e. those which vanish outside a ball of finite measure). These are dense by the DCT: if $p = 1$, for example, every function can be approximated by a function of bounded support by just cutting it off on $[-N, N]$. The other cases are similar.
- The bounded functions of bounded support (well, *essentially bounded* after applying the equivalence relation that you're allowed to change things on a null set). Again, approximate f by cutting it off at $y = N$, for increasing N.
- Simple functions $\sum_1^n a_i \chi_{E_i}$, where E_i are measurable subsets of finite measure
- The step functions $\sum a_i \chi_{R_i}$, where R_i are rectangles.
- The functions $\sum a_i \chi_{R_i}$ where $a_i \in \mathbb{Q}$ or $\mathbb{Q} + \sqrt{-1}\mathbb{Q}$, depending on which field we're using, and the step functions on rectangles R_i whose corners have rational (etc.) coordinates. This is a countable dense set, but is not a linear subspace. This example shows that L^p is separable for $1 \le p < \infty$. (But L^∞ is not separable: this is on the homework. For example, $\chi_{[-\sqrt{2},\sqrt{2}]}$ is approximated by $\chi_{[-1.4,1.4]}$, etc., but the L^∞ norm of this difference is 1.)
- In L^p the infinitely differentiable, or C^∞, functions of bounded support are dense in L^p also. For example, take step functions with rounded corners.

21. October 26

DEFINITION 21.1. Let X, Y be normed vector spaces over $\mathbb{R}$ or $\mathbb{C}$. A linear transformation, a.k.a. a linear operator $T : X \to Y$ is *bounded* if there exists $M \ge 0$ such that for all $x \in X$:

$$\|Tx\|_Y \le M \cdot \|x\|_X$$

The *operator norm* of T is the least such M; that is,

$$\begin{aligned}\|T\| &= \inf\{M \ge 0 : \|Tx\| \le M\|x\| \ \ \forall x\} \\ &= \sup\{\frac{\|Tx\|}{\|x\|} : x \ne 0\} \\ &= \sup\{\|Tx\| : \|x\| = 1\}\end{aligned}$$

EXAMPLE 21.2. If $X = \mathbb{R}^n$ then $\|x\| = x \cdot x = x^t x$. All linear maps $T : X \to X$ are bounded:

$$\begin{aligned}\|Tx\|^2 &= (Tx)^t(Tx) = x^t(T^tT)x \\ &= x^t \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_2 \end{pmatrix} x \\ &= \sum_i \lambda_i |x_i|^2 \leq (\max\{\lambda_i\})\|x\|^2 \\ \|T\| &= \sqrt{\max\{\lambda_i\}}\end{aligned}$$

EXAMPLE 21.3. The multiplication operator $T : L^p \to L^p$ where $T : f \mapsto gf$ for some fixed $g \in L^\infty$. Since g is essentially bounded, there is some M such that $|g(x)| \leq M$ for a.a.x, and so $\|Tf\|_p \leq M \cdot \|f\|_p$. This is the statement that $\int |fg|^p \leq (ess.sup.|f|)^p \int |f|^p$. If T is bounded, then $\|T\| \leq \|g\|_{L^\infty}$.

LEMMA 21.4. *Let X, Y be normed spaces, with Y a Banach space. Let $X' \subset X$ be a dense linear subspace; this inherits a norm from X. Suppose $T' : X' \to Y$ is bounded. Then there exists a unique way to extend this to a bounded operator $T : X \to Y$. This has the property that $\|T\| = \|T'\|$.*

PROOF. Given $x \in X$, choose a sequence $x'_n \in X'$ for $n \in \mathbb{N}$, where $x'_n \to x$. Claim: if (x'_n) is Cauchy, then $(T'x'_n)$ is Cauchy in Y. We want $\|T'x'_n - T'x'_m\| \to 0$ as $n, m \to \infty$. But this is true, because $\|T'x'_n - T'x'_m\| \leq M\|x'_n - x'_m\|$. So there is $y \in Y$ with $T'x'_n \to y$, because Y is complete. Define $Tx = y$.

We need to check that this is well-defined: y depends only on x, not the chosen sequence $x'_n \to x$. Checking linearity is straightforward: $T(x_1 + x_2) = T(x_1) + T(x_2)$. We also need to check that $\|Tx\| \leq M\|x\|$ where $M = \|T'\|$.

$$\|Tx\| = |y| = lim\|T'x'_n\| \leq \lim M\|x'_n\| = M \cdot \|x\|$$

$\boxminus$

The Fourier transform is an operator on $X = L^2(\mathbb{R}, \mathbb{C})$. Remember the Schwartz space $\mathcal{S}$ contains functions that decay faster than any polynomial. The Fourier transform is an operator on the Schwartz space.

LEMMA 21.5 (Plancherel identity). *For $f \in \mathcal{S}$*

$$\|\mathcal{F}f\|_{L^2} = \|f\|_{L^2}$$

PROOF. Remember $\|f\|_{L^2}^2 = \int |f|^2 = \int \overline{f} f$. For $f, g \in \mathcal{S}$:

$$\begin{aligned}\int (\overline{\mathcal{F}} f) g &= \int \overline{\mathcal{F} f(y)} g(Y)\,\mathrm{d}y \\ &= \int\int \overline{f(x)} e^{2\pi i x y} g(y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int\int \overline{f(x)} e^{2\pi i x y} g(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int \overline{f}(x)(\mathcal{F}^{-1} g)(x)\,\mathrm{d}x\end{aligned}$$

In the case where $g = \mathcal{F} f$, remember that the inverse Fourier transform is actually the inverse in Schwartz space, so we get

$$\|\mathcal{F} f\|_{L^2} = \|f\|_{L^2}$$

$\boxminus$

$\mathcal{S}$ is a dense subset of L^2, and $\mathcal{F} : \mathcal{S} \to L^2$ is a bounded linear operator with norm 1 (by Plancherel). By Lemma 21.4, there is a unique extension $\mathcal{F} : L^2 \to L^2$ that is also bounded, and of norm 1. The Schwartz space is dense in L^2: given $f \in L^2$ there exist $f_n \in \mathcal{S}$ such that $f_n \to f$ in the L^2 norm. According to the lemma, to define $\mathcal{F} f$, we do $\mathcal{F} f = \lim \mathcal{F} f_n$ where the limit means convergence in L^2. This is important because you can't just use the formula

$$\int f(x) e^{-2\pi i x \xi}\,\mathrm{d}x$$

because the inside might not be integrable.

We can also define $\mathcal{F}^{-1} : L^2 \to L^2$ in the same way: by using the definition of this inverse on Schwartz space. $\mathcal{F}^{-1} \circ \mathcal{F} = Id$ on $\mathcal{S}$; by a straightforward limit argument similar to the uniqueness statement, $\mathcal{F}^{-1} \cdot \mathcal{F} = Id$ on L^2. (This is because it i a bounded operator on L^2.)

EXAMPLE 21.6. If $f = \chi_{[-\frac{1}{2}, \frac{1}{2}]}$ then $(\mathcal{F} f)(x) = \frac{\sin(\pi x)}{\pi x} := g(x) \in L^2$. But we're not allowed to use the formula for the Fourier transform; you have to check that this is the same as what you get by taking the limit of Schwartz functions.

THEOREM 21.7. *The Fourier transform on $\mathcal{S}$ extends to an isometry (i.e. it preserves the norm)*

$$\mathcal{F} : L^2(\mathbb{R}; \mathbb{C}) \to L^2(\mathbb{R}; \mathbb{C})$$

with inverse $\mathcal{F}^{-1}$.

There is a similar story for Fourier series. Let $X = L^2(\mathbb{T}; \mathbb{C})$. Let $X' \subset X$ be the set of functions (trigonometric polynomials) whose Fourier series are finite:

$$f = \sum_{-n}^{n} a_m e_m(x) \text{ where } e_m(x) = e^{2\pi i m x}$$

So $a_m = \widehat{f}(m)$ here. There is a map $X' \to \ell^2(\mathbb{Z};\mathbb{C}) = \{(a_m)_{m\in\mathbb{Z}} : \sum |a_m|^2 < \infty\}$ given by

$$\mathcal{F}^{\mathbb{T}} : f = \sum_{-n}^{n} a_m e_m \mapsto (a_m)_{m\in\mathbb{Z}}$$

(So you're just mapping the function to the sequence of its Fourier coefficients.) Since we were considering finite sums,

$$\|f\|^2_{L^2(\mathbb{T})} = \|\mathcal{F}_{\mathbb{T}} f\|^2$$

i.e.

$$\begin{aligned}\int_0^1 |f|^2 &= \int_0^1 \sum_{m_1=-n}^{n} \sum_{m_2=-n}^{n} \overline{a}_{m_1} a_{m_2} \overline{e}_{m_1} e_{m_2} \\ &= \sum\sum \overline{a}_{m_1} a_{m_2} \int_0^1 \overline{e}_{m_1} e_{m_2} \\ &= \sum_{m=-n}^{n} |a_m|^2 = \|(a_m)\|^2_{\ell^2}\end{aligned}$$

So $\mathcal{F} : X' \to \ell^2$ is bounded. Furthermore, $X' \subset L^2(\mathbb{T})$ is dense: we have seen that $C^\infty \subset L^2(\mathbb{T})$ is dense. If $f \in C^\infty$, then the Fourier coefficients decay rapidly, and so the partial sums $s_n(f) \to f$ uniformly, hence in L^2. This gives a new conclusion:

PROPOSITION 21.8. *The map $\mathcal{F}_{\mathbb{T}} : X' \to \ell^2(\mathbb{Z};\mathbb{C})$ extends to all of $X = L^2(\mathbb{T};\mathbb{C})$. So there is a map*

$$\mathcal{F}_{\mathbb{T}} : L^2(\mathbb{T};\mathbb{C}) \to \ell^2(\mathbb{Z};\mathbb{C})$$

However, there is no need to do this: we could have define $f \mapsto (a_m)_{m\in\mathbb{Z}}$, and f being square integrable allows us to write

$$a_m = \int_0^1 f(x) e^{-2\pi i m x}\,\mathrm{d}x$$

If it is square integrable on $[0,1]$, then it is simply integrable on $[0,1]$ (see problem set). Unlike the case earlier, this formula simply works. This map has an inverse, by the same argument as in the previous case. In particular, $\mathcal{F}_{\mathbb{T}}$ is an isometry of Banach spaces.

22. October 28

Let X and Y be normed spaces. Then define $L(X,Y)$ to be the space of all bounded linear operators. Then this is a vector space. Let $T_1, T_2 \in L(X,Y)$.

$$(T_1+T_2)(x) = T_1(x) + T_2(x)$$
$$(\lambda T_1)(x) = \lambda(T_1 x)$$

are also in $L(X,Y)$. Sums and scalar multiples are also bounded:

$$\|T_1+T_2\| \le \|T_1\| + \|T_2\|$$
$$\|\lambda T_1\| = |\lambda|\|T_1\|$$

This shows that the operator norm is a norm on $L(X, Y)$.

As a special case, consider $Y = \mathbb{F}$.

DEFINITION 22.1. The *dual space* is

$$X^* = L(X, \mathbb{F})$$

For a *bounded linear functional* $\alpha : X \to \mathbb{F}$ in X^* use the operator norm:

$$\|\alpha\| = \sup_{x \neq 0} \frac{|\alpha(x)|}{\|x\|}$$

PROPOSITION 22.2. *X^* is a Banach space: it is complete. In fact, for any complete space Y, $L(X, Y)$ is complete.*

PROOF. Let $T_n \in L(X, Y)$ for $n \in \mathbb{N}$ be a Cauchy sequence. For $\varepsilon > 0$ there is some $n(\varepsilon)$ such that $\|T_n - T_m\| \leq \varepsilon$ for all $n, m \geq n(\varepsilon)$. We want to find a limit; i.e., find $T \in L(X, Y)$ such that $\|T_n - T\| \to 0$. Fix $x \in X$. For $n \in \mathbb{N}$, $T_n x \in Y$ is a Cauchy sequence in Y, because

$$(T_n - T_m)x = \|T_n x - T_m x\|_Y \leq \|T_n - T_m\|\|x\| \leq \varepsilon\|x\|$$

for $n, m \geq n(\varepsilon)$. So we can say that

$$T_n x - T_m x \to 0$$

as $n, m \to \infty$. Because Y is complete, there is a limit: define $Tx \in Y$ as $Tx = \lim_{n\to\infty} T_n x$. We need to check that T is linear: $T(x_1 + x_2) = Tx_1 + Tx_2$, and similarly for scalar multiples (this is straightforward). We also need that T is bounded: this is because $\|T_n\|$ is a Cauchy sequence, and hence bounded. Finally, we need to show that $\|T - T_n\| \to 0$ (we've already shown that this converges pointwise, but we need convergence in norm). Look at $\|(T - T_n)(x)\|_Y$. For $n \geq n(\varepsilon)$

$$\|(T_n - T_{n(\varepsilon)})\| \leq \varepsilon\|x\|$$

which implies

$$(T - T_{n(\varepsilon)})(x) \leq \varepsilon\|x\|$$

And similarly, $\|(T - T_m)(x)\| \leq \varepsilon\|x\|$ for $m \geq n(\varepsilon)$ i.e. $\|T - T_m\| \leq \varepsilon$ for $m \geq n(\varepsilon)$. (The point is that the $n(\varepsilon)$ do not depend on x.) $\boxminus$

EXAMPLE 22.3. If $X = \mathbb{R}^N$ or $\mathbb{C}^N$, every linear map $X \to \mathbb{R}$ or $\mathbb{C}$ is bounded. So X^* contains all linear maps $\mathbb{R}^N \to \mathbb{R}$ or $\mathbb{C}^N \to \mathbb{C}$.

Let $X = c_0$ (the space of sequences that converge to zero). Claim: X^* is isometrically isomorphic to ℓ^1 (the space of sequences where $\sum |a_n| < \infty$). (There is a bijective 1-1 linear map that preserves norms.) Given $a \in \ell^1$, we get $\alpha : c_0 \to \mathbb{R}$ given by $\alpha(b) = \sum_1^\infty a_n b_n$. Note

$$\begin{aligned}\sum |a_n b_n| &\leq (\sup_n |b_n| \sum_1^\infty |a_n|) \\ &= \|b\|_{\ell^\infty} \|a\|_{\ell^1}\end{aligned}$$

So $\alpha(b)$ makes sense and $\alpha : c_0 \to \mathbb{R}$ is bounded. So $|\alpha(b)| \leq \|a\|_{\ell^1} \cdot \|b\|_{\ell^\infty}$. We've shown that $\alpha \in X^* = c_0^*$, and $\|\alpha\| \leq \|a\|_{\ell^1}$. In fact, equality holds, because

$$\|a\|_{\ell^1} = \sup \frac{|\alpha(b)|}{\|b\|_\infty} = \frac{\sup |\sum a_n b_n|}{\|b\|_\infty}$$

as we see by taking b such that

$$b_n = \begin{cases} sign(a_n) & \text{if } n \leq N \\ 0 & \text{else} \end{cases}$$

Summary: for all b, $|\alpha(b)| \leq \|a\|_{\ell^1} \|b\|_{\ell^\infty}$, and you can achieve the worst case by taking b as defined above.

Given $a = (a_n) \in \ell^1$ we've defined α in c_0^* :

$$\alpha_a : c_0 \to \mathbb{R} \text{ where } b \mapsto \sum a_n b_n$$

and we showed that $\|\alpha_a\| = \|a\|_{\ell^1}$

Now do the converse. Given $\alpha \in c_0^*$ we must show that there is some $a \in \ell^1$ such that $\alpha = \alpha_a$; that is, $\alpha(b) = \sum a_n b_n$ for all $b \in c_0$. Given α set $a_n = \alpha(e_n)$ where e_n is the standard n^{th} basis vector. Then

$$\begin{aligned}
\sum_1^N |a_n| &= \sum_1^N a_n \cdot sign(a_n) \\
&= \alpha(\sum_1^N sgn(a_n) e_n) \\
&= \alpha((\underbrace{\pm 1, \cdots, \pm 1}_{N}, 0, \cdots, 0)) \\
&\leq \|a\| \cdot 1
\end{aligned}$$

So $a = (a_n)$ is in ℓ^1. Is $\alpha = \alpha^a$?

$$\alpha(e_n) = \alpha_a(e_n) \implies \alpha(b) = \alpha_a(b) \quad \forall b \in (c_0)_{fin}$$

$(c_0)_{fin} \subset c_0$, where $(c_0)_{fin}$ are the sequences which are zero eventually. Now $(c_0)_{fin}$ is dense in c_0, hence $\alpha = \alpha_0 a$ for all $b \in c_0$. Given $b \in c_0$, define $b^{(N)}$ as the sequence

$$\begin{cases} b_n & \text{if } n \leq N \\ 0 & \text{else} \end{cases}$$

$$\begin{aligned}
\|b - b^{(N)}\| &= \sup_n |b_n - b_{n^{(N)}} \\
&= \sup_{n > N} |b_n| \to 0
\end{aligned}$$

This is the only part that fails for ℓ^∞ instead of c_0; $\ell^\infty \supset c_0$, and so its dual space is much larger.

EXAMPLE 22.4. For $1 \leq p < \infty$, I claim that the dual space of ℓ^p is ℓ^q, where q is the dual exponent. (If $p = 1$ then $q = \infty$, and otherwise, $\frac{1}{p} + \frac{1}{q} = 1$.) In particular, $(\ell^1)^* \cong \ell^\infty$ but $(\ell^\infty)^*$ is something bigger than ℓ^1.

23. October 31

For $1 < p < \infty$ abbreviate $L^p = L^p(\mathbb{R}^d; \mathbb{F})$, where $\mathbb{F}$ refers to either $\mathbb{R}$ or $\mathbb{C}$. With $1 = \frac{1}{p} + \frac{1}{q}$ and $a \in L^q$ we get an element $\alpha \in \alpha_a : L^p \to \mathbb{F}$ which maps $b \mapsto \int ab$ for $b \in L^p$ (the integrability is Minkowski's inequality). α is certainly linear, because the integral is linear. Also, α is bounded, because $|\alpha(b)| \leq \|a\|_{L^q}\|b\|_{L^p}$ (this is Hölder's inequality). So there is a map $I : L^q \to (L^p)^*$ that sends $a \mapsto \alpha$. I is linear; this is just because of the linearity of the integral:

$$I(\lambda_1 a_1 + \lambda_2 a_2) = \lambda_1 I(a_1) + \lambda_2 I(a_2)$$

since

$$\alpha_{\lambda_1 a_1 + \lambda_2 a_2}(b) = \lambda_1 \alpha_{a_1}(b) + \lambda_2 \alpha_{a_2}(b)$$

THEOREM 23.1 (Riesz representation theorem for L^p). *I is an isometric isomorphism from L^q to $(L^p)^*$.*

PROOF. The first claim is that I preserves norms. If $a \mapsto \alpha \in (L^p)^*$ we must show that $\|a\|_{L^q} = \|\alpha\|_{dual}$. Define the dual space norm

$$\begin{aligned} \|\alpha\| &= \sup_{b \in L^p, b \neq 0} \frac{|\alpha(b)|}{\|b\|_p} \\ &= \sup \frac{|\int ab|}{\|b\|} \\ &\leq \|a\|_q \qquad \text{Hölder} \end{aligned}$$

The thing we're trying to prove is that this last $\leq$ can be replaced with equality.

CLAIM 23.2 (Converse to Hölder's inequality). *If $a \in L^q$ then*

$$\sup_b \frac{|\int ab|}{\|b\|_p} = \|a\|_q$$

where the supremum is taken over all b measurable, bounded, and of bounded support (not a.e. zero). If $a \notin L^q$ then the supremum on the left is ∞.

(We know the case $\leq$ above.) Fix N large, and take $\mathbb{F} = \mathbb{R}$. Set $\widetilde{a} = a \cdot \chi_{B(N)} \cdot \chi_{E(n)}$ where $E(N) = \{|a| \leq N\}$ and $B(N)$ is a ball of radius N. Set $\widetilde{b} = sign(\widetilde{a})|\widetilde{a}|^{q-1}$.

$$\begin{aligned}
\frac{|\int a\widetilde{b}|}{\|\widetilde{b}\|_p} &= \frac{|\int \widetilde{a}\widetilde{b}|}{\|\widetilde{b}\|_p} \\
&= \int |\widetilde{a}^q| / \left(\int |\widetilde{a} \right)^{\frac{1}{p}} \\
&= \frac{\int |\widetilde{a}|^q}{\left(\int |\widetilde{a}^q|\right)^{1-\frac{1}{q}}} \\
&= \|\widetilde{a}\|_q
\end{aligned}$$

In the claim, the supremum is $\geq \sup_N \|\widetilde{a}\|_q = \|a\|_q$ (or infinity, if $a \notin L^q$).

CLAIM 23.3. *I is surjective: given bounded $\alpha \in (L^p)^*$, we can find $a \in L^q$ such that $\alpha(b) = \int ab$ for all $b \in L^p$.*

We will do this with $\mathbb{F} = \mathbb{R}$ and p an even integer. We will take $p = 4$ and $q = \frac{4}{3}$. If we've proven it for all even integers p, then you can show it works for anything $< p$, and hence it works everywhere. Given $\alpha : L^4 \to \mathbb{R}$; we want to find $a \in L^{\frac{4}{3}}$. Consider the function (not linear!) which takes $\Gamma : L^4 \to \mathbb{R}$ via $u \mapsto \|u\|_4^4 - \alpha(u)$. We can get rid of the absolute value signs: $\Gamma(u) : \int u^4 - \alpha(u)$. We'll show that Γ is bounded below, and there exists some $U_* \in L^4$ where Γ achieves its infimum. The derivative is zero here (think critical points, etc.) Given this, consider any $v \in L^4$ and the path $u_t = u_* + tv$ where the parameter is $t \in \mathbb{R}$. We have $\Gamma(u_t) \geq \Gamma(u_0) = \Gamma(u_*)$ for all t.

$$\Gamma(u_t) = \int u_*^4 + 4t \int u_*^3 v + 6t^2 \int u_*^2 v^2 + 4t^3 \int u_* v^3 + t^4 \int v^4 - \alpha(u_*) - t\alpha(v) := P(t)$$

$P(t)$ is a quartic polynomial in t. Achieving the minimum at $t = 0$ means $P'(0) = 0$ so the linear term $4t \int u_*^3 v - t\alpha(v)$ has to be zero. So we need

$$4 \int u_*^3 v - \alpha(v) = 0$$

for all $v \in L^4$. Write $a = 4u_*^3$. Then

$$\left(\int av \right) - \alpha(v) = 0$$

for all $v \in L^4$. That is, $I(a) = \alpha$. (As a sanity check, note that $a \in L^{\frac{4}{3}}$: it goes as the cube of something in L^4, a.k.a. $|a|^{\frac{4}{3}} = 4^{\frac{4}{3}} |u_*|^4$.)

Why is Γ bounded below? $\Gamma(u) = \|u\|_4^4 - \alpha(v) \geq X^4 - M \cdot X$ where $X = \|u\|_4$ and $M = \|\alpha\|$. It achieves its minimum at $\left(\frac{M}{4}\right)^{\frac{1}{3}}$ by one-variable calculus. So it has some finite infimum.

Let $\Gamma_0 = \inf_{u \in L^4} \Gamma(u)$. There's a sequence $u_n \in L^4$ with $\Gamma(u_n) \searrow \Gamma_0$ (the point is it approaches Γ_0 and is not below it). If $(u_n)_n$ is Cauchy in L^4 then we're done, because we can set u_* to be the limit. Γ is continuous, so $\Gamma(u_*) = \lim \Gamma(u_n) = \Gamma_0$.

So the claim is that u_n is Cauchy; this is a property of convexity. Given $\varepsilon > 0$ we have for $n, m \geq n(\varepsilon)$

$$\Gamma_0 \leq \Gamma(u_n) \text{ and } \Gamma(u_m) \leq \Gamma_0 + \varepsilon$$

Given u_0, u_1 and knowing $\Gamma(u_0)$ and $\Gamma(u_1)$ in $\Gamma_0, \Gamma_0 + \varepsilon$ we will show what $\|u_0 - u_1\|_{L^4}$ goes to zero as $\varepsilon \to 0$. Look at the path from u_0 to u_1 in L^4:

$$u_t = u_0 + t(u_1 - u_0)$$

We have $\Gamma_0 \leq \Gamma(u_1) \leq \Gamma_0 + \varepsilon$ for all $t \in [0, 1]$. (The first inequality is the fact that Γ_0 is the infimum; the second inequality is by convexity of $\Gamma(u_t)$: $\frac{d^2}{dt^2}\Gamma(u_t) = 12 \int u_0^2 u_1^2 \geq 0$). $\Gamma(u_t)$ we have seen is a degree 4 polynomial in t, and its values on $[0, 1]$ lie in a narrow interval $[\Gamma_0, \Gamma_0 + \varepsilon]$.

Next time: $\|u_0 - u_1\|_{L^4}$ is small.

$\boxminus$

24. November 2

$\Gamma : L^4 \to \mathbb{R}$, $\|u\|_4^4 - \alpha(u)$, $\Gamma_0 = \inf_u \Gamma(u)$. There is some u_* such that $\Gamma_0 = \Gamma(u_*)$. If $u_0, u_1 \in L^4$ have $\Gamma(u_0) - \Gamma_0$ and $\Gamma(u_1) - \Gamma_0$ both less than $\varepsilon > 0$ then $\|u_0 - u_1\|_4$ is small. We were showing that $\|u_0 - u_1\|_4 \leq (C\varepsilon)^{\frac{1}{4}}$. Set

$$u_t = u_0 + t(u_1 - u_0) \in L^4 \text{ for } t \in [0, 1]$$

and $P(t) = \Gamma(u_t) - \Gamma_0$. If

$$P(t) = \int u_0^4 + \cdots + t^4 \int (u_1 - u_0)^4 - \alpha(u_0) - t\alpha(u_1 - u_0)$$

is a quartic polynomial. Turns out that $P''(t) \geq 0$ (proof last time sort of wrong); the L^4 norm is a convex function, and $\alpha(u)$ is a linear thing. $0 \leq P(0), P(1) \leq \varepsilon$ so $0 \leq P(t) \leq \varepsilon$ for all t, by convexity. This implies that all the coefficients are small (these are determined by five points). The coefficient of t^4 is

$$\frac{4^4}{4!}(P(0) - 4P(\frac{1}{4})) + 6P(\frac{1}{2}) - 4P(\frac{3}{4}) + P(1) \leq \frac{3}{4}\varepsilon$$

This coefficient is $\|u_0 - u_1\|_4^4 \leq \frac{3}{4}\varepsilon$.

This completes the proof for L^4; for L^{2n}, the same thing works:

$$\Gamma(u) = \|u\|_{2n}^{2n} - \alpha(u) \text{ for } u \in L^{2n}$$

Claim: If we can do L^p then we can prove Riesz representation theorem for $L^{p'}$ for $p' \leq p$. Start with some ball $B \subset \mathbb{R}^d$ a ball, so $L^{p'}(B)$ contains functions on $\mathbb{R}^d$ supported on B, with the $L^{p'}$ norm. We showed on the HW that $L^p(B) \subset L^{p'}(B)$. So given $\alpha \in (L^{p'})^*$ seek $a \in L^{q'}$ as follows. α gives also a bounded linear functional $L^p(B) \to \mathbb{R}$, by restriction. There exists $a \in L^q$ (where q is the dual exponent) such that $\alpha(b) = \int ab$ for all $b \in L^p$. We have $|\int_B ab| \leq M \cdot \|b\|_{p'}$ as long as $b \in L^p(B)$; in particular, it works if $b \in L^\infty(B)$. By the converse to Hölder's inequality, $a \in L^q$ and $\|a\|_{q'} \leq M$. We've only done this on bounded balls, but you can take larger and larger balls, and then claim that the limit is an actual $L^{q'}$ function. See Kronheimer's notes.

24.1. Hahn-Banach theorem. We have seen that $(L^p)^* \cong L^q$, and taking the dual again gets something isomorphic to L^p. You don't always get back to where you start: $(c_0)^{**} \cong \ell^\infty$. These are not isomorphic because one is separable and the other is not. Given a normed space X, do there exist *any* bounded linear functionals on X?

THEOREM 24.1 (Hahn-Banach theorem). *Let X be a normed (not necessarily complete) space, and $V \subset X$ a linear subspace; so V inherits a norm. Let $\alpha : V \to \mathbb{R}$ be a bounded linear functional. Then there exists $\overline{\alpha} : X \to \mathbb{R}$ which extends α to all of X, and such that $\|\overline{\alpha}\| = \|\alpha\|$.*

PROOF. Start with a basic case: X is spanned by V, together with a single extra vector x_1. Assume we have $\alpha : V \to \mathbb{R}$. Then $\overline{\alpha} : X \to \mathbb{R}$ is $\overline{\alpha}(\lambda x_1 + v) = \lambda c + \alpha(v)$, where c is a real number we have to choose: $c = \overline{\alpha}(x_1)$. We have to choose this carefully to make the norm work out. We already know $\|\overline{\alpha}\| \geq \|\alpha\|$. Without loss of generality $\|\alpha\| = 1$. We need $|\lambda c + \alpha(v)| \leq \|\lambda x_1 + v\|$ for all $\lambda \in \mathbb{R}$ and $v \in V$. Scaling by λ doesn't do anything; we just need $|c + \alpha(v)| \leq \|x_1 + v\|$, which is really

$$-\|x_1 + v\| + \alpha(v) \leq \|x_1 + v\|$$
$$-\|x_1 + w\| - \alpha(w) \leq c \leq \|x_1 + v\| - \alpha(v)$$

This is OK as long as $-\|x_1 + w\| - \alpha(w) \leq \|x_1 + v\| - \alpha(v)$ for all $v, w \in V$. This is a consequence of $\|\alpha\| = 1$ and the triangle inequality: it says that $\alpha(v) - \alpha(w) \leq \|x_1 + v\| + \|x_1 + w\|$ is true for all v, w. But the LHS is $\leq \|v - w\|$ becuase $\|\alpha\| = 1$; now use the triangle inequality.

Iterating this process gets us the Hahn Banach theorem in the finite-dimensional case. It is also enough to deal with the case where X is separable: choose $x_1, \cdots, x_N, \cdots$ whose span is dense in X, and extend one step at a time: $X_0 = V$, $X_1 = \mathit{span}(V \cup \{x_1\})$, $X_n = \mathit{span}(X_{n-1} \cup \{x_n\})$, and $X_\infty = \cup X_n$. Extend α to $\alpha_n : X_n \to \mathbb{R}$ for all n. $\alpha_\infty : X_\infty \to \mathbb{R}$ extends to all of X, because X_∞ was assumed to be dense in X. The norm is preserved at every step along the way, and the extension of α_∞ to all of X also preserves norm, by lemma 21.4.

For the non-separable case, use Zorn's lemma. Given V, α define a partial extension to be a pair (W, β) with $\beta \in W^*$ and $\beta|_V = \alpha$ and $\|\beta\| = \|\alpha\|$. We get a partial order on these pairs: we say that $(W, \beta) \leq (W', \beta')$ if $W' \supset W$ and $\beta'|_W = \beta$. To check the hypotheses of Zorn's lemma, we need every chain in this poset (i.e. every collection of (W_i, β_i) for $i \in I$ on which $\leq$ is a total order) has an upper bound. In this case, this is $W = \bigcup_{i \in I} W_i$ and β defined by $\beta|_{W_i} = \beta_i$. Zorn's lemma now says that there is a maximal partial extension:

one that is not $<$ any other partial extension. This must be $W = X$, because otherwise you could throw in another vector and extend it further. $\boxminus$

COROLLARY 24.2. *If $x \in X$ (a normed space) and $x \neq 0$ then there exists $\alpha \in X^*$ such that $\|\alpha\| = 1$ and satisfying $\alpha(x) = \|x\|$.*

PROOF. Take $V = span(x)$. Define $\alpha_0 : V \to \mathbb{R}$, where $\alpha_0(\lambda x) = \lambda\|x\|$. This is bounded, since $\|\alpha_0\| = 1$. Apply Hahn-Banach, and we are done, because we can extend α_0 to something of norm 1 that still maps $\lambda x \mapsto \lambda\|x\|$. $\boxminus$

COROLLARY 24.3. *For $x \in X$, $x \neq 0$,*

$$\|x\| = \sup \frac{|\alpha(x)|}{\|\alpha\|} = \sup_{\alpha \in X^*, \|\alpha\|=1} |\alpha(x)|$$

25. November 4 – from Ben Whitney's notes

25.1. Corollaries of Hahn-Banach Theorem.

- If $x \in X$ (a normed space) and $x \neq 0$, then there exists $\alpha \in X^*$ with $\|\alpha\| = 1$ and $\alpha(x) = \|x\|$. (Proof last time.)
- $\|x\| = \sup_{\alpha \in X^*, \|\alpha\|=1} |\alpha(x)|$. Compare this with the definition $\|\alpha\| = \sup_{x \in X, \|x\|=1} |\alpha(x)|$.
- If X is a normed space, $V \subset X$ is a closed linear subspace, and $x_1 \notin V$, then there exists $\alpha \in X^*$ with $\alpha|_V = 0$, $\alpha(x_1) = 1$, and $\|\alpha\| = \frac{1}{\delta}$, where $\delta = \inf_{v \in V}\{\|x_1 - v\|\} > 0$. Start by defining $\alpha_1 : \text{span}(\{x_1\} \cup V) \to \mathbb{F}$ by $\alpha(\lambda x_1 + v) = \lambda$. This is bounded in $(\text{span}(\{x_1\} \cup V))^*$. Indeed,

$$\begin{aligned}
\|\alpha_1\| &= \sup_{\substack{\lambda \in \mathbb{F}\setminus\{0\} \\ v \in V}} \left\{ \frac{|\alpha_1(\lambda x_1 + v)|}{\|\lambda x_1 + v\|} \right\} \\
&= \sup_{v \in V} \left\{ \frac{|\alpha_1(x_1 + v)|}{\|x_1 + v\|} \right\} \\
&= \sup_{v \in V} \left\{ \frac{1}{\|x_1 + v\|} \right\} \\
&= \frac{1}{\delta}
\end{aligned}$$

  By the Hahn-Banach theorem, there exists $\alpha \in X^*$ with $\alpha|_{\text{span}(\{x_1\} \cup V)} = \alpha_1$ and $\|\alpha\| = \|\alpha_1\| = \frac{1}{\delta}$.

Let X be a normed space. Compare X to X^{**}. An element of X^{**} is a bounded linear functional $\xi : X^{**} \to \mathbb{F}$. There's a map $X \to X^{**}$ sending $x \mapsto \breve{x}$ defined by $\breve{x} : \alpha \mapsto \alpha(x)$, i.e. $\breve{x}(\alpha) := \alpha(x)$ for $\alpha \in X^*$.

$\breve{x}$ is linear:

$$\begin{aligned}\breve{x}(\alpha_1+\lambda\alpha_2) &= (\alpha_1+\lambda\alpha_2)(x)\\ &= \alpha_1(x)+\lambda\alpha_2(x)\\ &= \smile\alpha_1+\lambda\breve{x}(\alpha_2)\end{aligned}$$

Thus $\breve{x} : X^* \to \mathbb{F}$ is a linear map. We must show that it is bounded. We want an M with $|\breve{x}| \leq M\|\alpha\|$ for all $\alpha \in X^*$. But

$$\begin{aligned}|\breve{x}(\alpha)| &= |\alpha(x)|\\ &\leq \|x\|\cdot\|\alpha\|\\ &\leq M\|\alpha\|\end{aligned}$$

where $M = \|x\|$. So $\breve{x}$ is bounded.

$$\begin{aligned}\|\breve{x}\| &= \sup_{\substack{\alpha\in X^*\\ \|\alpha\|=1}}\{|\breve{x}(\alpha)|\}\\ &= \sup_{\substack{x\in X^*\\ \|\alpha\|=1}}\{|\alpha(x)|\}\\ &= \|x\|\end{aligned}$$

using one of the corollaries to the Hahn-Banach theorem we did earlier today. Then map $\smile: X \to X^{**}$ is linear. We need to check that $(x_1+\lambda x_2)^{\smile} = \breve{x}_1 + \lambda\breve{x}_2$.

$$(x_1+\lambda x_2)^{\smile}(\alpha) = \ldots = \breve{x}_1(\alpha)+\lambda\breve{x}_2(\alpha)$$

which is easy to check. Summary: there's a linear map (called the canonical map) $\smile: X \to X^{**}$ sending $x \to \breve{x}$ which preserves norms (which implied injectivity and boundedness). But it's possible that $\smile$ is not surjective.

DEFINITION 25.1. X is **reflexive** is $\smile: X \to X^{**}$ is surjective. (Then, based on what we've seen before, $\smile$ is an isometric isomorphism. Then X is, like X^{**}, complete.)

c_0 is not reflexive: the dual of c_0 is ℓ^1, and the dual of ℓ^1 is ℓ^∞. c_0 is separable but ℓ^∞ is not. Now consider $\mathrm{L}^p(\mathbb{R}^d;\mathbb{F})$ with $1 < p < \infty$. If q is the dual exponennt, then the dual space of L^p $(\mathrm{L}^p)^* \simeq \mathrm{L}^q$. Similarly, the dual of L^q is $(\mathrm{L}^q)^* \simeq \mathrm{L}^p$. Is it reflexive? We must show that every bounded linear functional $\xi : (\mathrm{L}^q)^* \to \mathbb{F}$ has the form $\breve{b}$ for some $b \in \mathrm{L}^p$, i.e. has the form $\alpha \mapsto \alpha(b)$ ($\alpha \in (\mathrm{L}^p)^*$) for some b.

There is an isometric isomorphism $\mathrm{L}^q \to (\mathrm{L}^p)^*$ sending $a \mapsto \alpha_a$, where $\alpha_a(c) = \int ac$ ($c \in \mathrm{L}^p$). To veryify for the α_as, we check that every bounded linear functional $\mathrm{L}^q \to \mathbb{F}$ has the form $a \mapsto \int ab$ for some b. This is the Riesz Representation Theorem for L^q.

We will show on the problem set that if X^* is reflexive, then X is reflexive, and if X is not reflexive, then neither is X^{**}.

25.2. Dual Transformations. Let $T : X \to Y$ be a bounded linear operator between normed vector spaces. There there is a dual (or "transpose") operator $T^* : Y^* \to X^*$. Given $\beta \in Y^*$, i.e. $\beta : Y \to \mathbb{F}$ bounded and linear, define $T^*(\beta) \in X^*$, i.e. $T^*(\beta) : X \to \mathbb{F}$, by $T^*(\beta) = \beta \circ T$. As the composite of two bounded linear operators, $T^*(\beta)$ is bounded and linear, and therefore included in X^*. Unwrapping: $T^*(\beta)(x) = (\beta \circ T)(x) = \beta(T(x))$. We have ourselves a map of sets $T^* : Y^* \to X^*$. We need to check that T^* is linear and bounded.

$$T^*(\beta_1 + \lambda\beta_2)(x) = (\beta_1 + \lambda\beta_2)(T(x)) = \beta_1(T(x)) + \lambda\beta_2(T(x)) = T^*(\beta)(x) = \lambda T^*(\beta_2)(x)$$

We now want to show that T^* is bounded, and that furthermore $\|T^*\| = \|T\|$.

$$\begin{aligned}
\|T^*\| &= \sup_{\substack{\beta\in Y^*\\ \|\beta\|=1}} \{\|T^*\beta\|\} \\
&= \sup_{\substack{\beta\in Y^*\\ \|\beta\|=1}} \left\{ \sup_{\substack{x\in X\\ \|x\|=1}} \{|(T^*(\beta))(x)|\} \right\} \\
&= \sup_{\substack{\beta\in Y^*\\ \|\beta\|=1}} \left\{ \sup_{\substack{x\in X\\ \|x\|=1}} \{|\beta(T(x))|\} \right\} \\
&= \sup_{\substack{x\in X\\ \|x\|=1}} \left\{ \sup_{\substack{\beta\in Y^*\\ \|\beta\|=1}} \{|\beta(T(x))|\} \right\} \\
&= \sup_{\substack{x\in X\\ \|x\|=1}} \{\|T(x)\|\}
\end{aligned}$$

by the second corollary to the Hahn-Banach theorem. This is $\|T\|$ by definition.

26. November 7

On $X = L^p(\mathbb{R}^d)$, $1 < p < \infty$, define $T : X \to X$ which takes $(Tf)(x) = f(Ax)$ where A is an invertible $d \times d$ matrix. There's a transpose operator $T^* : X^* \to X^*$. Via Riesz representation, this is $S : L^q \to L^q$, $S = I^{-1} \circ T^* \circ I$ (where $I \cdot L^q \cong (L^p)^*$.) We have, for $g \in L^q$ and $f \in L^p$,

$$\int g(Tf) = \int g(x)f(Ax)dx = \frac{1}{|\det A|}\int g(A^{-1}y)f(y)dy = \int (Sg)f$$

So $(Sg)(y) = \frac{1}{|\det A|}g(A^{-1}y)$. Fix $G \in L^1(\mathbb{R}^d)$.

Consider $T : L^1 \to L^1$, with $f \mapsto G * f$. Then the dual linear transformation $T^* : (L^1)^* \to (L^1)^*$ is, via the Riesz representation theorem, an operator $S : L^\infty \to L^\infty$. For $a \in L^\infty$,

$b \in L^1$,

$$\begin{aligned}
\int a(Tb) &= \int a(G * b) \\
&= \int \int a(x) G(x-y) b(y) dy dx \\
&= \int \int a(x) G(x-y) b(y) dx dy \\
&= \int (\widetilde{G} * a) b \qquad\qquad \widetilde{G}(x) = G(-x)
\end{aligned}$$

So the operator S takes $b \mapsto \widetilde{G} * b$. If we define $K(x,y) = G(x-y)$ then

$$(G * b)(x) = \int K(x,y) b(y) dy$$
$$(\widetilde{G} * b)(x) = \int K(y,x) b(y) dy$$

The swapping of the variables is reminiscent of the swapping of indices in the transpose of a matrix.

26.1. Hilbert Spaces.

X is an inner product space over $\mathbb{R}$ or $\mathbb{C}$: i.e. a vector space X with $(\ ,\ ) : X \times X \to \mathbb{F}$ such that

- $x \mapsto (x,y)$ is linear in x for y fixed
- $(x,y) = \overline{(y,x)}$, which implies that $(x,x) \in \mathbb{R}$
- $(x,x) \geq 0$ with equality iff $x = 0$

Define $\|x\|$ by $\|x\|^2 = (x,x)$. This is indeed a norm: it satisfies the triangle inequality. If $x \perp y$ (i.e. $(x,y) = 0$) then we have "Pythagoras":

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2$$

DEFINITION 26.1. A Hilbert space is an inner product space X which is complete as a normed space (with norm as above).

In Stein and Shakarchi, Hilbert spaces are also required to be separable.

EXAMPLE 26.2.

- $\mathbb{C}^n$ with the usual inner product $(x,y) = \sum_{i=1}^n x_i \overline{y_i}$
- ℓ^2 with $(x,y) = \sum_{i=1}^\infty x_i \overline{y_i}$ (the previously defined norm comes from this inner product)
- $L^2(\mathbb{R}^d)$ with $(a,b) = \int a\bar{b}$

26.2. Orthonormal systems.

DEFINITION 26.3. $\{e_i : i \in I\}$ is orthonormal if $\|e_i\| = 1$ and $(e_i, e_j) = 0$ if $i \neq j$.

Example 26.4. X will be a Hilbert space.

- In $L^2(\mathbb{T})$, $e_n(x) = e^{2\pi inx}$ for $n \in \mathbb{Z}$ is an orthonormal system
- In ℓ^2, if δ_n is the sequence with 1 in the n^{th} spot and zero elsewhere, then these form an orthonormal system

If $a_n \in \mathbb{F}$ is a series of coefficients, and e_n are orthonormal and if $\sum_1^\infty |a_n|^2 < \infty$ then $s_N = \sum_1^N a_n e_n$ is Cauchy in X, hence convergent. For $N \geq M$ then

$$\|s_N - s_M\|^2 = \|\sum_{M+1}^N a_n e_n\|^2 = \sum_{M+1}^N |a_n|^2 \leq \sum_{M+1}^\infty |a_n|^2 \to 0 \text{ as } M \to \infty$$

by Pythagoras. Cauchy sequences have a limit here; so if $s = \sum_1^\infty a_n e_n$ then

$$\|s\|^2 = \lim \|s_N\|^2 = \sum_1^\infty |a_n|^2$$

Note that the statement $\sum_1^\infty |a_n|^2 < \infty$ is independent of the order of the terms. So the limit s is characterized by two properties:

- $s \in \widetilde{X} :=$ closure of the span of the e_n (because it is a limit of things in the span)
- $(s, e_n) = a_n$ for all n (because $(s_N, e_n) = a_n$ for $N \geq n$)

These properties uniquely characterize s: if s, s' satisfied both then $(s - s', e_n) = 0$ for all n, so $(s - s', y) = 0$ for all $y \in$ the span of e_n's; now $(s - s', x) = 0$ for all $x \in \widetilde{X}$. But $s - s' \in \widetilde{X}$ so $(s - s', s - s') = 0$. Let e_n for $n \in \mathbb{N}$ be an orthonormal system in X, a Hilbert space. Take $x \in X$. Take $x \in X$. Set $a_n = (x, e_n)$ and look at $\sum_1^\infty a_n e_n$. Does it converge? Look at the partial sums $s_N = \sum_1^N a_n e_n$ and $\|s_N\|^2 = \sum_1^N |a_n|^2$; we said convergence happens iff $\sum |a_n|^2$ converges. We have $(x - s_N) \perp e_n$. So $(x - s_N) \perp s_N$ because s_N is a linear combination of the e_N. Now $x = (x - s_N) + s_N$, and the parts are orthogonal, so

$$\|x\|^2 = \|x - s_N\|^2 + \|s_N\|^2$$

So $\|s_N\|^2 \leq \|x\|^2$ with equality iff $x = s_N$, i.e. $\sum_1^N |a_n|^2 \leq \|x\|^2$ with equality iff $x = s_N$. We have just proven

Proposition 26.5 (Bessel's inequality, finite case).

$$\sum_1^N |(x, e_n)|^2 \leq \|x\|^2$$

with equality iff $x = \sum_1^N (x, e_n) e_n$.

So our original sum $\sum_1^\infty a_n e_n$ converges, because $\sum_1^\infty |a_n|^2 \leq \|x\|^2 < \infty$.

Let $\widetilde{x} = \sum_1^\infty a_n e_n \in \widetilde{X}$. We have $(x - \widetilde{x}) \perp e_n$ for all n, which implies that $(x - \widetilde{x}) \perp y$ for all y in the span of the e_n, and hence $(x - \widetilde{x}) \perp y$ for all $y \in \widetilde{X}$. So this is the projection to $\widetilde{X}$. I claim that $x = \widetilde{x}$ iff $x \in \widetilde{X}$: if $x \in \widetilde{X}$ then $(x, e_n) = (\widetilde{x}, e_n)$ for all n. (Now use an argument like before to show $x = \widetilde{x}$.)

PROPOSITION 26.6 (Bessel's inequality).

$$\sum_1^\infty |(x, e_n)|^2 \le \|x\|^2$$

with equality iff $\|x - \widetilde{x}\|^2 = 0$*, iff* $x \in \widetilde{X}$*, iff* $x = \sum_1^\infty (x, e_n) e_n \in X$.

All these statements are independent of the ordering of the orthonormal system.

DEFINITION 26.7. An orthonormal system e_n for $n \in \mathbb{N}$ is *complete* if the closure of the span of the e_n is all of X. Equivalently, for all $x \in X$, $x = \sum_1^\infty (x, e_n) e_n$; also equivalently, no $x \in X$, $x \neq 0$ is $\perp$ to e_n for all n (use the previous expression for x).

27. November 9

27.1. Orthonormal systems. Example: in $L^2(\mathbb{T})$, $e_n(x) = e^{2\pi i n x}$ form a complete orthonormal system: $f \perp e_n \implies f = 0$ in $L^2(\mathbb{T})$ (i.e. $\widehat{f}(n) = 0\ \forall n \implies f = 0$). With our new notation, $\widehat{f}(n) = \langle f, e_n \rangle$ and $f = \sum_{n=-\infty}^\infty \widehat{f}(n) e_n$, with the sum converging in $L^2(\mathbb{T})$. Parseval's identity says $\|f\|_2^2 = \sum |\widehat{f}(n)|^2$ for $f \in L^2(\mathbb{T})$.

One can always find orthonormal systems by the Gram-Schmidt process. Start with a linearly independent set $f_0, f_1, \cdots$ in a Hilbert space X. We can find orthonormal $e_0, \cdots, e_n$ with the same span as $f_0, \cdots, f_n$ and (f_i, e_i) real and > 0. Define $e_0 = \frac{f_0}{\|f_0\|}$, and define $e_n = \frac{E_n}{\|E_n\|}$ where $E_n = f_n - \sum_{i=1}^{n-1} (f_n, e_i) e_i$.

In $L^2([-1, 1])$ take $f_n(x) = x_n$. These are linearly independent; using Gram-Schmidt, we can get polynomials e_n of degree n that are orthonormal. We end up with

$$e_n = P_n \cdot \sqrt{\frac{2}{2n+1}}$$

where the polynomials P_n are called the Legendre polynomials. These e_n are a complete orthonormal system. Now look at the orthonormal system $e_n = e^{2\pi i n x}$. If $\langle f, e_n \rangle_{L^2[-1,1]} = 0$ for all n then since Fourier transform can be written

$$\langle f, \sum_0^n \frac{1}{m!} (2\pi i c x)^m \rangle = 0$$

$\int_{-1}^1 f(x) e^{-2\pi i c x}\,\mathrm{d}x = 0$ for all c implies that the fourier transform is zero, so $f = 0$ in $L^2([-1, 1])$.

Now look at $X = L^2(D)$ where D is the unit disk in $\mathbb{R}^2$. If $f_n = x^n$ then

$$\langle f_n, f_m \rangle = \int x^{n+m} 2\sqrt{1 - x^2}\,\mathrm{d}x$$

Gram Schmidt produces a series of orthonormal e_n. These are related to the Chebyshev polynomials U_n of the second kind:

$$e_n = \frac{U_n}{\sqrt{\pi}}$$

$\|1\|_{L^2} = \sqrt{\pi}$; $U_0 = 1, U_1 = 2x, U_2 = 4x^2 - 1$

Apply Gram-Schmidt to $f_n = x^n e^{-\pi x^2/2}$ in $X = L^2(\mathbb{R})$. For example, $\|f_0\|^2 = \int e^{-\pi x^2} = 1$. We get $e_n(x) = h_n(x)e^{-\pi x^2/2}$ where h_n are the Hermite functions. These form a complete linear system: if $\langle f, e_n\rangle = 0$ for all n then $\langle f, polyn(x)e^{-\pi x^2/2}\rangle = 0$ for all polynomials; this implies $\langle f, e^{2\pi i\xi x}e^{-\pi x - 2/2}\rangle = 0$ for all ξ. So

$$\int f(x)e^{-\pi x^2/2}e^{-2\pi i\xi x}\,\mathrm{d}x = 0$$

Since $\widehat{f(x)e^{-\pi x^2/2}} = 0$ then $f(x)e^{-\pi x^2/2} = 0$, which implies that $f = 0$.

In general, if X is a separable Hilbert space, then X contains a complete orthonormal system: take any countable set (f_n) whose span is dense, throw away elements until the elements are linearly independent, and then apply Gram-Schmidt.

In the non-separable case, we can have an orthonormal system e_i where the indices i might be uncountable. If $x \in X$ then $\langle x, e_i\rangle \neq 0$ only for countably many indices. This is a consequence of Bessel's inequality

$$\sum_{n=1}^{N} |\langle x, e_{i_n}\rangle|^2 \leq \|x\|^2$$

which implies that the number of indices $i \in I$ with $|\langle x, e_n\rangle|^2 \geq \varepsilon$ is $\leq \|x\|^2/\varepsilon$. (Union together the finite sets corresponding to $\varepsilon = \frac{1}{n}$, and you get a countable set.) As last time, if $x \in X$ we get $\widetilde{x} = \sum_{i\in I}(x, e_i)e_i$. Then $\|\widetilde{x}\| \leq \|x\|$.

In a complete orthonormal system, if $x \perp e_i$ for all i then $x = 0$, and $span(e_i : i \in I)$ is dense. Either of these criteria are equivalent to having equality in Bessel's inequality. In general, to get a complete orthonormal system, we have to use Zorn's lemma. The first condition says that you can't take a complete orthonormal system and make it bigger. So, complete means "maximal," in the sense of not being contained in any strictly larger orthonormal system. Now Zorn's lemma just works. (But mainly, we only care about separable Hilbert spaces, and we don't need Zorn's lemma.)

Theorem 27.1 (Riesz representation theorem for Hilbert spaces). *Given a Hilbert space X and a bounded linear functional $\alpha : X \to \mathbb{C}$ there exists $a \in X$ such that $\alpha(x) = (x, a)$*[1] *for all $x \in X$.*

$|\alpha(x)| \leq M \cdot \|x\|$ where $M = \|a\|$. This uses Cauchy-Schwartz, which says $|\langle x, y\rangle| \leq \|x\|\|y\|$; this is a special case of Bessel's inequality: $\sum_{n=1}^{1} |\langle x, e_n\rangle|^2 \leq \|x\|^2$.

[1]Our inner products are linear in the first spot and conjugate linear in the second.

PROOF. Use an orthonormal system… or define

$$\Gamma : X \to \mathbb{R} \text{ where } \Gamma(x) = \|x\|^2 - Re(\alpha(x))$$

This is bounded below: $\Gamma \geq \|x\|^2 - M \cdot \|x\| \geq -\frac{M}{4}$ where $M = \|\alpha\|$. We have to show that it achieves its infimum, Γ_0. Take a sequence $x_n \in X$ with $\Gamma(x_n) \to \Gamma_0$ and show that this is a Cauchy sequence. If $\Gamma(x_n)$ and $\Gamma(x_m)$ are both in $[\Gamma_0, \Gamma_0 + \varepsilon]$

$$\frac{1}{4}\|x_n - x_m\|^2 = \frac{1}{2}(\|x_n\|^2 + \|x_m\|^2) - \|\frac{x_n + x^m}{2}\|^2$$

$$= \frac{1}{2}(\Gamma(x_n) + \Gamma(x_m)) - \Gamma\left(\frac{x_n + x_m}{2}\right) \quad \alpha \text{ terms cancel beca}$$

$$\leq (\Gamma_0 + \varepsilon) - \Gamma_0 = \varepsilon$$

So $\|x_n - x_m\| \leq \sqrt{4\varepsilon}$ once $\Gamma(x_n), \Gamma(x_m) \leq \Gamma_0 + \varepsilon$. Let $x_* = \lim x_n$. Look at $x_* + tv$ for $t \in \mathbb{R}$, $v \in X$. In the infimum, the derivative is zero:

$$0 = \frac{d}{dt}\Gamma(x_* + tv)|_0$$
$$= (x_*, v) + (v, x_*) - Re(\alpha(v)))$$

That is, for all $v \in X$, $Re(2(v, x_*) - \alpha(v)) = 0$. Apply this to iv as well, getting that the imaginary part is also zero. That is, $\alpha(v) = (v, 2x_*)$ for all v. So $a = 2x_*$, and this is the Riesz representation theorem. ⊟

We get a map $I : X \to X^*$, with $a \mapsto \alpha_a$, where $\alpha_a(x) = (x, a)$. This is injective; the Riesz representation theore m says that it is surjective. This is not a complex linear map: $I(\lambda a) \neq \lambda\alpha_a$; instead, $I(\lambda a) = \overline{\lambda}\alpha_a$ (this is because the λ ends up in the second spot of (x, a), which is the conjugate linear one). So the natural map $I : X \to X^*$ is conjugate linear.

28. November 14

28.1. More about Hilbert spaces.

PROPOSITION 28.1. *If X is a Hilbert space, and $V \subset X$ is a closed linear subspace, then $X = V \oplus V^\perp$, where $V^\perp = \{w : (w, v) = 0 \ \forall v \in V\}$.*

PROOF. It is certainly true that $V^\perp$ is a vector space, and $V \cap V^\perp = \{0\}$. The interesting part is to show that every element $x \in X$ can be written $x = v + w$ for $v \in V$ and $w \in V^\perp$. Equivalently, there is some $v \in V$ such that $x - v$ is $\perp$ to all elements of V. Use the Riesz representation theorem for the Hilbert space V: consider the map $\alpha : V \to \mathbb{C}$ which sends $u \mapsto (u, x)$ (where x is fixed). But Riesz says that there is some $v \in V$ such that $\alpha(u) = (u, v)$ for all $u \in V$. That is, $(u, x - v) = 0$ for all $u \in V$. So $x - v \in V^\perp$. ⊟

REMARK 28.2. If X is a Banach space and $V \subset X$ is a closed linear subspace, there may be no closed subspace $W \subset X$ such that $X = V \oplus W$. If V does have such a W, V is called *complemented*.

28.2. Baire Category theorem.

LEMMA 28.3. *Let X be a complete metric space. Let $S_1 \supset S_1 \supset \cdots$ be nested closed sets in X with $\varepsilon_n = diam(S_n) \to 0$ as $n \to \infty$. The intersection $\bigcap S_n$ is one point. (Recall the diameter is $\sup_{x,y \in S_n} d(x,y)$.)*

PROOF. It's pretty clear that there can't be two points in the intersection, because otherwise there would be a positive distance between them. Choose $x_n \in S_n$ for all n. For $m, m' \geq n$ then $d(x_m, x_{m'}) \leq \varepsilon_n$. So these are Cauchy. Let $x = \lim x_m$. Then $x \in S_n$ because S_n is closed and $x_m \in S_n$ for $m \geq n$. So $x \in \bigcap S_n$. ⊟

THEOREM 28.4 (Baire category theorem). *Let X be a complete metric space. Let G_n for $n \in \mathbb{N}$ be a collection of open and dense subsets. Then $G = \bigcap G_n$ is also dense in X.*

PROOF. Given $E \subset X$ open, we must show that $E \cap G$ is nonempty. $E \cap G_1$ is nonempty because G_1 is dense, and it's open because G_1 is open. So $G_1 \cap E$ contains an open metric ball $S_1 = B(x_0, \varepsilon) = \{x : d(x_0, x) < \varepsilon\}$. By making ε smaller, we can arrange for the closed ball $S_1^* \subset G_1 \cap E$. Now look at $S_1 \cap G_2$, which is open and nonempty as in the previous intersection. There is an open ball $S_2 = B(x_1, \varepsilon_2)$. We can arrange for $\varepsilon_2 \leq \frac{\varepsilon_1}{2}$. We can also arrange for $S_2^* \subset S_1 \cap G_2$. Continuing this process, we get $S_2 \subset S_2^* \subset S_{n-1} \cap G_n$ for each n, with $diam(S_n^*) \to 0$. By the lemma, there is some $x \in \bigcap S_n^* \subset \bigcap S_{n-1}$, so $x \in \bigcap G_n$ and $x \in E$. So $x \in E \cap G$. ⊟

DEFINITION 28.5. In X a set K is called *nowhere dense* if for all balls $B \subset X$, $K \cap B$ is not dense in B (i.e. it's not dense in any ball). Equivalently, $closure(K)$ does not contain any ball. Equivalently, $X \backslash closure(K)$ is open and dense.

THEOREM 28.6 (Baire category theorem, restated). *If K_n for $n \in \mathbb{N}$ is a nowhere dense set for all n, then the complement of $\bigcup K_n$ is dense in X.*

Set $K_n^* = closure(K_n)$, and set $G_n = X \backslash K_n^*$. The G_n are open and dense, by one of the definitions above. Then $\bigcap G_n$ is dense, i.e. $X \backslash \bigcup K_n^*$ is dense, so $X \backslash \bigcup K_n$ is dense.

DEFINITION 28.7. $K \subset X$ is meager if it is a countable union of nowhere dense subsets (i.e. looks like the $\bigcup K_n$ above).

THEOREM 28.8 (Baire category theorem, again). *The complement of a meager set is nonempty.*

For example, $\mathbb{R}^n$ is not a countable union of proper, affine subspaces. (So you can't fill up all of $\mathbb{R}^2$ with countably many lines.)

28.3. Uniform boundedness principle.

THEOREM 28.9. *Let X be a Banach space, Y a normed space, and $E \subset L(X,Y)$ be a subset of bounded linear operators. Suppose that , for every $x \in X$, there exists $M_x \geq 0$ such that $\|Tx\| \leq M_x$ for all $T \in E$. Then there exists some M such that $\|T\| \leq M$ for all $T \in E$.*

Proof. For each $n \in \mathbb{N}$ let $K_n \subset X$ be the set

$$K_n = \{x \in X \,:\, \|Tx\| \leq n \;\forall T \in E\}$$

This is closed: $\|Tx\| \leq n$ is a closed condition, and $\forall T \in E$ implies an intersection. If $n \geq M_x$ then $x \in K_n$. So for all x there is some n such that $x \in K_n$. That is, $X = \cap_1^\infty K_n$. By the Baire category theorem, there is some n such that K_n is not nowhere dense. That is, the closure (well, K_n itself) contains a ball – a closed ball, if we like. So $K_n \supset B^*(x_0;\varepsilon) = \{x \,:\, \|x - x_0\| \leq \varepsilon\}$. For any $x \in X$ look at $x_0 + \varepsilon\frac{x}{\|x\|}$. It is in K_n because it is in the ball $B(x_0, \varepsilon)$. So for all $T \in E$,

$$\left\|T(x_0 + \varepsilon\frac{x}{\|x\|})\right\| \leq n$$

So by the triangle inequality,

$$\left\|T(\varepsilon\frac{x}{\|x\|})\right\| \leq \|Tx_0\| + n \leq M_{x_0} + n$$

So for all $T \in E$, for all $x \in X$ $\|Tx\| \leq \frac{1}{\varepsilon}(M_{x_0} + n)\|x\|$. So

$$\|T\| \leq \frac{1}{\varepsilon}(M_{x_0} + n)$$

$\boxminus$

29. November 16

29.1. Uniform Boundedness.

We were considering a subspace $E \subset L(X, Y)$.

Theorem 29.1 (Uniform boundedness principle). *If X is a Banach space and $\alpha_n \in X^*$ for $n \in \mathbb{N}$, and if for all x, $(|\alpha_n(x)|)_n$ is bounded by M_x, then $\|\alpha_n\|$ is bounded. That is, if for all n there is some $x_n \in X$ with $\|x_n\| \leq 1$ and $|\alpha_n(x_n)|$ unbounded then there is some x with $|\alpha_n(x)|$ is unbounded as $n \to \infty$.*

Application 29.2. There exists a continuous periodic function $f \in C(\mathbb{T})$ such that the Fourier series $\sum \widehat{f}(n)e^{2\pi in}$ does not converge at $x = 0$. Indeed, the partial sums

$$s_n(f) = \sum_{-n}^{n} \widehat{f}(n)e^{2\pi inx}$$

have $|s_n(f)(0)|$ an unbounded sequence.

Remark 29.3. When we regard $C(\mathbb{T})$ as a Banach space, we usually do so by embedding it in $L^\infty(\mathbb{T})$ so the norm is $\|f\| = \sup|f|$.

Proof. Define $\alpha^n : C(\mathbb{T}) \to \mathbb{C}$ a bounded linear functional by $\alpha_n(f) = s_n(f)(0)$. We want to find $f_n \in C(\mathbb{T})$ with $\sup|f_n| \leq 1$ and $|s_n(f_n)(0)| \to \infty$. (Once we have found such, the uniform boundedness principle implies that there is some f with $|\alpha_n(f)|$ unbounded, which is what we want.) Recall that $s_n(f) = D_n *_{\mathbb{T}} f$. So $s_n(f) = \int_0^1 f(x)D_n(x)dx$. Let $g_n(x) = sign(D_n(x))$. (The domain of D_n can be broken up into five places where the sign is alternating.)

$$\begin{aligned}
\int_0^1 g_n D_n = \int_0^1 |D_n| &= \int_0^1 \frac{|\sin(\pi(n+1)x)|}{\sin(\pi x)} dx \\
&\geq \int_0^1 \frac{|\sin(\pi(n+1)x)|}{\pi x} \\
&= \int_0^{n+1} \frac{|\sin(\pi y)|}{\pi y} dy \\
&\geq \left(\frac{1}{\pi}\int_0^1 \sin(\pi y) dy\right) \times \sum_0^{n+1} \frac{1}{m} \\
&= \frac{2}{\pi^2} \sum_1^{n+1} \frac{1}{m} \geq \frac{2}{\pi^2} \log(n+2) \to \infty
\end{aligned}$$

We were trying to find continuous functions with $\sup|f_n| \leq 1$ such that $\int_0^1 f_n D_n \to \infty$; just approximate g_n by continuous functions. ⊟

This isn't all that rare: the functions with $(s_n f)(0)$ divergent are a comeager set in $C(\mathbb{T})$ (that is, the complement of a meager set).

29.2. Weak convergence. If X is a Banach space, $x_n \to x$ usually means $\|x_n - x\| \to 0$; this is "strong" convergence, a.k.a. "convergence in norm." What about other notions of convergence?

DEFINITION 29.4 (Weak convergence). A sequence $(x_n)_n$ converges weakly to $x \in X$ if for every $\alpha \in X^*$, $\alpha(x_n) \to \alpha(x)$. In this case write $x_n \stackrel{w}{\to} x$.

EXAMPLE 29.5. If X is a Hilbert space and x_n is an orthonormal sequence, then $x_n \not\to 0$ in norm. But $x_n \stackrel{w}{\to} 0$ because for all $a \in X$, $(x_n, a) \to 0$ as $n \to \infty$. Indeed, from Bessel's inequality $\sum_1^\infty |(x_n, a)|^2 \leq \|a\|^2$.

DEFINITION 29.6 (Weak-* convergence). Let α_n for $n \in \mathbb{N}$ be a sequence in the dual space X^* where X is a Banach space. We say that $\alpha_n \stackrel{w*}{\to} \alpha$ if for all $x \in X$, $\alpha_n(x) \to \alpha(x)$ as $n \to \infty$.

REMARK 29.7. For a sequence $\alpha_n \in X^*$, $\alpha_n \stackrel{w*}{\to} \alpha$ iff for all $\varphi \in X^{**}$, $\varphi(\alpha_n) \to \varphi(\alpha)$ as $n \to \infty$. If X is reflexive, then this is equivalent to $\alpha_n \stackrel{w}{\to} \alpha$.

PROPOSITION 29.8. *Let X be a Banach space and suppose $\alpha_n \stackrel{w*}{\to} \alpha$ in X^*. Then $\|\alpha_n\|$ is a bounded sequence.*

PROOF. Use the uniform boundedness principle. If $(\alpha_n)_n$ is not a bounded sequence, then there exists x such that $|\alpha_n(x)|$ is not a bounded sequence. So $\alpha_n(x) \not\to \alpha(x)$, which implies $\alpha_n \stackrel{w*}{\not\to} \alpha$. ⊟

PROPOSITION 29.9. *Let X be a separable Banach space. Let $\alpha_n \in X^*$ be a bounded sequence: $\|\alpha_n\| \leq M$ for all n. Then there is some $\alpha \in X^*$ and a subsequence n_k with $\alpha_{n_k} \stackrel{w*}{\to} \alpha$.*

PROOF. Choose in X a countable dense set x_k. By hypothesis $\alpha_n(x_1)$ is a bounded sequence. So there is some subset $\mathbb{N}^{(1)} \subset \mathbb{N}$ with

$$\lim_{\substack{n \in \mathbb{N}^{(1)} \\ n \to \infty}} \alpha_n(x_1) = \lambda_1$$

Find a convergent subsequence: there is some $\mathbb{N}^{(1)} \subset \mathbb{N}^{(1)}$ such that $\alpha_n(x_2) \to \lambda^2$ as $n \in \mathbb{N}^{(2)}$, $n \to \infty$. Keep doing this. Diagonalize to get $\mathbb{N}' \subset \mathbb{N}$ such that for all k, $\alpha_n(x_k)$ is convergent as $n \in \mathbb{N}'$. It follows that $\alpha_n(x)$ is convergent for $n \in \mathbb{N}'$, and all $x \in X$. Indeed, it's Cauchy: given $\varepsilon > 0$ find k such that $\|x - x_k\| \leq \frac{\varepsilon}{3M}$. Find n_0 such that $\forall n', \geq n_0$ where $n', n \in \mathbb{N}'$ we have $|\alpha_n(x_k o - \alpha_{n'}(x_k)) \leq \frac{\varepsilon}{3}$. $\|\alpha_n\| \leq M$ so

$$|\alpha_n(x_k) - \alpha_n(x) \leq M\|x_k - x\| \leq \frac{\varepsilon}{3}$$

So $|\alpha_n(x) - \alpha_{n'}(x)| \leq \varepsilon$ for $n, n' \geq n_0$.

Now define

$$\alpha(x) = \lim_{\substack{n \to \infty \\ n \in \mathbb{N}}} \alpha_n(x)$$

Need to check that this is a bounded linear functional. And $\alpha_n \stackrel{w*}{\to} \alpha$ because $\alpha_n(x) \to \alpha(x)$ for all x. ⊟

COROLLARY 29.10. *Let X (hence X^*) be a reflexive, separable Banach space. If $x_n \in X$ is a bounded sequence, then there is a weakly-convergent subsequence $x_{n'} \stackrel{w}{\to} x$ for $n' \in \mathbb{N}'$, $n' \to \infty$.*

EXAMPLE 29.11. Look at ℓ^1 as the dual space of c_0. Let $e_n = (0, \cdots, 0, 1, 0, \cdots) \in \ell^1$ be the sequence with a single 1 in the n^{th} spot. The e_n do not converge to zero in the strong sense, because $\|e_n - 0\| = 1$. But $e_n \stackrel{w*}{\to} 0$, because for all $a \in c_0$, $e_n(a) = a_n \to 0$ as $n \to \infty$.

$e_n \stackrel{w}{\not\to} 0$: the dual of ℓ^1 is isomorphic to ℓ^∞. Consider in ℓ^∞ the sequence $b = (1, \cdots, 1)$. As an element of $(\ell^1)^*$, $b(e_n)$ is the n^{th} term of b (recall the $\sum a_n b_n$ definition...). But $b(e_n) \not\to 0$ as $n \to \infty$, so the sequence is not weakly convergent.

30. November 18

THEOREM 30.1. *Let X, Y be Banach spaces, let $T \in L(X, Y)$ be a surjective bounded linear operator. Then T is an "open mapping": i.e. if $G \subset X$ is an open set, then $T(G)$ is open in Y.*

PROOF. Let $U_r \subset X$ be the open ball $\{x : \|x\| < r\}$ and $V_r = \{y : \|y\| < r\} \subset Y$. We will start by proving that $\overline{TU_1}$ contains V_ε for some $\varepsilon > 0$. $\overline{TU_n}$ are an increasing sequence of closed subsets. Because $X = \bigcup U_n$, $Y = T(X)$ implies $Y = \bigcup_n TU_n$, and $Y = \bigcup_n \overline{TU_n}$ because it is closed. By the Baire Category theorem, some $\overline{TU_n}$ contains an open ball $V = \{y : \|y - y_0\| < 2\varepsilon\} \subset Y$. By scaling, without loss of generality we can assume that this ball is contained in TU_1. Define $V_{2\varepsilon} = V - \{y_0\} \subset Y$ (subtraction, not set minus). Then

$$V_{2\varepsilon} \subset \overline{TU_1} - \overline{TU_1} \subset \overline{TU_2}$$

so $V_\varepsilon \subset \overline{TU_1}$.

Now I clam that $\overline{TU_1} \subset TU_2$. Let $y \in \overline{TU_1}$. There exists $x_1 \in U_1$ such that $\|y - Tx_1\| < \frac{\varepsilon}{2}$. So $y - Tx_1 \in V_{\frac{\varepsilon}{2}} \subset \overline{TU_{\frac{1}{2}}}$. Using the same argument, there is some $x_2 \in U_{\frac{1}{2}}$ such that $\|y - Tx_1 - Tx_2\| < \frac{\varepsilon}{4}$ so $y - Tx_1 - Tx_2 \in V_{\frac{\varepsilon}{4}}$. Continuing this for all n we get

$$x_n \in U_{2^{-(n-1)}} \qquad \text{with } \|y - Tx_1 - \cdots - Tx_n\| < 2^{-n}\varepsilon \tag{6}$$

Set $x = \sum x_n$ which exists because X is complete and $\sum \|x_n\| < 2 = 1 + \frac{1}{2} + \cdots$ Indeed $x \in U_2$ and $\|y - Tx\| = 0$ from (6). So $y \in TU_2$. Now see $TU_2 \supset V_\varepsilon$. By scaling, $TU_\delta \supset V_{\varepsilon \cdot \frac{\delta}{2}}$. So if U is a neighborhood of $0 \in X$, then TU contains some $V_{\varepsilon'}$.

So now let $G \subset X$ be open. Consider $y = Tx$ for some $x \in G$. We must show that TG contains a ball around y. $U = G - x$ (again group minus, not set minus!) is a neighborhood of 0 in X. So $TU = TG - Tx$ contains $V_{\varepsilon'}$, a ball around the origin in Y. So $TG \supset V_{\varepsilon'} + Tx = V_{\varepsilon'} + y$. ⊟

COROLLARY 30.2. *If X, Y are Banach spaces and $T : X \to Y$ is bounded, linear, and bijective, then the inverse map $T^{-1} : Y \to X$ is also bounded.*

PROOF. The open mapping theorem says that T^{-1} is continuous; but continuous $\iff$ bounded for linear operators on Banach spaces. We saw that $TU_1 \supset V_{\frac{\varepsilon}{2}}$, so $T^{-1}(V_{\frac{\varepsilon}{2}}) \subset TU_1$, a.k.a. $T^{-1}(V_1) \subset U_{\frac{2}{\varepsilon}}$. This means that T^{-1} sends things of norm ≤ 1 to things of norm $\leq \frac{2}{\varepsilon}$:

$$\|T^{-1}\| \leq \frac{2}{\varepsilon}$$

⊟

EXAMPLE 30.3. Let $\mathcal{F} : L^1(\mathbb{T}) \to c_0(\mathbb{Z})$ be the Fourier series operator sending $f \mapsto (\widehat{f}(n))_n$. This is a bounded linear map: $\|\mathcal{F}\| \leq 1$ because $\sup_n |\widehat{f}(n)| \leq \|f\|_{L^1}$, which means $\|\mathcal{F}(f)\|_{c_0} \leq \|f\|_{L^1}$. This is injective because the function can be recovered a.e. from its coefficients.

We will show that $\mathcal{F}$ is not surjective. If $\mathcal{F}$ were bijective, then $\mathcal{F}^{-1}$ would be bounded by Corollary 30.2. So

$$\|\mathcal{F}^{-1} a\|_{L^1} \leq c \cdot \|a\|_{c_0} = c \cdot \sup_n |a(n)|$$

for $a \in c_0(\mathbb{Z})$ and some constant c. For $f \in L^1(\mathbb{T})$,

$$\|f\|_{L^1(\mathbb{T})} \le c \cdot \sup_n |\widehat{f}(n)|$$

Take $f = D_k$ (the k^{th} Dirichlet kernel). It turns out that $\|f\|_{L^1(\mathbb{T})} \sim c' \cdot \log k$, but $\widehat{f}(n)$ is $(\cdots, 0, 1, \cdots, 1, 0, \cdots)$ where the ones are from indices $-k$ to k. So $\sup_n |\widehat{f}(n)| = 1$. There is no such constant c that is independent of f, so we have a contradiction.

If $T : X \to Y$ is bijective with $T^{-1} \in L(Y, X)$, then X and Y are isomorphic Banach spaces. But they might not be isometric spaces.

QUESTION 30.4. Is $L^1(\mathbb{T}) \cong c_0(\mathbb{Z})$ as Banach spaces?

No. $L^1(\mathbb{T})^* = L^\infty(\mathbb{T})$ but $c_0(\mathbb{Z})^* = \ell^1(\mathbb{Z})$. $\ell^1(\mathbb{Z})$ is separable, while $L^\infty(\mathbb{T})$ is not. $L^1(\mathbb{T})$ cannot be isomorphic to $c_0(\mathbb{Z})$ are not isomorphic, because their duals are not isomorphic.

Here is a reformulation of the open mapping theorem:

THEOREM 30.5 (Closed graph theorem). *Let X and Y be Banach spaces, and $T : X \to Y$ be a linear map (not necessarily bounded). Suppose T has closed graph: $\Gamma(T) := \{(x, Tx) : x \in X\}$ is closed in $X \times Y$. Then T is bounded.*

If $x_n \to x$ and $y_n \to y$, and $y_n = Tx_n$ for all n (i.e. $(x_n, y_n) \in \Gamma(T)$) then closed-ness shows that $(x, y) \in \Gamma(T)$, i.e. $y = Tx$. Note that this is not quite the same as continuity: we had to assume the y_n converged to something to begin with. (Think about $f : \mathbb{R} \to \mathbb{R}$ where $x \mapsto \frac{1}{x}$ and $0 \mapsto 0$. f has closed graph but is not continuous.)

$X \times Y$ is a Banach space when given the norm $\|(x, y)\|_{X\times Y} = \|x\| + \|y\|$. (Convergence of $(x_n, y_n)_n$ means $(x_n)_n$ and $(y_n)_n$ both converge.)

PROOF. $\Gamma(T) \subset X \times Y$ is a closed linear subspace of a Banach space, so it is a Banach space. Let $\pi : \Gamma(T) \to X$ be the projection $(x, Tx) \mapsto x$. This is a bounded, linear, bijective map. (To see boundedness, note that $\|x\| \le \|(x, y)\|$ so $\|\pi\| \le 1$.) By Corollary 30.2, $\pi^{-1} : X \to \Gamma(T)$ is bounded too: this is the map $x \mapsto (x, Tx)$. So there is some M such that $\|(x, Tx)\| \le M \cdot \|x\|$ for all x, or equivalently

$$\|x\| + \|Tx\| \le M \cdot \|x\|$$

and there is some M' such that $\|Tx\| \le M \cdot \|x\|$. ⊟

31. November 21

31.1. Measure theory. Let X be any set. A collection of subsets $\mathscr{M} \subset \mathscr{P}(X)$ (where $\mathscr{P}(X)$ is the collection of all subsets) is a σ-algebra if:

(1) $\emptyset \in \mathscr{M}$
(2) $\mathscr{M}$ is closed under complements (so $X \in \mathscr{M}$)
(3) $\mathscr{M}$ is closed under countable unions.

EXAMPLE 31.1.
- $\mathscr{M} = \{\emptyset, X\}$
- $\mathscr{M} = \mathscr{P}(X)$
- $\mathscr{M} = \mathcal{L}(\mathbb{R}^d)$ (Lebesgue-measurable subsets)
- $\mathscr{M} =$ all subsets $E \subset X$ such that either E is countable or $X \backslash E$ is countable.
- If $\mathscr{M}_1$ and $\mathscr{M}_2$ are σ-algebras, then $\mathscr{M}_1 \cap \mathscr{M}_2$. Actually, any intersection of $\mathscr{M}_\alpha$ works.

Given any collection $S \subset \mathscr{P}(X)$ of subsets of X, we can look at the intersection of all σ-algebras $\mathscr{M}$ containing S. This is called the σ-algebra generated by S. For example, in $\mathbb{R}$ if you take $S = ([0,2],[1,3])$ you end up with a finite set of intervals like $[0,1), [1,2], (2,3]$ etc.

EXAMPLE 31.2. $B(\mathbb{R}^d)$ is the σ-algebra generated by the open sets (this is the same as the σ-algebra generated by the rectangles). We know $B \subset \mathcal{L}$, but the inclusion is strict.

What is B? You have

(0) all open sets and all closed sets,
(1) all countable unions of sets from step (0), and all countable intersections of sets from step (0)
(2) etc. from steps zero and one...

But maybe you could have a set from step (0), a set from step (1), etc. and then union them together, or intersect them; this set is not in any step (n). This suggests considering a step (ω), and keep going.

DEFINITION 31.3. A measurable space is a pair $(X, \mathscr{M})$ where $\mathscr{M} \subset \mathscr{P}(X)$ is a σ-algebra.

A measure space is a triple $(X, \mathscr{M}, \mu)$ where X and $\mathscr{M}$ are as above, and $\mu : \mathscr{M} \to [0, \infty]$ satisfies

(1) $\mu(\emptyset) = 0$
(2) (*Countable additivity*) If $E \in \mathscr{M}$ and $E = \bigcup_1^\infty E_k$ is a countable union of disjoint sets, then $\mu(E) = \sum \mu(E_k)$

EXAMPLE 31.4.
- $(\mathbb{R}^d, \mathcal{L}, \lambda)$, where λ is Lebesgue measure. Or, consider $(\mathbb{R}^d, B, \lambda)$.
- $(X, \mathscr{P}(X), \gamma)$ is a measure space, where γ is the counting measure:

$$\gamma(E) = \begin{cases} \# \text{ of elements if } E \text{ if } E \text{ is finite} \\ \infty \text{ otherwise} \end{cases}$$

- There are other Borel measures on $\mathbb{R}$: $(\mathbb{R}, B, \mu)$

DEFINITION 31.5. In $(X, \mathscr{M}, \mu)$ call $E \in \mathscr{M}$ a null set if $\mu(E) = 0$. Then $(X, \mathscr{M}, \mu)$ is complete if every subset E' of a null set is measurable (i.e. is in $\mathscr{M}$). (This implies $\mu(E') = 0$.)

$(\mathbb{R}^d, \mathcal{L}, \lambda)$ is complete, but $(\mathbb{R}^d, B, \lambda)$ is not: the Cantor set is a Borel null set, but "most" of its subsets are not Borel sets. (The cardinality of the subsets of the Borel sets is larger than the cardinality of the Borel sets)

WARNING 31.6. In the usual measure, we did not distinguish between open and closed intervals; these had the same measure because points had measure zero. But that might not be the case in an arbitrary measure space.

Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function, and tweak it by defining

$$\widetilde{F} = \lim_{\varepsilon \to 0} F(x + \varepsilon)$$

(So if you have a jump discontinuity, in $\widetilde{F}$ you get the higher point in the jump.) For $(a, b] \subset \mathbb{R}$ define $\mu(a, b] = \widetilde{F}(b) - \widetilde{F}(a)$. This defines a measure on B. What is the measure of a point?

$$\mu(\{b\}) = \mu(\bigcap_{\varepsilon > 0} (b - \varepsilon, b]) = \lim_{\varepsilon > 0} (b - \varepsilon, b] = \widetilde{F}(b) - \lim_{\varepsilon > 0} \widetilde{F}(b - \varepsilon)$$

So this is the size of the jump discontinuity on F, if there is one, at this point.

An outer measure on a set X is a map

$$\mu_* : \mathscr{P}(X) \to [0, \infty]$$

satisfying

- $\mu_*(\emptyset) = 0$
- (*Countable subadditivity*): If $E \subset \bigcup_{k=1}^{\infty}$ then $\mu_*(E) \leq \sum_{k=1}^{\infty} \mu_*(E_k)$.

Every outer measure (X, μ_*) gives rise to a measure space $(X, \mathscr{M}, \mu)$, by Caratheodory's construction (i.e. the same one we used to get Lebesgue measure out of the original outer measure). $E \in \mathscr{M}$ if, for all $A \subset X$, we have

$$\mu_*(A \cap E) + \mu_*(A \cap E^c) = \mu_*(A)$$

In this case, we say that $\mu(E) = \mu_*(E)$. We need to check that $\mathscr{M}$ is a σ-algebra, and that μ is countably additive. This involves the same proofs as the rectangle-outer-measure case.

FACT 31.7. $(X, \mathscr{M}, \mu)$ as constructed by Caratheodory's construction is complete.

But how do you construct an outer measure? Let $S \subset \mathscr{P}(X)$ be any collection of subsets of X, with $\emptyset \in S$.[2] Suppose $\mu_0 : S \to [0, \infty]$ is any map taking $\emptyset \mapsto 0$. For any $E \subset X$

[2]This is in analogy with S being the set of rectangles used to construct outer measure before.

define

$$\mu_*(E) = \inf_{\substack{E=\cup E_k \\ E_k \in S,\ k\in\mathbb{N}}} \sum \mu_0(E_k)$$

There may be no covers $\{E_k\}$; in this case, take the infimum to be ∞. (i.e. use the convention $\inf\{\} = \infty$) This is an outer measure, by the proof of countable subadditivity in the Lebesgue measure case.

PROPOSITION 31.8. *μ_* is an outer measure.*

Does this agree with the original μ_0? No; elements of S might not belong to $\mathscr{M}$, and if they do, the measure might not be the same as before.

32. November 28

RECALL $(X, \mathscr{M}, \mu)$ was a measure space, where $\mathscr{M}$ was a σ-algebra and μ was a countably additive measure. We also had an outer measure space (X, μ_*) where $\mu_* : \mathscr{P}(X) \to [0, \infty]$ is defined on all subsets of X (whereas μ is only defined on the σ-algebra). We required it to have countable subadditivity:

$$\mu_*(E) \leq \sum_1^\infty \mu_*(E_i) \text{ if } E \subset \cup_1^\infty E_i$$

Let $\mathcal{S}$ be any collection $\mathcal{S} \subset \mathscr{P}(X)$ containing the empty set, and specify any function $\mu_0 : \mathcal{S} \to [0, \infty]$ that takes $\emptyset \mapsto 0$ (where $\mathcal{S}$ is analogous to the collection of rectangles in our original construction of Lebesgue measure).

CONSTRUCTION 32.1. Given μ_0 and a collection $\mathcal{S}$, define for $E \subset X$,

$$\mu_*(E) = \inf_{E=\bigcup R_i} \sum_{i=1}^\infty \mu_0(R_i)$$

This is an outer measure.

From a measure, we can construct a measure from Caratheodory's construction. However, by the time we've gotten a real measure μ, there might not be any measurable sets, and maybe $\mu_0(E) \neq \mu(E)$.

Why did this work before? Intersecting a rectangle with a half-space makes two disjoint sets where the volume is additive. Ditto volume is additive when you consider intersections of rectangles. We also proved:

PROPOSITION 32.2. *If $R \subset \mathbb{R}^n$ and $R \subset \bigcup_{i=1} \infty R_i$ then $vol(R) \leq \sum vol(R_i)$. (It is enough to check this when $\subset$ is $=$. This required Heine-Borel.)*

DEFINITION 32.3. A collection of sets $\mathcal{A} \subset \mathscr{P}(X)$ is a *ring* of sets if they are closed under finite unions and intersections, and under relative complements (that is, if $R_1, R_2 \in \mathcal{A}$ then $R_1 \backslash R_2 \in \mathcal{A}$). Equivalently, if $R_1, R_2 \in \mathcal{A}$ we need $R_1 \cap R_2 \in \mathcal{A}$ and $(R_1\backslash R_2) \cup (R_2 \backslash R_1) \in \mathcal{A}$.

(Think of intersection as multiplication and symmetric differences as addition; then this is actually a ring, where zero is the empty set, and 1 is the whole set X if it is in the ring).

DEFINITION 32.4. A *pre-measure* is a ring of sets $\mathcal{A}$ and a map $\mu_0 : \mathcal{A} \to [0, \infty]$ with $\mu_0(\emptyset) = 0$ that is finitely additive under disjoint unions, and satisfies:

$$\text{if } R \subset \textstyle\bigcup_1^\infty R_i \text{ for } R, R_i \in \mathcal{A} \text{ then } \mu_0(R) \leq \textstyle\sum_1^\infty \mu_0(R_i)$$

In our original setting, $\mathcal{A}$ is the finite unions and relative complements of rectangles.

THEOREM 32.5. *If $(X, \mathcal{A}, \mu_0)$ is a space with a pre-measure, and if μ_* is the resulting outer measure, and $(X, \mathscr{M}, \mu)$ is the resulting measure space by Caratheodory's construction, then $\mathcal{A} \subset \mathscr{M}$ and $\mu(R) = \mu_0(R)$ for all $R \in \mathcal{A}$.*

This works like before with the rectangles.

EXAMPLE 32.6. Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing, and right-continuous (i.e. $\lim_{y \searrow x} F(y) = F(x)$ for all x). Let $\mathcal{A}$ consist of all finite unions of half-open intervals $(a, b]$. Define $\mu_0((a, b]) = F(b) - F(a)$. Finite additivity comes from the fact that $(a, b] \cup (b, c] = (a, c]$, and countable additivity works as expected. If $(a, b] \subset \bigcup_1^\infty (a_k, b_k]$ then we need to check that $F(b) - F(a) \leq \sum (F(b_k) - F(a_k))$.

Given $\varepsilon > 0$, and using semi-continuity choose $\delta > 0$ such that $F(a + \delta) \leq F(a) + \varepsilon$. Choose $\delta_k > 0$ such that $F(b_k + \delta_k) \leq F(b_k) + 2^{-k}\varepsilon$. Then

$$[a + \delta, b] \subset \bigcup_1^\infty (a_k, b_k + \delta_k) \implies [a + \delta, b] \subset \bigcup_1^\mathbb{N} (a_k, b_k + \delta_k)$$

for some N. So

$$(a + \delta, b] \subset \bigcup_1^N (a_k, b_k + \delta_k]$$

which implies

$$F(b) - F(a + \delta) \leq \sum_1^N (F(b_k + \delta_k) - F(a_k))$$

$$F(b) - F(a) \leq \sum_1^\infty (F(b_k) - F(a_k)) + 2\varepsilon$$

So we get a measure space $(\mathbb{R}, \mathscr{M}, \mu_F)$ with $\mathscr{M} \supset B$ (the Borel subsets). Hence this restricts to a Borel measure space. We call μ_F the *Lebesgue-Stieltjes measure* corresponding to F.

EXAMPLE 32.7. If $F = \chi_{[0,\infty)}$ then

$$\mu_F = \begin{cases} 1 & \text{if } 0 \in E \\ 0 & \text{else} \end{cases}$$

If f is locally integrable over $\mathbb{R}$ and nonnegative, then define F such that $F(b) - F(a) = \int_a^b f$. In that case, $\mu_F(E) = \int f \cdot \chi_E$. If $f(x) = e^{-\pi x^2}$ then $\mu_F(\mathbb{R}) = 1$.

Given a measure space $(X, \mathscr{M}, \mu)$ as before we can define the integral $\int_X f \, \mathrm{d}\mu$ for $f : X \to \mathbb{R}$ a measurable function: that is, a function such that $f^{-1}(U) \in \mathscr{M}$ if U is open. For elementary functions $g = \sum_1^N a_i \chi_{E_i}$, for $E_i \in \mathscr{M}$ of finite measure, define

$$\int g \, \mathrm{d}\mu = \sum_1^N a_i \mu(E_i)$$

For $f \geq 0$ define

$$\int f \, \mathrm{d}\mu = \sup_{\substack{g \leq f \\ g \text{ elementary}}} \int g \, \mathrm{d}\mu$$

For all f, define

$$\int f = \int f_+ - \int f_-$$

We have the monotone convergence theorem and the dominated convergence theorem in this case. We get a Banach space $L^1(X, \mathscr{M}, \mu)$ of integrable functions f modulo a.e. zero functions.

EXAMPLE 32.8. If $X = \mathbb{N}$ and μ is the counting measure $\mu(E) = \#E$ and $\mathscr{M} = \mathscr{P}(\mathbb{N})$, then $L^1(X, \mathscr{M}, \mu) = \ell^1$. (Functions on $\mathbb{N}$ are just sequences, and we need them to have finitely many nonzero values, else the sum is infinite!)

DEFINITION 32.9. A measure space $(X, \mathscr{M}\mu)$ is called σ-finite if X can be covered by countably many $E \subset \mathscr{M}$ with $\mu(E) < \infty$.

An example of a space that is not σ-finite is $\mathbb{R}$ with the counting measure (you can't cover it with countably many finite sets).

EXAMPLE 32.10. Let $(X, \mathscr{M}, \sigma)$ arise from a pre-measure $(X, \mathcal{A}, \mu_0)$. Let $\mathcal{A}_\sigma$ be all the sets obtained as countable unions of elements of $\mathcal{A}$, and let $\mathcal{A}_{\sigma\sigma}$ be all countable intersections of $\mathcal{A}_\sigma$ sets. Let $E \in \mathscr{M}$ be of finite measure. Given $\varepsilon > 0$ we can find $\mathcal{O}_\varepsilon \supset E$ with $\mu(\mathcal{O}_\varepsilon) \leq \mu(E) + \varepsilon$ that is an element of $\mathcal{A}_\sigma$ (in the Lebesgue case, a countable union of rectangles), and such that $\mu(\mathcal{O}_\varepsilon \backslash E) \leq \varepsilon$. If $(\mathscr{M}\mu)$ is σ-finite, then for any $E \in \mathscr{M}$ there exists $\mathcal{O}_\varepsilon \in \mathcal{A}_\sigma$ such that $\mu(\mathcal{O}_\varepsilon \backslash E) \leq \varepsilon$. (Same proof as in Lebesgue case: use $2^{-k}\varepsilon$ argument.)

PROPOSITION 32.11. *Assume our space is σ-finite. For $E \in \mathscr{M}$ there is some $G \in \mathcal{A}_{\sigma\sigma}$ with $G \supset E$ and $\mu(G \backslash E) = 0$.*

PROOF. Put $G = \bigcap_{n=1}^\infty \mathcal{O}_{\frac{1}{n}}$ so $G \in \mathcal{A}_{\sigma\sigma}$. Then $G \supset E$ and $\mu(G \backslash E) \leq \frac{1}{n}$ for any n. $\boxminus$

33. November 30

33.1. Product measure. Let $(X_1, \mathscr{M}_1, \mu_1)$ and $(X_2, \mathscr{M}_2, \mu_2)$ be two measure spaces. Let $X = X_1 \times X_2$. On X, define a rectangle to be a product $E_1 \times E_2$, where $E_i \in \mathscr{M}_i$. Let $\mathcal{A}$ contain all finite unions of such rectangles. Note $\mathcal{A}$ is a ring, because it is closed under intersections, unions, and differences. Define $\mu_0 : \mathcal{A} \to \mathbb{R}$ by $\mu_0(E_1 \times E_2)$. Check that this is a well-defined, finitely additive map. Furthermore, it is a pre-measure: if $E \subset X_1 \times X_2$ is a rectangle, and $E_1 \times E_2 = E = \bigcup_{i=1}^\infty E_{1,i} \times E_{2,i}$ is a decomposition into disjoint rectangles, then we want this to be countably additive:

$$\mu_1(E_1)\mu_2(E_2) = \sum \mu_1(E_{1,i})\mu_2(E_{2,i})$$

See Stein-Shakarchi; the proof uses the monotone convergence theorem on both X_i. (Consider the measure on cross-sections $\{x\} \times E_2$.)

So this gives rise to a measure space $(X, \mathscr{M}, \mu)$, called the *product measure space.* If you start with Lebesgue measure on $\mathbb{R}$, this gives rise to Lebesgue measure on $\mathbb{R}^2$.

Suppose $X = \prod_{i\in I} X_i$, where I is a perhaps uncountable indexing set. Do not consider arbitrary products $\prod_{i\in I} E_i$, for $E_i \in \mathscr{M}$. You run into problems defining $\mu_0(\prod E_i) = \prod \mu_i(E_i)$: you can't take an infinite product of real numbers. To fix this, require each $(X_i, \mathscr{M}_i, \mu_i)$ to be a *probability space*: a measure space where $\mu_i(X_i) = 1$. (Euclidean space is not a probability space; the unit cube is.) Instead of considering arbitrary products, say that a rectangle will be a product $\prod_{i\in I} E_i$ where $E_i = X_i$ for all but finitely many i. So the infinite product makes sense, because all but finitely many of the factors are 1. This is a finite intersection of half-spaces. Let $\mathcal{A}$ be the collection of all finite unions of rectangles. Having defined $\mu_0(\textit{rectangle})$, we can define $\mu_0 : \mathcal{A} \to \mathbb{R}$ by linearity. (Check this is well-defined.) We need to check that this is a pre-measure (i.e. countable additivity). Get from μ_0 a measure space $(X, \mathscr{M}, \mu)$.

33.2. Push-forward. Let $(X, \mathscr{M}, \mu)$ be a measure space, and Y be a set. Let $X \xrightarrow{\pi} Y$ be any map, and consider the collection $\mathscr{N} \subset \mathscr{P}(Y)$ of subsets $F \subset Y$ with measurable preimage in X. $\mathscr{N}$ is a σ-algebra. Define

$$\nu : \mathscr{N} \to [0, \infty] \qquad \nu(F) = \mu(\pi^{-1}(F))$$

Then $(Y, \mathscr{N}, \nu)$ is a measure space. If $(X_i, \mathscr{M}_i, \mu_i)$ are probability spaces and $(X, \mathscr{M}, \mu)$ is the product, look at $\pi_i : X \to X_i$. The push-forward of μ by π_i is a measure on X_i, and this is an extension of μ_i: you get some $\mathscr{N}$ and ν as defined above, such that $\mathscr{N} \supset \mathscr{M}_i$ and $\nu|_{\mathscr{M}_i} = \mu_i$. If you apply this to the Borel sets, you get the Lebesgue measurable spaces back.

33.3. The language of probability.

THEOREM 33.1 (Wiener). *If b_m, for $m \in \mathbb{Z}$ are random variables that are $\mathbb{C}$-valued, independent, and each standard normal (mean = 0, standard deviation =1), then with* *probability 1, the series*

$$\sum_{m\neq 0} \frac{b_m}{m} e^{2\pi i m t}$$

is the Fourier series of a continuous function $\sigma(t)$. Also, with probability 1, $\sigma(t)$ satisfies Dini's criterion for all f, so the Fourier series converges.

Let $(\Omega, \mathscr{M}, \mathbb{P})$ be a probability space: a measure space such that $\mathbb{P}(\Omega) = 1$. A measurable subset of Ω (i.e. an element of $\mathscr{M}$) is called an event. The probability of E is its measure $\mathbb{P}(E)$. If $\mathbb{P}(E) = 1$, then we say that E is almost sure.

A random variable is a measurable function $a : \Omega \to \mathbb{R}$ or $\mathbb{C}$. So $a^{-1}(U) \in \mathscr{M}$ if U is open. If $a \in L^1$, then the *expected value* $\mathbb{E}(a)$ of a is $\int_\Omega a$.

If a random variable a has mean μ, then the *variance* of a is $\mathbb{E}((a-\mu)^2)$. Not every a has a mean, and not every a has a variance. In order for this to exist, we need $a \in L^2$. The standard deviation is $\sqrt{variance}$.

If $a_1, a_2 \in L^2$ (hence in L^1 because the space has finite measure) with $\mathbb{E}(a_i) = \mu_i$ then the covariance is

$$Cov(a_1, a_2) = \mathbb{E}((a_1 - \mu_1)(a_2 - \mu_2))$$

If $a_1, \cdots, a_n$ all have expected value zero and $Var(a_i) = 1$ and $Cov(a_i, a_j) = 0$, then the a_i are orthonormal in $L^2(\Omega, \mathbb{R})$.

34. December 2

Recall we had a probability space Ω. For a random variable a, the distribution function is a function $F : \mathbb{R} \to \mathbb{R}$ where $F(\lambda) = \mathbb{P}(a \leq \lambda)$, where $\lambda \in \mathbb{R}$. For example, a is a standard normal random variable if its distribution function is $F(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{x^2}{2}}\, dx$. If $F(\lambda) = \int_{-\infty}^{\lambda} f$ for some $f \in L^1(\mathbb{R})$ then f is called the *density function* for a. If a is a standard normal random variable, then $\mathbb{E}(a) = 0$ and $Var(a) = 1$. For every probability measure $\mathbb{P}$ we have a Lebesgue-Stiltjies measure on $\mathbb{R}$.

Random variables $a_1, \cdots, a_N$ are *independent* if for all $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$,

$$\mathbb{P}\left(\bigcap_1^N \{a_i \leq \lambda_i\}\right) = \prod_1^N \mathbb{P}(a_i \leq \lambda_i)$$

LEMMA 34.1. *Suppose $a_1, \cdots, a_N$ are independent standard normal random variables (ISNRV) and $b_m = \sum_1^N \lambda_{nm} a_n$ where (λ_{mn}) is an orthogonal matrix (so the covariance is zero). (ISNRV implies that the a_i are orthonormal in L^2. The b_n are also an orthonormal set, because they are a transformation by an orthogonal matrix.) Then the $b_1, \cdots, b_N$ are also ISNRV.*

ISNRV means: for all Borel sets $U \subset \mathbb{R}^N$, denote $x = (x_1, \cdots, x_N)$. Then

$$\mathbb{P}((a_1, \cdots, a_N) \int U) = \frac{1}{(2\pi)^{\frac{N}{2}}} \int_U e^{-|x|^2/2} dx_1 \cdots dx_N$$

This is rotation-invariant.

LEMMA 34.2. *If a_i for $i \in \mathbb{N}$ are ISNRV (i.e. any finite subcollection is), let $H \subset L^2(\Omega)$ be the closure of their span. Let $(b_i)_i$ be another complete orthonormal system in H. Then the b_i are ISNRV.*

(Proof in Kronheimer's notes. Take a limit of the previous case.) If a is a standard normal random variable, then

$$\mathbb{P}(a \geq M) = \int_M^\infty \frac{e^{-x^2}2}{\sqrt{2\pi}}\,\mathrm{d}x \leq e^{-M^2/2}$$

For a sequence of random variables, you might want a function that is $\geq$ all but finitely many of them.

PROPOSITION 34.3. *Fix $\beta > 1$. Let a_n be standard normal random variables. Then with probability 1,*

$$|a_n| \leq \sqrt{2\beta \log(n)}$$

for all but finitely many n.

LEMMA 34.4 (Borel-Cantelli). *If $\sum_1^\infty \mathbb{P}(E_n) < \infty$ then*

$$\bigcap_{N=1}^{\infty} \bigcup_{n \geq N} E_n = 0$$

That is, with probability 1, only finitely many occur.

PROOF OF PROPOSITION. Let E_n be the event

$$|a_n| > \sqrt{2\beta \log(n)}$$

$\mathbb{P}(E_n) \leq 2n^{-\beta}$ so $\sum 2n^{-\beta} < \infty$. Apply the Borel-Cantelli lemma. $\boxminus$

We were considering $\sum_{m \neq 0} b_m(\omega) \frac{e^{2\pi i m t}}{m}$, and wanted this to converge for all ω to a continuous function. $\frac{e^{2\pi i m t}}{m}$ looks like the antiderivative of $e^{2\pi i m t}$, which is an orthonormal collection. Let

$$L^2(\mathbb{T})' = \{f \in L^2 \,:\, \int_0^1 f = 0\}$$

We have an orthonormal system e_m where $e_m(t) = e^{2\pi i m t}$. Let ε_m be the antiderivative of e_m with constant chosen so that $\int e_m = 0$. Let $(d_n)_n$ be the complete orthonormal system in $L^2(\mathbb{T})'$

$$d_1(x) = \chi_{[0,\frac{1}{2}]} - \chi_{[\frac{1}{2},1]}$$
$$d_2 = \sqrt{2}\chi_{[0,0.25]} - \sqrt{2}\chi_{[0.25,0.5]}$$
$$d_3 = a\chi_{[0.5,0.75]} - a\chi_{[0.75,1]}$$

$$d_4 = 2\chi_{[0,0.125]} - \chi_{[0.125,0.25]}$$

The indefinite integrals $\widetilde{\delta}_n = \int_0^t d_n(s)ds$, $\delta_n(t) = \widetilde{\delta}_n(t) - c_n$ where $c_n = \int_0^1 \widetilde{\delta}_n$ For n large, $\widetilde{\delta}_n$ is just a small triangle.

PROPOSITION 34.5. *If a_n are ISNRV then with probability 1,*

$$\sum_1^\infty a_n \delta_n(t)$$

converges uniformly to a continuous function of t. (Actually, the function is Hölder continuous.)

With probability 1, a_n satisfies the relation above. For $2^m \le n < 2^{m+1}$, the support of the δ_n do not overlap. Do stuff; it's not hard.

Say that b is a $\mathbb{C}$-valued SNRV if $\Re(b)$ and $\Im(b)$ are ISNRV. In Lemma 34.3, take a_n to be independent and $\mathbb{C}$-valued. For a.a. $\omega \in \Omega$,

$$\sum_1^\infty a_n(\omega)\delta_n(t)$$

converges to $\sigma(t) : [0,1] \to \mathbb{R}$. Look at the Fourier coefficients $\widehat{\sigma}(m)$. (Remember that $\sigma(t)$ depends on ω.) By uniform convergence, and using the fact that δ_n is an antiderivative, write

$$\widehat{\sigma(m)} = \sum_{n=1}^\infty a_n(\omega)\widehat{\delta}_n(m)$$

$$\sum_{n=1}^\infty a_n(\omega)\frac{\widehat{d}_n(m)}{2\pi i m}$$

If $\lambda_{nm} = \widehat{d}_n(m)$ is a change of basis matrix between two complete orthonormal systems in $L^2(\mathbb{T})'$.

$$\widehat{\sigma}(m) = \sum_{n=1}^\infty \lambda_{nm}\frac{a_n(\omega)}{2\pi i m}$$

Because this is Hölder continuous, it satisfies Dini's criterion. So its Fourier series converges pointwise. (Recall, σ depends on ω, so this is "with probability 1" – there is some null set in Ω where this does not work.) So with probability 1,

$$\begin{aligned}\sigma(t) &= \sum \widehat{\sigma}(m)e^{2\pi i m t} \\ &= \sum_{n=1}^\infty \sum_{m\neq 0} \lambda_{nm}\frac{a_n(\omega)}{2\pi i m}e^{2\pi i m t} \\ &= \sum_{m \neq 0} b_m(\omega)\frac{e^{2\pi i m t}}{2\pi i m}\end{aligned}$$

where $b_m = \sum \lambda_{nm}a_n$. Because the a_n are ISNRV, so are the b_m (it's just a change of basis).

Take a linear combination of δ_n with random coefficients; you get a Brownian path.