MATH 205A NOTES

ARUN DEBRAY
AUGUST 21, 2015

These notes were taken in Stanford's Math 205A class in Fall 2014, taught by Lenya Ryzhik. I live-TEXed them using `vim`, and as such there may be typos; please send questions, comments, complaints, and corrections to `a.debray@math.utexas.edu`.

Contents

Part 1. Measure Theory

1. Outer Measure: 9/23/14

"The most important thing about this class is, there's no class Thursday."

Most books on real analysis are very boring, since the subject is kind of dry. Terry Tao's books (the course books) are more verbose, but they're nicer to read. Don't treat them as textbooks *per se*, but if you're ever on a beach and need something to read, try these books.[1] The books are *Introduction to Measure Theory* and *An Epsilon of Room*, both by Terry Tao.[2] Another book to look into is Evans and Gariepy, *Measure Theory and Fine Properties of Functions*; this is astoundingly precise and concise, with no motivation. Finally, on Fourier analysis, refer to the book *Fourier Analysis*, by Duoandikoetxea, and to M. Pinsky's *Introduction to Fourier Analysis and Wavelets*. There are misprints, but a wealth of useful information and interesting things.

In this class, we will spend a lot of time on measure and integration, then a small amount of Fourier analysis. Finally, we'll discuss a small amount of Brownian motion.

Now let's start talking about the Lesbegue measure. We want to extend the notion of the length of an interval to other sets. In two dimensions, for example, we calculated the area of a circle in elementary geometry by approximating it by refining the areas of polygons until we believed the limit existed. This is approximately what we'll do with the Lesbegue measure.

[1] The professor learned undergrad complex analysis this way, apparently.
[2] "Buy them, find them on a Chinese website, whatever."

We want the measure to meet the following four requirements:

(0) The measure of any set is defined.
(1) If I is an interval, then its measure $m(I) = \ell(I)$, its length.
(2) If $E_1, E_2, \ldots$ are disjoint, then
$$m\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{n} m(E_i).$$
(3) If $E \subseteq \mathbb{R}$ and $x \in \mathbb{R}$, then define $E + x = \{y \in \mathbb{R} : y = e + x, e \in E\}$. Then, we would like to have $m(E + x) = m(E)$, so that m is translation-invariant.

This turns out to be impossible. Oops.

Let's look at the half-open interval $[0, 1)$ as if it were the circle: define
$$x \oplus y = \begin{cases} x + y, & \text{if } x + y < 1 \\ x + y - 1, & \text{if } x + y \geq 1 \end{cases}$$

Claim. Assuming postulates (0) – (3), then $m(E \oplus x) = m(E)$.

Proof. Let $E_1 = \{y \in E : y < 1 - x\}$ and $E_2 = \{y \in E : 1 - x \leq y < 1\}$, so that $E \oplus x = (E_1 \oplus x) \cup (E_2 \oplus x)$, $E_1 \oplus x = E_1 + x$, and $E_2 \oplus x = E_1 + (x - 1)$. Furthermore, $(E_1 \oplus x) \cap (E_2 \oplus x) = \emptyset$, and therefore using the postulates we laid out,
$$\begin{aligned} m(E \oplus x) &= m((E_1 + x) \cup (E_2 + x - 1)) \\ &= m(E_1) + m(E_2) = m(E). \end{aligned}$$
$\boxtimes$

Define $x \in y$ if $x = y \oplus q$ for some $q \in \mathbb{Q}$; it's easy to prove this is an equivalence relation, and partitions $[0, 1)$ into equivalence classes. Then, use the axiom of choice to choose one element from each equivalence class (this doesn't work without the axiom of choice), and call the set of these elements P, and let $P_j = P \oplus q_j$ for $q_j \in \mathbb{Q}$; these P_j are all disjoint, because if $y \in P_j \cap P_k$, then $y = p_j \oplus q_j = p_k \oplus q_k$, so $p_j \sim p_k$, and therefore $j = k$ (since there's exactly one element from each equivalence class).

Thus, $[0, 1)$ is decomposed into a countable collection of P_j (indexed by $\mathbb{Q}$), which are all translations of each other! Thus,
$$\begin{aligned} m([0,1]) &= m\left(\bigcup_{j=1}^{\infty} P_j\right) \\ &= \sum_{j=1}^{\infty} m(P_j) \\ &= \sum_{j=1}^{\infty} m(P). \end{aligned}$$
Thus, the sum is either 0 or ∞, depending on whether P has zero or positive measure. But we want it to be equal to 1, which means our postulates are wrong. We'll throw out postulate (0); the other three come from physics, so it would be strange to throw them out.

That means we restrict the measure to good sets, which will be called *measurable*.

Definition. An *outer measure* $m^*(A)$ is defined on any set A as
$$m^*(A) = \inf\left\{\sum_{j=1}^{\infty} |I_j| \mid A \subset \bigcup_{j=1}^{\infty} I_j\right\},$$
such that each I_j is an open[3] interval. It is allowed that $m^*(A) = \infty$.

It is thus pretty clear that $m^*(\emptyset) = 0$ and if $A \subseteq B$, then $m^*(A) \leq m^*(B)$.

Exercise 1. What changes if we only allow finite covers, rather than countable covers?

Terry Tao devotes a few pages to this idea.

Proposition 1.1. *If I is an interval (open or closed), then $m^*(I) = |I|$.*

[3]It doesn't matter whether you choose open or closed intervals, since one can add any ε of length and go between the two. It comes down to personal preference.

Proof. Let $I = [a, b]$, so that $I \subset (a - \varepsilon, b + \varepsilon)$ for any $\varepsilon > 0$. Thus, $m^*(I) \leq b - a + 2\varepsilon$, so $m^*([a, b]) \leq b - a$.

In the other direction, take I_j open such that $[a, b] \subseteq \bigcup I_j$; then, use the Heine-Borel lemma to choose $I_1, \ldots, I_N$ such that $[a, b] \subseteq \bigcup_{j=1}^{N} I_j$, and therefore (which is a bit of a laborious exercise) $\sum_{j=1}^{n} |I_j| \geq b - a$.

Thus, $b - a \leq m^*([a, b]) \leq b - a$, so $m^*([b - a]) = b - a$.

For the open interval, it's clear that $m^*((a, b)) \leq m^*([a, b]) = b - a$, but also that $m^*((a, b)) \geq m^*([a + \varepsilon, b - \varepsilon]) = b - a - 2\varepsilon$ for any $\varepsilon > 0$, so $m^*(a, b) = b - a$ as well. ⊠

Proposition 1.2 (Countable sub-additivity)**.**

$$m^*\left(\bigcup_{j=1}^{\infty} E_j\right) \leq \sum_{j=1}^{\infty} m^*(E_j).$$

Proof. Let $A = \bigcup_{j=1}^{\infty} E_j$ and take I_{jk} to be such that

$$E_j \subseteq \bigcup_{k=1}^{\infty} I_{jk}$$

and

$$m^*(E_j) \leq \sum_{j=1}^{\infty} |I_{jk}| + \frac{\varepsilon}{2^j}.$$

Since $A \subset \bigcup_{j,k} I_{jk}$, then

$$\begin{aligned}
m^*(A) &\leq \sum_{j,k=1}^{\infty} |I_{jk}| \\
&= \sum_{j=1}^{\infty} \left(\sum_{k=1}^{\infty} |I_{jk}| \right) \\
&\leq \sum_{j=1}^{\infty} \left(m^*(E_j) + \frac{\varepsilon}{2^j} \right) \\
&= \varepsilon + \sum_{j=1}^{\infty} m^*(E_j).
\end{aligned}$$

⊠

Corollary 1.3. *The outer measure of a countable set is zero.*

Define the *distance* between two sets E and F to be

$$\operatorname{dist}(E, F) = \inf\{|x - y| : x \in E, y \in F\}.$$

Proposition 1.4. *Assume* $\operatorname{dist}(E, F) > 0$*; then,* $m^*(E \cup F) = m^*(E) + m^*(F)$.

Proof. We already know that $m^*(E \cup F) \leq m^*(E) + m^*(F)$, so we need to check that $m^*(E) + m^*(F) \leq m^*(E \cup F)$.

Choose I_j such that

$$m^*(E \cup F) \geq \sum_{j=1}^{\infty} |I_j| - \varepsilon \qquad \text{and} \qquad E \cup F \subseteq \bigcup_{j=1}^{\infty} I_j.$$

Exercise 2. Show that we can take

$$|I_j| \leq \frac{\operatorname{dist}(E, F)}{24}.$$

Then, each I_j intersects one of E or F; call those that intersect E the I'_j, and those that intersect F as I''_j. Thus, E is covered by the I'_j, and F by the I''_j. Now

$$\begin{aligned}
m^*(E) &\leq \sum |I'_j| \\
m^*(F) &\leq \sum |I''_j| \\
m^*(E) + m^*(F) &\leq \sum (|I'_j| + |I''_j|) \\
&\leq m^*(E \cup F) + \varepsilon.
\end{aligned}$$

⊠

Note that if the assumption that $\text{dist}(E,F) > 0$ is relaxed, the proposition is not true; for example, consider the sets P and $P + 1/2$ that we constructed earlier today.

We want to generalize this to higher dimensions, so we'll consider *open boxes*[4] $B = \{a_1 < x_1 < a_1', a_2 < x_2 < a_2', \dots, a_n < x < a_n''\}$. Then, we define the outer measure of a set $E \subseteq \mathbb{R}^n$ to be

$$m^*(E) = \inf \left\{ \sum_{j=1}^{\infty} |B_j| : E \subset \bigcup_{j=1}^{\infty} B_j \right\}.$$

Exercise 3. The results we saw in the one-dimensional case pass almost identically to the case of $\mathbb{R}^n$; redo these arguments for this case.

Exercise 4. If $B_1, \dots . B_N$ is a finite collection of disjoint boxes, then show that

$$m^* \left(\bigcup_{j=1}^{N} B_j \right) = \sum_{j=1}^{N} |B_j|.$$

This generalizes trivially to countable collections.

Proposition 1.5. *Let $\{B_j\}$ be a countable collection of disjoint boxes; then,*

$$m^* \left(\bigcup_{j=1}^{\infty} B_j \right) = \sum_{j=1}^{\infty} |B_j|.$$

Proof. Let $E = \bigcup_{j=1}^{\infty} B_j$, so that $m^*(E) \le \sum_{j=1}^{\infty} |B_j|$. But for any N, $E \supseteq \bigcup_{j=1}^{N} B_j$, and so $\sum_{j=1}^{N} |B_j| \le m^*(E)$. Thus, take $n \to \infty$, and equality holds. ⊠

Now, we want to extend this to more general open sets.

Exercise 5. Any open set in $\mathbb{R}$ is an at most countable collection of disjoint open intervals.

The idea is to try to find the maximal interval(s), and then induct.

The generalization in higher dimensions is much more useful. Consider the lattice $\mathbb{Z}^n$, and scale it by 2^k, for any $k \in \mathbb{Z}$; call this lattice $\mathcal{Q}_k$. Thus, $|Q| = 1/2^{nk}$ for $Q \in \mathcal{Q}_k$. These cubes are known as *closed dyadic cubes*, and are useful in many places in measure theory, as well as image processing and electrical engineering, because each cube is contained in a cube of the next level, and doesn't intersect any of the cubes on the previous level.

Take a nonempty open bounded[5] $U \subseteq \mathbb{R}^n$; any point in U is covered by a closed dyadic cube contained in U; these cubes will not be disjoint. Let C be the collection C of all closed dyadic cubes which are contained in U and C_M be the collection of maximal dyadic cubes in C, those not contained in any other dyadic cube in C. These cubes' interiors cannot intersect, since they're both maximal, so neither can contain the other, and any point lives in a maximal cube. Thus U is the union of the dyadic cubes in C_M, which have non-overlapping interiors.

Terry Tao gives a nice, intuitive definition for Lesbegue measure, which is not the standard definition. We'll be given both, and then later show that they're equivalence.

Proposition 1.6. *Given any $A \subseteq \mathbb{R}^n$ and any $\varepsilon > 0$, there exists an open set U such that $A \subseteq U$ and $m^*(A) \ge m^*(U) - \varepsilon$.*

This follows directly from the definition (give the right covering).

Proposition 1.7. $m^*(A) = \inf\{m^*(U) : A \subseteq U$ *and U is open*$\}$.

Philosophically, this means that U approximates A well; it's contained in U, and the measures are very similar. But this is only true in some cases; there are some sets A and opens $U \supset A$ such that the measure of U is very close to that of A, but the measure of $U \setminus A$ is not small. This is counterintuitive.

However, we can just take the measureable sets to be those that are well approximated by open sets; this is exactly what Terry Tao does.

Definition. A set A is *Lesbegue measurable* if for all $\varepsilon > 0$ there exists an open set $U \supseteq A$ such that $m^*(U \setminus A) < \varepsilon$.

[4]Once again, it doesn't matter whether they're closed or open.

[5]To generalize to the unbounded case, consider only dyadic cubes up to a certain level, and then the same argument works. We're worrying about the small cubes, not the large ones.

"A lot of this is improvisation which is not necessarily correct."

Suppose $A \subseteq \mathbb{R}^n$; then, we defined the *outer measure* of A as

$$m^*(A) = \inf \sum_{j=1}^{\infty} |\mathcal{D}_j|,$$

where the infimum is over all countable collections $\{\mathcal{D}_j\}$ of open boxes that cover A; we then proved that:

(1) if E and F are such that $\operatorname{dist}(E, F) > 0$, then $m^*(E \cup F) = m^*(E) + m^*(F)$,
(2) there is no countable collection of disjoint sets E_i such that

$$m^*\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} m^*(E_j),$$

and
(3) any open set in $\mathbb{R}^n$ is an at most countable union of almost disjoint (i.e. except on the boundary) closed boxes.

Finally, we defined a set E to be *Lesbegue measurable* if for any $\varepsilon > 0$ there is an open set $U \supset E$ such that $m^*(U \setminus E) < \varepsilon$.

Fact. Any open set is measurable.

This is pretty much by definition. We also have another fact:

Observation 2.1. *Any set of measure zero is measurable.*

Proof. Suppose $m^*(E) = 0$, and for any $\varepsilon > 0$ let $\{\mathcal{D}_j\}$ be a countable collections of boxes covering E and such that $\sum |\mathcal{D}_j| < \varepsilon$. Then, let $\mathcal{U} = \bigcup_{j=1}^{\infty} \mathcal{D}_j$; then $m^*(\mathcal{U} \setminus E) \le m^*(\mathcal{U}) < \varepsilon$. ⊠

Observation 2.2. *A countable union of measurable sets is measurable.*

Proof. Let $E_1, E_2, \dots$ be measurable and $E = \bigcup_{j=1}^{\infty} E_j$. Let $\mathcal{U}_j$ be open sets such that $\mathcal{U}_j \supseteq E_j$ and $m^*(\mathcal{U}_j \setminus E_j) < \varepsilon/2^j$; then, setting

$$\mathcal{U} = \bigcup_{j=1}^{\infty} \mathcal{U}_j \supset E,$$

one can sum the individual inqualities and check the definition. ⊠

Observation 2.3. *Every closed set is measurable.*

Proof. We can assume without loss of generality that E is a closed, bounded set (writing it as the union of disjoint, measurable sets and using countable additivity). Then, take a bounded open set U such that $m^*(U) < m^*(E) + \varepsilon$. Then, $U \setminus E$ is an open set, so there's a countable union of disjoint closed boxes Q_j such that $\operatorname{dist}(Q_j, E) > 0$ (since E is closed) and such that

$$\begin{aligned} m^*\left(E \cup \bigcup_{j=1}^{m} Q_j\right) &= m^*\left(\bigcup_{j=1}^{m} Q_j\right) + m^*(E) \\ &\ge m^*\left(\bigcup_{j=1}^{m} Q_j\right) + m^*(U) - \varepsilon, \end{aligned}$$

so

$$m^*(U) < \sum_{j=1}^{\infty} |Q_j| < \varepsilon.$$

Notice that boundedness is necessary because $m^*(U)$ cannot be infinite. ⊠

Observation 2.4. *If E is measurable, then so is its complement $E^c = \mathbb{R}^n \setminus E$.*

Proof. For each $n \in \mathbb{N}$, let U_n be an open set containing E such that $m^*(U_n \setminus E) < 1/n$. Then, let $F_n = U_n^c$, which is closed. Then, $F_n \subseteq E_n^c$ and $E^c \setminus F_n = U_n \setminus E$, so $m^*(E^c \setminus F_n) < 1/n$.

Write

$$E^c = \left(\bigcup_{n=1}^{\infty} F_n\right) \cup S,$$

so that S is everything else, so to speak. Then, $S \subseteq E^c \setminus F_n$ for each n, so $m^*(S) < m^*(E^c \setminus F_n) = 1/n$ for each n, and therefore $m^*(S) = 0$. Thus, E^c is a union of countably many measurable sets (S and the F_n), so it is measurable. ⊠

Observation 2.5. *A countable intersection of measurable sets is measurable.*

Proof. A countable intersection is the complement of a countable union of complements of measurable sets, and each of these operations preserves measurability, so countable intersections must as well. ⊠

Now we can speak a little more generally.

Definition. A collection of sets $\mathcal{F}$ is an *algebra* if:

(1) $\emptyset \in \mathcal{F}$.
(2) Whenever $A_1, A_2 \in \mathcal{F}$, then $A_1 \cup A_2 \in \mathcal{F}$.
(3) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

If in addition we have

(4) If $A_1, A_2, \dots \in \mathcal{F}$, then
$$\bigcup_{j=1}^{\infty} A_j \in \mathcal{F},$$

then $\mathcal{F}$ is also called a *σ-algebra*.

Thus, the observations above establish the following theorem.

Theorem 2.6. *The collection of all Lesbegue measurable sets in $\mathbb{R}^n$ forms a σ-algebra.*

Definition. The *Borel σ-algebra $\mathcal{B}$* is the smallest σ-algebra that contains all open sets. A set is called *Borel* if it is in $\mathcal{B}$.

Exercise 6. Show that not every Lesbegue-measurable set is Borel.

This is a hard exercise which will eventually be easier.

Though we have seen Tao's definition of measurability, it isn't the standard one. We will present this definition as well, and prove the two are equivalent.

Definition. A set E is *Carathéodory measurable* or *C-measurable* if for every set $A \subseteq \mathbb{R}^n$, $m^*(A) = m^*(A \cap E) + m^*(A \cap E^c)$.

Now, Observation 2.4 becomes a triviality in the case of C-measurability.

Observation 2.7. *It is always true that $m^*(A) \le m^*(A \cap E) + m^*(A \cap E^c)$, so to show that A is measurable it is sufficient to show that $m^*(A) \ge m^*(A \cap E) + m^*(A \cap E^c)$.*

Observation 2.8. *If $m^*(E) = 0$, then E is C-measurable.*

Proof. Since $m^*(A \cap E) = 0$, since $A \cap E \subseteq A$ for all A, and $m^*(A \cap E^c) \le m^*(A)$, then $m^*(A) \ge m^*(A \cap E^c) + m^*(A \cap E)$. ⊠

Observation 2.9. *Any* open corner
$$E = \{x_1 > a_1, x_2 > a_2, \dots, x_n > a_n\}$$
is C-measurable.

Proof. Take any $A \subseteq \mathbb{R}^n$ and cover it with boxes $\mathcal{D}_n$ such that
$$m^*(A) + \varepsilon \ge \sum_{n=1}^{\infty} |\mathcal{D}_n|,$$
and let
$$E_n = \left\{x_1 \ge a_1 + \frac{\varepsilon}{2^n}, x_2 \ge a_2 + \frac{\varepsilon}{2^n}, \dots, x_n \ge a_n + \frac{\varepsilon}{2^n}\right\},$$
which is a slightly smaller corner. We also define $\mathcal{D}_j' = \mathcal{D}_j \cap E$ and $\mathcal{D}_j'' = D_j \cap E_n^c$, which is a finite union of boxes. Finally, define
$$A_1 = A \cap E \subseteq \bigcup_{j=1}^{\infty} \mathcal{D}_j', \text{ and}$$
$$A_2 = A \cap E^c \subseteq \bigcup_{j=1}^{\infty} \mathcal{D}_j''.$$

Thus, $m^*(A_1) \leq \sum|\mathcal{D}'_j|$ and $m^*(A_2) \leq \sum|\mathcal{D}''_j|$, so

$$m^*(A_1) + m^*(A_2) \leq \sum_{j=1}^{\infty}\big(|\mathcal{D}'_j| + |\mathcal{D}''_j|\big).$$

We'd like to be done here, but there's some extra that is counted twice. We can account for it in the inequality:

$$\begin{aligned} &\leq \sum_{j=1}^{\infty}\left(|\mathcal{D}_j| + \frac{C\varepsilon}{2^j}\right) \\ &\leq m^*(A) + \varepsilon + C\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, then $m^*(A \cap E) + m^*(A \cap E^c) \leq m^*(A)$. ⊠

Observation 2.10. *If E_1 and E_2 are C-measurable, then so is $E_1 \cup E_2$.*

Proof.

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c) = m^*(A \cap E_1) + m^*((A \cap E_1^c) \cap E_2) + m^*(A \cap E_1^c \cap E_2^c),$$

but $E_1^c \cap E_2^c = (E_1 \cup E_2)^c$, and thus $(A \cap E_1) \cup ((A \cap E_1^c) \cap E_2) = A \cap (E_1 \cup E_2)$, so plugging these back into the above inequality shows the union is measurable. ⊠

It seems like we're closely studying trivialities over and over in order to gain insights. The Hebrew word for this is *pilpul.*

Observation 2.11. *If E_1 and E_2 are C-measurable, then so is $E_1 \cap E_2$.*

This is because $(E_1 \cap E_2)^c = E_1^c \cup E_2^c$, and we know unions and complements are measurable.

Lemma 2.12. *Let A be any set and $E_1, \dots, E_n$ be pairwise disjoint, C-measurable sets. Then,*

$$m^*\left(A \cap \bigcup_{j=1}^{n} E_j\right) = \sum_{j=1}^{n} m^*(A \cap E_j).$$

Proof sketch. Prove by induction on n, which makes it pretty straightforward.

Theorem 2.13. *The collection of C-measurable sets is a σ-algebra.*

Proof. We've already done a lot of the proof, but we need to check that countable unions of C-measurable sets are C-measurable.

Let $E_1, E_2, \dots$ be a countable collection of C-measurable sets, and let $\widetilde{E}_1 = E_1$, $\widetilde{E}_2 = E_2 \setminus E_1$, and so on; in general,

$$\widetilde{E}_n = \left(\bigcup_{j=1}^{n} E_j\right) \setminus \left(\bigcup_{j=1}^{n-1} \widetilde{E}_j\right).$$

Then,

$$E = \bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} \widetilde{E}_n,$$

but the $\widetilde{E}_n$ are pairwise disjoint.

We know that $F_n = \bigcup_{j=1}^{n} \widetilde{E}_n$ is C-measurable, and

$$\begin{aligned} m^*(A) &= m^*(A \cap F_n) + m^*(A \cap F_n^c) \\ &\geq m^*(A \cap F_n) + m^*(A \cap E^c) \\ &= \sum_{j=1}^{n} m^*(A \cap \widetilde{E}_j) + m^*(A \cap E^c). \end{aligned}$$

Letting $n \to \infty$,

$$m^*(A) \geq \sum_{j=1}^{\infty} m^*(A \cap \widetilde{E}_j) + m^*(A \cap E^c)$$

and

$$\bigcup_{j=1}^{\infty}(A \cap \widetilde{E}_j) = A \cap \bigcup_{j=1}^{\infty} E_j = A \cap E,$$

so $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c)$. ⊠

Corollary 2.14.

(1) *All open boxes are C-measurable.*
(2) *All closed boxes are C-measurable (since they're complements).*
(3) *All open sets and closed sets are countable unions of open or closed boxes, and are therefore measurable.*
(4) *All Borel sets are C-measurable, since they're generated by open and closed sets.*

Now we can show that the two notions of measurability are the same.

Proposition 2.15. *A set is Lesbegue measurable iff it is C-measurable.*

Proof. Let E be a Lesbegue-measurable set and A be any set. We want that $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c)$. Take $\mathcal{U} \supseteq E$ such that $m^*(\mathcal{U} \setminus E) < \varepsilon$; then,

$$m^*(A) = m^*(A \cap \mathcal{U}) + m^*(A \cap \mathcal{U}^c)$$
$$\geq m^*(A \cap E) + m^*(A \cap \mathcal{U}^c).$$

Since $E^c = (\mathcal{U} \setminus E) \cup \mathcal{U}^c$, then $m^*(A \cap \mathcal{U}^c) \geq m^*(A \cap E^c) - m^*(\mathcal{U} \setminus E) \geq m^* A \cap E^c) - \varepsilon$.

Thus, $m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c) - \varepsilon$, so E is C-measurable.

Conversely, if E is C-measurable, then there is an open set $\mathcal{U}$ such that $\mathcal{U} \supseteq E$, and $m^*(\mathcal{U}) \leq m^*(E) + \varepsilon$ (difference of measure, not measure of differences), and $\mathcal{U} \setminus E = \mathcal{U} \cap E^c$, so since $\mathcal{U} \cap E^c$ and E are disjoint, then $m^*(\mathcal{U}) = m^*(\mathcal{U} \setminus E) + m^*(E)$, so $m^*(\mathcal{U} \setminus E) < \varepsilon$; thus, E is Lesbegue measurable. ⊠

Thus, we'll just use the term "Lesbegue measurable" to describe these two equivalent notions; sometimes, the Carathéodory notion is more convenient.

But we can speak yet more generally about measures.

Definition. A mapping $\mu^* : 2^X \to \mathbb{R}_+ \cup \{\infty\}$ is an *outer measure*[6] on X if:

(1) $\mu^*(\emptyset) = 0$ and
(2) whenever $A \subseteq \bigcup_{k=1}^{\infty} A_k$,

$$\mu^*(A) \leq \sum_{k=1}^{\infty} \mu^*(A_k).$$

If $\mu^*(X)$ is finite, then μ^* is also called *finite*.

Definition. If μ^* is an outer measure on X, then for any $A \subseteq X$, the outer measure restricted to A is $\mu^*|_A(B) = \mu^*(A \cap B)$.

Example 2.16.

(1) The Lesbegue measure is an outer measure, as we have shown.
(2) $\mu_{\#}(A) = \#A$ (the number of elements of A), the *counting measure*.
(3) The δ-*measure* on $\mathbb{R}^n$:

$$\mu(A) = \begin{cases} 1, & \text{if } 0 \in A \\ 0, & \text{if } 0 \notin A. \end{cases}$$

Definition. A set E is *measurable* if for any set $A \subseteq X$, $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c)$.

If E is measurable, we write $\mu(E) = \mu^*(E)$.

Theorem 2.17. *The collection of measurable sets forms a σ-algebra.*

The proof is exactly the same as we did above, since that didn't depend on the specifics of the Lesbegue measure; the words are all the same.

3. Borel and Radon Measures: 10/2/14

Last time, we defined an *outer measure* on a set X to be a function $f : 2^X \to \mathbb{R}_+ \cup \{\infty\}$ such that $\mu^*(\emptyset) = 0$ and for all countable collections $E_1, \dots \subseteq X$,

$$\mu^*\left(\bigcup_{j=1}^{\infty} E_j\right) \leq \sum_{j=1}^{\infty} \mu^*(E_j).$$

[6] This is not the best name for it; there's nothing outer about an outer measure. But it's used to connote that it's not truly a measure yet.

The idea is that we only defined the measure on some sets, called (very creatively) *measurable sets*; these were defined to be the sets $A \subseteq X$ such that for all $B \subseteq X$, $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$, and proved that the collection of measurable sets forms a σ-algebra (Theorem 2.17).

We'll continue on this path today, talking about relatively abstract measures.

Proposition 3.1 (Countable additivity)**.** *Let $E_1, E_2, \dots$ be a countable collection of disjoint, measurable sets; then,*

$$\mu\left(\sum_{j=1}^{\infty} E_j\right) = \sum_{j=}^{\infty} \mu(E_j).$$

Proof. By countable subadditivity,

$$\mu\left(\sum_{j=1}^{\infty} E_j\right) \leq \sum_{j=}^{\infty} \mu(E_j),$$

so we just need to go in the other direction. For any m,

$$\mu\left(\bigcup_{j=1}^{\infty} E_j\right) \geq \mu\left(\bigcup_{j=1}^{m} E_j\right) = \sum_{j=1}^{m} \mu(E_j),$$

since the measure is finitely additive; thus, we can let $m \to \infty$. ⊠

We can also talk about nested sequences of measurable sets.

Proposition 3.2. *Let $E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots$ be measurable and such that $\mu(E_1)$ is finite, then*

$$\lim_{j\to\infty} \mu(E_j) = \mu\left(\bigcap_{j=1}^{\infty} E_j\right).$$

Proof. Let $F_j = E_j \setminus E_{j+1}$; then, all of the F_j are disjoint, and $\bigcup_{j=1}^{\infty} F_j = E_1 \setminus E$ (where E is the intersection of all of the E_j), $\mu(E_1 \setminus E) = \mu(E_1) - \mu(E)$. Then,

$$\mu\left(\bigcup_{j=1}^{\infty} F_j\right) = \sum_{j=1}^{\infty} \mu(F_j) = \sum_{j=1}^{\infty} (\mu(E_j) - \mu(E_{j+1})).$$

This is a telescoping sequence, so it's easier to evaluate.

$$\begin{aligned} &= \lim_{n\to\infty} (\mu(E_1) - \mu(E_{n+1})) \\ &= \mu(E_1) - \lim_{n\to\infty} \mu(E_n). \end{aligned}$$

Thus, $\mu(E) = \lim_{n\to\infty} \mu(E_n)$. (Here, all of the subtractions work because we have real numbers, not infinities.) ⊠

Notice that the finite hypothesis is necessary: if $E_n = (n, \infty)$, then each has infinite measure on $\mathbb{R}^n$, but their intersection is empty.

There's a corresponding result for increasing sequences of sets.

Proposition 3.3. *Let $E_1 \subseteq E_2 \subseteq \cdots$ be an increasing sequence of measurable sets. Then,*

$$\mu\left(\bigcup_{j=1}^{\infty} E_j\right) = \lim_{j\to\infty} \mu(E_j).$$

Proof. We can assume the measures are finite, because if any set has infinite measure, then of course their union does.

$$\begin{aligned} \mu(E_{k+1}) &= \mu(E_1) + \sum_{j=1}^{k} (\mu(E_{j+1}) - \mu(E_j)) \\ &= \mu(E_1) + \sum_{j=1}^{k} \mu(E_{j+1} \setminus E_j. \end{aligned}$$

Thus, when we pass to the limit,

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

and since these are disjoint sets, then countable additivity can be used to pass to the union. ⊠

Note that these proofs are often simplifications of things that appear in print, and as a result could be wrong.

Borel and Radon measures. Borel and Radon measures are notion of measure that are almost like the Lesbegue measure, but not quite (e.g. Radon measures might not have translation-invariance).

Definition.

(1) A measure μ on $\mathbb{R}^n$ is *Borel* if every Borel set is measurable.
(2) A Borel measure is *Borel regular* if every set can be approximated by Borel sets: for every set A, there exists a Borel set B such that $A \subseteq B$ and $\mu^*(A) = \mu(B)$.
(3) A Borel measure is *Radon* if for every compact set K, the measure of K is finite.

The Lesbegue measure is a good example of all of these notions.

It's pretty easy to construct a counterexample to the notion of Borel measure; we'll have to carefully construct the Lesbegue integral later, but we can see that

$$\mu(A) = \int_A \frac{\mathrm{d}x}{x}$$

on $\mathbb{R}$ is not Radon (and the measure of any set containing 0 is infinite). Radon probability measures include those with densities, e.g. $\mu(A) = \int_A f(x)\,\mathrm{d}x$ for $f \geq 0$; then, f is the density function. The δ-measure (recall Example 2.16) is not like that, but it is Radon.

In chemistry (and medicine), the Radon transform is a process akin to the Fourier transform. You'd think it's about the element, but apparently not. Clearly Radon was a broad scholar.

We'll be able to show that every Radon measure is an outer and an inner measure, which means measurable sets can be well approximated by open and by closed sets.

Theorem 3.4. *Let μ be a Radon measure. Then,*

(1) for each set $A \subseteq \mathbb{R}^n$ we have

$$\mu^*(A) = \inf\{\mu(U) : A \subseteq U, U \text{ open}\},$$

(2) and for each μ-measurable set A,

$$\mu(A) = \sup\{\mu(K) : A \supseteq K, K \text{ compact}\}.$$

This says that the Radon measure, like the Lesbegue measure, is an outer measure and an inner measure.

Exercise 7. Show that (2) may fail for a non-measurable set.

Lemma 3.5. *Let B be a Borel set and μ a Borel measure.*

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

Proof. This proof is terrible, because the concrete statement is proved by abstract nonsense, but such is life.

For part (1), set $\nu = \mu|_B$.

Exercise 8. Let μ be a regular Borel measure and $\mu(A)$ be finite for some μ-measurable set A. Then $\nu = \mu|_A$ (i.e. $\nu(B) = \mu(A \cap B)$) is a Radon measure.

This exercise is Theorem 1.35 in the lecture notes, but isn't too difficult or interesting. All that needs to be checked (which is nontrivial) is that we don't lose Borel regularity, which makes sense: we gain finiteness, but don't lose the goodness of the measure.

Thus, ν in our proof is a finite Radon measure.

Claim. If ν is a finite Radon measure, then for any Borel set B' and any $\varepsilon > 0$, there exists a closed set $C \subseteq B'$ such that $\nu(B' \setminus C) < \varepsilon$.

Proof. When you want to prove something true for all Borel sets, show it for open sets and then use the fact that the things satisfying the claim form a σ-algebra. In this case, we need to start with closed sets, which is unusual.

Let $\mathcal{F}$ be the collection of measurable sets A such that, for all $\varepsilon > 0$, there exists a closed set C such that $\nu(A \setminus C) < \varepsilon$. Then:

(1) Clearly, $\mathcal{F}$ contains all closed sets.

(2) We'll show that $\mathcal{F}$ contains countable intersections. Let $A_1, A_2, \dots \in \mathcal{F}$, and for each j, choose a closed set $C_j \subseteq A_j$ such that $\nu(A_j \setminus C_j) < \varepsilon/2^j$, and let $C = \bigcap_{j=1}^\infty C_j$. Then,

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

(3) We want to do this for countable unions, but the same trick doesn't work: a countable union of closed sets is not always closed. Nonetheless, let $A_1, A_2, \dots \in \mathcal{F}$ and take $C_j \subseteq A_j$ to be closed, with $\nu(A_j \setminus C_j) < \varepsilon/2^j$.

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

Thus, there exists an N such that

$$\nu\left(A \setminus \bigcup_{j=1}^N C_j\right) < \varepsilon;$$

Let $C_\varepsilon = \bigcup_{j=1}^N C_j$, which is closed, and thus $\mathcal{F}$ has infinite unions.

(4) Let $\mathcal{G}$ be the collection of all sets A such that $A \in \mathcal{F}$ and $A^c \in \mathcal{F}$ (so $\mathcal{G} \subseteq \mathcal{F}$); then, we'll show that G contains all open sets.

If U is open, then $U^c \in \mathcal{F}$, and U is a countable union of closed boxes (as we showed a couple lectures ago), so $U \in \mathcal{F}$. Thus, $U \in \mathcal{G}$.[7]

(5) Next, we'll show that $\mathcal{G}$ is a σ-algebra.
   (a) Of course, complements come for free: if $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$.[8]
   (b) Countable unions are in $\mathcal{G}$: if $A_1, A_2, \dots \in \mathcal{G}$, then their union is in $\mathcal{F}$ by 3, and the complement is the intersection of $A_1^c, A_2^c, \dots$, which are all in $\mathcal{F}$ (since $A \in \mathcal{G}$), and thus their intersection is as well. Thus, the union is in $\mathcal{G}$.

Thus, $\mathcal{G}$ is a σ-algebra containing all open sets, so it contains all Borel sets. Thus, so does $\mathcal{F}$, so the one Borel set we're looking for has the right property. ⊠

Now, on to part 2: that a Borel set can be well approximated from the outside by an open set. Let B be a Borel set. It seems natural to pass to complements, and use things we've already proven in part 1, but this only works if $\mu(B^c)$ is finite. In this case, though, we can pick a $C \subseteq B^c$ such that $\mu(B^c \setminus C) < \varepsilon$, and let $U = C^c$, so that $\mu(U \setminus B) = \mu(B^c \setminus C) < \varepsilon$.

Thus, let's restrict to balls: for every $m \in \mathbb{N}$, let $U_m = U(0, m)$ be the ball of radius m around the origin. Then, $\mu(U_m \setminus B)$ is finite, so there is a closed set $C_m \subseteq U_m \setminus B$ and such that $\mu((U_m \setminus B) \setminus C_m) < \varepsilon/2^m$. Then, let

$$\begin{aligned}
U &= \bigcup_{m=1}^\infty (U_m \setminus C_m)\\
B &= \bigcup_{m=1}^\infty (U_m \cap B) \subseteq U.
\end{aligned}$$

Whenever we do this argument by restricting to both, the argument is the same, so do it as an exercise, or see the lecture notes. ⊠

Now, we (finally!) have the lemma, so let's prove the theorem.

[7] This might seem like a bit of silly extra detail, but it's nice to make sure nobody's missing it.

[8] I saw a sign in White Plaza once that said, "Free Compliments!" — is this what they meant?

Proof of Theorem 3.4. First, for part (1), if $\mu^*(A)$ is infinite, then there is nothing to prove (just take $U = \mathbb{R}^n$); thus, without loss of generality, assume $\mu^*(A)$ is finite.

Since μ is Borel regular, then we can find a Borel set $B \supseteq A$ such that $\mu^*(A) = \mu(B)$, and by the definition of the outer measure,

$$\mu(B) = \inf\{\mu(U) : U \supseteq B \text{ is open}\}.$$

We have a similar construction for $\mu^*(A)$, but since more sets contain A than contain B, then the infimum for A is at most that for the infimum for B, which is what we wanted for the theorem.

For part (2), first assume that $\mu(A)$ is finite. Then, $\nu = \mu|_A$ is Radon, so apply part (1) to A^c, which has measure zero. Thus, there exists an open set U such that $\nu(U) < \varepsilon$ and $U \supseteq A^c$, i.e. $\mu(U \cap A) < \varepsilon$. Let $C = U^c$, so that $C \subseteq A$ and $\mu(A \setminus C) = \mu(U \cap A) < \varepsilon$. Thus, the theorem is true for closed sets, though we'll need to do some kind of diagonal argument (outlined in the lecture notes) to show that it also works for compact sets.

If instead A has infinite measure under μ, then look at the annuli

$$\mathcal{D}_k = \{x : k - 1 \leq |x| \leq k\},$$

so that $A = \bigcup_k (A \cap \mathcal{D}_k)$ and

$$\infty = \mu(A) = \sum_{k=1}^{\infty} \mu(A \cap \mathcal{D}_k).$$

Since μ is Radon, then $\mu(A \cap \mathcal{D}_k)$ is finite for each k. Then, choose $C_k \subseteq A \cap \mathcal{D}_k$ such that $\mu((A \cap \mathcal{D}_k) \setminus C_k) < \varepsilon$. If $G_n = \bigcup_{k=1}^n C_k$, then

$$\mu(G_n) = \sum_{k=1}^{n} \mu(C_k) \geq \sum_{k=1}^{n} \left(\mu(A \cap D_k) - \frac{1}{2^k}\right).$$

But the latter series diverges, since $\mu(A)$ is infinite, and thus $\mu(G_n) \to \infty$ as $n \to \infty$, and $A \supseteq G_n$ with each G_n closed and bounded (and therefore compact), so we're done. ⊠

4. Measurable Functions: 10/7/14

Recall Theorem 3.4, which demonstrates that Radon measures are those that can be well approximated on open sets containing a given $A \subseteq \mathbb{R}^n$ and by compact sets within A.

Today we'll talk about measurable functions. These are basically only ever used for integration, but everyone except for Terry Tao decided to talk about them separately for some reason.

Proposition 4.1. *The following are equivalent:*

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

Proof. Clearly, (1) $\iff$ (4) and (2) $\iff$ (3). Since

$$\{f(x) > \alpha\} = \bigcup_{m=1}^{\infty} \left\{f(x) \geq \alpha + \frac{1}{m}\right\},$$

then (2) $\implies$ (1). Similarly,

$$\{f(x) < \alpha\} = \bigcup_{m=1}^{\infty} \left\{f(x) < \alpha - \frac{1}{m}\right\},$$

so (4) $\implies$ (3). ⊠

Now, we can define measurable functions (those which will be integrable). In some sense, they generalize continuous functions, which can be helpful for intuition but isn't everything.

Definition. Let X be a space with a measure μ, and Y be a topological space. Then, $f : X \to Y$ is *measurable* if for any open set $U \subseteq Y$, its preimage $f^{-1}(U)$ is μ-measurable.

Notice that for a continuous f, the preimage of an open set is open (so for Borel measures, continuous functions are measurable).

Here's a bureaucratic proposition.

Proposition 4.2. *If f and g are real-valued measurable functions and $c \in \mathbb{R}$, then cf, $c + f$, f^2, $f + g$, and fg are all measurable.*

Proof. Without loss of generality assume $c > 0$; then, $\{cf > \alpha\} = \{f > \alpha/c\}$ and $\{c + f > \alpha\} = \{f > \alpha - c\}$, so we get sets with measurable preimage.

If $f + g < \alpha$, then $f < \alpha - g(x)$, so there's a $p \in \mathbb{Q}$ such that $f(x) < p < \alpha - g(x)$, so

$$\{f + g < \alpha\} = \bigcup_{p \in \mathbb{Q}} \{x : f(x) < p, g(X) < \alpha - p\},$$

so it's a countable union of measurable sets.

For f^2, we have

$$\left\{f^2 > \alpha\right\} = \left\{f > \sqrt{\alpha}\right\} \cup \left\{f < -\sqrt{\alpha}\right\}$$

if $\alpha > 0$ (and if not, it's trivial). Thus, this is once again a union of measurable sets.

For fg, it's possible to write this in terms of a sum of $(f + g)^2$ and $(f - g)^2$, so it follows from the previous results. ☒

Theorem 4.3. *Suppose $\{f_n\}$ is a sequence of measurable functions. Then, so are the following:*

- $g_n(x) = \sup_{1 \le j \le n} f_j(x)$, *and* $g(x) = \sup_j f_j(x)$.
- $q_n(x) = \inf_{1 \le j \le n} f_j(x)$, *and* $q(x) = \inf_j f_j(x)$.
- $s(x) = \limsup_{n \to \infty} f_n(x)$ *and* $w(x) = \liminf_{n \to \infty} f_n(x)$.

Notice this is completely false for continuous functions; the limit is not expected to be continuous! For example, continuous functions can converge to step functions; there are many examples.

Proof. Once again, we'll write the sets that should be measurable in terms of sets that are already measurable. Specificallly,

$$\{g_n(x) > \alpha\} = \bigcup_{j=1}^{b} \{f_j(x) > \alpha\}$$

$$\{g(x) > \alpha\} = \bigcup_{j=1}^{\infty} \{f_j(x) > \alpha\}.$$

Then, q_n and q are very similar, and for s (and w, which is similar), we have

$$\limsup_{n \to \infty} f_n(x) = \inf_n (\sup_{k \ge n} f_k(x)),$$

so it follows from the first part of the proof. ☒

Definition. A measurable function $f(x)$ is *simple* if it takes at most countably many values.

Any simple function can be expressed as the form

$$f(x) = \sum_{k=1}^{\infty} \alpha_k \chi_{A_k}(x),$$

where

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, x \notin A, & \end{cases}$$

where the A_k are measurable and disjoint.

If we drop the requirement that the A_k are disjoint, then we obtain many, many more functions.

Theorem 4.4. *Any nonnegative measurable function f can be written*

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{k} \chi_{A_k}(x),$$

where the A_k are measurable, though not necessarily disjoint.

Proof. Begin by taking $A_1 = \{f(x) > 1/1\}$ (taking everything above the line $y = 1$), and so forth, so that in general

$$A_{j+1} = \left\{x : f(x) \ge \frac{1}{j+1} + \sum_{k=1}^{j} \frac{1}{k} \chi_{A_k}(x)\right\}.$$

Then, $f(x) = \sum_{k=1}^{\infty} \frac{1}{k}\chi_{A_k}(x)$, because if $f(x)$ is infinite, then $x \in A_k$ for all k, and the sum diverges, so we're good. If $f(x)$ is finite, then $x \notin A_k$ for infinitely many A_k (or the series would diverge, implying $f(x)$ does also). Thus, there are infinitely many j such that

$$\sum_{k=1}^{j} \frac{1}{k}\chi_{A_k}(x) \leq f(x) \leq \frac{1}{j+1} + \sum_{k=1}^{j} \frac{1}{k}\chi_{A_k}(x)$$
$$\implies \left| f(x) - \sum_{k=1}^{j} \frac{1}{k}\chi_{A_k}(x) \right| < \frac{1}{j}.$$

Since this is true for infinitely many j, then the two must be equal. ⊠

Of course, functions that have negative values can still be approximated with simple functions.

Lusin's theorem tells us that any measurable function is identical to a continuous function on a very large set (a set of full measure). This set may be very complicated, e.g. if $f(x) = \chi_{\mathbb{Q}}(x)$: this is equal to the continuous $f(x) = 0$ on a set of full measure, but that set isn't so well-behaved.

In analysis, there are some notions of extensions and restrictions. Since measurable functions are often defined only up to a set of measure zero, then restricting to a set of measure zero is pretty unhelpful. We also have a theorem for extension of continuous functions.

Theorem 4.5. *Let K be compact and $f : K \to \mathbb{R}^m$ be continuous; then, there is a function $\overline{f} : \mathbb{R}^n \to \mathbb{R}^m$ such that $\overline{f}$ is continuous on $\mathbb{R}^n$ and $\sup_{y\in\mathbb{R}^n}|\overline{f}(y)| = \sup_{x\in K}|f(x)|$.*

Proof. Without loss of generality, assume $m = 1$ (if not, it's just componentwise).

Let $U = \mathbb{R}^n \setminus K$; for every $x \in U$, we'd like to assign $\overline{f}(x)$ to be a weighted average of points near it in K. First, define

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

First, $0 \leq u_s(x) \leq 1$, and if s is fixed, then $u_s(x) \to 1$ as $|x| \to \infty$, since the weights look more or less the same. For a fixed x close to K, $u_s(x) = 0$ unless s is close to a point $s_x \in K$ such that $\operatorname{dist}(x, s_k) = \operatorname{dist}(x, K)$.

We would want to integrate this, but we don't have that yet, so take a dense subset $\{s_j\} \subset K$ and set

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

where $x \in U = \mathbb{R}^n \setminus K$. By the Weierstrauss test, $\sigma(x)$ can be bounded termwise by $1/2^j$, so it's continuous. Thus, for every $x \in U$, there exists an s_j such that $|x - s_j| \leq |0| \operatorname{dist}(x, K)$, so $u_{s_j}(x) > 0$. Thus, $\sigma(x) > 0$. This will be very useful; it means we can divide by it.

The weights are now pretty simple: let

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

Then, $\sum v_j(x) = 1$ everywhere, and the weight functions v_j are 1 far away from K.

Now, set

$$\overline{f}(x) = \begin{cases} f(x), & x \in K \\ \sum_{j=1}^{\infty} f(s_j) v_{s_j}(x), & x \notin K. \end{cases}$$

On U,

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

Since f is continuous on a compact set, it's bounded by some M, and $|f(s_j)| \leq M$, so by the Weierstrass test, $\overline{f}$ is continuous on U.

On K, let $\varepsilon > 0$ and choose a δ such that $|f(x) - f(x')| < \varepsilon$ if $|x - x'| < \delta$ and $x, x' \in K$. Fix a $y \in K$ and take x such that $|x - y| < \delta/4$. Without loss of generality, assume $x \in U$, and choose an s_j such that $|y - s_j| > \delta$, so that

$$\delta < |y - s_j| < |y - x| + |x - s_j|,$$

so $|x - s_j| > 3\delta/4$ and $|x - y| < \delta/4$, so $|x - s_j| > 3\operatorname{dist}(x, K)$, which means $u_{s_j}(x) = 0$. Thus,

$$\begin{aligned}|f(x) - f(y)| &= \left|\sum_{j=1}^{\infty}(f(s_j) - f(y))v_{s_j}(x)\right| \\ &\le \varepsilon \sum_{j=1}^{\infty} v_{s_j}(x) = \varepsilon.\end{aligned}$$

Thus, $\overline{f}$ is continuous on K as well. ⊠

This doesn't use much about $\mathbb{R}^n$, so it can be generalized a bit.

Exercise 9. If f has nicer properties, what happens to $\overline{f}$? Specifically, determine what happens when f is Lipschitz or Holder continuous.

Now let's state Lusin's theorem about extensions.[9]

Theorem 4.6 (Lusin)**.** *Let μ be a Borel regular measure on $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^m$ be μ-measurable. If $A \subseteq \mathbb{R}^n$ has a finite μ-measure, then for any > 0 there exists a compact $K_\varepsilon \subseteq A$ such that $\mu(A \setminus K_\varepsilon) < \varepsilon$ and f is continuous on K_ε.*

Proof. We'll construct a compact set K_ε and a sequence of continuous functions $g_n(x)$ such that $g_n(x) \to f(x)$ uniformly on K_ε. Let's start with the mesh $B_{pj} = [j/2^p, (j+1)/2^p)$ for $j \in \mathbb{N}$, and let $A_{pj} = f^{-1}(U_{pj})$.

Since μ is Borel regular and $\mu(A)$ is finite, then there exist compact $K_{pj} \subseteq A_{pj}$ such that $\mu(A_{pj} \setminus K_{pj}) < \varepsilon/2^{p+j}$. This is a sort of coarse splitting of A into subsets, with p controlling the size of the refinement.

Thus,

$$\mu\left(A \setminus \bigcup_{j=1}^{\infty} K_{pj}\right) < \frac{\varepsilon}{2^p}.$$

This can't be our K, since the infinite union of compact sets may not be compact, so we can cut it off: there must be an $N(p)$ such that $\mu(A \setminus \mathcal{D}_p) < \varepsilon/2^p$, where $\mathcal{D}_p = \bigcup_{j=1}^{N(p)} K_{pj}$.

All of the K_{pj} are at a finite distance from each other, so define $g_p(x) = j/2^p$ for $x \in K_{pj}$. This is continuous on $\mathcal{D}_p$ (since it's constant on each connected component), and $|f(x) - g_p(x)| < 1/2^p$. Finally, take $K_\varepsilon = \bigcap_{p=1}^{\infty} \mathcal{D}_p$, so that

$$\mu(A \setminus K_\varepsilon) < \sum_{k=1}^{\infty} \mu(A \setminus \mathcal{D}_k) < \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon.$$

Furthermore, K_ε is compact, and $f(x) = \lim_{p\to\infty} g_p(x)$ on K_ε; since each g_p is continuous and the limit is uniform, then f is also continuous. ⊠

Corollary 4.7. *Let f and A be as above; then, there exists a continuous function $\overline{f} : \mathbb{R}^n \to \mathbb{R}^m$ such that $\mu(\{x \in A : \overline{f}(x) \ne f(x)\}) < \varepsilon$.*

This is proven by combining the previous two theorems.

We'll eventually want to discuss the convergence of integrals, so let's formulate a convergence theorem. The man behind this theorem, Dmitri Egorov, was Lusin's advisor.

Theorem 4.8 (Egorov)**.** *Let μ be a measure and $f : \mathbb{R}^n \to \mathbb{R}$ be μ-measurable. Let A be a set with finite μ-measure, and suppose $f_k \to g$ almost everywhere on A. Then, for any $\varepsilon > 0$, there exists a measurable set $B - \varepsilon$ such that $\mu(A \setminus B_\varepsilon) < \varepsilon$ and $f_k \to g$ uniformly on B_ε.*

Proof. We'll construct some "not good-by-j sets:"

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

Then, $C_{i,j+1} \subseteq C_{i,j}$ and $\bigcap_{j=1}^{\infty} C_{i,j} = \emptyset$. Thus, there exists an N_i such that $\mu(C_i, N_i) < \varepsilon/2^i$. If $x \notin C_{i,N_i}$, then $|f_k(x) - g(x)| < 1/2^i$ for all $k \ge N_i$.

[9] Lusin had a very interesting personal history. Recall that in the 1930s in the Soviet Union, there were purges, in which people had to confess to crimes they probably didn't do. Lusin was forced to condemn publicly at Moscow State University, for apparently producing bad papers in Russian and better papers abroad , in order to destroy Russian mathematics. Everyone was forced to condemn him, but he was never arrested… and none of the people who condemned him ever made it to the academy before he died (since he was a member). There were probably religious factors in his denouncement, since he was very religious.

Set
$$B_\varepsilon = A \setminus \bigcup_{i=1}^{\infty} C_{i,N_i}.$$
Then, $\mu(B_\varepsilon) \leq \sum_i \varepsilon/2^i = \varepsilon$ and $|f_k(x) - g(x)| < 1/2^i$ for all $k > N_i$ and $x \in B_\varepsilon$; thus, $f_k \to g$ uniformly on B_ε. ⊠

Being able to throw away a set of arbitrary small measure to achieve uniform convergence will be very useful when we discuss difference of integrals.

Convergence in probability was supposed to be discussed today, but we ran out of time.

5. Convergence in Probability: 10/9/14

"The idea [of the homeworks] is not to kill you. Just have fun with it."

Definition. A sequence $f_n(x)$ *converges in probability* to $f(x)$ if for all $\varepsilon > 0$ there exists an N such that for any $n \geq N$ we have $\mu(\{x : |f_n(x) - f(x)| > \varepsilon\} < \varepsilon$.

This does not imply convergence: let f_n be the step function on $[0, 1)$ seen as the circle, with starting point $\sum_{j=1}^{n-1} 1/j \bmod 1$ and width $1/n$. Then, the support of each f_n is $1/n$ and shifts by $1/n$, so f_n converges to probability in convergence, but at no point converges to zero.

Proposition 5.1. *Let $f_n \to f$ in probability on a set E. Then, there exists a subsequence $f_{n_k} \to f$ almost everywhere on E.*

Proof. Choose N such that
$$\mu\left\{x : |f(x) - f_n(x)| > \frac{1}{2^j}\right\} < \frac{1}{2^j},$$
and look at $f_{N_j}(x)$: let $E_j = \{x : |f_{N_j}(x) - f(x)| > 1/2^j\}$, so if $\mathcal{D}_k = \bigcup_{j=k}^{\infty} E_j$ and $x \notin \mathcal{D}_k$ for all k, then $f_{N_j}(x) \to f(x)$. However,
$$\mu(\mathcal{D}_k) \leq \sum_{j=k}^{\infty} \mu(E_j) \leq \frac{1}{2^k},$$
so as more and more k are considered, in the limit the set has measure zero. ⊠

To see why convergence in probability can be established computationally, we'll need to define the integral.

Let f be a simple function, i.e. f can be written
$$f(x) = \sum_{k=1}^{\infty} y_k \chi_{A_k}(x),$$
where the A_k are disjoint measurable sets. If f is a nonnegative simple function, we set
$$\int_E f \, \mathrm{d}\mu = \sum_{k=1}^{\infty} y_k \mu(A_k \cap E).$$
In general, if f is simple, write $f = f^+ - f^-$ for nonnegative f^+ and f^-, and define f to be *integrable* if one of $\int f^+ \, \mathrm{d}\mu$ and $\int f^- \, \mathrm{d}\mu$ is finite; in this case, we set
$$\int_E f \, \mathrm{d}\mu = \int_E f^+ \, \mathrm{d}\mu - \int_E f^- \, \mathrm{d}\mu.$$
We'll gradually extend this to more complicated functions; first, bounded functions on sets of bounded measure.

Definition. Let f be a bounded function defined on a set E of finite measure. Then, define the *upper integral* to be
$$\int^* f \, \mathrm{d}\mu = \inf_{\substack{f \leq \psi \\ \psi \text{ simple}}} \int \psi \, \mathrm{d}\mu$$
and the *lower integral to be*
$$\int_* f \, \mathrm{d}\mu = \sup_{\substack{f \geq \psi \\ \psi \text{ simple}}} \int \psi \, \mathrm{d}\mu.$$

Proposition 5.2. *The upper and lower integrals agree iff f is measurable.*

Proof. First, assume f is measurable, and set

$$B_{nk} = \left\{x : \frac{k-1}{n} \leq f(x) \leq \frac{k}{n}\right\}$$

and

$$\psi_n(x) = \sum_k \frac{k}{n}\chi_{B_{nk}}(x)$$

$$\varphi_n(x) = \sum_k \frac{k-1}{n}\chi_{B_{nk}}(x).$$

Thus, $\psi_n(x) \geq f(x) \geq \phi_n(x)$, and ψ_n and ϕ_n are simple. However,

$$\int_E \psi_n \,\mathrm{d}\mu - \int_E \varphi_n \,\mathrm{d}\mu = \sum_k \frac{1}{n}\mu(B_{nk}) \leq \frac{\mu(E)}{n},$$

so it goes to 0 as $n \to \infty$.

Conversely, suppose $\int^* f \,\mathrm{d}\mu = \int_* f \,\mathrm{d}\mu$. We'll squish f between measurable functions, and thus conclude that it is also measurable. Choose $\psi_n \geq f$ and $\varphi_n \geq f$ such that

$$\int \psi_n \,\mathrm{d}\mu \leq \int \varphi_n \,\mathrm{d}\mu + \frac{1}{n}.$$

Set $\psi^*(x) = \liminf_{n\to\infty} \psi_n(x)$ and $\varphi_*(x) = \limsup_{n\to\infty} \varphi_n(x)$, so that $\psi^*(x) \geq \varphi_*(x)$. Let $A = \{x : \psi^*(x) \geq \varphi_*(x)\}$, which means that if

$$A + k = \left\{x : \psi^*(x) > \varphi^*(x) + \frac{1}{k}\right\},$$

then $A = \bigcup_{k=1}^\infty A_k$.

For large enough n, $\psi_n(x) > \varphi_n(x) + 1/k$ on A_k, so

$$\int_E \psi_n - \int_E \varphi_n \geq \int_{A_k} (\psi_n - \varphi_n) > \frac{\mu(A_k)}{k}.$$

As $n \to \infty$, $\mu(A_k) \to 0$, so $\mu(A) = 0$. Thus, $\psi^*(x) = f(x) = \varphi^*(x)$ almost everywhere on E, and thus, since ψ^* is measurable, then so is f. ⊠

Now, we can extend this a bit more. There are three more definitions, which are equivalent on the intersection of sets where they're defined.

Definition 1. If f is a bounded measurable function defined on a set E of finite measure, then set

$$\int_E f \,\mathrm{d}\mu = \inf_{\substack{\psi \geq f \\ \psi \text{ simple}}} \int_E \psi \,\mathrm{d}\mu.$$

Now, we generalize to sets of possibly infinite measure.

Definition 2. If $f \geq 0$ is measurable, set

$$\int_E f \,\mathrm{d}\mu = \sup_{h \in \mathcal{H}} \int_E h \,\mathrm{d}\mu,$$

where $\mathcal{H}$ is the set of bounded measurable functions which vanish outside of a set of finite measure.

Now, f doesn't even have to be bounded.

Definition 3. A measurable function is integrable if $\int |f| \,\mathrm{d}\mu$ is finite.

The Markov and Chebyshev Inequalities. These inequalities are somewhat silly to prove, but very useful.

Theorem 5.3 (Markov inequality)**.** *Let $f \geq 0$ be measurable. Then, for any $\lambda > 0$,*

$$\mu(\{x : f(x) \geq \lambda\}) < \frac{1}{\lambda}\int f \,\mathrm{d}\mu.$$

Proof. $f(x) \geq \lambda\chi_{E_\lambda}(x)$, where $E_\lambda = \{x : f(x) \geq \lambda\}$. Then, integrate. ⊠

Theorem 5.4 (Chebyshev inequality)**.** *With the same conditions on f as in Theorem 5.3,*

$$\mu(\{x : f(x) \geq \lambda\}) \leq \frac{1}{\lambda^2}\int f^2 \,\mathrm{d}\mu.$$

The proof is essentially the same. In probability theory, Theorem 5.4 means that if P is a probability distribution, $P(|X| \geq \lambda) \leq (1/\lambda^2)E(X^2)$.

One quick application of this is the Law of Large Numbers.

Proposition 5.5 (Law of Large Numbers)**.** *Suppose X_n are independently and identically distributed, $E(x_k) = 0$, and $E(X_k^2)$ is finite. Then, if $S_N = (X_1 + \cdots + X_n)/N$, then $S_n \to 0$.*

The Strong Law of Large Numbers requires this to be true almost everywhere; the Weak Law merely requires convergence in probability.

Proof. We know that $E(S_N) = 0$ and

$$\begin{aligned} E(S_N^2) &= \frac{1}{N^2} E\left(\sum_{i,j=1}^{N} X_i X_j \right) \\ &= \frac{1}{N^{2p}} \sum_{i,j=1}^{N} E(X_i X_j) = \frac{1}{N^2} \sum_{i=1}^{N} E(X_i^2) \\ &= \frac{mN}{N^{2p}}, \end{aligned}$$

which goes to 0 as $N \to \infty$ so long as $p > 1/2$. For the strong law, a little more work is needed. ⊠

Similar proofs allow one to prove the Central Limit Theorem, and so forth.

Convergence Theorems. The goal of the convergence theorems is to understand, given a sequence $f_n \to f$, when $\int f_n \to \int f$. This does not hold true in general.

Example 5.6. Let $f_n = \chi_{[n,n+1]}$. Then, f_n converges to 0 everywhere, but $\int f_n = 1$ for all n. This fails, in some sense, because f_n "floats away to infinity" horizontally.

Example 5.7. Let $f_n = n\chi_{[-1/n,1/n]}$; then $f_n \to 0$ everywhere but 0, but $\int f_n = 2$ for all n. This leads to the Dirac delta "function," and fails because it floats vertically off to infinity.

Example 5.8. Consider $f_n = (1/n)\chi_{[-n,n]}$. This also floats to infinity horizontally, has $f_n \to 0$ everywhere, and $\int f_n = 2$ for all n.

The first reasonable idea is to prevent the sequence from floating away to infinity.

Theorem 5.9 (Bounded Convergence Theorem)**.** *Assume $f_n \to f$ almost everywhere on E, $|f_n| \leq M$, and $\mu(E)$ is finite. Then,*

$$\int_E f_n \, \mathrm{d}\mu \longrightarrow \int_E f \, \mathrm{d}\mu.$$

Proof. If $f_n \to f$ uniformly on E, then for all $\varepsilon > 0$ there's an N such that when $n < N$, then $|f_n - f| < \varepsilon$, and thus

$$\left| \int_E f_n \, \mathrm{d}\mu - \int_E f \, \mathrm{d}\mu \right| \leq \int_E |f_n - f| \, \mathrm{d}\mu < \varepsilon \mu(E).$$

By Egorov's Theorem, for all $\varepsilon > 0$ there's an $A_\varepsilon \subseteq E$ such that $f_n \rightrightarrows f$ (i.e. uniformly) on A_ε and $\mu(E \setminus A_\varepsilon) < \varepsilon$. Then, choose an N such that $|f_n - f| < \varepsilon$ on A_ε when $n \geq N$, and

$$\begin{aligned} \left| \int_E f_n \, \mathrm{d}\mu - \int_E f \, \mathrm{d}\mu \right| &\leq \int_{A_\varepsilon} |f_n - f| \, \mathrm{d}\mu + \int_{E \setminus A_\varepsilon} |f_n - f| \, \mathrm{d}\mu \\ &\leq \varepsilon \mu(A_\varepsilon) + 2M\mu(E \setminus A_\varepsilon) \\ &\leq (2M + 1)\varepsilon \mu(E), \end{aligned}$$

which goes to 0 as $\varepsilon \to 0$. ⊠

This is very pretty, but it's too good to be true: not everything is bounded this nicely. Fatou's lemma, however, is used daily by millions of mathematicians.

Theorem 5.10 (Fatou's lemma)**.** *Let $f_n \geq 0$ and $f_n \to f$ almost everywhere on E. Then,*

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

Proof. Let h, such that $0 \le h \le f$ be a bounded simple function which vanishes outside of a set of finite measure, and set $h_n(x) = \min(h(x), f_n(x))$, so that $h_n \to h$ and h is bounded and vanishes outside a fixed set of finite measure. Thus, Theorem 5.9 implies that

$$\begin{aligned}\int_E h \,\mathrm{d}\mu &= \lim_{n\to\infty} \int h_n \,\mathrm{d}\mu \\ &\le \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu.\end{aligned}$$

Take the supremum over h, which implies that

$$\int f \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu. \qquad \boxtimes.$$

The nonnegativity of f is used to construct h; if f isn't nonnegative, negative parts may cause it to lose mass, and the theorem isn't always true.

Theorem 5.11 (Monotone Convergence Theorem)**.** *Suppose $f_n(x)$ is an increasing sequence and $f_n \to f$ almost everywhere on E. Then,*

$$\lim_{n\to\infty} \int_E f_n \,\mathrm{d}\mu = \int_E f \,\mathrm{d}\mu.$$

Proof. Apply Fatou's lemma (Theorem 5.10) to the differences $f_{n+1} - f_n$, which are nonnegative because the f_n are an increasing sequence. $\boxtimes$

Corollary 5.12. *If $u_n(x) \ge 0$ for each n and $S(x) = \sum_{n=1}^\infty u_n(x)$, then*

$$\int_E S(x) \,\mathrm{d}\mu = \sum_{n=1}^\infty \int_E u_n(x) \,\mathrm{d}\mu.$$

The idea is that nonnegative functions can more or less be integrated pointwise. This will relate to an analogue of Fubini's theorem.

The Lesbegue Dominated Convergence Theorem is a more adult version of the Bounded Convergence Theorem: it provides a slightly more sophisticated bound that is used all the time. Basically, if we can bound f_n by a nice bump function, then it can't go completely wrong.

Theorem 5.13 (Lesbegue Dominated Convergence Theorem)**.** *Let f_n be a sequence of measurable functions defined on a measurable set E. Assume that $f_n(x) \to f(x)$ and that there exists a $g(x)$ such that $\int_E g(x) \,\mathrm{d}\mu$ is finite and $|f_n(x)| \le g(x)$ almost everywhere on E. Then,*

$$\int_E f_n(x) \,\mathrm{d}\mu \longrightarrow \int_E f \,\mathrm{d}\mu$$

as $n \to \infty$.

The most general statement has to do with uniform integrability, which will (of course) appear on the homework.

Proof. $g - f_n \ge 0$ almost everywhere on E, and $g - f_n \to g - f$, so by Fatou's lemma,

$$\int (g - f) \,\mathrm{d}\mu \le \liminf_{n\to\infty} \left(\int (g - f_n) \,\mathrm{d}\mu \right).$$

Since $|f| \le g$, then f is integrable, so

$$\int g \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu \le \int g \,\mathrm{d}\mu - \limsup_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$

Thus,

$$\limsup_{n\to\infty} \int f_n \,\mathrm{d}\mu \le \int f \,\mathrm{d}\mu.$$

That's one direction; the other is given by taking $g + f_n$, which is nonnegative almost everywhere on E, so by Fatou's lemma we have

$$\int (g + f) \,\mathrm{d}\mu \le \liminf_{n\to\infty} \int (g + f_n) \,\mathrm{d}\mu.$$

Thus, just as before, f is integrable and

$$\int g \,\mathrm{d}\mu + \int f \,\mathrm{d}\mu \le \int g \,\mathrm{d}\mu + \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu.$$

Thus,

$$\int f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

Thus, both sides hold, so

$$\int f \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu. \qquad \boxtimes$$

In the case where g is a constant function, one recovers Theorem 5.9.

It seems funny to have one lecture on Chebyshev's inequality, Fatou's lemma, and the Lesbegue Dominated Convergence Theorem, but we have still five minutes left, so let's talk about the absolute continuity of the integral.

The idea of this theorem is that δ-functions don't actually exist. These are functions defined physically as supported at 0 and with total integral 1.

Theorem 5.14 (Absolute Continuity of the Integral)**.** *Let $f \geq 0$ and assume $\int_E f \, \mathrm{d}\mu$ is finite. Then, for all $\varepsilon > 0$, there's a $\delta > 0$ such that if $\mu(A) < \delta$, then $\int_A f \, \mathrm{d}\mu < \varepsilon$.*

Proof. Assume not; then, there exists an A_k such that $\mu(A_k) < 1/2^k$ but $\int_{A_k} f \, \mathrm{d}\mu > \varepsilon_0$. Then, restrict f_n to A_n, via $g_n(x) = f(x)\chi_{A_n}(x)$. Then, $g_n(x) \to 0$ except for

$$x \in A = \bigcap_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right).$$

Then, for any n,

$$\mu(A) \leq \mu\left(\bigcup_{k=n}^{\infty} A_k \right) \leq \sum_{k=n}^{\infty} \frac{1}{2^k} = \frac{1}{2^n}.$$

Thus, $\mu(A) = 0$ and $g_n(x) \to 0$ almost everywhere on E. By Fatou's lemma,

$$\liminf_{n\to\infty} \int_E (f - g_n) \, \mathrm{d}\mu \geq \int_E f \, \mathrm{d}\mu,$$

so $\int_E (f - g_n) \, \mathrm{d}\mu \leq \int_E f - \varepsilon_0$. $\boxtimes$

Of course, this doesn't mean that δ-functions don't exist, just that they're not integrable.

Next time: differentiation theorems and, relatedly, the covering lemma.

6. Back to Calculus: The Newton-Leibniz Formula: 10/14/14

"Does anyone know an elementary proof of this? I mean — a more elementary proof?"

Today, we'll ask two calculus-like questions related to the Newton-Leibniz[10] formula. These are,

(1) Is it true that

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_a^b f(t) \, \mathrm{d}t = f(x)?$$

(2) Is it true that

$$F(b) - F(a) = \int_a^b F'(x) \, \mathrm{d}x?$$

This is the *Newton-Leibniz formula.*

If we go back to real calculus,[11] so $f \in C^1([a, b])$, then both of these are true. But to answer these questions in generality, we need to know what the derivative of a measurable function is.

[10] We had a hard time spelling "Leibniz" in class. Praise be to the Internet.

[11] Relevant song lyrics: "So I thought back to Calculus. Way back to Newton and to Leibniz, and to problems just like this." `https://www.youtube.com/watch?v=P9dpTTpjymE`

First, there are the upper and lower derivatives, which lead to four pieces of notation.

$$\begin{aligned}
\mathcal{D}^+ f(x) &= \limsup_{h\downarrow 0} \frac{f(x+h)-f(x)}{h} \\
\mathcal{D}_+ f(x) &= \liminf_{h\downarrow 0} \frac{f(x+h)-f(x)}{h} \\
\mathcal{D}^- f(x) &= \limsup_{h\downarrow 0} \frac{f(x)-f(x-h)}{h} \\
\mathcal{D}_- f(x) &= \liminf_{h\downarrow 0} \frac{f(x)-f(x-h)}{h}.
\end{aligned}$$

Definition. f is *differentiable* if these are equal: $\mathcal{D}^+ f(x) = \mathcal{D}_+ f(x) = \mathcal{D}^- f(x) = \mathcal{D}_- f(x)$.

Theorem 6.1. *Let $f \in L^1([a,b])$ and set $F(x) = \int_a^x f(t)\,\mathrm{d}t$. Then, $F'(x)$ exists and $F'(x) = f(x)$ almost everywhere.*

Definition. A function f is *absolutely continuous* on $[a,b]$ if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for any finite collection of intervals (x_i, y_i) such that $\sum_{i=1}^N |y_i - x_i| < \delta$, we have $\sum_{i=1}^N |f(x_i) - f(y_i)| < \varepsilon$.

A typical example of a continuous function that's not absolutely continuous is $\sin(1/x)$ on $(0,1)$; it's not even uniformly continuous, so it can't be absolutely continuous. Absolute continuity is stronger than uniform continuty; Lipschitz continuity is stronger still.

Theorem 6.2. *A function $F(x)$ has the form $F(x) = F(a) + \int_a^x f(t)\,\mathrm{d}t$ with $f \in L^1([a,b])$ iff $F(x)$ is absolutely continuous on $[a,b]$, and then $F'(x) = f(x)$ almost everywhere.*

Remark. If $f \in L^1([a,b])$, then $F(x) = \int_a^x f(t)\,\mathrm{d}t$ is absolutely continuous by what we proved last time, since

$$\sum_{j=1}^N |F(x_i) - F(y_i)| \le \int_{\bigcup [x_i, y_i]} |f(t)|\,\mathrm{d}t < \varepsilon$$

when $m\left(\bigcup_{j=1}^N [x_i, y_i]\right) < \delta$. Thus, one direction is pretty straightforward.

The proofs of differentiation theorems are all quite related, and use covering lemmas. Here's an example.

Theorem 6.3. *If f is continuous at x, then*

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(t)\,\mathrm{d}t = f(x).$$

This actually will generalize to $f \in L^1([a,b])$, but we'll address that later. Onto[12] the covering lemma.

Definition. A cover $\mathcal{J}$ of a set A by closed balls is *fine* if for every $x \in A$ and $\varepsilon > 0$ there exists a ball $B \in \mathcal{J}$ such that $\operatorname{diam}(B) < \varepsilon$ and $x \in B$.

Now, we'll cover Vitali's lemma.

Lemma 6.4 (Vitali)**.** *Let $E \subseteq \mathbb{R}$ with $m^*(E)$ finite, and let $\mathcal{J}$ be a fine covering of E. Then, for any $\varepsilon > 0$, there exists a finite subcollection of disjoint intervals $I_1, \dots . I_N \in \mathcal{J}$ such that*

$$m^*\left(E \setminus \bigcup_{j=1}^N I_j\right) < \varepsilon.$$

This generalizes pretty easily to $\mathbb{R}^n$; the proof is very similar. The idea is that in the Lesbegue measure, if onw increases the volume of a ball slightly, the measure increases slightly, and we cover everything. In other measures, it might not work, as the measure could blow up.

This is related to one of our homework problems, which requires rewriting a fine cover into two countable disjoint sets that collectively cover the required set in $\mathbb{R}$. As the dimension grows, the number of sets needed increases; in $\mathbb{R}^2$, it's 19 sets!

Proof of Lemma 6.4. Take another open set $\mathcal{U} \supseteq E$ such that $m^*(\mathcal{U})$ is finite. Then, without loss of generality, all intervals in $\mathcal{J}$ are contained in $\mathcal{U}$ (or just ignore them, since we only need to care about these intervals).

[12] No pun intended.

Choose I_1 as you please, and then inductively choose the rest: if we have already chosen $I_1, \dots, I_n$ disjoint, then set

$$k_n = \sup\left\{ |I| : I \in \mathcal{J}, I \cap \bigcup_{j=1}^{n} I_j = \emptyset \right\}.$$

Then, if $k_n = 0$, $E \subseteq \bigcup_{j=1}^n I_j$, so we're done. If $k_n > 0$, then choose I_{n+1} disjoint from $I_1, \dots, I_{k-1}$ such that $|I_{n+1}| \geq k_n/2$.

If we stop, then we win, so let's assume we continue forever. Then, since these are disjoint intervals contained in $\mathcal{U}$, then

$$\sum_{j=1}^{\infty} |I_j| \leq m(\mathcal{U}),$$

which is finite, so $|I_j| \to 0$. Given an $\varepsilon > 0$, choose an N such that

$$\sum_{i=N+1}^{\infty} |I_j| < \frac{\varepsilon}{5}.$$

Claim. Given an I_j, define $\widehat{I}_j$ to be the interval with the same center as I_j, but with five times the width. Then,

$$E \setminus \bigcup_{j=1}^{N} I_j \subseteq \bigcup_{j=N+1}^{\infty} \widehat{I}_j.$$

Proof. Choose an $x \notin \bigcup_{j=1}^N I_j$; then, there exists an I with $x \in I$ and

$$I \cap \bigcup_{j=1}^{N} I_j \neq \emptyset.$$

As long as I is disjoint from $I_1, \dots, I_n$, then $|I| \leq k_n$, because $|I| \leq 2|I_{n+1}|$, which goes to 0 as $n \to \infty$ (because the next I_{n+1} is chosen to reduce the amount of area not covered). Thus, there exists an $n_0 > N$ such that I intersects one of $I_1, \dots, I_{n_0}$. If we take the smallest such n_0, then I doesn't intersect $I_1, \dots, I_{n_0 - 1}$, and also $|I| \leq k_n$, $|I_{n_0}| \geq kn_0/2$, and $I \cap I_{n_0} \neq \emptyset$, so $|I_{n_0}| \geq |I|/2$ and they intersect.

If we have intersecting intervals with comparable length and they intersect, and we blow one of them up by a factor of 5, then we certainly consume the other one. Thus, $I \subseteq \widehat{I}_{n_0}$, so

$$E \setminus \bigcup_{j=1}^{N} I_j \subseteq \bigcup_{j=N+1}^{\infty} \widehat{I}_j. \qquad \boxtimes$$

This allows us to finish the overall proof:

$$m^*\left(E \setminus \bigcup_{j=1}^{N} I_j \right) \leq 5 \sum_{j=N+1}^{\infty} |I_j| < \varepsilon. \qquad \boxtimes$$

Corollary 6.5. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open and have finite measnre, and $\delta > 0$. Then, there exists a countable collection of* disjoint *closed balls $B_1, B_2, \dots \subseteq \mathcal{U}$ such that $\operatorname{diam}(B_j) < \delta$ for all j and*

$$m\left(\mathcal{U} \setminus \bigcup_{j=1}^{\infty} B_j \right) = 0.$$

Proof. Take a fine cover of $\mathcal{U}$ by balls of diameter less than δ. Then, choose from it any B_1, and inductively construct the rest: if $B_1, \dots, B_N$ already exist, then set $\mathcal{U}' = \mathcal{U} \setminus \bigcup_{j=1}^n B_j$, and repeat. $\boxtimes$

Now we can go back to differentiation. A lot of what we need to say can be discussed using monotone functions as a start.

Theorem 6.6. *Let f be increasing on $[a, b]$; then, $f'(x)$ exists almost everywhere (with respect to the Lesbegue measure) on $[a, b]$, and $f'(x)$ is measurable.*

Proof. Consider, for instance, the set $E = \{x : \mathcal{D}^+ f(x) > \mathcal{D}^- f(x)\}$. (The other examples will be very similar.) If $E_{rs} = \{x : \mathcal{D}^+ f(x) > r > s > \mathcal{D}^- f(x)\}$, then $E = \bigcup_{r,s \in \mathbb{Q}} E_{rs}$. We'll show that $m(E_{rs}) = 0$ for all $r, s \in \mathbb{Q}$.

Let $\ell = m^*(E_{rs})$ and given an $\varepsilon > 0$, enclose E in an open set $\mathcal{U}$ such that $m^*(\mathcal{U}) < \ell + \varepsilon$. Then, for each $x \in E_{rs}$, there exist $h_n \downarrow 0$ such that $f(x + h_n) - f(x_n) < sh_n$. Then, E is covered by such intervals, and in fact this is a fine covering. Thus, by Lemma 6.4, one can choose a finite subcollection $I_1, \dots, I_N$ such that if

$$A = \left(\bigcup_{j=1}^{N} I_j^0\right) \cap E,$$

then $\ell - \varepsilon < m^*(A) < \ell + \varepsilon$, and also

$$\sum_{j=1}^{N} (f(x_j + h_j) - f(x_j)) < s(\ell + \varepsilon).$$

For each $x \in A$, choose h_n such that $(x, x + h_n)$ is inside one of the I_k. In this case, $f(x + h_n) - f(x) > rh_n$. Now, choose a finite subcollection of those intervals I_p' (there will be N' of them); each I_p' is contained in one of the I_k and $f(x_p + h_p) - f(x_p) > rh_p$ and

$$m^*\left(A \setminus \bigcup_{j=1}^{N'} I_j'\right) < \varepsilon.$$

Thus, the total drop over all I_j' is greater than $r(\ell - 2\varepsilon)$, and the total drop over I_j is less than $s(\ell + \varepsilon)$, but since ε is arbitrary, this can force $r \le s$, which is a contradiction unless $\ell = 0$.

Thus, $\mathcal{D}^+ f(x) = \mathcal{D}^- f(x)$ almost everywhere, and the other $\binom{4}{2} - 1$ cases are pretty much identical. Thus, $f'(x)$ exists almost everywhere.

The second part of the claim is that f' is measurable. If

$$q_n(x) = \frac{f(x + 1/n) - f(x)}{1/n},$$

where $f(x) = f(b)$ for some $x > b$, then $q_n(x) \to f'(x)$ almost everywhere as $n \to \infty$, so $f'(x)$ is measurable. ⊠

Now, we want to determine if the Newton-Leibniz formula holds. In mathematics, it seems often that if a formula is true on a nice class of functions and both sides make sense on a larger class of functions, then it generalizes nicely.

A good counterexample to the Newton-Leibniz formula for general monotone functions, however, is a step function with a single jump; the Cantor function is a more interesting counterexample.

Theorem 6.7 (Newton-Leibniz formula for monotone functions)**.** *If $F(x)$ is increasing on $[a, b]$, then $F'(x)$ is finite almost everywhere, and*

$$F(b) - F(a) \ge \int_a^b F'(x)\,\mathrm{d}x.$$

Proof. We already know $F'(x)$ is measurable by Theorem 6.6, so set

$$g_n(x) = \frac{F(x + 1/n) - F(x)}{1/n},$$

where $F(x) = F(b)$ for some $x > b$ (or this is an easier special case). Then, $g_n(x) \ge 0$ and $g_n(x) \to F'(x)$ almost everywhere, so by Fatou's lemma,

$$\begin{aligned}
\int_a^b F'(x)\,\mathrm{d}x &\le \liminf_{n\to\infty} \int_a^b g_n(x)\,\mathrm{d}x \\
&= \liminf_{n\to\infty} n \int_a^b \left(F\left(x + \frac{1}{n}\right) - F(x)\right)\mathrm{d}x \\
&= \liminf_{n\to\infty} n \left(\int_b^{b+1/n} F(x)\,\mathrm{d}x - \int_a^{a+1/n} F(x)\,\mathrm{d}x\right) \\
&\le F(b) - F(a).
\end{aligned}$$

⊠

Functions of Bounded Variation. The notion of a function of bounded variation is a slight generalization of that of a monotonic function.

Definition. Let f be a real-valued function on $[a, b]$, and let $\mathcal{P}$ be the set of partitions of $[a, b]$. Then for any partition $P = \{a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b\}$, define

$$t = \sum_{k=0}^{n} |f(x_{k+1}) - f(x_k)|$$
$$p = \sum_{k=0}^{n} (f(x_{k+1}) - f(x_k))_+$$
$$n = \sum_{k=0}^{n} (f(x_{k+1}) - f(x_k))_-.$$

Then, f has *bounded variation*, denoted $f \in \mathrm{BV}(a, b)$, if $T_a^b[f] = \sup_{P\in\mathcal{P}} t$ is finite.

Similarly, define $N_a^b[f] = \sup_{P\in\mathcal{P}} n$, and $P_a^b[f] = \sup_{P\in\mathcal{P}} p$.

Theorem 6.8. *$f \in \mathrm{BV}(a, b)$ iff $f = g_1 - g_2$, where g_1 and g_2 are increasing.*

Proof. $f(x) - f(a) = p - n$, so
$$p = n + f(x) - f(a) \le N_a^x(f) + f(x) - f(a).$$
Similarly, $P_a^x(f) \le N_a^x(f) + f(x) - f(a)$. Similarly,
$$n = p - f(x) + f(a) \le P_a^x[f] + f(a) - f(x),$$
so $N_a^x[f] \le P_a^x[f] + f(a) - f(x)$. Thus
$$P_a^x[f] - N_a^x[f] \le f(x) - f(a) \le P_a^x - N_a^x[f],$$
so $f(x) = f(a) + P_a^x[f] - N_a^x[f]$, and the last two terms are increasing functions. ⊠

This is a very romantic notion, which belies a kind of sad proof; all the functions that were looked at in the eighteenth century were of bounded variation.

Corollary 6.9. *If $f \in \mathrm{BV}(a, b)$, then f is differentiable almost everywhere (with respect to the Lesbegue measure).*

7. Two Integration Theorems and Product Measures: 10/16/14

> *"Even when I was at Chicago, I had no idea who was understanding well, because all the first-year graduate students would have identical scores, because they worked together, and everyone else did less well, because they didn't work together."*

Recall that a function is *absolutely continuous* on $[a, b]$ if for all $\varepsilon > 0$ there's a $\delta > 0$ such that any finite collection of disjoint intervals $I_1, \ldots, I_N$, with $I_j = (x_j, y_j)$, if $\sum |I_j| < \delta$, then $\sum_{j=1}^{N} |f(x_i) - f(y_i)| < \varepsilon$.

We wanted to prove theorems about differentiation and integration, specifically Theorems 6.1 and 6.2. We did prove that a monotonic function is differentiable almost everywhere, however, as well as Vitali's lemma, which states that any fine covering $\mathcal{J}$ of a set E of finite measure contains a disjoint finite subcollection capturing at meast $m(E) - \varepsilon$ for any $\varepsilon > 0$. Finally, we defined a function to have *bounded variation* on $[a, b]$ if, for a partition $P = (x_0 = a, x_1, \ldots, x_n = b)$ and $t(P) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|$, then the supremum of this over all partitions is finite. We saw that if a function has bounded variation, then it's the difference of two nondecreasing functions, and therefore is differentiable almost everywhere.

Now, we can return to proving Theorems 6.1 and 6.2.

Proof of Theorem 6.1. First of all, observe that $F(x)$ has bounded variation, because for any partition $a = x_0 < x_1 < \cdots < x_n = b$, then
$$\begin{aligned}\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| &= \sum_{i=1}^{n} \left| \int_{x_{i-1}}^{x_i} f(t)\,\mathrm{d}t \right| \\ &\le \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} |f(t)|\,\mathrm{d}t \\ &= \int_a^b |f(t)|\,\mathrm{d}t,\end{aligned}$$
which is finite. Thus, $F \in \mathrm{BV}([a, b])$, so $F'(x)$ exists almost everywhere.

Lemma 7.1. *Let $f \in L^1([a,b])$ and $\int_a^x f(t)\,\mathrm{d}t = 0$ for all x. Then, $f(x) = 0$ almost everywhere.*

Proof. This is one of those obvious things, so it's a little tricky to prove. Let $E = \{f(x) > 0\}$, and asume $m(E) > 0$. Thus, as we've proven somewhere in the past, there's a compact set $K \subseteq E$ such that $m(K) > 0$, and $\mathcal{U} = [a,b] \setminus K$ is open. Thus,

$$0 = \int_a^b f\,\mathrm{d}x = \int_K f + \int_{\mathcal{U}} f,$$

so $\int_{\mathcal{U}} f$ must be negative (since f is positive K, so its integral must also be positive there). Thus, since $\mathcal{U}$ is a countable union of open intervals, there's an interval I on which $\int_I f(x)\,\mathrm{d}x \neq 0$. This is a contradiction. ⊠

First, assume f is bounded, so that $|f| \le K$. Let

$$f_n(x) = \frac{F(x+1/n) - F(x)}{1/n} = n\int_x^{x+1/n} f(t)\,\mathrm{d}t.$$

Thus, $f_n(x) \to F'(x)$ almost everywhere, and $|f_n(x)| < n(1/n)K = K$, so we can apply the Bounded Convergene Theorem, and get that

$$\begin{aligned}\int_a^x F'(x)\,\mathrm{d}x &= \lim_{n\to\infty}\int_a^x f_n(t)\,\mathrm{d}t\\ &= \lim_{n\to\infty} n\int_a^x \left(F\left(x+\frac{1}{n}\right) - F(x)\right)\mathrm{d}t\\ &= \lim_{n\to\infty} n\left(\int_x^{x+1/n} F(t)\,\mathrm{d}t - \int_a^{a+1/n} F(t)\,\mathrm{d}t\right)\\ &= F(x) - F(a) = F(x),\end{aligned}$$

since $F(a) = 0$ and $F(x)$ is continuous, so the second-to-last step is just like in your first calculus class. Then, since f is bounded and in $L^1([a,b])$, then it's integrable, so

$$\int_a^x F(t)\,\mathrm{d}t = \int_a^x f(t)\,\mathrm{d}t,$$

so by Lemma 7.1, $F'(t) = f(t)$ almost everywhere.

So now we ned to generalize to where f might not be bounded on $[a,b]$. Assume without loss of generality that $f \ge 0$ (if not, write it as the difference of its positive part and negative part); then, let $g_n(x) = \min(f(x), n)$, so that $f - g_n \ge 0$, so $G_n(x) = \int_a^x (f - g_n)\,\mathrm{d}t$ is a non-decreasing function, so G_n' exists almost everywhere and

$$F'(x) = G_n'(x) + \frac{\mathrm{d}}{\mathrm{d}x}\int_a^x g_n(t)\,\mathrm{d}t = G_n' + g_n,$$

since g_n is bounded. But since G_n is non-decreasing, then $G_n' \ge 0$, so $F'(x) \ge g_n(x)$ for any n almost everywhere, and thus $F'(x) \ge f(x)$ almost everywhere. Since $F(x)$ is increasing, then

$$\int_a^x F'(x)\,\mathrm{d}x \ge \int_a^x f(t)\,\mathrm{d}t = F(x),$$

and on the other hand,

$$\int_a^x F'(x)\,\mathrm{d}x \le F(x) - F(a) = F(x).$$

Thus, $F(x) = \int_a^x F'(x)\,\mathrm{d}x$, so we once again invoke Lemma 7.1. ⊠

Proof of Theorem 6.2. One direction is obvious: if $F(x) = F(a) + \int_a^x f(t)\,\mathrm{d}t$, so by absolute continuity of the integral, $F(x)$ is absolutely continuous.

In the other direction, let $F(x)$ be absolutely continuous, so that $F(x) = F_1(x) - F_2(x)$, where F_1 and F_2 are nondecreasing. Additionally, $F'(x) = F_1'(x) - F_2'(x)$, so $|F'(x)| \le F_1'(x) + F_2'(x)$. Thus,

$$\begin{aligned}\int_a^b |F'(x)|\,\mathrm{d}x &\le \int_a^b F_1'(x)\,\mathrm{d}x + \int_a^b F_2'(x)\,\mathrm{d}x\\ &\le F_1(b) - F_1(a) + F_2(b) - F_2(a),\end{aligned}$$

which is finite, so $F' \in L^1([a,b])$ (we can make these transformations on F_1' and F_2' using the previous theorem!). Let $G(x) = \int_a^x F'(x)\,\mathrm{d}x$, so that $G'(x) = F'(x)$ and G is absolutely continuous (by the absolute continuity of the integral), and thus $R(x) = F(x) - G(x)$ is absolutely continuous, but $R'(x) = 0$ almost everywhere.

Lemma 7.2. *If $f(x)$ is absolutely continuous and $f'(x) = 0$ almost everywhere, then $f(x) = f(a)$ for all $x \in [a,b]$.*

Notice that this is not true of continuous functions: on the homework, we constructed a function that has derivative 0 almost everywhere, but is not globally constant. The idea is that the places where f can jump are restricted to smaller and smaller sets.

Proof of Lemma 7.2. Let $A \subseteq [a,b]$ be the set $\{x : f'(x) = 0\}$, so $m(A) = b - a$. For each $x \in A$, choose $h_n(x) < 1/n$ such that $|f(x + h_n(x)) - f(x)| < \varepsilon h_n$.

This is a fine cover of A, so chose a subcollection $I_1, \dots, I_N$ such that

$$m\left(A \setminus \bigcup_{j=1}^N I_j\right) < \delta,$$

so it's also true that

$$m\left([a,b] \setminus \bigcup_{j=1}^N I_j\right) < \delta.$$

Write $[a.b] \setminus \bigcup_{j=1}^N I_j = \bigcup_{k=1}^{N+1} J_k$ with $\sum_{k=1}^{N+1} |J_k| < \delta$, so if $I_j = (x_j, x_j + h_j)$ and $J_k = (x_k + h_k, x_{k+1})$, we get a partition, so

$$\sum_j |f(x_j + h_j) - f(x_j)| < \varepsilon \sum h_j < \varepsilon(b - a).$$

Thus, we can get this less than ε if δ is chosen according to the definition of absolute continuity of f. ⊠

Returning to the outermost proof, $R(x) = F(x) - G(x)$ is absolutely continuous and $R'(x) = 0$ almost everywhere, so $F(x) = G(x)$. ⊠

Thus, the answer to the question, *when does the Newton-Leibniz formula hold?* is *for absolutely continuous functions.*

Now, let's go back to Italy and talk about Fubini's Theorem.

The notion of product measure goes back to Archimedes and such, thousands of years ago. The more recent notion backed by integration was first analyzed in Italy, with Cavalieri,[13] in the first half of the 17$^{\text{th}}$ century. Not just the English and Germans cared about integration! Fubini came around later, so was carrying on a tradition: analysis is really an Italian subject.

We've probably all seen or done the following exercise in a multivariable calculus class, and it's useful to remember.

Exercise 10. Find a function $f(x,y)$ on $[-1,1] \times [-1,1]$ such that

$$\int_{-1}^1 \left(\int_{-1}^1 f(x,y)\,\mathrm{d}y\right)\mathrm{d}x \qquad \text{and} \qquad \int_{-1}^1 \left(\int_{-1}^1 f(x,y)\,\mathrm{d}x\right)\mathrm{d}y$$

exist, but are not equal.

Now, Cavalieri would be offended by such a thing, because these are both ways of measuring volume, and thus they ought not to be different.

This exercise leads to the more general question of when one can change the order of integration in general measure-theoretic integration. But to do this, we must make precise the notion of a product measure.[14]

Definition. Let μ be a measure on X and ν be a measure on Y. Then, the *product measure* is

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

Next, for which sets can the iterated integral be defined? Say that $S \in \mathcal{F}$ if $S(y) = \int_X \chi_S(x,y)\,\mathrm{d}\mu_x$ is defined and measurable for almost all $y \in Y$. If $S \in \mathcal{F}$, then define $\rho(S) = \int_Y s(y)\,\mathrm{d}\nu_y$.

What comes next is a bunch of tautologies, thankfully only ten minutes (leading to a proof of Fubini's Theorem), but it may still put you to sleep. Science has discovered the cure to insomnia, I guess, and it's the road to Fubini's Theorem.

If $U \subseteq V$ and $U, V \in \mathcal{F}$, then $\rho(U) \le \rho(V)$, simply because $\chi_U(x,y) \le \chi_V(x,y)$.

Fact. If $S = A \times B \in \mathcal{F}$ for A and B measurable, then $\rho(A \times B) = \mu(A) \times \nu(B)$.

That is, every rectangle is "Cavalierizable."

[13]There's a Piazza Cavalieri in Pisa, but sadly it's not named after this Cavalieri.

[14]"Recall your geometry days in high school, or middle school... or elementary school, or kindergarten..."

Fact. Let
$$\mathcal{P}_1 = \left\{\bigcup_{j=1}^\infty A_j \times B_j : A_j \subset X \text{ is } \mu\text{-measurable}, B_j \subset Y \text{ is } \nu\text{-measurable}\right\}.$$
Then, $\mathcal{P}_1 \subseteq \mathcal{F}$.

This is true because if $S \in \mathcal{P}_1$, then it can be written as a pairwise disjoint union of these $A_j \times B_j$, so
$$\rho(S) = \sum_{j=1}^\infty \mu(A_j)\nu(B_j).$$

Fact. For every $S \subset X \times Y$,
$$(\mu\times\nu)^*(S) = \inf\{\rho(R) : R \in \mathcal{P}_1, S \subseteq R\}.$$

Proof. If $S \subseteq R = \bigcup_{j=1}^\infty A_j \times B_j$ and $R \in \mathcal{P}_1$, then
$$\begin{aligned}\rho(R) &= \int_Y \left(\int_X \chi_R(x,y)\,\mathrm{d}\mu_x\right)\mathrm{d}\mu_Y\\ &\le \int_Y \left(\int_X \sum_j \chi_{A_j\times B_j}(x,y)\,\mathrm{d}\mu_x\right)\mathrm{d}\nu_Y\\ &= \sum_{j=1}^\infty \mu(A_i)\nu(B_i).\end{aligned}$$
Thus, we can replace the sum in the definition of the product measure with ρ, i.e.
$$\begin{aligned}&\inf_{R'} \rho(R') \le \sum_{j=1}^\infty \mu(A_j)\nu(B_j)\\ \implies &\inf_{R\supset S} \rho(R) \le (\mu\times\nu)^*(S)\\ \implies &\inf_{R\supset S} \rho(R) = (\mu\times\nu)^*(S).\end{aligned}$$ ⊠

The last thing we'll say today will be a result on measurability.

Proposition 7.3. *Let A be μ-measurable and B be ν-measurable. Then, $A \times B$ is $(\mu\times\nu)$-measurable.*

Proof. Let $S = A \times B$ and T be any set. Let $R \supseteq T$ with $R \in \mathcal{P}_1$; then,
$$\begin{aligned}(\mu\times\nu)^*(T\cap(A\times B)^c) + (\mu\times\nu)^*(T\cap(A\times B)) &\le \rho(R\cap(A\times B)^c) + \rho(R\cap(A\times B))\\ &= \rho((R\cap(A\times B)^c \cup (R\cap(A\times B))) = \rho(R).\end{aligned}$$
Thus, this is bounded above by $(\mu\times\nu)^*(T)$, so $A\times B$ is $(\mu\times\nu)$-measurable. ⊠

8. Product Measures and Fubini's Theorem: 10/21/14

> *"This proof used techniques that wouldn't be unfamiliar in sixth grade. Now, in sixth grade, the theorem may have looked a little more nontrivial. . . "*

Recall that when we're talking about product measures, μ is a measure on a space X, ν is a measure on a space Y, and we defined the *product measure* on $X \times Y$ by
$$(\mu\times\nu)^*(S) = \inf\left\{\sum_{j=1}^\infty \mu(A_j)\nu(B_j) : S \subseteq \bigcup_{j=1}^\infty A_j \times B_j\right\}.$$
This is a generalization of a trick used in high school geometry: to divide a region into a lot of rectangles that closely approximate the region, and add them up.

We want to have a simpler formula, i.e.
$$(\mu\times\nu)(S) = \int_Y\left(\int_X \chi_S(x,y)\,\mathrm{d}\mu_x\right)\mathrm{d}\nu_y. \tag{1}$$
If S is measurable, we first want to show that the right-hand side is even defined; last time, we let $\mathcal{F}$ be the collection of sets S such that the right-hand side of (1) is defined, i.e. $\chi_S(x,y)$ is measurable for all $y \in Y$ and $\int_X \chi_S(x,y)\,\mathrm{d}\mu_x$ is ν-measurable.

Next, we proved several facts about these collections.

(1) Let $\mathcal{P}_0 = \{A \times B : A$ is μ-measurable and B is ν-measurable$\}$. Then, if $S \in \mathcal{P}_0$, then $S \in \mathcal{F}$ and $\rho(S) = \mu(A)\nu(B)$.
(2) If $\mathcal{P}_1$ is the set of countable unions of elements of $\mathcal{P}_0$, then if $S \in \mathcal{P}_1$, then $S \in \mathcal{F}$ and if its components are disjoint, then

$$\rho(S) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j).$$

(3) Let $\mathcal{P}_2$ be the set of countable intersections of sets in $\mathcal{P}_1$. Then, $\mathcal{P}_2 \subseteq \mathcal{F}$.

Then, we had several propositions about these, in particular that (1) makes sense for and holds in $\mathcal{P}_2$.

Now, we'll prove Fubini's Theorem.

Definition. A set A is *σ-finite* with respect to a measure μ if A is the union of a countable collection of sets each with finite measure.

Theorem 8.1 (Fubini)**.** *Let S be σ-finite with respect to the product measure $\mu \times \nu$. Then, the cross-section $S_y = \{x : (x, y) \in S\}$ is μ-measurable for almost all (with respect to ν) y, and the cross-section $S_x = \{y : (x, y) \in S\}$ is ν-measurable for almost all (with respect to μ) $x \in X$. Moreover, $\nu(S_x)$ is a μ-measurable function of X, $\mu(S_y)$ is a ν-measurable function of y, and in addition,*

$$(\mu \times \nu)(S) = \int_X \nu(S_x)\,\mathrm{d}x = \int_Y \mu(S_y)\,\mathrm{d}y.$$

Proof. If $(\mu \times \nu)(S) = 0$, then take an $R \in \mathcal{P}_2$ such that $S \subseteq R$ and $\rho(R) = 0$ (which we proved last time we can do). Then, $\chi_S(x, y) \le \chi_R(x, y)$, so $\mu(S_y) \le \mu(R_y) = 0$ for ν-almost all y, and thus

$$\int_Y \mu(S_y)\,\mathrm{d}\nu_y = 0 = (\mu \times \nu)(S).$$

If $(\mu \times \nu)(S)$ is finite, then instead take an $R \in \mathcal{P}_2$ such that $(\mu \times \nu)(S) = \rho(R)$. Thus (since ρ coincides with $\mu \times \nu$ on $\mathcal{P}_2$), $(\mu \times \nu)(R \setminus S) = 0$. Thus, by what we just did, $0 = \int_Y \mu((R \setminus S)_y)\,\mathrm{d}\nu_y$, so $\mu((R \setminus S)_y) = 0$ for ν-almost all $y \in Y$. Thus, for ν-almost all y, $\mu(R_y) = \mu(S_y)$, so $\mu(S_y)$ is ν-measurable and

$$\int_Y \mu(S_y)\,\mathrm{d}\nu_y = \int_Y \mu(R_y)\,\mathrm{d}\nu_y = \rho(R) = (\mu \times \nu)(S).$$

Finallt, if $(\mu \times \nu)(S)$ is infinite, then we use that S is σ-finite: write $S = \bigcup_{k=1}^{\infty}$, where each B_k is $(\mu \times \nu)$-measurable and has a finite measure, and the B_k are disjoint. Then,

$$\begin{aligned}(\mu \times \nu)(S) &= \sum_{j=1}^{\infty} (\mu \times \nu)(B_j) \\ &= \sum_{j=1}^{\infty} \int_Y \mu(\{x : (x, y) \in B_j\})\,\mathrm{d}\nu.\end{aligned}$$

Recall that we had a remarkable theorem that allowed integrals to commute with positive series.

$$\begin{aligned}&= \int_Y \sum_{j=1}^{\infty} \mu(\{x : (x, y) \in B_j\})\,\mathrm{d}\mu_y \\ &= \int_Y \mu\left(\left\{x : (x, y) \in \bigcup_{j=1}^{\infty} B_j\right\}\right)\mathrm{d}\nu = \int_Y \mu(S_y)\,\mathrm{d}\nu_y.\end{aligned}$$

⊠

Notice that this is just a reasonably straightforward application of definitions; there's little creativity here. The proof can sort of be ground out.

The following corollary is sometimes also called Fubini's Theorem.

Corollary 8.2. *Let $X \times Y$ be σ-finite and f be $(\mu \times \nu)$-integrable. Then, $p(y) = \int_X f(x, y)\,\mathrm{d}y$ is ν-integrable, $q(x) = \int_Y f(x, y)\,\mathrm{d}\nu_y$ is μ-measurable, and*

$$\int_{X \times Y} f(x, y)\,\mathrm{d}(\mu \times \nu) = \int_Y p(y)\,\mathrm{d}\nu_y = \int_X q(x)\,\mathrm{d}\mu_x.$$

Proof. Without loss of generality, assume $f \geq 0$ (if not, it's the difference of two nonnegative functions, so we're OK). If $f(x,y) = \chi_S(x,y)$ for some $(\mu \times \nu)$-measurable set S, then we're done; otherwise, write

$$f(x,y) = \sum_{k=1}^{\infty} \frac{1}{k} \chi_{S_k}(x,y),$$

so that

$$\begin{aligned} \int_{X \times Y} f(x,y) \, \mathrm{d}(\mu \times \nu) &= \sum_{k=1}^{\infty} \int_X \frac{1}{k} \nu(\{y : (x,y) \in S_k\}) \, \mathrm{d}\mu_x \\ &= \int_X \left(\int_Y f(x,y) \, \mathrm{d}\nu_y \right) \mathrm{d}\mu_x. \end{aligned}$$

⊠

The assumption that rules out the counterexamples is that $f \geq 0$ (or is the difference of two finite nonnegative functions), which implicitly relies on the measurability of f.

The Besikovitch Theorem. Besikovitch was a Soviet mathematician in the 1920s, who proved the theorem with his name that we're about to talk about. It's a covering theorem, sort of like Vitali's Theorem.

Theorem 8.3 (Besikovitch)**.** *There exists a constant $N(n)$ depending only on the dimension, such that the following property: if $\mathcal{F}$ is any collection of closed balls in $\mathbb{R}^n$,*

$$\mathcal{D} = \sup\{\operatorname{diam}(\overline{B}) : \overline{B} \in \mathcal{F}\}$$

is finite, and A is the set of centers of the balls, then there exists $\mathcal{J}_1, \dots, \mathcal{J}_{N(n)}$ such that each $\mathcal{J}_j$ is a countable collection of disjoint balls in $\mathcal{F}$ and

$$A \subset \bigcup_{j=1}^{N(n)} \bigcup_{\overline{B} \in \mathcal{J}_j} \overline{B}.$$

This looks like a problem on our problem set, and the corollary should also look familiar.

Corollary 8.4. *Let μ be a Borel measure on $\mathbb{R}^n$ and $\mathcal{F}$ be any collection of non-degenerate closed balls. Let A be the set of centers of the balls in $\mathcal{F}$, and assume $\mu(A)$ is finite and $\inf\{r : \overline{B}(a,r) \in \mathcal{F}\} = 0$ for all $a \in A$. Then, for each open $U \subseteq \mathbb{R}^n$, there exists a countable collection $\mathcal{J}$ of disjoint balls in $\mathcal{F}$ such that*

$$\bigcup_{\overline{B} \in \mathcal{J}} \overline{B} \subseteq U \qquad \textit{and} \qquad \mu\left((A \cap U) \setminus \bigcup_{\overline{B} \in \mathcal{J}} \overline{B} \right) = 0.$$

Note that, unlike Vitali's lemma, there's no doubling assumption on μ, and so this will be quite useful for various things, more so than Besikovitch's theorem itself.

Proof of Corollary 8.4. Let $N(n)$ be as in Theorem 8.3, and take a countable disjoint collection $\mathcal{J}$ of disjoint balls in $\mathcal{F}$ such that

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

One of these must exist, because we've put $A \cap U$ into $N(n)$ baskets, so one must contain at least as much as $\mu(A \cap U)/N(n)$ of the measure. Then, choose M_1 such that

$$\mu\left((A \cap U) \setminus \bigcup_{j=1}^{M_1} \overline{B} \right) \geq \left(1 - \frac{1}{2N(n)}\right) \mu(A \cap U).$$

Set

$$U_2 = (A \cap U) \setminus \bigcup_{j=1}^{M_1} \overline{B}_j$$

and repeat: find $B_{M_1} + 1, \dots, B_{M_2}$ such that

$$\mu\left((A \cap U_2) \setminus \bigcup_{j=M_1+1}^{M_2} \overline{B} \right) \geq \left(1 - \frac{1}{2N(n)}\right) \mu(A \cap U_2).$$

Then, we keep repeating; all of the chosen balls are disjoint, and so forth. ⊠

Proof sketch of Theorem 8.3. The first step is to reduce to a bounded set of centers. Choose very large annuli (i.e. a set of points lying between two circles); then, since the sizes of the balls are bounded, then this can be done such that each set intersects at most two of the annuli; then, we can partition into even and odd parts, and at each step, that part of A is bounded.

The next step is to choose the balls to include in $\mathcal{J}$; this is similar to, but not the same, as in Vitali's lemma. The third step is to show that these balls cover A, and then, the last step is to count the balls that intersect.

Let's choose a collection of balls $B_1, B_2, \ldots$ which cover all of A and such that each B_m intersects at most $N(n)$ balls out of $B_1, B_2, \ldots, B_{m-1}$. The reason for this is a middle-school problem: we can put the balls in baskets depending on which things they intersect, and since we're just dealing with closed balls, they can't intersect too many other balls.

Let's expand on why it's sufficient to consider bounded sets of centers. Assume that the theorem is proven for bounded sets, and let $\mathcal{D}$ be as in the problem statement, so it's finite. Let's take layers

$$A_j = \{x \in A : 25\mathcal{D}j \leq |x| < 25\mathcal{D}(j+1)\},$$

where $j = 0, 1, \ldots$: we're taking layers of thickness $25\mathcal{D}$. Now, we cover each A_j with $\mathcal{J}_{j_1}, \ldots, \mathcal{J}_{j_{N(n)}}$, so if $B_1 \in \mathcal{J}_{j+k}$ and $B_2 \in \mathcal{J}_{j+2,m}$ (i.e. it hits three annuli), then $B_1 = B(a_1, r_1)$ and $B_2 = (a_2, r_2)$, so $|a_1 - a_2| \geq 50\mathcal{D} \geq r_1 + r_2$, so $B_1 \cap B_2 = \emptyset$. Thus, the even and odd layers can be mixed however we want, and disjointness is still kept intact; specifically,

$$\mathcal{J}_k = \bigcup_{j=1}^{\infty} \bigcup_{B \in \mathcal{J}_{2j,k}} \overline{B} \qquad \text{and} \qquad \mathcal{J}'_k = \bigcup_{j=0}^{\infty} \bigcup_{B \in \mathcal{J}_{2j+1,k}} B$$

are both pairwise disjoint collections of balls. Thus, if $N(n)$ works for bounded sets of centers, then $2N(n)$ works for arbitrary sets of centers, so we can reduce.

Returning to the bounded case, which we still have to prove, assume A is bounded; now, how do we choose the balls? One again, take $\mathcal{D}$ as in the problem statement, and choose $B_1(a, r_1)$ such that $r_1 \geq (3/4)\mathcal{D}/2$.[15]

Assume $B_1, \ldots, B_{j-1}$ are chosen, and let

$$A_j = A \setminus \bigcup_{k=1}^{j-1} B_k.$$

One possibility is that A_j is empty, in which case we've covered everything. In this case, let $\mathcal{J} = j$ and stop. Otherwise, set

$$\mathcal{D}_j = \sup\{\operatorname{diam}(B(a,r)) : B(a,r) \in \mathcal{F}, a \in A_j\}.$$

Choose $B_j = B(a_j, r_j)$ such that $a \in A_j$ and $r_j \geq (3/4)\mathcal{D}_j/2$, an continue.

Even when $\mathcal{J}$ is finite, we're not done, since the balls may intersect; if $\mathcal{J}$ is not finite, set $\mathcal{J} = +\infty$. Intuitively, balls that come later can't be too much larger than those that came before them, because they were considered at each step. Thus, it's not true that the radii of the balls decreases, but it's almost true. Using this, one can show that $r_j \to 0$. Then, blowing the radii of the balls up or scaling them ha nice properties with regards to covering or disjointness, and so forth. ⊠

This is a typical argument in analysis: there are sets of two kinds, small and large, and each kind is dealt with differently.

9. Besikovitch's Theorem: 10/23/14

"Whenever you take a shortcut, it comes back to bite you, but you won't know when."

There's a shorter statement of Besikovitch's theorem, Theorem 8.3.

Theorem (Besikovitch). *Let $\mathcal{F}$ be a collection of non-degenerate closed balls $\overline{B}$ such that*

(1) $\sup_{B \in \mathcal{F}} \operatorname{diam} B$ *is finite, and*
(2) if A is the set of centers of the balls in $\mathcal{F}$, then for any $a \in A$ and $\varepsilon > 0$, there exists a $B(a,r) \in \mathcal{F}$ with $0 < r < \varepsilon$.

Then, there exist N_n countable subcollections of balls $\mathcal{J}_1, \ldots, \mathcal{J}_{N_n}$ of $\mathcal{F}$ such that if $B_1, B_2 \in \mathcal{J}_k$, then $\overline{B}_1 \cap \overline{B}_2 = \emptyset$ and

$$A \subset \bigcup_{k=1}^{N_n} \bigcup_{\overline{B} \in \mathcal{J}_k} \overline{B}.$$

The constant N_n depends only in the dimension n.

[15]"Of course, real men choose 99/100, but then more real men choose 999/1000, and so on until World War III. So we'll stay out of this and choose 3/4."

Recall also Corollary 8.4, which we proved last time; but this required assuming that we know that the set of centers is measurable. We'll be able to get around it, and will soon, and allow the set of centers to be unmeasurable, but this forces μ to be Borel regular for the corollary to hold.

Suppose $A_1 \subseteq A_2 \subseteq \cdots$ and μ is a regular measure; then, take a measurable $C_k \supseteq A_k$ such that $\mu^*(A_k) = \mu(C_k)$. This may not be an increasing sequence of sets, but $B_k = \bigcap_{j \geq k} C_j$, and $A_k \subseteq B_k$. Thus, $\mu(B_k) \leq \mu(C_k) = \mu^*(A_k)$, so we can reduce to the case of these B_k.

Proof of Theorem 8.3. The proof will go in several steps. Here's an outline.

Step 1. It is enough to find a sequence $\{B_j\}$ such that $A \subseteq \bigcup_{j=1}^\infty B_j$ and each B_j intersects at most $N_n - 1$ balls amongst $B_1, \dots, B_{k-1}$.

Step 2. As discussed last time, we can reduce to the case where A is a bounded set.

Step 3. Choose $B(a, r_1)$ such that $r_1 \geq (3/8)\sup_{B \in \mathcal{F}}(\operatorname{diam} B)$. After choosing $B_1, \dots, B_k$, set $A_k = A \setminus \bigcup_{j=1}^k B_k$. Then, take $B(a_k, r_k)$, where $a_k \in A_k$ and

$$r_k \geq (3/8)\sup\{\operatorname{diam} B(a,r) : B(a,r) \in F, a \in A_k\}.$$

if $r_k = 0$, then stop; otherwise, continue.

Step 4. Finally, verify this collection satisfies the theorem, specifically the bound on the number of sets.

This construction has a few nice properties.

Fact 1. If $j > i$, then $r_j > (3/4)r_i$.

Fact 2. If we shrink the balls by a factor of 3, they become pairwise disjoint, i.e. the balls $B(a_j, 2r_j/3)$ are pairwise disjoint.

Why is the second fact true? Let $j > i$, so that $a_j \notin B(a_i, r_i)$. For the two balls to be disjoint, we would need $|a_j - a_i| > r_i > (r_i + r_j)/3$, but this is equivalent to $2r_i > r_j$, which is true.

Fact 3. If we never stop choosing $B(a_k, r_k)$, then $\lim_{j\to\infty} r_j = 0$.

This is because if $S = \{x : \operatorname{dist}(x, A) < \mathcal{D}\}$, then

$$\bigcup_{j=1}^\infty B\Big(a_j, \frac{r_j}{3}\Big) \subseteq S,$$

and since A is bounded, this is too. Thus, $\sum r_j$ is finite, so $\lim_{j\to\infty} r_j = 0$.

Fact 4.

$$A \subseteq \bigcup_{j=1}^{\mathcal{J}} B_j,$$

where $\mathcal{J}$ is the number of $B(a_k, r_k)$ we considered in this process.

If $\mathcal{J}$ is finite, this is obvious, and if not, then assume there exists an $a \in A$ such that $a \notin \bigcup_{j=1}^\infty \overline{B}_k$; then, there exists a ball $B(a, r) \in \mathcal{F}$ which was a candidate at some point (since $r_k \to 0$), so had to have been chosen.

Now, we can move to the last step, counting the intersections of the balls.

Proposition 9.1. *There exists a number N_n such that each ball $\overline{B}_k$ intersects at most $N_n - 1$ balls out of $\overline{B}_1, \dots, \overline{B}_{k-1}$.*

We'll define three sets:

- $I_m = \{1 \leq j < m : B_j \cap B_m \neq \emptyset\}$, which is the set of bad balls, so to speak.
- $K_m = \{j \in I_m : r_j \leq 3r_m\}$, so the small bad balls.
- $P_m = \{j \in I_m : r_j > 3r_m\}$, the big bad balls.

Lemma 9.2. $|K_n| < 20^n$.

Proof. This is a very coarse estimate, but does the job for us.

We'll see that $B(a_m, 5r_m) \supseteq B(a_j, r_j/3)$ for each $j \in K_m$. Specifically, take an $x \in B(a_j, r_j/3)$, so

$$\begin{aligned} |x - a_m| &\leq |x + a_j| + |a_j - a_m| \leq \frac{r_j}{3} + r_j + r_m \\ &\leq \frac{4}{3} r_j + r_m \leq 4r_m + r_m = 5m. \end{aligned}$$

Thus, $B(a_j.r_j/3) \subseteq B(a_m, 5r_m)$, and by the construction of K_m, all of the $B(a_j, r_j/3)$ are disjoint. Thus, we can conclude that

$$5^n r_m^n \geq |K_m| \Big(\frac{1}{3} \cdot \frac{3}{4} r_m\Big)^n,$$

so $|K_m| < 20^n$. ⊠

Lemma 9.3. *Let $i, j \in P_m$; then, the angle θ between $a_j - a_m$ and $a_i - a_m$ is larger than* $\arccos(61/64)$*.*

Let r_0 be so small that if x, y lie on the unit sphere in $\mathbb{R}^n$ and $|x-y| < r_0$, then the angle between them is smaller than $\arccos(61/64)$; then, if we can cover the unit sphere with L_n balls of radius r_0, then $|P_n| \le L_n$.

Lemma 9.4. *If $|a_i| \le |a_j|$ and $\cos\theta > 5/6$, then $a_i \in \overline{B}_j$.*

Lemma 9.5. *If $a_i \in \overline{B}_j$ and $|a_i| \le |a_j|$, then $\cos\theta < 61/64$ (where θ is as in Lemma 9.3).*

Proof of Lemma 9.3. Set $a_m = 0$. Then, $r_i < |a_i|$ and $r_j \le |a_j|$, since $m > i, j$. Since $B_m \cap B_j \ne \emptyset$ and similarly with B_i, then $|a_i| < r_i + r_m$ and $|a_j| < r_j + r_m$, so $r_i > 3r_i$ and $r_j > 3r_m$, because $i, j \in P_m$.

Putting this all together, $3r_m < r_i < |a_i| < r_i + r_m$, and similarly with j. So we want to show that if $\cos\theta > 5/6$, then $|a_i - a_j| \le |a_j|$. Assuming $|a_i - a_j| \ge |a_j|$, then

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

which is a contradiction. ⊠

Proof of Lemma 9.4. Assume $a_i \in \overline{B}_j$, so $r_j < |a_i - a_j|$. Thus,

$$\begin{aligned}\cos\theta &= \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} \\ &= \frac{|a_i|}{2|a_j|} + \frac{(|a_j| - |a_i - a_j|)(|a_j| + |a_i - a_j|)}{2|a_i||a_j|} \\ &\le \frac{1}{2} + \frac{(|a_i| - |a_i - a_j|)2|a_j|}{2|a_i||a_j|} \\ &\le \frac{1}{2} + \frac{|a_j| - |a_i - a_j|}{2|a_i|}.\end{aligned}$$

Here, we actually use the assumption on $a_i \in \overline{B}_j$, though there are a few other places it's used. After this point, plugging in r_j causes $5/6$ to fall out, albeit somewhat magically, and the needed contradiction is reached. ⊠

Proof. We want to show that if $a_i \in \overline{B}_j$, then $\cos\theta < 61/64$. We do have that

$$\frac{|a_j|}{8} \le |a_i - a_j| + |a_i| - |a_j| \le \frac{8}{3}|a_j|(1 - \cos\theta).$$

First, $|a_i - a_j| + |a_i| - |a_j| \ge r_i + r_i - r_j + r_m$. Since $a_i \in \overline{B}_j$, then $i < j$ and $a_i \in \overline{B}_i$, so this value is also

$$\begin{aligned}|a_i - a_j| + |a_i| - |a_j| &\ge 2r_i - r_j - r_m \\ &\ge 2 \cdot \frac{3}{4} r_j - r_j - \frac{1}{3} r_j \ge \frac{r_j + r_m}{8} \ge \frac{|a_j|}{8}.\end{aligned}$$

Now, the other direction. The Triangle Inequality implies that

$$\begin{aligned}0 &\le \frac{|a_i - a_j| + |a_i| - |a_j|}{|a_j|} \\ &\le \frac{|a_i - a_j| + |a_i| - |a_j|}{|a_j|} \cdot \frac{|a_i - a_j| + |a_j| - |a_i|}{|a_i - a_j|} \\ &= \frac{|a_i - a_j|^2 - (|a_i| - |a_j|)^2}{|a_j||a_i - a_j|} \\ &= \frac{2|a_i||a_j|(1 - \cos\theta)}{|a_j||a_i - a_j|} \\ &= \frac{2|a_i|(1 - \cos\theta)}{|a_i - a_j|} \\ &= \frac{2(r_i + r_m)(1 + \cos\theta)}{r_i} \\ &\le \frac{2(r_i + r_i/3)(1 - \cos\theta)}{r_i} = \frac{8}{3}(1 - \cos\theta).\end{aligned}$$

⊠

Phew… I think we're finished now. ⊠

This will be the longest proof in the class. Since it took a lecture and a half to prove this, we're going to milk it for whatever we can, e.g. the Radon-Nikodym Theorem in $\mathbb{R}^n$ (though it does also hold in infinite-dimensional spaces, which has a longer proof that doesn't depend on the Besikovitch Theorem) and a wealth of other things.

10. Differentiation of Measures: 10/28/14

"I've taught the Radon-Nikodym theorem many times, and still know nothing about Nikodym."

Differentiation of measures is a way of looking at local properties of measures, specifically their relative sizes.

If μ and ν are two Radon measures, we can compute $\nu(A)$ by Riemann integration with respect to ν: dividing A into many small boxes $B_1, \dots, B_N$; then,

$$\nu(A) \approx \sum_{j=1}^{N} \frac{\nu(B_j)}{\mu(B_j)} \mu(B_j).$$

$\nu(B_j)/\mu(B_j)$ is really a local property, so call that ratio $f(x)$. Thus, this can further be approximated by

$$\approx \sum_{j=1}^{N} f(x)\mu(B_j) \approx \int_A f \, \mathrm{d}\mu,$$

where the integral is in the sense of Riemann.

This is all fine in theory, but there are some issues with it that we'll address.

(1) Do we have any reason to believe the limit exists? This applies both to defining f to be a local property and to the limit as $N \to \infty$.
(2) What if $\mu(B_j) = 0$?

First perhaps we should formally define what we're talking about. In today's lecture, all balls are closed.

Definition. Given two Radon measures μ and ν, the *upper and lower Radon-Nikodym derivatives* of μ and ν are respectively

$$\overline{\mathcal{D}}_\mu \nu(x) = \begin{cases} \limsup\limits_{r \to 0} \dfrac{\nu(\overline{B}(x,r))}{\mu(\overline{B}(x,r))}, & \text{if } \mu(\overline{B}(x,r)) > 0 \text{ for all } r > 0 \\ +\infty, & \text{if } \mu(\overline{B}(x,r)) = 0 \text{ for some } r > 0; \end{cases}$$

$$\underline{\mathcal{D}}_\mu \nu(x) = \begin{cases} \liminf\limits_{r \to 0} \dfrac{\nu(\overline{B}(x,r))}{\mu(\overline{B}(x,r))}, & \text{if } \mu(\overline{B}(x,r)) > 0 \text{ for all } r > 0 \\ +\infty, & \text{if } \mu(\overline{B}(x,r)) = 0 \text{ for some } r > 0. \end{cases}$$

If $\overline{\mathcal{D}}_\mu \nu = \underline{\mathcal{D}}_\mu \nu$, then ν is said to be *differentiable* with respect to μ, and $\mathcal{D}_\mu \nu$ is the *Radon-Nikodym derivative* or *density* of ν with respect to μ.

Example 10.1.

(1) Suppose μ is the Lesbegue measure on $\mathbb{R}^n$ and $\nu(A) = \int_A f(x) \, \mathrm{d}\mu$, where f is continuous. Then, $\mathcal{D}_\mu \nu(x) = f(x)$ and
$$\nu(A) = \int_A \mathcal{D}_\mu \nu(x) \, \mathrm{d}\mu.$$
(2) Suppose $\mu(A) = m(A \cap [0,1])$ and $\nu(A) = m(A \cap [2,3])$, so that $\mathcal{D}_\mu \nu(x) = 0$ on $[0,1]$. Thus, $\mathcal{D}_\mu \nu = 0$ for μ-almost everywhere x, and therefore $\int_A \mathcal{D}_\mu \nu(x) \, \mathrm{d}\mu = 0$ for all A. In particular,
$$\nu(A) \neq \int_A \mathcal{D}_\mu \nu \, \mathrm{d}\mu$$
whenever $\nu(A)$ is positive.

This leads to two important questions:

(0) Even before the first question, when is it possible to integrate $\mathcal{D}_\mu \nu(x)$ with respect to μ?
(1) Then, for what μ and ν is the following true?
$$\nu(A) = \int_A \mathcal{D}_\mu \nu(x) \, \mathrm{d}\mu.$$

Theorem 10.2. *Let μ and ν be Radon measures on $\mathbb{R}^n$. Then, $\mathcal{D}_\mu \nu$ exists for μ-almost everywhere x, is a nonnegative, μ-measurable function, and is finite μ-almost everywhere.*

Proof. $\mathcal{D}_\mu \nu(x)$ is a local quantity, so it doesn't change if we restrict μ and ν to $B(x, R)$ for some $R > 0$. Therefore, assume without loss of generality therefore that $\nu(\mathbb{R}^n)$ and $\mu(\mathbb{R}^n)$ are finite.

Lemma 10.3. *Let μ and ν be finite Radon measures on $\mathbb{R}^n$; then,*

(1) *$A \subseteq \{\underline{\mathcal{D}}_\mu \nu \le s\}$ implies that $\nu(A) \le s\mu(A)$.*
(2) *$A \subseteq \{\overline{\mathcal{D}}_\mu \nu \ge s\}$ implies that $\nu(A) \ge s\mu(A)$.*

Proof. We'll prove (1); then, (2) is almost exactly the same. Notice also that A doesn't need to be μ- or ν-measurable (in which case the measures should be replaced with outer measures μ^* and ν^*).

Take an open set $U \supset A$ and cover each $x \in A$ by a ball of radius $r_n(x) \le 1/n$ such that

$$\nu(\overline{B}(x, r_n(x))) \le (s+\varepsilon)\mu(\overline{B}(x, r_n(x))),$$

and, by the magical corollary to the Besikovitch Theorem, we can chooe a countable collection of disjoint balls $\overline{B}_j$ such that

$$\nu\left(A \setminus \bigcup_{j=1}^\infty \overline{B}_j\right) = 0.$$

Then,

$$\begin{aligned}
\nu^*(A) &\le \nu^*\left(A \setminus \bigcup_{j=1}^\infty \overline{B}_j\right) + \nu\left(\bigcup_{j=1}^\infty \overline{B}_j\right) \\
&= \sum_{j=1}^\infty \nu(\overline{B}_j) \\
&\le \sum_{j=1}^\infty (s+\varepsilon)\mu(\overline{B}_j) \\
&= (s+\varepsilon)\sum_{j=1}^\infty \mu(\overline{B}_j) \le (s+\varepsilon)\mu(U).
\end{aligned}$$

If we proceed to take the infimum over all such open sets U, then since μ and ν are Radon, then $\nu^*(A) \le (s+\varepsilon)\mu^*(A)$. ⊠

Now, armed with the lemma, let's look at the set where $\overline{\mathcal{D}}_\mu \nu > \underline{\mathcal{D}}_\mu \nu$, as well as the set where $\overline{\mathcal{D}}_\mu \nu$ is infinite (at positive infinity).

(1) Let $I = \{x : \overline{\mathcal{D}}_\mu \nu = +\infty\}$, so that $\nu^*(I) \ge s\mu^*(I)$ for all $s > 0$; since $\nu(\mathbb{R}^n)$ is finite, then $\mu(I) = 0$.
(2) Let $R_{ab} = \{\underline{\mathcal{D}}_\mu \nu(x) \le a < b \le \overline{\mathcal{D}}_\mu \nu(x)\}$. Then, since $\underline{\mathcal{D}}_\mu \nu \le a$, then $\nu(R_{ab}) \le a\mu(R_{ab})$ and $\overline{D}_\mu \nu \ge b$, so $\nu(R_{ab}) \ge b\mu(R_{ab})$ (both by Lemma 10.3), so $\mu(R_{ab}) = 0$.

Thus, $\mathcal{D}_\mu \nu$ exists and is finite μ-almost everywhere. ⊠

Since the Radon-Nikodym derivative is a local quantity, the restriction to small balls is a very useful trick; this isn't the last time we'll see it today.

Lemma 10.4. *Let μ and ν be Radon measures; then, for each $x \in \mathbb{R}^n$ and $r > 0$,*

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

Proof. Let $y_k \to x$, and set $f_k(z) = \chi_{\overline{B}_k}(z)$. The idea is to relate moving small amounts with now much the mass of the measure can change.

Claim. $\limsup_{k\to\infty} f_k(z) \le \chi_{\overline{B}(x,r)}(z)$.

Proof. For the claim, we only need to check that if $\chi_{\overline{B}(x,r)}(z) = 0$, then $f_k(z) = 0$ for large k (otherwise, it's immediate). But $\liminf_{k\to\infty}(1 - f_k(z)) \ge 1 - \chi_{\overline{B}(x,r)}(z)$, so we can integrate and apply Fatou's lemma:

$$\int_{\overline{B}(x,2r)} (1 - \chi_{\overline{B}(x,r)}(z))\,\mathrm{d}\mu \le \liminf_{k\to\infty} \int_{\overline{B}(x,2r)} (1 - f_k(z))\,\mathrm{d}\mu.$$

TODO I didn't get to write down this either… what is this? ⊠

⊠

But we still need to deal with the counterexample in Example 10.1. In fact, let's just exclude it.

Definition. A Borel measure ν is *absolutely continuous* with respect to another Borel measure μ, written $\nu \ll \mu$, if for any set A such that $\mu(A) = 0$, we have $\nu(A) = 0$.

Theorem 10.5 (Radon-Nikodym)**.** *If $\nu \ll \mu$, then for any μ-measurable set A we have that*

$$\nu(A) = \int_A \mathcal{D}_\mu \nu \, \mathrm{d}\mu.$$

For this to even make sense, we have to show that μ-measurability implies ν-measurability.

Claim. If $\nu \ll \mu$, then any μ-measurable set is ν-measurable.

Proof. Take a μ-measurable set A and a Borel set B such that $A \subseteq B$ and $\mu(A) = \mu(B)$, and therefore $\mu(B \setminus A) = 0$, so $\nu(B \setminus A) = 0$; in partciular, $B \setminus A$ is ν-measurable, and so is $A = B \setminus (B \setminus A)$. ⊠

Proof of Theorem 10.5. Set $Z = \{\mathcal{D}_\mu \nu = 0\}$ and $I = \{\mathcal{D}_\mu \nu \text{ is infinite}\}$. Then, by Theorem 10.2, $\mu(I) = 0$.

For any $R > 0$, set $Z_R = Z \cap B(0, R)$, so that for all $s > 0$, $\nu(Z_R) \le s \cdot \mu(Z_R)$, and therefore $\nu(Z_R) = 0$ for all $R > 0$. In particular, $\nu(Z) = 0$. Thus,

$$\nu(Z) = \int_Z \mathcal{D}_\mu \nu \, \mathrm{d}\mu \qquad \text{and} \qquad \nu(I) = \int_I \mathcal{D}_\mu \nu \, \mathrm{d}\mu.$$

Take any set A, so that

$$A = (A \cap Z) \cup (A \cap I) \cup \bigcup_{m=-\infty}^{\infty} A_m,$$

where $A_m = \{x \in A : t^m \le \mathcal{D}_\mu \nu(x) < t^{m+1}\}$ for some $t > 1$. Then,

$$\begin{aligned}
\nu(A) &= \sum_{m=-\infty}^{\infty} \nu(A_m) \\
&\le \sum_{m=-\infty}^{\infty} t^{m+1} \mu(A_m) \le t \sum_{m=-\infty}^{\infty} t^m \mu(A_m) \\
&\le t \sum_{m=-\infty}^{\infty} \int_{A_m} \mathcal{D}_\mu \nu \, \mathrm{d}\mu \le t \int_A \mathcal{D}_\mu \nu \, \mathrm{d}\mu.
\end{aligned}$$

Thus, $\nu(A) \le \int_A \mathcal{D}_\mu \nu \, \mathrm{d}\mu$.

The proof in the other direction is not complicated; see the lecture notes. ⊠

But what if ν isn't absolutely continuous with respect to μ? Maybe we can try to make it as bad as possible, and then throw that part away.

Definition. If μ and ν are Borel measures, then they are said to be *mutually singular* if there exists a Borel set B such that $\mu(\mathbb{R}^n \setminus B) = \nu(B) = 0$ (i.e. their supports are split by a Borel set). This is written $\mu \perp \nu$.

Theorem 10.6. *Let μ and ν be Radon measures; then, it is possible to write $\nu = \nu_a + \nu_s$, where $\nu_a \ll \mu$, $\nu_s \perp \mu$, and $\mathcal{D}_\mu \nu = \mathcal{D}_\mu \nu_a$ for μ-almost all x, and for each Borel set A we have*

$$\nu(A) = \int_A \mathcal{D}_\mu \nu \, \mathrm{d}\mu + \nu_s(A).$$

Proof. The hardest part of this proof is splitting ν; once this is done, the rest follows fairly quickly.

Once again, we can assume without loss of generality that $\mu(\mathbb{R}^n)$ and $\nu(\mathbb{R}^n)$ are finite. The goal is to find a Borel set B such that $\nu|_B \ll \mu$ and $\nu|_{B^c} \perp \mu$. We'll want to pick B such that $\mu(B^c) = 0$.

Let's look at candidates for B, collected in $\mathcal{F} = \{A : A \text{ is Borel and } \mu(A^c) = 0\}$. Given a $A \in \mathcal{F}$, $\nu|_{A^c}$ is certainly mutually singular with μ, since $\nu|_{A^c}$ is 0 on A, where μ takes measure. But we also have to get $\nu|_A \ll \mu$. But if we chose A too large, there may be a subset of it on which μ is 0, but ν is positive; we have to address this part, by choosing B such that this isn't possible.

Take a $B_k \in \mathcal{F}$ such that $\nu(B - k) \le \inf_{A \in \mathcal{F}} \nu(A) + 1/k$, and set $B = \bigcap_{k=1}^{\infty} B_k$. Then,

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

so $B \in \mathcal{F}$ and $\nu(B) = \inf_{A \in \mathcal{F}} \nu(A)$, and once again we get $\nu|_{B^c} \perp \mu$.

Now, we need to show that $\nu|_B \ll \mu$. Assume not, so that there is a set $A \subseteq B$ such that $\mu(A) = 0$ but $\nu(A)$ is positive. Pick a Borel set $A' \supseteq A$ such that $\nu(A') = 0$ and $\nu|_B(A') = 0$. Set $B' = B \setminus A'$, so that $\mu(B'^c) \leq \mu(B^c) = 0$ and $\nu(B') < \nu(B)$, but this is a contradiction to $\nu(B) = \inf_{A\in\mathcal{F}} \nu(A)$ for B Borel. Thus, $\nu|_B \ll \mu$.

Now, we can write $\nu_a = \nu|_B$ and $\nu_s = \nu_{B^c}$, so that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$, and $\nu_s \perp \mu$. We just need to check that $\mathcal{D}_\mu \nu_s = 0$ μ-almost everywhere.

Let $C_z = \{x : \mathcal{D}_\mu \nu_s(x) \geq S\}$ for an $S > 0$. Then, $C_z = (C_z \cap B) \cup (C_z \cup B^c)$, but $\mu(C_z \cap B^c) = 0$ by choice of B, so $S\mu(C_z \cap B) \leq \nu_s(C_z \cap B) = 0$, and so $\mu(C_z \cap B) = 0$ as well; thus, $\mu(C_z) = 0$, so $\mathcal{D}_\mu \nu = \mathcal{D}_\mu \nu_a$ μ-almost everywhere. ⊠

Next time, we'll prove a much cuter theorem.

Theorem 10.7. *Let E be a measurable set. Then,*

$$\frac{m(E \cap B(x,r))}{m(B(x,r))} \longrightarrow \begin{cases} 1, & \text{if } x \in E \\ 0, & \text{if } x \notin E \end{cases}$$

μ-almost everywhere.

It is a nice high-school problem, which, unfortunately, uses the Besikovitch theorem.

11. Local Averages and Signed Measures: 10/30/14

"Of course, the way mathematicians are brought up is to think of horrible counterexamples."

Recall that last time, we defined the derivative of a Radon measure ν with respect to another Radon measure μ is

$$\mathcal{D}_\mu \nu = \begin{cases} \lim_{r\to 0} \nu(\overline{B}(x,r))/\mu(\overline{B}(x,r)), & \text{if } \mu(\overline{B}(x,r)) > 0 \text{ for all } r > 0 \\ +\infty, & \text{otherwise,} \end{cases}$$

provided the limit exists. We showed this exists and is finite μ-almost everywhere, and that it is a μ-measurable function. Then, we asked when it's true that $\nu(A) = \int_A \mathcal{D}_\mu \nu(x)\,\mathrm{d}\mu$, and saw this was true when ν is absolutely continuous with respect to μ, denoted $\nu \ll \mu$ (i.e. if A is such that $\mu(A) = 0$, then $\nu(A) = 0$). Then, we defined $\nu \perp \mu$, i.e. mutual singularity, as when ν and μ take on zero measure on sets that are complements of each other, and proved the Lesbegue decomposition of any ν into an absolutely continuous part and a mutually singular part.

Today, we're going to milk this to talk about local averages, where we average the value of a function f over a small ball.

Definition. The *average* of a measurable function f over a measurable set E is

$$\fint_E f\,\mathrm{d}\mu = \frac{1}{\mu(E)} \int_E f\,\mathrm{d}\mu.$$

So in general, how close is $\fint_{B(x,r)} f\,\mathrm{d}\mu$ to $f(x)$? One answer might be "as far as you want," since there are all sorts of monstrosities and counterexamples that come out on Halloween night. Yet if f is continuous, then of course the average converges to $f(x)$ as $r \to 0$. But what if f is not continuous?[16]

To make sense of the question, we need $f \in L^1(\mathbb{R}^n, \mathrm{d}\mu)$. But L^1 functions aren't that weird, and in fact despite their discontinuities, they're very well-behaved.

Theorem 11.1. *Let $f \in L^1(\mathbb{R}^n, \mathrm{d}\mu)$ and μ be a Radon measure; then,*

$$\lim_{r\downarrow 0} \fint_{B(x,r)} f\,\mathrm{d}\mu = f(x)$$

μ-almost everywhere.

Proof. Unfortunately, said the professor, the proof is very easy.

For a Borel set B, define two measures by

$$\nu_\pm(B) = \int_B f_\pm\,\mathrm{d}\mu,$$

where $f_+ = \max(f, 0)$ and $f_- = f_+ - f$. If μ is Radon and $f \in L^1$, then these measures are Radon and $\nu_+, \nu_- \ll \mu$. Then, we'll extend $\nu_\pm$ to all sets as an outer measure as usual.

[16] This is an excellent question to ask, say, eleventh-graders who have reasonable ideas what functions and continuity are. The professor did this with a friend once, but with Fermat's Last Theorem!

The absolute continuity implies that

$$\nu_+(A) = \int_A \mathcal{D}_\mu \nu \, \mathrm{d}\mu, \text{ and}$$
$$\nu_-(A) = \int_A \mathcal{D}_\mu \nu_- \, \mathrm{d}\mu.$$

But we can show that $\mathcal{D}_\mu \nu_\pm = f_\pm$ almost everywhere. Specifically, let $S_q = \{f_+ - \mathcal{D}_\mu \nu_+ > q\}$ for $q > 0$ and $q \in \mathbb{Q}$. Then,

$$0 = \int_{S_q} (f_+ - \mathcal{D}_\mu \nu_+) \, \mathrm{d}\mu \geq q\mu(S_q),$$

and therefore $\mu(S_q) = 0$. Thus,

$$f(x) = \mathcal{D}_\mu \nu_+(x) - \mathcal{D}_\mu \nu_-(x) = \lim_{r \to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f \, \mathrm{d}\mu. \qquad \boxtimes$$

Definition. A point $x \in \mathbb{R}^n$ is a *Lesbegue point* for $f \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$ if the oscillation of f around x vanishes, i.e.

$$\lim_{r \to 0} \fint_{B(x,r)} |f(y) - f(x)|^p \, \mathrm{d}\mu = 0.$$

Corollary 11.2. *If μ is a Radon measure and $f \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$, then μ-almost all $x \in \mathbb{R}^n$ are Lesbegue points of f.*

Proof. Take a countable dense set $S \subset \mathbb{R}$, and for each $\xi_i \in S$, the function $f(y) - \xi_i \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$. Thus, there exists a set $A_j \subseteq \mathbb{R}^n$ such that $\mu(A_j) = 0$ and for all $x \notin A_j$,

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

Let $A = \bigcup_{j=1}^\infty A_j$ and $G = A^c$, so that $\mu(G^c) = 0$.

For all $x \in G$, we have that

$$\lim_{r \to 0} \fint_{B(x,r)} |f(y) - \xi_j|^p \, \mathrm{d}\mu_y = |f(x) - \xi_j|^p$$

for all j. Thus, take ξ_j such that $|f(x) - \xi_j| < \varepsilon$, so that

$$\begin{aligned}\fint_{B(x,r)} |f(y) - f(x)|^p \, \mathrm{d}\mu &\leq C_p \fint_{B(x,r)} |f(y) - \xi_j|^p \, \mathrm{d}\mu_y + C_p \fint_{B(x,r)} |\xi_j - f(x)|^p \, \mathrm{d}\mu_y \\ &\leq I(r, \xi_j) + C\varepsilon^p,\end{aligned}$$

when the ξ_j are such that $|f(x) - \xi_j| < \varepsilon$. Thus, the proof is finished when r is chosen such that

$$\left| \fint_{B(x,r)} |f(y) - \xi_j|^p \, \mathrm{d}\mu_y - |f(x) - \xi_j|^p \right| < \varepsilon. \qquad \boxtimes$$

This corollary states that integrable functions don't oscillate too much locally, which is actually quite nice.

Now, we'll also be able to prove Theorem 10.7.

Proof of Theorem 10.7. Apply the Lesbegue-Besicovitch theorem to $\chi_E(x)$; then, the theorem statement is just that when we average this function on small balls, it converges to the function itself. $\boxtimes$

This requires knowing very little, even if the buildup to it is a little involved.

Signed Measures and the Riesz Representation Theorems.

Definition. A *signed measure* on a σ-algebra $\mathcal{F}$ is a function $\nu : \mathcal{F} \to \mathbb{R}$ such that:

(1) $\nu(\emptyset) = 0$.
(2) For pairwise disjoint sets $E_1, E_2, \ldots \in \mathcal{F}$,

$$\nu\left(\bigcup_{j=1}^\infty E_j\right) = \sum_{j=1}^\infty \nu(E_j),$$

and the series converges absolutely.
(3) ν takes on only one value out of $+\infty$ and $-\infty$.

A typical example of a signed measure is the difference of two unsigned measures. The restriction on convergence are so that the infinite sum is well-defined.

Definition. If ν is a signed measure, then a set A is *positive* if $\nu(E) \geq 0$ for any subset E of A. *Negative* sets are defined in the analogous way.

Proposition 11.3. *If E is a measurable set such that $\nu(E) > 0$, but is finite, then there exists a positive subset $A \subseteq E$ with $\nu(A) > 0$.*

Proof. If E is not positive, choose n_1 to be the smallest integer such that E contains a subset A_1 such that $\nu(A_1) < -1/n_1$, and let $E_1 = E \setminus A_1$, and keep going: we get an $A_1 \subseteq E_2$ with $E_2 = E_1 \setminus A_2$, and so on.

If this process stops, then we're done, but if it doesn't stop, then look at

$$E' = E \setminus \left(\bigcup_{j=1}^{\infty} A_j \right).$$

Note that $\sum |\nu(A_j)|$ is finite, because $\nu(A) = 0$, so $n_k \to 0$ as $k \to +\infty$.

Suppose that E' contains a subset S such that $\nu(S) < 0$; then, for k large enough, $\nu(S) < -1/(n_k - 1)$, so why did we pick n_k and not $n_k - 1$? Contradiction. ⊠

Theorem 11.4. *Let ν be a signed measure on a space X; then, $X = A \cup B$, where A is a positive set and B is a negative set.*

Proof. Assume ν omits $+\infty$ and define $\lambda = \sup\{\nu(A) : A \text{ is a positive set}\}$. Then, for each j choose a positive A_j such that $\nu(A_j) \geq \lambda - 1/2^j$, and define $A = \bigcup_{j=1}^{\infty} A_j$. Then:

- A must be positive.
- $\nu(A) \geq \nu(A_j)$ for all j, and thus $\nu(A) \geq \lambda$, so $\nu(A) = \lambda$.

If $B = A^c$, then B cannot contain a subset E of positive measure, because if $\nu(E) > 0$ and we have a positive $E' \subseteq B$ with nonzero measure, then $E' \cup A$ is a positive set with $\mu(E' \cup A) > \lambda$, which is a contradiction. ⊠

This means that every signed measure is a difference of two measures.

Corollary 11.5. *Any signed measure $\mu = \mu^+ - \mu^-$, where $\mu^{\pm}$ are two measures and $\mu^+ \perp \mu^-$.*

Definition.

- The *total variation* of a signed measure ν is $|\nu| = \nu^+ + \nu^-$.
- If ν is a signed measure and μ is an unsigned measure, then ν is absolutely continuous with respect to μ, denoted $\nu \ll \mu$, if $\nu^+, \nu^- \ll \mu$.

The Radon-Nikodym Theorem holds for signed measures as well: if $\nu \ll \mu$ and μ, ν are Radon, then $\nu(A) = \int_A f \, \mathrm{d}\mu$, where $f = f^+ - f^-$ is given by $f^+ = \mathcal{D}_\mu \nu^+$ and $f^- = \mathcal{D}_\mu \nu^-$.

The Riesz Representation Theorem classifies all bounded linear functionals on $L^p(\mathbb{R}^n, \mathrm{d}\mu)$ (except for L^∞), where μ is a Radon measure. First recall the following definition.

Definition. If X is a normed vector space, $F : X \to \mathbb{R}$ is a *bounded linear functional* if $|F(x)| \leq C\|x\|_X$. Then, the *norm* of F is $\|F\| = \sup_{\|x\|_X = 1} |F(x)|$.

Recall also Hölder's inequality: that if $1/p + 1/q = 1$, then

$$\left| \int fg \, \mathrm{d}\mu \right| \leq \left(\int |f|^p \right)^{1/p} \left(\int |g|^q \right)^{1/q}.$$

We want to know what the bounded linear functionals are on $L^p(\mathbb{R}^n, \mathrm{d}\mu)$.

Example 11.6. Suppose $g \in L^q(\mathbb{R}^n, \mathrm{d}\mu)$, where $1/p + 1/q = 1$, and define

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

Then, $|F(f)| \leq \|f\|_p \|g\|_q$, so it's bounded, and $\|F\| \leq \|g\|_q$.

The Riesz representation theorem says in some sense that these are the only bounded linear functionals.

Theorem 11.7 (Riesz Representation)**.** *Let μ be a Radon measure, $1 \leq p < \infty$, and F be a bounded linear functional on $L^p(\mathbb{R}^n, \mathrm{d}\mu)$. Then, there exists a $g \in L^q(\mathbb{R}^n, \mathrm{d}\mu)$ such that $F(f) = \int_{\mathbb{R}^n} fg \, \mathrm{d}\mu$ for all $f \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$, and $\|F\| = \|g\|_{L^q}$.*

Proof. First, we'll want to try to guess what g is, and then show it's in L^q; then, the last step is to compute the norm of F.

Assume that μ is a finite measure, so that $\chi_E \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$ for any measurable set E. If g exists, then $F(\chi_E) = \int_E g \,\mathrm{d}\mu$, so define $\nu(E) = F(\chi_E)$, which is a signed measure; in particular because μ is finite, so sums of characteristic functions converge:

$$\nu\left(\bigcup_{j=1}^{\infty} E_j\right) = F\left(\sum_{j=1}^{\infty} \chi_{E_j}\right) = \sum_{j=1}^{\infty} F(\chi_{E_j}).$$

Thus, $|\nu(E)| \le \|F\|\|\chi_E\|_p = \|F\|(\mu(E))^{1/p}$, so $\nu \ll \mu$. Thus, set $g = \mathcal{D}_\mu \nu$, since we've just shown this is the only option.

Thus, if $f(x) = \sum_{j=1}^N a_j \chi_{E_j}(x)$ with the E_j pairwise disjoint, then

$$F(f) = \sum_{j=1}^{N} a_j f(\chi_{E_j}) = \int_E f(x) g(x) \,\mathrm{d}\mu.$$

Take ψ to be a simple function of the form

$$\psi(x) = \sum_{j=1}^{\infty} a_j \chi_{A_j}(x),$$

where the A_j are pairwise disjoint, and let ψ_N be the N^{th} partial sum. Thus,

$$\begin{aligned} \|\psi - \psi_N\|_{L^p}^p &= \sum_{i=N+1}^{\infty} |a_j|^p \mu(A_j), \text{ and} \\ \|\psi\|_{L^p}^p &= \sum_{j=1}^{\infty} |a_j|^p \mu(A_j), \end{aligned}$$

and the latter is finite, since the series converges. Thus, $\|\psi - \psi_N\|_{L^p} \to 0$ as $N \to \infty$. Thus, F acts as $f \mapsto \int fg \,\mathrm{d}\mu$ for all $f \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$.

The next step is to show that $g \in L^q(\mathbb{R}^n, \mathrm{d}\mu)$. We'll approximate it by simple functions; let ψ_n be a pointwise non-decreasing sequence of simple functions such that $\psi_n^{1/q} \chi_{B(0,n)} \to |g|$, and let $\varphi_n = \psi_n^{1/p} \operatorname{sign}(g)$. Then, $\|\varphi_n\|_{L^p} = \int \psi_n \,\mathrm{d}\mu$, and

$$\begin{aligned} \int \psi_n \,\mathrm{d}\mu = \int \psi_n^{1/p} \psi_n^{1/q} \,\mathrm{d}\mu &= \int |\psi_n|^{1/q} |\varphi_n| \,\mathrm{d}\mu \\ &\le \int |g| \chi_{B(0,n)} |\varphi_n| \,\mathrm{d}\mu = \int g \chi_{B(0,n)} \varphi_n \,\mathrm{d}\mu \\ &= F(\varphi_n \chi_{B(0,n)}) \\ &\le \|F\| \|\varphi_n \chi_{B(0,n)}\|_p. \end{aligned}$$

Sadly, we ran out of time here, but will continue next lecture.

12. The Riesz Representation Theorem: 11/4/14

> *"I don't know, I'm just making this up, but Franz Riesz's brother was proving theorems that are much more fun."*

Last time, we defined *signed measures*, i.e. measures ν such that $\nu = \nu^+ - \nu^-$ for unsigned, mutually singular measures ν^+ and ν^-, such that at most one of ν^+ and ν^- takes on infinite values. (This wasn't the definition, but we were able to prove it.) Furthermore, if ν is a signed measure on X, then $X = A \cup B$, such that ν^+ is supported on A and ν^- is supported on B.

Then, we turned to the Riesz representation theorem, Theorem 11.7, which states that any bounded linear functional $F : L^p(\mathbb{R}^n, \mathrm{d}\mu) \to \mathbb{R}$, with $1 \le p < \infty$, is given by a $g \in \text{Ł}^q(\mathbb{R}^n, \mathrm{d}\mu)$, where $1/p + 1/q = 1$, i.e. $L(f) = \int fg \,\mathrm{d}\mu$ for all f.

Continuation of the proof of Theorem 11.7. First, we assumed that μ is a finite Radon measure, so that $\chi_E \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$ for any μ-measurable set E. Then, we defined $\nu(E) = F(\chi_E)$; so we want to show that $F(\chi_E) = \int_E g \,\mathrm{d}\mu$, so that g would be equal to $\mathcal{D}_\mu \nu$.

Note that ν is a signed measure, which is fairly easy to check (countable additivity, for example, follows from F being a continuous function), and

$$|\nu(E)| = |F(\chi_E)| \leq \|F\| \|\chi_E\|_p = \|F\|(\mu(E))^{1/p},$$

i.e. ν is absolutely continuous with respect to μ. Thus, there's a measurable g such that $\nu(E) = \int g \, \mathrm{d}\mu$, $\mathcal{D}_\mu \nu^+ = g^+$, and $\mathcal{D}_\mu \nu^- = g^-$.

Since $|\nu(E)| \leq \|F\|(\mu(\mathbb{R}^n))^{1/p}$ (which makes sense because μ is finite), then $\nu^+(\mathbb{R}^n) + \nu^-(\mathbb{R}^n) \leq \|F\|(\mu(\mathbb{R}^n))^{1/p}$, i.e. $\|g\|_{L^1} \leq \|F\|(\mu(\mathbb{R}^n))^{1/p}$. Thus, $g \in L^1(\mathrm{d}\mu)$, which is nice, but we need to bootstrap this to showing $g \in L^q(\mathbb{R}^n, \mathrm{d}\mu)$ and that $\|g\|_{L^q} = \|F\|$.

So far, we know that $F(\psi) = \int \psi g \, \mathrm{d}\mu$ for any simple function ψ which takes on finitely many values. We want to do calculations in L^q to show that $\|F\| = \|g\|_{L^q}$, but since we don't know that $g \in L^q(\mathrm{d}\mu)$, we can't do that yet, so we'll have to approximate g. Specifically, approximate $|g|^q$ by simple functions ψ_n as follows:

$$\psi_n(x) = \begin{cases} \dfrac{j}{2^n}, & \text{if } |x| \leq n \text{ and } \dfrac{j}{2^n} \leq |g|^q < \dfrac{j+1}{2^n}, 0 \leq j \leq 2^{2n-1} \\ 0, & \text{if } |x| \geq n \text{ or } |g|^q \geq 2^n. \end{cases}$$

Thus, $\psi_n^{1/q}$ approach g from below, but they're compactly supported, simple functions taking on finitely many values.

$\psi_n(x)$ lies in all $L^p(\mathbb{R}^n, \mathrm{d}\mu)$, with $1 \leq p \leq \infty$, so set $\varphi_n = \psi_n^{1/p} \operatorname{sign}(g)$. Then, φ_n is also a simple function taking on finitely many values and

$$\|\varphi_n\|_{L^p} = \left(\int \psi_n \, \mathrm{d}\mu \right)^{1/p}.$$

Thus,

$$\begin{aligned} \int \psi_n \, \mathrm{d}\mu = \int \psi_n^{1/p + 1/q} \, \mathrm{d}\mu &= \int |\psi_n|^{1/q} |\varphi_n| \, \mathrm{d}\mu \\ &\leq \int |g| |\varphi_n| \, \mathrm{d}\mu = \int g \varphi_n \\ &\leq \|F\| \|\varphi_n\|_p = \|F\| \left(\int \psi_n \, \mathrm{d}\mu \right)^{1/p}. \end{aligned}$$

Thus, $\int \psi_n \, \mathrm{d}\mu \leq \|F\|^q$, so $\int |g|^q \, \mathrm{d}\mu \leq \|F\|^q$, so $g \in L^q(\mathrm{d}\mu)$ and $\|g\|_{L^q} \leq \|F\|$.

Since $g \in L^q(\mathrm{d}\mu)$, then the linear functional $G(f) = \int fg \, \mathrm{d}\mu$ is bounded on L^p by $|G(f)| \leq \|g\|_q \|f\|_p$, and $G(\psi) = F(\psi)$ for all simple functions which take on finitely many values. Since these are dense in L^p when $1 \leq p < \infty$,[17] then $G(f) = F(f)$ on a dense set and therefore for all $f \in L^p$.

If μ is not finite and $1 < p < \infty$, then look at $F_R(f) = F(f\chi_R)$, where $\chi_R(x) = \chi_{B(0,R)}(x)$. The above argument applied to $\mu_R = \mu|_{B(0,R)}$ shows that $F_R(f) = \int g_R f \, \mathrm{d}\mu$, where $g_R(x) = g(x)\chi_R(x)$.

Additionally, $\|f\chi_R - f\|_{L^p} \to 0$ as $R \to +\infty$ and $g_R(x) = g_{R'}(x)$ if $R > R'$ and $|x| < R'$, so there exists a $g(x)$ such that $g_R(x) = g(x)$ for $|x| \leq R$. Since $\|g_R\|_{L^q} \leq \|F_R\|$, then $\|g_R\|_{L^q} \leq \|F\|$, and we know that

$$|F_R(f)| = |F(f\chi_R)| \leq \|F\| \|f\chi_R\|_p \leq \|F\| \|f\|_p,$$

and therefore $\|g\|_{L^q} \leq \|F\|$.

Next,

$$F(f) = \lim_{R \to +\infty} F(f\chi_R) = \lim_{R \to +\infty} \int fg\chi_R \, \mathrm{d}\mu = \int fg \, \mathrm{d}\mu,$$

because $F \in L^p$ and $g \in L^q$. Thus, by Hölder's inequality, $|F(f)| \leq \|f\|_p \|g\|_q$, so $\|F\| \leq \|g\|_{L^q}$, so $\|F\| = \|g\|_{L^q}$.

Now we're done in the case $p > 1$. If $p = 1$ and μ is a finite measure, then $F(\psi) = \int \psi g \, \mathrm{d}\mu$ for any simple ψ that takes on only finitely many values, and therefore for any measurable E,

$$\left| \int_E g \, \mathrm{d}\mu \right| \leq \|F\| \mu(E). \tag{2}$$

Take $E = \{x : g(x) \geq \|F\| + \varepsilon\}$; then, (2) implies that $\mu(E) = 0$, and the same argument works to show that $E' = \{x : g(x) \leq -\|F\| - \varepsilon\}$ has measure zero. Thus, $\|g\|_{L^\infty} \leq \|F\|$. Now, we can show that $F(f) = \int fg \, \mathrm{d}\mu$ for all $f \in L^1(\mathrm{d}\mu)$, so $\|F\| \leq \|g\|_\infty$, and thus the two are equal. Then, generalizing to μ having infinite measure is exactly the same. ⊠

[17] Pay particular attention to this statement, because it's the reason the Riesz Representation Theorem doesn't hold in L^∞, though when we set $\varphi_n = \psi_n^{1/p} \operatorname{sign}(g)$, the calculations made with φ also require $p > 1$.

For L^∞, it used to be top secret, but nowadays anybody can go on Wikipedia and learn that the dual space to L^∞ is the space of finitely additive measures, which is pretty weird.

Instead, we'll talk about the Riesz Representation Theorem for compactly supported continuous functions, denoted $C_c(\mathbb{R}^n, \mathbb{R}^m)$. The reason we would do this is because the trick in the proof is to approximate by simple functions, which are dense in L^p. They're not dense in $C_c(\mathbb{R}^n, \mathbb{R}^m)$, since they're not even members of that space, but they can still approximate continuous functions arbitrarily well.

$C_c(\mathbb{R}^n, \mathbb{R}^m)$ is a nice linear space (i.e. vector space), but it's not complete, which is a little annoying.

Theorem 12.1. *Let $L : C_c(\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}$ be a linear functional such that for any compact set $K \subseteq \mathbb{R}^n$,*

$$\sup\{L(f) : f \in C_c(\mathbb{R}^n, \mathbb{R}^m), |f| \leq 1, \operatorname{supp} f \subseteq K\}$$

is finite. Then, there exists a Radon measure μ and a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ such that $|\sigma| = 1$ μ-almost everywhere and for any $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$, $L(f) = \int f\sigma \, \mathrm{d}\mu$.

Proof. Notice that when $m = 1$, this implies that $\sigma = \pm 1$ and therefore $L(f) = \int f \, \mathrm{d}\nu$ for a signed measure ν. Specifically, in this case, we want the signed measure to be $\nu(E) = L(\chi_E)$, but this can't be done yet, because $\chi_E \notin C_c(\mathbb{R}^n, \mathbb{R}^m)$. Since this fails, we'll define for any open set V

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

Thus, for any set A, $\mu^*(A) = \inf\{\mu^*(V) : V \text{ open }, A \subseteq V\}$.

Now, we need to show that this μ and some σ work.

(1) First, we'll show that μ is a Radon measure, which is relatively straightforward.
(2) Then, for any $f \in C_c^+(\mathbb{R}^n, \mathbb{R}^m)$, let
$$\lambda(f) = \sup\{L(g) : g \in C_c(\mathbb{R}^n, \mathbb{R}^m), |g| \leq f\}.$$
We'll need to show that λ is a bounded linear functional and $\lambda(f) = \int f \, \mathrm{d}\mu$.
(3) Finally, to get σ, let $\lambda_e(f) = L(fe)$, with $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$; then, we'll see that $\lambda_e(f) = \int f\sigma_e \, \mathrm{d}\mu$, so let $\sigma = \sum \sigma_{e_j} e_j$ for some choice of e_j we'll explain when we get to this point in the proof.

For part 1, take any open set V and open sets V_j such that

$$V \subseteq \bigcup_{j=1}^\infty V_j.$$

Then, choose a $g \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ such that $\|g\| \leq 1$; let $K_g = \operatorname{supp} g \subseteq V$. Since K_g is a compact set, then $K_g \subseteq \bigcup_{j=1}^R V_j$.

Exercise 11. Show that there exists a partition of unity for g, i.e. ζ_j for $1 \leq j \leq k$ such that:

(1) $\operatorname{supp} \zeta_j \subseteq V_j$, $\sum \zeta_j = 1$ on K_g,
(2)
$$\sum_{j=1}^k \zeta_j = 1 \text{ on } K_g,$$
(3)
$$g = \sum_{j=1}^k g\zeta_j,$$
(4) and $|g\zeta_j| \leq 1$ on V_j.

Armed with this partition of unity, we see that

$$|L(g)| \leq \sum_{j=1}^k |L(g\zeta_j)| \leq \sum_{j=1}^k \mu^*(V_j) \leq \sum_{j=1}^\infty \mu^*(V_j).$$

Thus,

$$\mu^*(V) \leq \sum_{j=1}^\infty \mu^*(V_j).$$

Now, for any $A \subseteq_{j=1}^\infty A_j$, with A and A_j not necessarily open, choose open $V_j \supset A_j$ such that $\mu^*(A_j) \geq \mu^*(V_j) - \varepsilon/2^j$. Thus, approximating the A_j by the V_j shows that countable subadditivity holds for all sets.

To go further, we'll need a purely measure-theoretic lemma, which we'll state now and prove later.

Lemma 12.2 (The Carathéodory Criterion)**.** *Let μ be a measure. If $\mu^*(A \cup B) = \mu * (A) + \mu^*(B)$ for all sets A and B such that* $\operatorname{dist}(A, B) > 0$*, then μ is a Borel measure.*

First, we should check that this criterion actually applies to μ. Let U_1 and U_2 be open sets such that $\operatorname{dist}(U_1, U_2) > 0$; then, $\mu^*(U_1 \cup U_2) = \mu^*(U_1) + \mu^*(U_2)$, so for any A_1 and A_2 such that $\operatorname{dist}(A_1, A_2) > 0$, take open sets $V_1 \supseteq A_1$ and $V_2 \supseteq A_2$ such that $\operatorname{dist}(V_1, V_2) > 0$. Choose $V \supseteq A_1 \cup A_2$ and let $U_1 = V \cap V_1$ and $U_2 = V \cap V_2$. Thus,

$$\mu^*(V) \geq \mu^*(U - 1 \cup U_2) = \mu^*(U_1) + \mu^*(U_2) \geq \mu^*(A_1) + \mu^*(A_2).$$

Thus, $\mu^*(A_1 \cap A_2) = \mu^*(A_1) + \mu^*(A_2)$, so Lemma 12.2 applies, and μ is a Borel measure.

For part 2, suppose $f \in C_c^+(\mathbb{R}^n)$ and define λ as above. First, we want to show that λ is linear: if $f_1, f_2 \in C_c(\mathbb{R}^n)$, choose g_1 and g_2 such that $|g_1| \leq f_1$ and $|g_2| \leq f_2$. Then, let $g_1' = g_1 \operatorname{sign}(L(g_1))$ and $g_2' = g_2 \operatorname{sign}(L(g_2))$.

Thus, $|g_1' + g_2'| \leq f_1 + f_2$, so $|L(g_1)| < |L(g_2)| = L(g_1') + L(g_2')$, and in particular $L(g_1' + g_2') \leq \lambda(f_1 + f_2)$, so $\lambda(f_1) + \lambda(f_2) \leq \lambda(f_1 + f_2)$ (i.e. it's *superlinear*).

To show that it's also *sublinear*, take a $g \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ such that $|g| \leq f_1 + f_2$, and let

$$g_1 = \begin{cases} f_1 g/(f_1 + f_2), & f_1(x) + f_2(x) > 0 \\ 0, & f_1 + f_2 = 0, \end{cases}$$

and define g_2 in the analogous way. Then, $|g_1| \leq f_1$, $|g_2| \leq f_2$, and $g_1, g_2 \in C_c(\mathbb{R}^n, \mathbb{R}^m)$, so $|L(g)| \leq |L(g_1)| + |L(g_2)| \leq \lambda(f_1) + \lambda(f_2)$. Thus, $\lambda(f_1 + f_2) \leq \lambda(f_1) + \lambda(f_2)$.

We'll finish the proof on Thursday.

13. The Riesz Representation Theorem for $C_c(\mathbb{R}^n, \mathbb{R}^m)$: 11/6/14

> *"The Fourier transform on the circle is like playing the piano, but the Fourier transform for the whole space is like a modern symphony. There are a finite number of strings on the piano, but composers can do more or less what they want, and most functions can be approximated by those 88 trigonometric polynomials. On the line, though, we have more frequencies, and so we can approximate any function, and if you listen to modern music, you've probably noticed every function being approximated."*

Recall that last time we proved the Riesz Representation Theorem (Theorem 11.7) for L^p, $1 \leq p < \infty$, i.e. that if $f : L^p \to \mathbb{R}$ is a bounded linear functional, then there exists a $g \in L^q(\mathbb{R}^n, \mathrm{d}\mu)$ with $1/p + 1/q = 1$ such that $L(f) = \int fg \,\mathrm{d}\mu$ for all $f \in L^p(\mathbb{R}^n, \mathrm{d}\mu)$.

For the L^∞ case, we are proving another theorem (Theorem 12.1), that if $L : C_c(\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}$ is a linear functional such that $\sup\{|L(g)| : g \in C_c(\mathbb{R}^n, \mathbb{R}^m), \operatorname{supp} g \in K\}$ is finite on compact sets $K \subseteq \mathbb{R}^n$, then there exists a Borel measure μ and μ-measurable function σ such that $|\sigma| = 1$ μ-almost everywhere, such that $L(f) = \int f \cdot \sigma \,\mathrm{d}\mu$ for all $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$.

Continuation of the proof of Theorem 12.1. First, we defined the *variation measure* on open sets V as

$$\mu(V) = \sup\{|L(g)| : \operatorname{supp} g \subseteq V, |g| \leq 1\}$$

and then extending as an outer measure μ^* everywhere else. We have yet to show that μ is Borel.

Then, we defined a functional λ, and showed that it is linear. We want to show that $\lambda(f) = \int f \,\mathrm{d}\mu$. The idea is this would be true for characteristic functions, but they're not in $C_c(\mathbb{R}^n, \mathbb{R}^m)$, even though they well approximate these functions.

Choose a partition $0 = t_0 < t_1 < \cdots < t_N = 2\|f\|_\infty$ such that $t_j - t_{j-1} < \varepsilon$ for each j and $\mu(\{f^{-1}(t_j)\}) = 0$. Let $V_j = f^{-1}(t_{j-1}, t_j)$, so that these are open and bounded sets, so $\mu(V_j)$ is finite. We'll choose h_j such that $\operatorname{supp} h_j \subset V_j$ and $\mu(V_j) - \varepsilon/N \leq \lambda(h_j) \leq \mu(U_j)$; specifically, to do that, choose a compact K_j such that $\mu(U_j \setminus K) < \varepsilon/N$ and choose a g_j such that $L(g_j) \geq \mu(U_j) - \varepsilon/N$ and $\operatorname{supp} g_j \subseteq U_j$. We also have that $\|g_j\| = 1$.

Then, take h_j such that $h_j = 1$ on $K_j \cup \operatorname{supp} g_j$ and $\operatorname{supp} h_j \subseteq U_j$, and such that $0 \leq h_j \leq 1$ (akin to a partition of unity). Then,

$$\mu(U_j) \geq \lambda(h_j) \geq L(g_j) \geq \mu(U_j) - \frac{\varepsilon}{N}.$$

Now, consider the set

$$A = \left\{ f(x) \left(1 - \sum_{j=1}^N h_j \right) > 0 \right\}.$$

Then,

$$\mu(A) \leq \sum_{j=1}^N \mu(U_j - K_j) \leq N \frac{\varepsilon}{N} = \varepsilon.$$

Now, let's evaluate λ:

$$\begin{aligned}\lambda\left(f - f\sum_{j=1}^{N} h_j\right) &= \sup\left\{|L(g)| : |g| \le f - f\sum_{j=1}^{N} h_j\right\} \\ &= \sup\{|L(g)| : |g| \le \|f\|_\infty \chi_A\} \\ &= \|f\|_\infty \mu(A) \le \varepsilon \|f\|_\infty.\end{aligned}$$

Thus, as we let $\varepsilon \to 0$, we can approximate

$$\begin{aligned}\lambda(f) &\approx \lambda\left(f \sum h_j\right) = \sum \lambda(f h_j) \\ &\approx \sum t_j \lambda(h_j) \approx \sum t_j \mu(U_j) \\ &\approx \int f \, \mathrm{d}\mu.\end{aligned}$$

That is, we approximated f by piecewise constant functions, which made calculating λ nicer. This wasn't rigorous, but it ca be made so by sprinking around some epsilons.

For the next step in this proof, take an $f \in C_c(\mathbb{R}^n)$ and fix an $e \in S^{n-1}$ (i.e. $|e| = 1$). Set $\lambda_e(f) = L(fe)$. Thus,

$$\begin{aligned}\lambda_e(f) &\le \sup\{|L(g)|, g \in C_c(\mathbb{R}^n, \mathbb{R}^m), |g| \le |f|\} \\ &= \lambda(|f|) = \int |f| \, \mathrm{d}\mu.\end{aligned}$$

Thus, λ_e can be extended to a bounded linear functional on $L^1(\mathbb{R}^n, \mathrm{d}\mu)$, because $C_c(\mathbb{R}^n)$ is dense in L^1.[18] Thus,[19] there exists a σ_e such that $\lambda_e(f) = \int (f \cdot \sigma_e) \, \mathrm{d}\mu$. Now, let $\{e_j\}$ be an orthonormal basis for $\mathbb{R}^m$ and

$$\sigma = \sum_{j=1}^{m} \sigma_{e_j} e_j.$$

This means that

$$\begin{aligned}L(f) &= L\left(\sum_{j=1}^{m} (f \cdot e_j) e_j\right) \\ &= \sum_{j=1}^{m} \lambda_{e_j}(f \cdot e_j) \\ &= \sum_{j=1}^{m} \int (f \cdot e_j) \sigma_{e_j} \, \mathrm{d}\mu = \int f \left(\sum_{j=1}^{m} \sigma_{e_j} e_j\right) \mathrm{d}\mu \\ &= \int (f \cdot \sigma) \, \mathrm{d}\mu.\end{aligned}$$

Now we need to calculate $|\sigma|$:

$$|\lambda_e(f)| = \left|\int f \sigma_e \, \mathrm{d}\mu\right| \le \int |f| \, \mathrm{d}\mu,$$

so $\|\lambda_e\| = \|\sigma_e\|_\infty \le 1$, so $\|\sigma\| \le m$.

Now, we'll prove that $|\sigma| = 1$ μ-almost everywhere. Let U be an open set and $\sigma' = \sigma/|\sigma|$ when $\sigma \ne 0$, and $\sigma' = 0$ if $\sigma = 0$. Choose a compact $K \subseteq U$ such that $\mu(U \setminus K_j) \le 1/j$ and σ' is continuous on K. Thus, this extends σ' to a continuous function f_j on $\mathbb{R}^n$ such that $|f_j| \le 1$.

Now, take a cutoff function h_j such that $0 \le h_j \le 1$, $h_j = 1$ on K_j and $|supph_j \subseteq U$. Define $g_j = h_j f_j$, so that $|g_j| \le 1$, $\operatorname{supp} g_j \subset U$, and $g_j \cdot \sigma \to |\sigma|$ in probability, since $g_j \cdot \sigma = |\sigma|$ on K_j.

But convergence in probability implies there exists a subsequence $j_k \to \infty$ such that $g_{j_k} \to |\sigma|$ μ-almost everywhere. Thus, by the Bounded Convergence Theorem,

$$\int_U |\sigma| \, \mathrm{d}\mu = \lim_{k \to +\infty} \int (g_{j_k} \cdot \sigma) \, \mathrm{d}\mu = \lim_{k \to +\infty} L(g_{j_k}) \le \mu(U).$$

[18] This is true for every L^p with p finite, but not L^∞.

[19] Technically, we didn't prove this result for vector-valued functions, but it still holds, just with a little more work.

If $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$, then $L(f) = \int (f \cdot \sigma)\,\mathrm{d}\mu$, so if $\operatorname{supp} f \subset I$ and $|f| \le 1$, then $|L(f)| \le \int_U |\sigma|\,\mathrm{d}\mu$. Thus, $\mu(U) \le \int_U |\sigma|\,\mathrm{d}\mu$, and so $\mu(U) = \int_U |\sigma|\,\mathrm{d}\mu$ for all open sets U with finite μ-measure, and so $|\sigma| = 1$ μ-almost everywhere. ⊠

Here is a very natural break in the class material, since we'll be switching to the Fourier transform soon. If a fire alarm had to go off today, we might as well put it here. But it didn't. In any case, we'll prove the Carathéodory criterion, Theorem 12.2, though we really could have done so back in September.

Proof of Theorem 12.2. Take a closed set C and any set A, so we need to show that $\mu^*(A) \ge \mu^*(A \setminus C) + \mu^*(A \cap C)$. Assume $\mu^*(A)$ is finite (if not, then we're done), and let

$$C_m = \left\{ x \in \mathbb{R} : \operatorname{dist}(x, C) \le \frac{1}{m} \right\},$$

which is sort a tube around C. Now,

$$\mu^*(A \setminus C_n) + \mu^*(A \cap C) = \mu^*((A \setminus C_n) \cup (A \cap C)) \le \mu^*(A).$$

Now, we want that $\lim_{n\to\infty} \mu(A \setminus C_n) = \mu(A \setminus C)$, so let's look at the annuli

$$R_k = \left\{ x \in A : \frac{1}{k+1} < \operatorname{dist}(x, C) \le \frac{1}{k} \right\}.$$

The idea is, if we look only at the even layers or at the odd layers, things work nicely.

$$\sum_{k=1}^{\infty} \mu^*(R_{2k}) = \mu^*\left(\bigcup_{k=1}^{\infty} R_{2k} \right) \le \mu^*(A)$$

In the same way,

$$\sum_{k=1}^{\infty} \mu^*(R_{2k-1}) \le \mu^*(A).$$

Now, write

$$A \setminus C = (A \setminus C_n) \cup \left(\bigcup_{k=n}^{\infty} R_k \right),$$

and therefore

$$\begin{aligned} \mu^*(A \setminus C_n) &\le \mu^*(A \setminus C) \\ &\le \mu^*(A \setminus C_n) + \sum_{k=n}^{\infty} \mu^*(R_k), \end{aligned}$$

so $\mu^*(A \setminus C_n) \to \mu^*(C)$, since the sum of the measures of the R_k is at most $2\mu^*(A)$, which is finite. ⊠

Part 2. The Fourier Transform

First, we will consider the Fourier transform on a circle, and then on the whole space.[20]

Definition. The *Fourier transform* of a 2π-periodic, integrable function (i.e. a function on the sphere) f is

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

Then, $|\widehat{f}(k)| \le \|f\|_{L^1}$, and therefore $\mathcal{F} : L^1(S^1) \to L^\infty(\mathbb{Z})$.

The following lemma, that the Fourier coefficients decay, is reasonably immediate, though for some reason follows the tradition that the most important results in mathematics are so often called lemmas. If f is Riemann integrable, there's a very nice proof using only elementary calculus, but we'll have to be more creative.

Lemma 13.1 (Riemann-Lesbegue)**.** *If $f \in L^1(S^1)$, then* $\lim_{k\to\infty} \widehat{f}(k) = 0$.

[20]The fire alarm went off right about now. Talk about convenient timing.

Proof.

$$\begin{aligned}\widehat{f}(k) &= \int_0^1 f(x)e^{-2\pi ik(x+1/2k)}\,\mathrm{d}x = -\int_0^1 f\left(x-\frac{1}{2k}\right)e^{-2\pi ikx}\,\mathrm{d}x \\ &= \frac{1}{2}\int_0^1\left(f(x)-f\left(x-\frac{1}{2k}\right)\right)e^{-2\pi ikx}\,\mathrm{d}x.\end{aligned}$$

Thus

$$|\widehat{f}(k)| \le \frac{1}{2}\int_0^1\left|f(x)-\left(x+\frac{1}{2k}\right)\right|\mathrm{d}x,$$

and we showed on a problem set this goes to 0 as $k\to\infty$. ☒

Corollary 13.2. *If $f\in L^1(S^1)$,*

$$\lim_{m\to\infty}\int_0^1 f(x)\sin(\pi mx)\,\mathrm{d}x = 0.$$

This leads to a couple questions.

(1) When is the following true?

$$f(x) = \sum_{k=-\infty}^{\infty}\widehat{f}(k)e^{-2\pi ikx}.$$

(2) The set $\{e^{2\pi ikx}\}$ is orthonormal in $L^1(S^1)$. Is it a basis for $L^1(S^1)$?
(3) When does convergence hold pointwise or in some other sense?

To clean up the notation, let

$$S_Nf(x) = \sum_{k=-N}^{N}\widehat{f}(k)e^{-2\pi ikx}.$$

Thus, this looks like a convolution with a kernel, which is pretty nicely behaved:

$$S_Nf(x) = \sum_{k=-N}^{N}\left(\int_0^1 f(y)e^{-2\pi iky}\,\mathrm{d}y\right)e^{2\pi ikx} = \int_0^1 f(t)D_N(x-t)\,\mathrm{d}t,$$

where the kernel is

$$\begin{aligned}D_N(t) &= \sum_{k=-N}^{N} e^{2\pi ikt} = e^{-2\pi iNt}\sum_{k=0}^{2N}e^{2\pi ikt} \\ &= e^{-2\pi iNt}\frac{1-e^{2\pi i(2N+1)t}}{1-e^{2\pi it}} \\ &= \frac{e^{2\pi i(N+1)t}-e^{-2\pi iNt}}{e^{2\pi it}-1} \\ &= \frac{e^{-2\pi i(N+1/2)t}-e^{-2\pi i(N+1/2)t}}{e^{\pi it}-e^{-\pi it}} \\ &= \frac{\sin((2N+1)\pi t)}{\sin\pi t}.\end{aligned}$$

This kernel D_N is called the *Dini kernel*. Lots of Italians in this branch of mathematics! Let's restate some facts:

$$\begin{aligned}S_Nf(x) &= \int_0^1 D_n(x-t)f(t)\,\mathrm{d}t \\ D_N(t) &= \frac{\sin((2N+1)\pi t)}{\sin(\pi t)}. \\ \int_0^1 & D_N(t)\,\mathrm{d}t = 1.\end{aligned}$$

Furthermore, if $|t|\ge\delta$, then $|D_N(t)|\le 1/\sin(\pi\delta)$.

Next, let's move the bounds a bit.

$$L_N = \int_{-1/2}^{1/2} |D_N(t)|\,\mathrm{d}t = 2\int_0^{1/2} \frac{|\sin((2N+1)\pi t)|}{|\sin(\pi t)|}\,\mathrm{d}t$$
$$\geq 2\int_0^{1/2} \frac{|\sin((2N+1)\pi t)|}{\pi t}\,\mathrm{d}t - 2\int_0^1 |\sin((2N+1)\pi t)|\left|\frac{1}{\pi t} - \frac{1}{\sin \pi t}\right|\mathrm{d}t.$$

The second term is bounded above by a constant, and the first part can be bounded below.

$$\int_0^{1/2} \frac{|\sin((2N+1)\pi t)|}{\pi t}\,\mathrm{d}t = \int_0^{(N+1/2)\pi} \frac{\sin t}{\pi t}\,\mathrm{d}t \geq C\sum_{k=1}^{N+1/2} \frac{1}{k} \geq c\log N,$$

and thus $\|D_N\|_{L^1} \geq c\log N$ for some constant c.

14. The Fourier Transform and Convergence Criteria: 11/13/14

"Steinhaus later told Ulam that the happiest day of his life was the day the Germans retreated from Lvov and the Russians hadn't gotten there yet. But I am not sure if such a day existed."

Recall that we defined the Fourier transform for $f \in L^1(S^1)$ as

$$\widehat{f}_k = \int_0^1 e^{-2\pi ikx} f(x)\,\mathrm{d}x$$
$$S_N f(x) = \sum_{k=-N}^{N} \widehat{f}_k e^{2\pi ikx}.$$

We want to know when it's true that $S_N f(x) \to f(x)$, and proved two results: the Riemann-Lesebgue lemma that $\widehat{f}_k \to 0$, and that

$$S_N f(x) = \int_{-1/2}^{1/2} D_N(x-t) f(t)\,\mathrm{d}t,$$

where $D_N(T) = \sin((2N+1)\pi t)/\sin(\pi t)$ is the *Dini kernel*, and $\int_{-1/2}^{1/2} D_N(t)\,\mathrm{d}t = 1$. Thus, S_N acts as convolution by this kernel, which is very useful for studying it.

If one uses this to obtain a bound, then

$$L_N = \int_{-1/2}^{1/2} |D_N(t)|\,\mathrm{d}t \longrightarrow +\infty$$

as $N \to \infty$, but it grows logarithmically, so unless there's some oscillation this won't converge.

Theorem 14.1 (Dini's criterion)**.** *If $f \in L^1(S^1)$ and there's a $\delta > 0$ such that*

$$\int_{-\delta}^{\delta} \left|\frac{f(x+t) - f(x)}{t}\right| \mathrm{d}t \tag{3}$$

is finite, then $S_N f(x) \to f(x)$ as $N \to \infty$.

Proof. The idea is that the condition on (3) imposes regularity conditions on $S_N f(x) - f(x)$. More specifically,

$$S_N f(x) - f(x) = \int_{-1/2}^{1/2} (f(x-t) - f(x)) D_N(t)\,\mathrm{d}t = \int_{-1/2}^{1/2} \frac{f(x-t) - f(x)}{\sin \pi t} \sin((2N+1)\pi t)\,\mathrm{d}t$$
$$= \underbrace{\int_{|t|\leq\delta} \frac{f(x-t) - f(x)}{\sin \pi t} \sin((2N+1)\pi t)\,\mathrm{d}t}_{A_N} + \underbrace{\int_{|t|\geq\delta} \frac{f(x-t) - f(x)}{\sin \pi t} \sin((2N+1)\pi t)\,\mathrm{d}t}_{B_N}.$$

This is as far as we're able to go, so now use the assumption: $A_N = \int_{-1/2}^{1/2} \sin((2N+1)\pi t)\,\mathrm{d}t$, so if

$$g_x(t) = \frac{f(x-t) - f(x)}{\sin \pi t} \chi_{\{|t|\leq\delta\}}(t),$$

then $g_x \in L^1(\mathrm{d}t)$, so by the Riemann-Lesbegue lemma, $A_N \to 0$ as $N \to \infty$.

For B_N, we can write $B_N = \int_{-1/2}^{1/2} q_x(t) \sin((2N+1)\pi t)\,\mathrm{d}t$, where

$$q_x(t) = \frac{f(x-t) - f(x)}{\sin \pi t} \chi_{\{|t|\geq\delta\}},$$

and $q_x \in L^1(S^1, \mathrm{d}t)$, so the same argument works here. $\boxtimes$

Note that if f is Lipschitz or Hölder, this is automatically true.

Theorem 14.2 (Jordan)**.** *Let $f \in \mathrm{BV}([x-\delta, x+\delta])$; then, $S_N f(x) \to (1/2)(f(x^+) + f(x^-))$.*

Proof. We'll make the following assumptions.

(0) $f \in \mathrm{BV}(S^1)$, rather than just on intervals. This can be dealt with by choosing a δ as in the proof of Theorem 14.1 and dealing with the exceptions in the same way.

(1) Without loss of generality, we may assume that f is increasing, $x = 0$, and $f(0^+) = 0$ (by the theorems we had on functions of bounded variation).

However, since f has bounded variation, it's the difference of two monotonic functions, so $f(x^+)$ and $f(x^-)$ certainly exist. Then,

$$S_N f(0) = \int_{-1/2}^{1/2} D_N(-t) f(t)\,\mathrm{d}t = \int_0^{1/2} D_N(t)(f(-t) + f(t))\,\mathrm{d}t.$$

We will show that $\int_0^{1/2} D_N(t) f(t)\,\mathrm{d}t \to 0$.

Choose a $\delta > 0$ such that $0 \le f(t) \le \varepsilon$ for $t \in (0, \delta)$, so we can write

$$\int_0^{1/2} f(t) D_N(t)\,\mathrm{d}t = \underbrace{\int_0^{\delta} f(t) D_N(t)\,\mathrm{d}t}_{I_N} + \underbrace{\int_{\delta}^{1/2} f(t) D_N(t)\,\mathrm{d}t}_{II_N}.$$

Then,

$$II_N = \int_0^{1/2} \frac{f(t)}{\sin \pi t} \chi_{\{|t| \ge \delta\}}(t) \sin((2N_1)\pi t)\,\mathrm{d}t,$$

which goes to 0 as $N \to \infty$ for a fixed δ, as this function is in $L^1(S^1)$.

I_N doesn't converge quite as easily:

$$I_N = \int_0^{\delta} f(t) D_N(t)\,\mathrm{d}t \le \varepsilon \int_0^{\delta} |D_N(t)|\,\mathrm{d}t \approx \varepsilon \log N.$$

We can approximate D_N with another function of the same order: if $\alpha, \beta \in [0, 1/2]$, then

$$\int_{\alpha}^{\beta} \frac{\sin((2N+1)\pi t)}{\sin \pi t}\,\mathrm{d}t \approx \int_{\alpha}^{\beta} \frac{\sin((2N+1)\pi t)}{\pi t}\,\mathrm{d}t.$$

But we can bound this puppy.

$$\int_{\alpha}^{\beta} \frac{\sin((2N+1)\pi t)}{\pi t}\,\mathrm{d}t = \int_{(2N+1)\pi\alpha}^{(2N+1)\pi\beta} \frac{\sin t}{\pi t}\,\mathrm{d}t,$$

which is bounded above by some M, which can be shown by integration by parts; it's a standard calculus exercise.

If h is an increasing function, we can approximate it by underestimating it for a while and then overestimating. More precisely, there exists a $c \in (a, b)$ such that

$$\int_a^b h(y) g(y)\,\mathrm{d}y = h(a^+) \int_a^c g(y)\,\mathrm{d}y + h(b^-) \int_c^b g(y)\,\mathrm{d}y.$$

Thus, applying this to I_N,

$$I_N = f(0^+) \int_0^c D_N(t)\,\mathrm{d}t + f(\delta^-) \int_c^{\delta} D_N(t)\,\mathrm{d}t = f(\delta^-) \int_c^{\delta} D_N(t)\,\mathrm{d}t.$$

Thus, $|I_N| \le \varepsilon M$, so $I_N \to 0$. $\boxtimes$

The last elementary property we'll prove is a localization principle. This is a bit of a surprise, because the Fourier series is at least on first glance a very global phenomenon.

Theorem 14.3 (Localization principle)**.** *If $f(x) = 0$ in $(x - \delta, x + \delta)$ for some $f \in L^1(S^1)$, then $S_N f(x) \to 0$ as $N \to \infty$.*

Proof. We can write

$$S_N f(x) = \int_{-1/2}^{1/2} D_N(t) f(x-t)\,\mathrm{d}t = \int_{|t| \ge \delta} \sin((2N+1)\pi t) \frac{f(x-t)}{\sin \pi t} \chi_{\{|t| \ge \delta\}}(t)\,\mathrm{d}t,$$

so we can apply the Riemann-Lesbegue lemma. $\boxtimes$

For a while in the 19th Century, Fourier transforms were a large focus of activity. One famous open problem was to prove that the Fourier series of a continuous function converges. This ended up not being true, proven in 1873 by du Bois-Raymond. A lot of this was formalized in Poland in the 1930s, as part of the Lvov School of Mathematics, e.g. Banach, Steinhaus, Ulam, Schauder and so forth.[21] In order to get this example, we'll need to detour through the following example.

Theorem 14.4 (Banach-Steinhaus)**.** *Let X be a Banach space and Y a normed vector space. Let T_α be a family of bounded linear operators $T_\alpha : X \to Y$. Then, either $\sup_\alpha \|T_\alpha\|$ is finite, or there exists an $x \in X$ such that $\sup_\alpha \|T_\alpha x\|_Y = \infty$.*

Proof. Define $\varphi_\alpha(x) = \|T_\alpha x\|_Y$, which is a continuous function on X because $|\varphi_\alpha(x) - \varphi_\alpha(x')| \le \|T_\alpha(x - x')\|_Y \le \|T_\alpha\|\|x - x'\|$.

Set $\varphi(x) = \sup_\alpha \varphi_\alpha(x)$ and

$$V_n = \{\varphi(x) > n\} = \bigcup_\alpha \{x : \varphi_\alpha(x) > n\}.$$

Each V_n is therefore open. Additionally, either one of the V_n isn't dense in X, or they're all dense in X.

Suppose V_N isn't dense in X, so that there exists an x_0 and a $\rho > 0$ such that $B(x_0, \rho) \cap V_N = \emptyset$. Thus, $\varphi_\alpha(x) \le N$ for all $x \in B(x_0, \rho)$, and therefore $\|T_\alpha(x + x_0)\| \le N$ if $\|x\| \le \rho$, so if $x \in B(0, 2\rho)$, $\|T_\alpha(x)\| \le \|T_\alpha(x_0)\| + N \le 2N$, and in partiuclar for all α, $\|T_\alpha\| \le 2N/\rho$.

Instead, if all of the V_N are dense in X and are open, then $\overline{V} = \bigcap_n V_n$ is nonempty. Take an $x \in \overline{V}$, so $x \in V_n$ for all n. Thus, there exists an α_n such that $\|T_{\alpha_n} x\| \ge n$ for any n, and therefore the supremum is infinite. ⊠

The idea is that if things don't converge, it's not that they just oscillate; the norms have to be unbounded too. But like every proof using the Baer Category Theorem, it's beautifully simple and completely non-instructive. It takes much more time to find a continuous, everywhere non-differentiable function than to prove almost all continuous functions are non-differentiable! But we can show there's a continuous function whose Fourier series diverges.

Theorem 14.5. *There exists an $f \in C(S^1)$ such that the Fourier series of f diverges at 0.*

Proof. Define $T_N : C(S^1) \to \mathbb{C}$ by $T_N(f) = S_N f(0) = \int_{-1/2}^{1/2} f(t) D_N(t)\,\mathrm{d}t$. Then, $|T_N(f)| \le \|D_N\|$, which goes to ∞ as $N \to \infty$. Additionally, there exist $f_j \in C(S^1)$ such that $|f_j| \le 1$ and $f_j \to \operatorname{sign} D_N$ almost everywhere as $j \to \infty$ (in some sense, a limit of continuous functions that becomes a step function).

Now, we know that $|T_N(f_j)| \to \int_{-1/2}^{1/2} |D_N|\,\mathrm{d}t$, and therefore $\|T_N\| = \|D_N\|_1$, which is unbounded above. Thus, by Theorem 14.4, there exists an $f \in C(S^1)$ such that $\sup_N |T_N f(0)| = \infty$, so $S_N f(0)$ doesn't converge. ⊠

Approximation by Trigonometric Polynomials. If the Fourier series doesn't converge, we may be able to improve it by averaging.

Definition. Define the *Cesaro averages* of $S_N f$ to be

$$\sigma_N f(x) = \frac{1}{N+1} \sum_{j=0}^{N} S_j f(x) = \frac{1}{N+1} \int_{-1/2}^{1/2} f(x - t) \sum_{j=0}^{N} D_j(t)\,\mathrm{d}t.$$

Let's see if we can simplify this.

$$\begin{aligned}
\sum_{j=0}^{N} D_j(t) = \sum_{j=0}^{N} \frac{\sin((2j+1)\pi t)}{\sin \pi t} &= \frac{1}{\sin^2(\pi t)} \sum_{j=0}^{N} \sin((2j+1)\pi t) \sin \pi t \\
&= \frac{1}{2\sin^2(\pi t)} \sum_{j=0}^{N} \cos((2j+1)\pi t - \pi t) - \cos((2j+1)\pi t + \pi t) \\
&= \frac{1}{2\sin^2(\pi t)} \sum_{j=0}^{N} (\cos(2j\pi t) - \cos((2j+2)\pi t)) \\
&= \frac{\sin^2((N+1)\pi t)}{\sin^2(\pi t)}.
\end{aligned}$$

[21] Of course, the whole school was destroyed by the war; Steinhaus hid from the Nazis, for example. But he was able to get access to Polish newspapers and make impressively accurate calculations about the German war losses, and he survived the war. Ulam emigrated to the States and Banach wasn't Jewish, so he survived. Schauder ended up dying in a concentration camp.

Thus, the kernel we get is

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

which is called the *Fejér kernel*. This is much better behaved than the Dini kernel, because it's positive and more easily bounded. Specifically, for any $\delta > 0$, $F_N(t) \le C(\delta)/N$ for $|t| \ge \delta$ and some constant C, and since $\int D_N = 1$ and we're averaging it, then $\int_{-1/2}^{1/2} F_N(t)\,\mathrm{d}t = 1$ as well.

Now that we've done all this work, we can restate the Cesaro coefficients as

$$\sigma_N f(x) = \int_{-1/2}^{1/2} F_N(x-t) f(t)\,\mathrm{d}t.$$

Theorem 14.6.

(1) Let $f \in L^p(S^1)$ for $1 \le p < \infty$; then, $\|\sigma_N f - f\|_p \to 0$ as $N \to \infty$.
(2) If $f \in C(S^1)$, then $\sup_x |\sigma_N f(x) - f(x)| \to 0$ as $N \to \infty$.

Proof. We'll prove (1); the other case is pretty similar.

Write

$$\begin{aligned}\|\sigma_N f - f\|_p &\le \int_{-1/2}^{1/2} F_N(t) \|f(x-\cdot) - f(\cdot)\|_p\,\mathrm{d}t \\ &= \underbrace{\int_{|t|\le\delta} F_N(t)\|f(x-\cdot)-f(\cdot)\|_p\,\mathrm{d}t}_{I_N} + \underbrace{\int_{|t|\ge\delta} F_N(t)\|f(x-\cdot)-f(\cdot)\|_p\,\mathrm{d}t}_{II_N}.\end{aligned}$$

Now, given an $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|f(x-\cdot) - f(\cdot)\|_p \le \varepsilon$ for all $|x| < \delta$, and using this shows that $I_N, II_N \to 0$. ⊠

Corollary 14.7. *Trigonometric polynomials are dense in $L^p(S^1)$, $1 \le p < \infty$.*

This is because the Cesaro partial sums are trigonometric polynomials.

Corollary 14.8. *The Fourier transform is an isometry $L^2(S^1) \to \ell^2$, i.e.*

$$\sum_{k\in\mathbb{Z}} |\widehat{f}_k|^2 = \int_{-1/2}^{1/2} |f(x)|^2\,\mathrm{d}x.$$

Proof. $\{e^{2\pi i k x}\}$ are orthonormal and span a dense set, so they must be a basis for $L^2(S^1)$. Thus, by Corollary 14.7,

$$\|f\|_{L^2} = \sum_{k\in\mathbb{Z}} |f_k e^{2\pi i k x}|^2.$$ ⊠

Corollary 14.9. *If $f \in L^1(S^1)$ and $\widehat{f}(k) = 0$ for all $k \in \mathbb{Z}$, then $f = 0$.*

Next, let's discuss the ergodicity of irrational rotations of the circle. What are the invariant sets of $T_\alpha(e^{2\pi i x}) = e^{2\pi i(x+\alpha)}$, i.e. the sets $R \subset S^1$ such that $T_\alpha(R) = R$?

Claim. If $\alpha \notin \mathbb{Q}$ and R is a measurable, T_α-invariant set, then either $m(R) = 1$ or $m(R) = 0$.

Proof. Let R be an invariant set and $\chi_R(x)$ be its characteristic function, so that $\chi_R(x+\alpha) = \chi_R(x)$. Then, define $\chi_\alpha(x) = \chi_R(x+\alpha)$. Its Fourier coefficients are

$$\begin{aligned}\widehat{\chi}_\alpha(k) &= \int_0^1 e^{2\pi i k x} \chi_R(x+\alpha)\,\mathrm{d}x \\ &= e^{-2\pi i k\alpha} \int_0^1 e^{2\pi i k x}\chi_R(x)\,\mathrm{d}x \\ &= e^{-2\pi i k \alpha}\widehat{\chi}_{R,k}.\end{aligned}$$

But $\chi_\alpha = \chi_R$, so $\widehat{\chi}_\alpha(k) = \widehat{\chi}_R(k)$, and thus $e^{2\pi i k\alpha}\widehat{\chi}_R(k) = \widehat{\chi}_R(k)$. Thus, either $\widehat{\chi}_R = 0$, so $\chi_R = 0$ and $m(R) = 0$, or $\widehat{\chi}_R(0) \ne 0$, in which case $\chi_R(x) = \chi_R(0)$, so $m(R) = 1$. ⊠

There are other proofs of this, but the Fourier series-flavored one is among the cleanest.

"I should talk much less."

Definition. Given an $f \in L^1(\mathbb{R}^n)$, we can define the *Fourier transform* of f as

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

This is really more like the orchestra: there is a continuous set of frequencies, not just a discrete set. One can think of x as time and ξ as the frequency, though this is a bit weirder outside of $\mathbb{R}^1$.

Two facts are immediately apparent.

(1) $\|\widehat{f}\|_{L^\infty} \le \|f\|_{L^1}$.
(2)
$$\widehat{f}(\xi) - \widehat{f}(\xi') = \int_{\mathbb{R}^n} f(x) \left(e^{-2\pi i x \cdot \xi} - e^{-2\pi i x \cdot \xi'} \right) \mathrm{d}x,$$
which goes to 0 as $\xi \to \xi'$, and therefore $\widehat{f}$ is continuous.

Suppose $f \in C_c^\infty(\mathbb{R}^n)$ (i.e. bounded, continuous, compactly supported). Then, one can calculate that

$$\begin{aligned}
\frac{\partial}{\partial \xi_j} \widehat{f}(\xi) &= -2\pi i \widehat{x_j f} f(\xi) \\
\widehat{\frac{\partial f}{\partial x_j}}(\xi) &= 2\pi i \xi_j \widehat{f}(\xi).
\end{aligned}$$

See the professor's lecture notes for full derivations.

Definition. Let

$$p_{\alpha\beta}(\varphi) = \sup_{x \in \mathbb{R}^n} |x|^\alpha \left| \frac{D^\beta \varphi}{\partial x^\beta} \right| = \sup_{x \in \mathbb{R}^n} |x_1|^{\alpha_1} |x_2|^{\alpha_2} \cdots |x_n|^{\alpha_n} \left| \frac{\partial_{x_1}^{\beta_1} \partial_{x_2}^{\beta_2} \cdots \partial_{x_n}^{\beta_n} \varphi}{\partial x_1^{\beta_1} \cdots \partial x_n^{\beta_n}} \right|.$$

Then, $p_{\alpha\beta}$ is finite for all α and β iff $p_{\beta\alpha}$ is; a function φ with this property is said to be in the *Schwarz class* $\mathcal{S}(\mathbb{R}^n)$.

This is the class of smooth functions which decrease faster than polynomially and whose derivatives also decrease faster than polynomially.

One says that $\varphi_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ if $p_{\alpha\beta}(\varphi_k) \to 0$ for all α and β.

Proposition 15.1. *If $\varphi_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$, then $\varphi_k \to 0$ in $L^p(\mathbb{R}^n)$ for $1 \le p \le +\infty$.*

Proof.

$$\begin{aligned}
\int |\varphi|^m \, \mathrm{d}x &= \int_{|x| \le 1} |\varphi|^m + \int_{|x| \ge 1} |\varphi|^m \\
&\le C_n |p_{0,0}|^m + 2 \int_{|x| \ge 1} \frac{|x|^{n+1} |\varphi|^m}{1 + |x|^{n+1}} \, \mathrm{d}x \\
&\le C_n |p_{0,0}|^m + C_n' |p_{(n+1)/m,0}|^m.
\end{aligned}$$
⊠

Here's the main theorem, which is why we care about Schwarz functions.

Theorem 15.2. *The Fourier transform is a continuous map $\mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ such that for all $f, g \in \mathcal{S}(\mathbb{R}^n)$,*

$$\int_{\mathbb{R}^n} f(x) \widehat{g}(x) \, \mathrm{d}x = \int_{\mathbb{R}^n} \widehat{f}(x) g(x) \, \mathrm{d}x, \tag{4}$$

and

$$f(x) = \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi. \tag{5}$$

(5) is of particular note as the inversion formula for recovering f from $\widehat{f}$.

Proof. Continuity follows from the relationship between the Fourier transform, differentiation, and multiplication by x. The rest of the proof follows from the next lemma.

Lemma 15.3. *Let $f(x) = e^{-\pi |x|^2}$; then, $\widehat{f}(\xi) = f(\xi)$.*

Proof. One could do a five-page contour integration, but instead notice that

$$\widehat{f}(\xi) = \int_{\mathbb{R}^n} e^{-\pi|x|^2+2\pi i x\cdot\xi}\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{-\pi x_1^2+2\pi i x_1\xi_1}\,\mathrm{d}x_1 \cdot \int_{-\infty}^{\infty} e^{-\pi x_2^2+2\pi i x_2\xi_2}\,\mathrm{d}x_2 \cdots .$$

Notice that $f' + 2\pi x f = 0$, $f(0) = 1$, but also $\widehat{f}' + 2\pi\xi\widehat{f} = 0$ and $\widehat{f}(0) = 1$. Thus, f and $\widehat{f}$ are smooth functions satisfying the same ODE with the same initial condition, so they must be the same. ⊠

This lemma is vital in probability theory, since Gaussians are ubiquitous there.

Now, for (4), we see that

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

The proof of the inversion formula is somewhat magical and not very instructive. Let $\lambda > 0$. Then,

$$\begin{aligned}\int f(x)\widehat{g}(x)\,\mathrm{d}x &= \int f(x)g(\xi)e^{-2\pi i\lambda x\cdot\xi}\,\mathrm{d}\xi\,\mathrm{d}x = \int \widehat{f}(\lambda\xi)g(\xi)\,\mathrm{d}\xi\\ &= \frac{1}{\lambda^n}\int \widehat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)\mathrm{d}\xi.\end{aligned}$$

Thus,

$$\begin{aligned}\lambda^n \int f(x)\widehat{g}(\lambda x)\,\mathrm{d}x &= \int \widehat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)\mathrm{d}\xi\\ \implies \int f\left(\frac{x}{\lambda}\right)\widehat{g}(x)\,\mathrm{d}x &= \int \widehat{f}(\xi)g\left(\frac{\xi}{\lambda}\right)\mathrm{d}\xi.\end{aligned}$$

Let $\lambda \to \infty$; then,

$$f(0)\int \widehat{g}(x)\,\mathrm{d}x = g(0)\int \widehat{f}(\xi)\,\mathrm{d}\xi,$$

so if we take $g(x) = e^{-\pi|x|^2}$, so $g = \widehat{g}$, then we conclude that $f(0) = \int \widehat{f}(\xi)\,\mathrm{d}\xi$. For a general $x \in \mathbb{R}^n$ set $f_x(y) = f(x+y)$, so that

$$\widehat{f_x}(\xi) = \int f(x+y)e^{-2\pi i\xi\cdot y}\,\mathrm{d}y = e^{2\pi i\xi\cdot x}\widehat{f}(\xi).$$ ⊠

We can also talk about Schwarz distributions, though they'll be covered much more thoroughly in 205B.

Definition. A *Schwarz distribution* T is a continuous linear functional on $\mathcal{S}(\mathbb{R}^n)$: $T(f_k) \to 0$ if $f_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$. The space of Schwarz distributions is called $\mathcal{S}'(\mathbb{R}^n)$.

For example, we have $\delta_0(f) = f(0)$. However, $g(x) = e^{|x|} \notin \mathcal{S}'(\mathbb{R}^n)$. Many such distributions are given by $T(f) = \int fg$ for some g, though δ_0 isn't.

Definition. If $T \in \mathcal{S}'(\mathbb{R}^n)$, then define its *Fourier transform* to be $\widehat{T}(f) = T(\widehat{f})$.

This is motivated by (4), because in the case where $T(f) = \int fg$, they end up saying the same thing. As an example, $\widehat{\delta}_0(f) = \delta_0(\widehat{f}) = \widehat{f}(0) = \int f(x)\,\mathrm{d}x$. Thus, $\widehat{\delta}_0(\xi) = 1$.

Exercise 12. Compute the Fourier transform of $1/(1+x^2)$.

Distributions aren't always $\mathbb{R}^n$-valued functions, but they can be differentiated. If $T(f) = \int gf$ for some $g \in \mathcal{S}(\mathbb{R}^n)$, then

$$T\left(\frac{\partial f}{\partial x_j}\right) = \int g\frac{\partial f}{\partial x_j}\,\mathrm{d}x = -\int \frac{\partial g}{\partial x_j}f\,\mathrm{d}x.$$

Thus, we make the following definition.

Definition. If $T \in \mathcal{S}'(\mathbb{R}^n)$, its *derivative* acts by

$$\frac{\partial T}{\partial x_j}(f) = -T\left(\frac{\partial f}{\partial x_j}\right).$$

Example 15.4. Suppose $g(x) = \operatorname{sgn}(x)$ on $\mathbb{R}$. Then,

$$\begin{aligned}\frac{\partial g}{\partial x}(f) = -g\left(\frac{\partial f}{\partial x}\right) &= \int_{-\infty}^{0}\frac{\partial f}{\partial x}\,\mathrm{d}x - \int_0^{\infty}\frac{\partial f}{\partial x}\,\mathrm{d}x\\ &= f(0) + f(0) = 2\delta_0(f).\end{aligned}$$

Thus, $\frac{\mathrm{d}}{\mathrm{d}x}\operatorname{sgn}(x) = 2\delta_0(x)$.

It turns out that probabilistic statements such as the Law of Large Numbers and the Central Limit Theorem use the Fourier Transform.

Let $x_1, \ldots, x_n$ be independent, identically distributed random variables, and let $s_n = x_1 + \cdots x_n$. For example, a particle moving randomly and continuously can be considered in several independent increments, where the motion in the i^{th} increment is x_i; then, the total motion is s_n. One can normalize the distribution such that $E(x_1) = 0$, and suppose $D = E(x_1^2)$ is finite.

Let $z_n = s_n/n$; we want to understand how this behaves as $n \to \infty$. If $x_i = x_1$, then this is very easy, but in general they're independent.

In the simplest case $n = 2$, let X and Y be independent and $Z = X + Y$. Let p_X and p_Y be the density functions for X and Y, so that

$$p_Z = p_X p_Y = \int p_X(x-y) p_Y(y)\,\mathrm{d}y.$$

In general, $p_{s_n}(x) = (p_x * p_x * \cdots * p_x)(x)$, since the densities are all the same.

For a $\lambda > 0$, let $x_\lambda = x/\lambda$, so that

$$P(x_\lambda \in A) = P(x \in \lambda A) = \int_A p_\lambda(x)\,\mathrm{d}x = \lambda \int_A p(\lambda z)\,\mathrm{d}z.$$

Thus, $p_{x_\lambda}(x) = \lambda p(\lambda x)$. In particular, $p_{z_n}(x) = n(p_x * p_x * \cdots * p_x)(nx)$.

Since there are convolutions floating around, we should use the Fourier transform, which handles these much more nicely:

$$\begin{aligned}
\widehat{f * g}(\xi) &= \int \mathrm{d}x\, e^{-2\pi i \xi \cdot x} \int \mathrm{d}y\, f(x-y) g(y) \\
&= \int \mathrm{d}x\,\mathrm{d}y\, e^{-2\pi i \xi \cdot (x-y) - 2\pi i \xi \cdot y} f(x-y) g(y) \\
&= \widehat{f}(\xi)\widehat{g}(\xi).
\end{aligned}$$

Thus, if $q_n(x) = (p_x * p_x * \cdots * p_x)(x)$, then $\widehat{q}_n(\xi) = (\widehat{p}(\xi))^n$, or $\widehat{p}_{z_n}(\xi) = (\widehat{p}(\xi/n))^n$, and

$$\begin{aligned}
\widehat{p}(0) &= \int p(x)\,\mathrm{d}x = 1. \\
\widehat{p}'(0) &= \int 2\pi i x p(x)\,\mathrm{d}x = 0. \\
\widehat{p}''(0) &= \int (2\pi i x)^2 p(x)\,\mathrm{d}x = -4\pi^2 D.
\end{aligned}$$

Hence, for a fixed ξ,

$$\widehat{p}_{z_n}(\xi) = \left(\widehat{p}\left(\frac{\xi}{n}\right)\right)^n \simeq \left(1 - 2\pi^2 \frac{D|\xi|^2}{n^2}\right)^n,$$

which goes to 0 as $n \to \infty$.

This implies that

$$E(f(z_n)) = \int f(x) p_{z_n}(x)\,\mathrm{d}x = \int \widehat{f}(\xi)\widehat{p}_{z_n}(\xi)\,\mathrm{d}\xi,$$

and therefore as $n \to \infty$, this approaches $\int \widehat{f}(\xi)\,\mathrm{d}\xi = f(0)$.

Corollary 15.5 (Weak Law of Large Numbers)**.**

$$\lim_{n\to\infty} E(f(z_n)) = f(0).$$

That is, $z_n \to 0$ in expectation. One can think of doing a large number of experiments; this law says that as more and more experiments are conducted, the true value does eventually approach the mean.

Sometimes, we might want to consider $R_n = S_n/n^\alpha$ for some α (if z_n doesn't converge fast enough, I guess). Then,

$$\begin{aligned}
E(R_n^2) &= \frac{1}{n^{2\alpha}} E(x_1 + x_2 + \cdots + x_n)^2 \\
&= \frac{1}{n^{2\alpha}} E\left(\sum_{k=1}^n X_k^2 + \sum_{i\neq j} x_i x_j\right) \\
&= \frac{nD}{n^{2\alpha}} + \frac{1}{n^{2\alpha}} \sum_{i\neq j} E(x_i, x_j) \\
&= \frac{nD}{n^{2\alpha}}.
\end{aligned}$$

This is $O(1)$ iff $\alpha = 1/2$, so we conclude $s_n \sim \sqrt{n}$.

Thus, let's consider $R_n = (x_1 + \cdots + x_n)/\sqrt{n}$, so for compactly supported ξ,

$$\widehat{p}_{R_n}(\xi) = \left(\widehat{p}\left(\frac{\xi}{\sqrt{n}}\right)\right)^n \simeq \left(1 - \frac{2\pi^2 D|\xi|^2}{n}\right)^n,$$

and this approaches $e^{-2\pi^2 D|\xi|^2}$, so

$$E(f(R_n)) = \int f(y) p_{R_n}(y)\,\mathrm{d}y = \int \widehat{f}(\xi)\widehat{p}_{R_n}(\xi)\,\mathrm{d}\xi,$$

which converges to

$$\int \widehat{f}(\xi) e^{-2\pi^2 D|\xi|^2}\,\mathrm{d}\xi = \int f(x)\frac{e^{-|x|^2/2D}}{\sqrt{2\pi D}}\,\mathrm{d}x.$$

That is:

Corollary 15.6 (Central Limit Theorem)**.** *As $n \to \infty$, R_n conveges to a Gaussian with the probability density $p(x) = e^{-|x|^2/2D}/\sqrt{2\pi D}$.*

16. Interpolation in L^p Spaces: 11/20/14

"These days, most math majors don't take any physics, so going from quantum mechanics, which you don't know, to classical mechanics, which you don't know, is the most absurd thing ever."

Consider the spaces $L^p(\mathbb{R}^n)$, for $1 \le p \le +\infty$; we want to talk about functions in multiple L^p spaces; specifically, consider p_0, p_1 and $f \in L^{p_0}(\mathbb{R}^n) \cap L^{p_1}(\mathbb{R}^n)$. Let $p_t = (1-t)p_0 + tp_1$; then,

$$\int |f|^{p_t}\,\mathrm{d}x = \int |f|^{(1-t)p_0}|f|^{tp_1}\,\mathrm{d}t \le \left(\int |f|^{p_0}\right)^{1-t}\left(\int |f|^{p_1}\right)^t.$$

In particular, the quick corollary is that if $f \in L^{p_0}(\mathbb{R}^n) \cap L^{p_1}(\mathbb{R}^n)$, then also $f \in L^p(\mathbb{R}^n)$ for all $p_0 \le p \le p_1$.

This is a nice result, but we will wish to generalize it to any measure spaces.

Theorem 16.1 (Riesz[22]-Thorin)**.** *Let (M, μ) and (N, ν) be measure spaces and $1 \le p, q \le +\infty$. Then, for any $t \in [0,1]$, there exists a bounded linear operator $A_t : L^{p_t}(M) \to L^{q_t}(N)$, where*

$$\frac{1}{p_t} = \frac{1-t}{p_0} + \frac{t}{p_1} \qquad \textit{and} \qquad \frac{1}{q_t} = \frac{1-t}{q_0} + \frac{t}{q_1},$$

such that it coincides with $A : L^{p_0}(M) \cap L^{p_1}(M) \to L^{p_0}(N) \cap L^{p_1}(N)$ and $\|A\|_{L^{p_t}\to L^{q_t}} \le k_0^{1-t}k_1^t$, where $k_0 = \|A\|_{L^{p_0}\to L^{q_0}}$ and $k_1 = \|A\|_{L^{p_1}\to L^{q_1}}$.

This is an essentially compelx-analytic theorem, and we'll prove it that way.

Example 16.2 (Hausdorff-Young Inequality)**.** If $f \in L^p(\mathbb{R}^n)$ and $1 \le p \le 2$, then $\widehat{f} \in L^p(\mathbb{R}^n)$ with $\|\widehat{f}\|_{L^{p'}} \le \|f\|_{L^p}$ when $1/p' + 1/p = 1$.

Example 16.3. $\|f * g\|_{L^r} \le \|f\|_{L^p}\|g\|_{L^q}$, where $1 + 1/r = 1/p + 1/q$.

[22]This is a scientific result, not a fun one, so it's by F. Riesz, not M. Riesz.

Proof. Since $(f * g)(x) = \int f(y)g(x-y)\,\mathrm{d}y$, then fix g and let $f \to f * g$. Then, $\|f * g\|_{L^\infty} \le \|g\|_{L^1}\|f\|_{L^\infty}$ and $\|f * g\|_{L^1} \le \|g\|_{L^1}\|f\|_{L^1}$, so therefore interpolating between p and 1, $\|f * g\|_{L^p} \le \|g\|_{L^1}\|f\|_{L^p}$. Furthermore, for the L^∞ norm, if $1/p + 1/p' = 1$, then

$$\begin{aligned}|f * g(x)| &\le \left(\int |f|^p \,\mathrm{d}x\right)^{1/p}\left(\int |g(x-y)|^{p'}\,\mathrm{d}y\right)^{1/p'}\\ &\le \|f\|_{L^p}\|g\|_{L^{p'}}.\end{aligned}$$

Thus, $\|f * g\|_{L^\infty} \le \|f\|_{L^p}\|g\|_{L^{p'}}$.

Now, we interpolate between L^p and L^∞, giving the desired result. ⊠

Example 16.4. Suppose $a \in \mathcal{S}(\mathbb{R}^n \times \mathbb{R}^n)$ and $\varepsilon \in (0,1)$. Define

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

This seems like a silly function, but the idea is to capture very fine oscillations of f on a certain scale ε. How should one get bounds on this operator? Rewrite it as

$$(a(x,\varepsilon D)f)(x) = \int e^{2\pi i x\cdot\xi}\widetilde{a}(x,y)e^{-2\pi i y\cdot\xi}\widehat{f}(\xi)\,\mathrm{d}\xi,$$

where $\widetilde{a}(x,y)$ is the Fourier transform of a with respect to y.

$$= \int \widetilde{a}(x,y)f(x+\varepsilon y)\,\mathrm{d}y.$$

Since $a \in \mathcal{S}(\mathbb{R}^n \times \mathbb{R}^n)$, then so is $\widetilde{a}$, so we can take suprema to get the bounds

$$\begin{aligned}\|a(x,\varepsilon D)f\|_{L^\infty} &\le C(a)\|f\|_\infty\\ \|a(x,\varepsilon D)f\|_{L^1} &\le C(a)\|f\|_{L^1}.\end{aligned}$$

Thus, after interpolating, we conclude that

$$\|a(x,\ eD)f\|_{L^p} \le C(a)\|f\|_{L^p}.$$

We care particularly about $p = 2$, especially in the context of differential equations; there are plenty of familites of *non-dissipative* systems (i.e. those that preserve the L^2-norm over all time) but that don't preserve the L^1-norm or the L^∞-norm.

For example, consider the *Schrödinger equation*

$$ih\frac{\partial\psi}{\partial t} + \frac{h^2}{2}\Delta\psi - V(x)\psi = 0,$$

where h is the *Planck constant.* This is by no means Schwarz, but if one lets $p(x,\xi) = |\xi|^2/2 + B(x)$, then the Schrödinger equation is

$$ih\frac{\partial\psi}{\partial t} + p(x,hD)\psi = 0.$$

Note that as $\varepsilon \to 0$, $\langle a(x,\varepsilon D)\psi,\psi\rangle \to \langle a, W\rangle$ for some $W \ge 0$ (which is particularly nontrivial) satisfying the equation

$$\frac{\partial W}{\partial t} + \{p, W\} = 0,$$

where the braces denote the *Poisson bracket*

$$\{p,W\} = \sum \frac{\partial p}{\partial x_i}\frac{\partial W}{\partial \xi_i} - \frac{\partial p}{\partial \xi_i}\frac{\partial W}{\partial x_i}.$$

Thus, one can conclude that

$$\frac{\partial W}{\partial t} + \xi\cdot\nabla_x W - \nabla V\cdot\nabla_\xi W = 0,$$

and from this one recovers Newton's law of motion: $\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\nabla V(x)$, since $\frac{\partial\xi}{\partial t} = -\nabla V$ and $\frac{\partial x}{\partial t} = \xi$. Thus, we've gone from quantum to classical.

This next theorem is not an example of the use of the Riesz-Thorin Theorem, but rather an ingredient in its proof. It comes from the neighboring nation of complex analysis.

Theorem 16.5 (The Three-Lines Theorem)**.** *Let $F(z)$ be a bounded analytic function in $\{0 \le \mathrm{Re}(z) \le 1\}$ with $|F(iy)| \le m_0$ and $|F(1+iy)| \le m_1$. Then, $|F(x+iy)| \le m_0^{1-x}m_1^x$.*

Proof. Let $F_1(z) = F(z)/(m_0^{1-z} m_1^z)$. Then,

$$|F_1(iy)| \le \frac{m_0}{|m_0^{1-iy}||m_1^{iy}|} = 1$$
$$|F_1(1+iy)| \le \frac{m_1}{|m_0^{-iy}||m_1|^{1+iy}} = 1,$$

and $|F(z)| \le m_0^{1-x} m_1^x |F_1(z)|$. Thus, it only remains to show $|F_1(x+iy)| \le 1$.

(1) If $|F(x+iy)| \to 0$ as $|y| \to \infty$ uniformly in x, then choose an M such that if $|y| \ge M$, $|F_1|(z) \le 1/2$; then, on the sides of this window, $|F_1| \le 1$, so by the maximum modulus princile, $|F_1(z)| \le 1$ in $\{0 \le \operatorname{Re} x \le 1\}$.

(2) If otherwise, set
$$G_n(z) = F_1(z) e^{(z^2-1)/n} = F_1(z) e^{(x^2-y^2+2ixy-1)/n}.$$
For a fixed n, $|G_n(z)| \to 0$ as $|y| \to \infty$, and furthermore, $|G_n(iy)| \le |F_1(iy)| e^{(-y^2-1)/n} \le 1$, and in a similar way, the other bound can be checked, so G_n satisfies the conditions needed for the previous case. Thus, it satisfies the theorem, and when $n \to \infty$, this also holds true for f. ⊠

Basically, we can relax slightly to allow a little bit of growth in the second case.

Proof of Theorem 16.1. First, we need to define A on L^{p_t}. This is even a useful exercise on its own.

Take an $f \in L^{p_t}$, and write it as $f = f_1 + f_2$, where $f_1(x) = f(x)\chi_{\{|f| \le 1\}}(x)$ and $f_2(x) = f(x)\chi_{\{|f| \ge 1\}}(x)$. Thus, $|f_1|^{p_t} \le |f|^{p_0}$, $|f_2|^{p_t} \le |f|^{p_1}$, $|f_1|^{p_1} \le |f|^{p_t}$, and $|f_2|^{p_0} \le |f|^{p_t}$, so we conclude that $f_1 \in L^{p_1}$ and $f_2 \in L^{p_0}$. In particular, we can thus define $Af = Af_1 + Af_2$ (since A was already defined on the tw starting spaces), and thus it clearly agrees with that definition.

Next, we will want to bound A. Given a $g \in L^p$¡ define $L_g : L^{p'} \to \mathbb{R}$ by $L_g(f) = \int gf$ (where $1/p + 1/p' = 1$). Thus, $\|g\|_{L^p} = \|L_g\|$. Then, we want to calculate the operator norm

$$\|A\|_{L^{p_t} \to L^{q_t}} = \sup_{\|f\|_{L^{p_t}} = 1} \|Af\|_{L^{q_t}} = \sup_{\substack{\|f\|_{L^{p_t}} = 1 \\ \|g\|_{L^{q_t'}} = 1}} \int (Af)\overline{g}\, \mathrm{d}\nu.$$

(Here, $\overline{g}$ denotes the complex conjugate.) This is what we want to estimate.

Now, we want to take "simple functions" f and g; in the real case, we just approximated the values of these functions, but here we'll approximate the absolute value and leave the argument alone. In particular, take

$$f(x) = \sum_{j=1}^n a_j e^{i\alpha_j(x)} \chi_{A_j}(x)$$
$$g(y) = \sum_{j=1}^m b_j e^{i\beta_j(x)} \chi_{B_j}(x).$$

Notice that since we're in between the endpoints, these don't live in $L^\infty(M)$ or $L^\infty(N)$; thus, A_j and B_j have finite measure. But $f \in L^{p_t}(M)$ and $g \in L^{q_t'}(N)$.[23]

Now, let's extend p_t, q_t, and q_t' to the strip $\{0 \le \operatorname{Re}(\zeta) \le 1\}$, as

$$\frac{1}{p(\zeta)} = \frac{1-\zeta}{p_0} + \frac{\zeta}{p_1}$$
$$\frac{1}{q(\zeta)} = \frac{1-\zeta}{q_0} + \frac{\zeta}{q_1}$$
$$\frac{1}{q'(\zeta)} = \frac{1-\zeta}{q_0'} + \frac{\zeta}{q_1'}.$$

Thus, $p(t) = p_t$ if $\zeta = t$ is real, and the same holds for q and q'.

Define two more functions

$$u(x,\zeta) = \sum_{j=1}^n a_j^{p_t/p(\zeta)} e^{i\alpha_j(x)} \chi_{A_j}(x)$$
$$v(y,\zeta) = \sum_{j=1}^m b_j^{q_t'/q'(\zeta)} e^{i\beta_j(x)} \chi_{B_j}(y).$$

[23] This proof originally came from a book full of misprints; this proof in particular hadn't a single completely correct sentence. *Caveat emptor* (though the professor did the proof slowly and carefully to be sure).

Thus, $u(x,t) = f(x)$ and $v(y,t) = g(y)$ when $\zeta = t$ is real. Finally, define

$$F(\zeta) = \int Au(y,\zeta)v(y,\zeta)\,\mathrm{d}y = \sum_{j=1}^n \sum_{k=1}^m a_j^{p_t/p(\zeta)} b_k^{q_t'/q'(\zeta)} \int (A\psi_j(y))e^{i\beta_k(y)}\chi_{B_k}(y)\,\mathrm{d}\nu.$$

Thus, we want to show that $\langle Af, g\rangle \le k_0^{1-t}k_1^t$, or $|F(t)| \le k_0^{1-t}k_1^t$. We can see the silhouette of Theorem 16.5 becoming more distinct...

Suppose $\zeta = i\xi$, where $\xi \in \mathbb{R}$; then, $1/p(\zeta) = (1-i\xi)/p_0 + i\xi/p_1$, so $|a_j|^{p_t/p(i\xi)} = |a-j|^{p_t/p_0}$. Then,

$$\begin{aligned}
\|u(x,i\xi)\|_{L^{p_0}} &\le \left(\int\left(\sum |a_j|^{p_t/p_0}\chi_{A_j}(x)\right)^{p_0}\right)^{1/p_0}\\
&= \left(\int \sum |a_j|^{p_t}\chi_{A_j}\right)^{1/p_0}\\
&= \|f\|_{L^{p_t}}^{p_t/p_0} = 1.\\
\|v(x,i\xi)\|_{L^{q_0'}} &\le \left(\int\left(\sum |b_j|^{q_t'/q_0'}\chi_{B_j}(x)\right)^{q_0'}\right)^{1/q_0}\\
&= \left(\int \sum |b_j|^{q_t'}\chi_{B_j}\right)^{1/q_0}\\
&= \|f\|_{L^{q_t'}}^{q_t'/q_0'} = 1.
\end{aligned}$$

Thus, we have a bound $|F(i\xi)| \le k_0$, and an absolutely identical proof shows that $|F(1+i\xi)| \le k_1$. Thus, $|F(\eta + i\xi)| \le k_0^{1-\eta}k_1^\eta$, and thus $|F(t)| \le k_0^{1-t}k_1^t$. ⊠

There is great power and beauty to complexifying the problem; though this looks kind of ugly, it's nowhere near as bad as a brute-force approach would be; the beauty is there, but hidden.

17. The Hilbert Transform: 12/2/14

Today, we've had a break from the course for over a week, so we will discuss things that aren't unrelated to the rest of the course, but stand on their own, as things in themselves (*ding an sich*).

The Hilbert Transform. The Hilbert transform was related to things done in Hilbert's thesis, and motivated by questions on integrable systems.

Suppose we have an $f \in \mathcal{S}(\mathbb{R})$; is there a way to extend it to an analytic function on $\mathbb{C}$? Well, we can write

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

and there's no reason we can't write

$$f(z) = \int_{-\infty}^{\infty} \widehat{f}(\xi)e^{2\pi z\cdot\xi}\,\mathrm{d}\xi. \tag{6}$$

Then, the derivatives are still Schwarz, and $\widehat{f} \in \mathcal{S}$ as well. But sometimes it's infinite: if $z = x+it$, then $z\xi = -t\xi + ix\xi$, so if t and ξ have opposite signs, the integral (6) might be infinite: f might not decrease exponentially.

Nonetheless, if $\widehat{f}$ is compactly supported, then $f(z)$ is well-defined, and if $|\widehat{f}(z)| \le e^{-\alpha|\xi|}$, then $f(x)$ is well-defined when $|\mathrm{Im}(z)| < \alpha$. However, this $f(z)$ grows to infinity, which is still suboptimal.

Maybe we just want to be modest and require f to be bounded, but extended as a harmonic function, and maybe we just want to extend to the upper half-plane. This we can do, according to the formula

$$g(x,t) = \int e^{-2\pi t|\xi| + 3\pi i x\xi}\widehat{f}(\xi)\,\mathrm{d}\xi$$

for $t > 0$. Then, $|g(x,t)| \le \|\widehat{f}\|_{L^1}$. Furthermore, we can calculate $\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial t^2} = 0$, and $g(x,t) \to f(x)$ as $t \to 0$. This is unique, since it determines its boundary value, and it's a convolution with the *Poisson kernel*, as we've mostly all seen in complex analysis courses. Specifically,

$$\begin{aligned}
g(x,t) &= \int e^{-2pit|\xi|+2\pi x\xi} f(y) e^{-2\pi iy\xi}\,\mathrm{d}y\,\mathrm{d}\xi \\
&= \int_{-\infty}^{\infty} f(y)\left(\int_{-\infty}^{0} e^{2\pi t\xi+2\pi i(x-y)\xi}\,\mathrm{d}\xi + \int_0^\infty e^{-2\pi t\xi+2\pi i(x-y)\xi}\,\mathrm{d}\xi\right)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f(y)\left(\frac{1}{2\pi t+2\pi i(x-y)} + \frac{1}{2\pi t-2\pi i(x-y)}\right)\mathrm{d}y \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} f(y)\frac{t}{t^2+(x-y)^2}\,\mathrm{d}y.
\end{aligned}$$

Thus, $g(x,t) = P_t \star f(x)$, where

$$P_t(x) = \frac{1}{\pi}\frac{t}{t^2+x^2} = \frac{1}{t\pi}\frac{1}{1+(x/t)^2} = \frac{1}{t}P_1\left(\frac{x}{t}\right).$$

Since $\int|P_1|\,\mathrm{d}x = 1$, then this is an approximation of identity (as $t\to 0$, this approaches $f(x)$).

Returning to the Hilbert problem, what is the harmonic conjugate to $g(x,t) = P_t\star f(x)$? We can write

$$g(x,t) = \underbrace{\int_0^\infty \widehat{f}(\xi)e^{-2\pi t\xi}e^{2\pi ix\xi}\,\mathrm{d}\xi}_{q(x)} + \int_{-\infty}^0 \widehat{f}(\xi)e^{2\pi t\xi+2\pi ix\xi}\,\mathrm{d}\xi.$$

Now, since $i(x+it)\xi = -t\xi + ix\xi$ when $\xi > 0$, then $q(z) = \int_0^\infty \widehat{f}(\xi)e^{2\pi iz\xi}\,\mathrm{d}\xi$ is analytic in the upper half-plane. Furthermore, let

$$\begin{aligned}
iv(x,t) &= \int_0^\infty \widehat{f}(\xi)e^{-2\pi t\xi+2\pi ix\xi}\,\mathrm{d}\xi - \int_0^\infty \widehat{f}(\xi)e^{2\pi it\xi+2\pi ix\xi}\,\mathrm{d}\xi \\
&= \int_{-\infty}^\infty \operatorname{sgn}(\xi)e^{-2\pi t|\xi|+2\pi ix\xi}\widehat{f}(\xi)\,\mathrm{d}\xi.
\end{aligned}$$

Then, $g(z)+iv(z) = 2\int_0^\infty \widehat{f}(\xi)e^{2\pi iz\xi}\,\mathrm{d}\xi$ is analytic on the upper half-plane.

So we're trying to raise f onto the upper half-plane and then take its harmonic conjugate, sending $f(x)\to g(x,t)\to v(x,t)\to v(x,0)$.

Definition. $v(x,0)$ constructed in the above way is the *Hilbert transform* of $f(x)$.

Let's take its Fourier transform:

$$\widehat{v}(\xi,t) = -i\operatorname{sgn}(\xi)\widehat{f}(\xi)e^{-2\pi t|\xi|}.$$

In particular, $v(x,t) = (Q_t\star f)(x)$, where $\widehat{Q}_t(\xi) = -i\operatorname{sgn}(t)e^{-2\pi t|\xi|}$, so therefore $\widehat{Q}_0(\xi) = -i\operatorname{sgn}(\xi)$. Thus,

$$v(x,0) = \int(-i\operatorname{sgn}(\xi))\widehat{f}(\xi)e^{2\pi ix\xi}\,\mathrm{d}\xi.$$

We want to know how slowly this decays; if f is discontinuous, we'll need lots of high frequencies to account for the jumps, so it decays more slowly.

Definition. The *principal value of* $1/x$ is a Schwarz distribution $\mathrm{PV}(1/x)\in\mathcal{S}'(\mathbb{R})$ given by

$$\mathrm{PV}\left(\frac{1}{x}\right)(\varphi) = \lim_{\varepsilon\to 0}\int_{|x|>\varepsilon}\frac{\varphi(x)}{x}\,\mathrm{d}x.$$

Since φ is Schwarz, this is well-defined at infinity, but why is it bounded as $\varepsilon\to 0$? It turns out that

$$\mathrm{PV}\left(\frac{1}{x}\right)(\varphi) = \int_{|x|>1}\frac{\varphi(x)\,\mathrm{d}x}{x} + \lim_{\varepsilon\to 0}\int_{\varepsilon<|x|<1}\frac{\varphi(x)-\varphi(0)}{x}\,\mathrm{d}x = \int_{|x|>1}\frac{\varphi(x)\,\mathrm{d}x}{x} + \int_{-1}^{1}\frac{\varphi(x)-\varphi(0)}{x}\,\mathrm{d}x.$$

Thus, it is in fact bounded.

Proposition 17.1. *Let* $Q_t(x) = (1/\pi)(x/(t^2+x^2))$*; then,* $\lim_{t\to 0}Q_t(x) = (1/\pi)\,\mathrm{PV}(1/x)$.

Proof. Let $\psi_t(x) = (1/t)\chi_{t<|x|}$, so that

$$\mathrm{PV}\left(\frac{1}{x}\right)(\varphi) = \lim_{t\to 0}\int_{-\infty}^{\infty}\psi_t(x)\varphi(x)\,\mathrm{d}x.$$

Then,

$$\begin{aligned}\int (\pi Q_t(x) - \psi_t(x))\varphi(x)\,\mathrm{d}x &= \int_{-\infty}^{\infty} \frac{x\varphi(x)}{x^2+t^2}\,\mathrm{d}x - \int_{|x|>t} \frac{\varphi(x)}{x}\,\mathrm{d}x \\ &= \int_{|x|<t} \frac{x\varphi(x)\,\mathrm{d}x}{x^2+t^2} + \int_{|x|>t} \left(\frac{x}{x^2+t^2} - \frac{1}{x}\right)\varphi(x)\,\mathrm{d}x \\ &= \int_{|z|<1} \frac{z\varphi(zt)}{z^2+1}\,\mathrm{d}z - \int_{|x|>t} \frac{t^2}{x(x^2+t^2)}\varphi(x)\,\mathrm{d}x \\ &= \int_{|z|<1} \frac{z\varphi(zt)\,\mathrm{d}z}{z^2+1} + \int_{|z|>1} \frac{\varphi(tz)}{z(z^2+1)}\,\mathrm{d}z.\end{aligned}$$

As $t \to 0$, this goes to 0 (the limit works because these are bounded, so we can use the Lesbegue Dominated Convergence Theorem). ⊠

The fact that $1/x$ cancels is very important — it's an odd function, so the average over any sphere is 0, and we couldn't do this with $1/|x|$. These can be generalized to other functions which are 0 averaged over spheres, leading to a class of operators called Calderón-Zygmund operators.

Now, we know the Hilbert transform is the convolution with the principal value of $1/x$:

$$Hf(x) = \lim_{t\to 0^+} Q_t \star f(x) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{|y|>\varepsilon} \frac{f(x-y)\,\mathrm{d}y}{y}.$$

Then, $\widehat{Hf}(\xi) = -i\,\mathrm{sgn}(\xi)\widehat{f}(\xi)$, so if $f \in \mathcal{S}(\mathbb{R})$, then $\|\widehat{Hf}\|_{L^2} = \|\widehat{f}\|_{L^2}$, and in particular $H : L^2 \to L^2$ is well-defined.

Another useful property is that this is skew-symmetric on L^2: $(Hf, g) = -(f, Hg)$, and in particular $H(Hf) = -f$. There's more than one way to see this, but since $Hf(x) = \int f(y)/(x-y)\,\mathrm{d}y$, then $H^*f(x) = \int (1/(y-x))f(y)\,\mathrm{d}y = -Hf$.

We wish to generalize this to other L^p spaces, and we'll be able to use Theorem 16.1 to do this.

Definition. The space of *weak L^1 functions* L^1_w is the space of functions f such that $m(\{x : |f(x)| > \lambda\}) < c/\lambda$ for some c. Similarly, a *weak L^p function* is one such that $f^p \in L^1_w$.

Theorem 17.2 (Kolmogorov (1925))**.** *The Hilbert transform satisfies* $m(\{x : |Hf(x)| > \lambda\}) \le c\|f\|_{L^1}/\lambda$.

This says it's weak L^1, which is a nice bound. We can do better sometimes, though; the Riesz-Thorin interpolation theorem allowed interpolation between operators $L^{p_1} \to L^{q_1}$ and $L^{p_2} \to L^{q_2}$; then, a theorem of Marcinkiewiecz[24] allows interpolation between operators $L^{p_1} \to L^{q_1}_w$ and $L^{p_2} \to L^{q_2}_w$. The proof is elementary, but takes a while, so will be omitted.

If $p \ne 1$, we can do better. (It's a nice exercise to see why this doesn't work in L^1.)

Theorem 17.3 (M. Riesz (1928))**.** *For $1 < p < \infty$, the Hilbert transform is a bounded linear operator $L^p(\mathbb{R}) \to L^p(\mathbb{R})$.*

Proof. Let

$$u(z) = \int_0^\infty e^{2\pi i z\xi}\widehat{f}(\xi)\,\mathrm{d}\xi = \int_0^\infty e^{2\pi i z\xi}\left(\widehat{f}(\xi) + i\widehat{Hf}(\xi)\right)\mathrm{d}\xi.$$

We'll look at $\xi = 0$ and $z \to +\infty$.

Let $\mathcal{S}_0 = \{f \in \mathcal{S} : \text{there exists } \varepsilon > 0 \text{ such that } \widehat{f}(\xi) = 0 \text{ for } |\xi| < \varepsilon\}$; then, if $f \in \mathcal{S}_0$, then $\widehat{Hf}(\xi) = -i\,\mathrm{sgn}(\xi)\widehat{f}(\xi) \in \mathcal{S}(\mathbb{R})$, so $Hf \in \mathcal{S}(\mathbb{R})$, which is nice.

Claim. $\mathcal{S}_0$ is dense in $L^p(\mathbb{R})$.

Proof. Given an $f \in \mathcal{S}$ and $p = 2$, take

$$\chi(\xi) = \begin{cases} 1, |\xi| > 1 \\ 0, |\xi| \le 1/2. \end{cases}$$

Then, set $g_n(\xi) = f(\xi)\chi(n\xi) \in \mathcal{S}_0$, since $g_n(\xi) = 0$ for $|\xi| \le 1/2n$, and so

$$\|g_n - f\|_{L^2} = \|\widehat{g}_n - \widehat{f}\|_{L^2} \le 2\int_{-1/n}^{1/n} |\widehat{f}(\xi)|^2\,\mathrm{d}\xi,$$

[24] Like several Eastern European mathematicians at around the time of the Second World War, Marcinkieviecz's work was interrupted by the Soviet invasion of Poland; since he served as an officer, he eventually died in a Soviet concentration camp.

which goes to 0 as $n \to \infty$. Thus, it converges in L^2, and we can also show it converges in L^∞: if $p = \infty$, then

$$\|g_n - f\|_{L^\infty} \le \|\widehat{g}_n - \widehat{f}\|_{L^1} \le 2\int_{-1/n}^{1/n} |\widehat{f}(\xi)|\,\mathrm{d}\xi,$$

which also goes to 0 as $n \to +\infty$.

Thus, $g_n \to f$ in all L^p for $2 \le p \le \infty$. ⊠

Now, define $p(x) = f(x) + iHf(x)$, and its analytic continuation is

$$p(z) = 2\int_0^\infty e^{2\pi i z\xi}\widehat{f}(\xi)\,\mathrm{d}\xi.$$

Assume $f \in \mathcal{S}_0(\mathbb{R})$ and $|\widehat{f}(\xi)| = 0$ for $|\xi| \le \varepsilon$. Then, $|p(z)| \le 2\|\widehat{f}\|_{L^1}e^{-2\pi\varepsilon|y|}$, if $z = x + iy$, and so we can integrate along the semicircle C_R around 0 with radius R in the upper half-plane. By Cauchy's theorem, $\oint_{C_R} p(z)^4\,\mathrm{d}z = 0$,[25] so $\int_{-\infty}^\infty (f + iHf)^4\,\mathrm{d}x = 0$. Taking the real part, $\int(f^4 - 6f^2(Hf)^2 + (Hf)^4)\,\mathrm{d}x = 0$, and therefore

$$\int (Hf)^4\,\mathrm{d}x = 6\int (f^2)(Hf)^2 - \int f^4\,\mathrm{d}x \le 6\cdot 1000\int f^4 + \frac{6}{1000}\int (Hf)^4 - \int f^4.$$

Thus, Hf is L^4! Now we see why the fourth power was used: because we remember the coefficients. One can do this with any power, so we know $\|Hf\|_p \le C\|f\|_p$ for $2 \le p < \infty$.[26]

Now, to get $1 < p < 2$, we rely on science, not tricks. Specifically, since $\|Hf\|_{L^p} \le C\|f\|_{L^p}$, then the joint operator H^*f satisfies the dual bound: $\|H^8 f\|_{L^{p'}} \le C\|f\|_{L^{p'}}$, where $L^{p'}$ is the dual to L^p, which by the Riesz Representation Theorem is when $1/p + 1/p' = 1$. But we know $H^* = -H$, so $\|Hf\|_p \le C\|f\|_p$ for all $1 < p < \infty$. ⊠

This is good for the Hilbert transform, but for more general integral linear operators the proof doesn't work quite as well.

18. Brownian Motion: 12/4/14

"We pretend it is continuous. You ask too many questions."

First, let's talk about what Brownian motion really is; then, we can construct it. This is advantageous because the formal construction is somewhat abstract.

Start with a random walk on $\mathbb{Z}$; at the starting position x, jump to the right with probability 1/2 and the left with probability 1/2; then, do the same thing (going forwards and backwards, each equally likely at each step). This creates a path from integers to integers, or a piecewise linear function $\mathbb{R} \to \mathbb{R}$. We want to rescale this so that it becomes a continuous time-space curve; specifically, send the spatial step $h = \delta x \to 0$ and the time step $\tau = \delta t \to 0$. How should we do this?

Before rescaling, $X(n) = Y_1 + \cdots + Y_n$ for Y_i independent random variables equal to each of ± 1 with probability 1/2. The right tools to deal with this will be the Central Limit Theorem and the Law of Large Numbers. Recall that the latter states that

$$\frac{X(n)}{n} = \frac{Y_1 + Y_2 + \cdots + Y_n}{b} \longrightarrow 0$$

as $n \to \infty$, so $X(n)/n = (1/n)(Y_1 + Y_2 + \cdots + Y_n) = Y_1/n + \cdots + Y_n/n \to 0$, which now has spatial step $h = 1/n$. However, the Central Limit Theorem tells us this stays near zero with very high probability, so we need to do more jumps, in some sense; thus, we need a shorter timestamp.

The Central Limit Theorem tells us that $X(n)/\sqrt{n}$ goes to a Gaussian with mean zero and variance 1, so that if the spatial step is $h = 1/\sqrt{n}$ and $\tau = 1$, then $Y_1/\sqrt{n} + \cdots + Y_n/\sqrt{n}$ converges to a Gaussian, which has mean zero, but is not identically zero. This is an interesting example, because the steps one takes get larger and larger; the hope is that when one passes to the limit as $n \to \infty$, the result is a continuous time-space process.

The reason this works is a little bit of abstract nonsense: the probability describes a measure on the space of continuous functions that is only supported on these piecewise linear functions, and since they have a modulus of continuity, one can prove that the limit point must exist.[27]

Now, we want to understand why $(\delta x)^2 = \delta t$. Let τ_{ab} denote the time spent within the interval $[a, b]$ (where $x \in [a, b]$). We want to know $g(x) = E_x[\tau_{ab}]$; if $x = a$ or $x = b$, then $g(x) = 0$. In particular,

$$g(x) = \frac{1}{2}g(x + \delta x) + \frac{1}{2}g(x - \delta x) + \delta t.$$

[25] Why the 4^{th} power? May the fourth be with you!

[26] This is much nicer than the standard proof, if a little dependent on trickery.

[27] See Billingsley, *Weak Convergence of Probability Measures.*

This is an exact equation.[28] Recall that we've all taken calculus classes, so let's expand out along a Taylor series (the analyticity works out, but is a little painful to prove):

$$g(x) = \frac{1}{2}\left(g(x) + g'(x)\delta x + \frac{g''(x)}{2}(\delta x)^2 + \cdots\right) + \frac{1}{2}\left(g(x) - g'(x)\delta x + \frac{g''(x)}{2}(\delta x)^2 - \cdots\right) + \delta t.$$

Thus, $-g''(x)/2 = \delta t/(\delta x^2) + \cdots$, so we must have a relation similar to $\delta t = (\delta x)^2$. Now we have an informal argument along with this scientific argument. When Brownian motion is done in higher dimensions, we get an elliptic partial differential equation, which is reasonably nice: $-\Delta g/2 = 1$ in Ω, and $g = 0$ on $\partial\Omega$.

We will constrct the Brownian motion using *Haar functions.* Let

$$\psi(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ -1, & 1/2 \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we'll generate lots of different functions by scaling and translating these: we want ψ_{jk} to have scale j and location k; specifically, let $\psi_{jk}(x) = 2^{j/2}(2^j x - k)$, which is located around $x = k/2^j$. Additionally, the ψ_{jk} are normalized:

$$\int_{-\infty}^{\infty} \psi_{jk}^2(x)\,\mathrm{d}x = 2^j \int \psi^2(2^j x - k)\,\mathrm{d}x = \int \psi^2(y-k)\,\mathrm{d} = 1.$$

Intuitively, to preserve the L^2 norm for things which are scaled, they have to be compressed a lot (this is what the 2^j is doing). There are also two more normalization properties:

$$\int \psi_{jk}(x)\,\mathrm{d}x = 2^{j/2} \int \psi(2^j x - k)\,\mathrm{d}x = \frac{1}{2^{j/2}} \int \psi(y-k)\,\mathrm{d}y = 0.$$
$$\int |\psi_{jk}|\,\mathrm{d}x = \frac{1}{2^{j/2}}.$$

Finally, a calculation in the lecture notes (though there is some intuition behind it) shows that

$$\int \psi_{jk}(x)\psi_{j'k'}(x)\,\mathrm{d}x = \delta_{jj'}\delta_{kk'}.$$

Thus, the ψ_{jk} are orthonormal! Thus, for an $f \in L^2(\mathbb{R})$, we can obtain Fourier coefficients for them: let

$$c_{jk} = \int_{-\infty}^{\infty} f(x)\psi_{jk}(x)\,\mathrm{d}x.$$

Since $\langle \psi_{jk}, \psi_{j'k'} \rangle = \delta_{jj'}\delta_{kk'}$, then $\sum c_{jk}^2 \le \|f\|^2$.

Claim. $\{\psi_{jk}(x) : j, k \in \mathbb{Z}\}$ forms a basis for $L^2(\mathbb{R})$.

Proof. Let $I_{mk} = ((m-1)/2k, m/2^k)$ (dyadic intervals). We want to approximate the function on these intervals: the best approximation is

$$P_m f(x) = \frac{1}{|I_{mk}|} \int_{I_{mk}} f\,\mathrm{d}y.$$

Thus, in probabilistic terms this corresponds to the conditional expectation of f!

We want to show that

$$P_{n+1}f - P_n f = \sum_{k \in \mathbb{Z}} c_{nk}\psi_{nk}(x).$$

If one refines the dyadic intervals, this corresponds to splitting each interval into two. The averaging property means that $\int_{I_{nk}} P_{n+1}f = \int_{I_{nk}} P_n f$, and there's a coefficient α_{nk} such that $P_{n+1}f - P_n f = \alpha_{nk}\psi_{nk}$.

Exercise 13. Show that $\alpha_{nk} = \int f\psi_{nk} = c_{nk}$.

Then,

$$P_{n+1}f - P_{-mf} = \sum_{j=-m}^{m} \sum_{k} c_{jk}\psi_{jk}$$

[28] "You can explain it to your sibling, though you may need to pick a sibling based on age. I probably can't explain it to my daughter until she is 24."

and

$$\begin{aligned}\int_{\mathbb{R}} |P_n f|^2 \,\mathrm{d}x &= \sum_{k\in\mathbb{Z}} \int_{I_{nk}} |P_n f|^2 \\ &= \sum 2^{-n} 2^{2n} \left| \int_{I_{nk}} f(y)\,\mathrm{d}y \right|^2 \\ &\le \sum 2^n 2^{-n} \int_{I_{nk}} |f(y)|^2 \,\mathrm{d}y = \|f\|_2^2,\end{aligned}$$

so these are all bounded linear operators. Then (this requires some care about which norms one uses), if $f \in C_c(\mathbb{R})$, then averaging it on a very large interval makes it very small (since it's compactly supported): as $m \to \infty$, $P_{-mf} \le (1/2^m) \int |f| \to 0$, and additionally, $P_n f \to f$. Thus, letting $m, n \to \infty$, we see that

$$f(x) = \sum_{j,k\in\mathbb{Z}} c_{jk} \psi_{jk}. \qquad \boxtimes$$

Now, we will relate this back to Brownian motion, which will be constructed in terms of these Haar functions. First, though, what do we want from Brownian motion $B(t)$?

(1) $B(t)$ should be a continuous process almost surely.
(2) Increments of $B(t)$ should be independent: specifically, $B(t_1) - B(t_2)$ should be independent of $B(t_3) - B(t_4)$ when $t_1 > t_2 > t_3 > t_4$.
(3) The increments $B(t) - B(s)$ (with $0 \le s < t$) should be Gaussian.
(4) We would want the variance to be $E((B(t) - B(s))^2) = t - s$. This comes from the idea that

$$E((Y_1 + \cdots + Y_n)(Y_1 + \cdots + Y_n)) = \sum_{i,j} E(Y_i Y_j) = \sum_{j=1}^{n} E(Y_i^2) = n.$$

Now we'll be able to set up the measure-theoretic construction. For $0 \le t < 1$¡ consider $\psi_{jk}(x)$ where $j \ge 0$, $k = 0, 2^j - 1$. Specifically, let $\varphi_{2^j+k}(t) = \psi_{jk}(t)$. These are piecewise linear "hat functions:" first they go up, then they go down. The result is a triangle.

Next, let $z_n(\omega)$ be independently and identically distributed random Gaussian variables such that $E(z_n) = 0$, $E(z_n^2) = 1$, so that

$$P(z_n > y) = \int_y^\infty e^{-y^2} \frac{\mathrm{d}y}{\sqrt{2\pi}}.$$

Claim. The function

$$X(t,\omega) = \sum_{n=1}^{\infty} z_n(\omega) \int_0^t \varphi_n(s)\,\mathrm{d}s$$

is the Brownian motion, i.e. it satisfies the constraints outlined above.

Proof. First, we need to check that this series even converges in $L^2(\Omega)$ (where $\omega \in \Omega$, which is the probability space), since the sum of Gaussians could diverge. Let's take the *Cauchy tail* of this:

$$\begin{aligned}E\left(\sum_{k=1}^{m} z_n(\omega) \int_0^t \varphi_k(s)\,\mathrm{d}s\right)^2 &= \sum_{k=n}^{m} \sum_{k'=n}^{m} E(z_k z_{k'}) \int_0^t \varphi_k \int_0^t \varphi_{k'} \\ &= \sum_{k=n}^{m} \left(\int_0^t \varphi_k(s)\,\mathrm{d}s \right)^2 \\ &= \sum_{k=n}^{m} \langle \chi_{[0,t]}, \varphi_k \rangle^2,\end{aligned}$$

which is just the sum of the wavelet coefficients. Thus, this goes to 0 as $m, n \to \infty$.

Now, we can do the same thing with the increments. Let $s < t$; then,

$$\begin{aligned}
E(X(t,\omega) - X(s,\omega)) &= \sum_{k=1}^\infty \sum_{k'=1}^\infty E(z_k z_{k'}) \int_s^t \varphi_k(\tau)\,\mathrm{d}\tau \int_s^t \varphi_{k'}(\tau')\,\mathrm{d}\tau' \\
&= \sum_{k=1}^\infty \left(\int_s^t \varphi_k(\tau)\,\mathrm{d}\tau \right)^2 \\
&= \sum_{k=1}^\infty \langle \chi_{[s,t]}, \varphi_k \rangle^2 \\
&= \|\chi_{[s,t]}\|^2 = t - s.
\end{aligned}$$

Thus, the increments have the right variance. Let's show they're independent. We know that all of the finite sums are Gaussians, and the limit of Gaussians is Gaussian, so the increments are all Gaussian. Thus, to check their independence, it suffices to check that the covariance matrix is diagonal. Let's check it. Suppose $t_1 < t_2 < t_3 < t_4$; then,

$$\begin{aligned}
E((X_{t_4} - X_{t_3})(X_{t_2} - X_{t_1})) &= \sum_{k=1}^\infty \left(\int_{t_3}^{t_4} \varphi_k(s)\,\mathrm{d}s \int_{t_1}^{t_2} \varphi_k(s)\,\mathrm{d}s \right) \\
&= \langle \chi_{[t_3,t_4]}, \chi_{[t_1,t_2]} \rangle = 0.
\end{aligned}$$

Now, we want to show that $B(t)$ is a continuous process almost surely. We won't be able to get uniform continuity in ω, but that's all right. Here's a useful fact about Gaussian variables.

Claim. The random variable

$$M(\omega) = \sup_n \frac{|z_n(\omega)|}{\sqrt{\log n}}$$

is finite almost surely.

Proof. We can calculate

$$P\left(|z_n(\omega)| \geq 2\sqrt{\log n}\right) \leq e^{-(2\sqrt{\log n})^2/2} = \frac{1}{n^2}.$$

Thus, the sum of all of these probabilities is finite, so by the Borel-Cantelli lemma, almost surely only a finite number of them happen. In particular, almost surely we have that $|z_n(\omega)| \geq 2\sqrt{\log n}$ only finitely many times.[29] ⊠

This is useful outside of this proof, since sequences of i.i.d. Gaussians tend to be useful.

Here, though, we use it to say that there exists an $M(\omega)$ such that $|z_n(\omega)| \leq M(\omega)\sqrt{\log n}$ almost surely. In particular, for each j, $\int_0^t \varphi_{2^j+k}(s)\,\mathrm{d}s$ is nonzero for exactly one k (which does have a little intuition for how φ_{2^j+k} looks like triangular hats with height $(1/2^{j/2})$). Thus, in the left-hand side of the following equation, there's only one term to really sum over:

$$\begin{aligned}
\left| \sum_{k=0}^{2^j-1} z_{2^j+k}(\omega) \int_0^t \varphi_{2^j+k}(s)\,\mathrm{d}s \right| &\leq \frac{1}{2^{j/2}} M(\omega) \sqrt{\log(2^j + 2^j)} \\
&\leq \frac{M(\omega)\sqrt{j}\sqrt{\log 2}}{2^{j/2}}.
\end{aligned}$$

Thus, if one looks at the series

$$X(t,\omega) = \sum_{n=1}^\infty z_n(\omega) \int_0^t \varphi_n(s)\,\mathrm{d}s,$$

it's bounded above, and therefore uniformly convergent in t for each fixed ω! Since each term is continuous, this implies that $X(t,\omega)$ is almost surely continuous. ⊠

Though we ran out of time, one can also prove that there's no control over the modulus of continuity, and therefore that $X(t,\omega)$ is nowhere differentiable almost surely. (See the lecture notes for a proof.) In particular, this relatively explicit series gives a nice example of a continuous, but nowhere differentiable function. And it's not even much harder to show that it's Hölder with exponent less than 1/2, but is not Hölder with exponent 1/2 or more. These are somewhat useful functions, but people quested for them for a long time in the 19th Century (which makes one wonder why we study it in the 21th Century, but we digress).

[29] The number of times it happens is random, but the important point is that it's finite almost surely.