MATH 321-3: Real Analysis
Northwestern University, Lecture Notes

Written by Santiago Cañez

These are notes which provide a basic summary of each lecture for MATH 321-3, the third quarter of "MENU Real Analysis", taught by the author at Northwestern University. The book used as a reference is *A (Terse) Introduction to Lebesgue Integration* by Franks. Watch out for typos! Comments and suggestions are welcome.

Contents

Lecture 1: Introduction **1**
Lecture 2: Contractions **3**
Lecture 3: Inverse Functions **8**
Lecture 4: Implicit Functions **8**
Lecture 5: Step Functions **8**
Lecture 6: Regulated Integrals **8**
Lecture 7: Outer Measure **8**
Lecture 8: Measurability **13**
Lecture 9: More on Measurability **13**
Lecture 10: Sigma Algebras **14**
Lecture 11: Regularity **14**
Lecture 12: Simple Functions **14**
Lecture 13: Measurable Functions **14**
Lecture 14: Lebesgue Integration **14**
Lecture 15: Bounded Convergence **14**
Lecture 16: Unbounded Functions **14**
Lecture 17: Absolute Continuity **14**
Lecture 18: Dominated Convergence **14**
Lecture 19: General Functions **14**
Lecture 20: L^p Spaces **14**
Lecture 21: Hölder and Minkowski **17**
Lecture 22: L^p Convergence **21**
Lecture 23: Completeness **24**
Lecture 24: Dual Spaces **27**
Lecture 25: Hilbert Spaces **33**

Lecture 1: Introduction

The final quarter of this year-long sequence will focus on developing a better notation of integration. This better notion is known as *Lebesgue integration* and, as we'll see, has some crucial properties that makes it more suitable for various applications than Riemann integration, which you developed last quarter. The Lebesgue integral depends on the notion of the *Lebesgue measure*, which provides a way to measure the "length" of a wide variety of subsets of $\mathbb{R}$.

As one main use of the Lebesgue integral, in the second half of the quarter we will further the develop the theory of the space L^2 you were introduced to last quarter. Indeed, we will give the *real* definition of this space (the one you saw last quarter was only a glimpse!), and along the way we will also give a brief introduction to the subject of functional analysis.

To start, we describe some of the limitations of the Riemann integral, and hint at how the Lebesgue integral will overcome them.

Riemann and areas under graphs. The standard example of a non-Riemann integrable function is the function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q}. \end{cases}$$

This is called the *indicator* function of $\mathbb{Q} \cap [0, 1]$ since it "indicates" which elements are in this set by assigning 1 to those elements and 0 to others. Recall that this function is not Riemann integrable since its infimum over any subinterval of $[0, 1]$ is 0 (since the irrationals are dense) and its supremum is 1 (since the rationals are dense), so all lower Riemann sums are 0 and all upper Riemann sums are 1.

But, in fact we claim that this function *does* have a well-defined area under its graph, only one that the Riemann integrable is not strong enough to detect. Indeed, this area is zero. The region under the graph of this function consists of vertical line segments of height 1 lying above each rational in $[0, 1]$ on the x-axis. To see that the "area" of the union of these segments is zero, fix $\epsilon > 0$. Enumerate the (countable) elements of $\mathbb{Q} \cap [0, 1]$ as

$$r_1,\ r_2,\ r_3,\ r_4,\ \ldots.$$

Pick an interval I_1 around r_1 of length $\epsilon/2$, an interval I_2 around r_2 of length $\epsilon/2^2$, an interval around r_3 of length $\epsilon/2^3$, and in general an interval around r_n of length $\epsilon/2^n$. Then for each n take a rectangle R_n with base I_n and height 1. This rectangle covers the vertical line segment under the graph of f corresponding to r_n on the x-axis. Thus the union of these segments is covered by the union of these rectangles, so we should have:

$$\text{"area" under the graph} \leq \text{area of the union of the } R_n.$$

The area of R_n is the length of I_n times the height 1, so this area is $\epsilon/2^n$, and thus we have

$$\text{"area" under the graph} \leq \text{area of the union of the } R_n \leq \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon.$$

But $\epsilon > 0$ was arbitrary, so we conclude that the "area" under the graph of f must be zero as claimed. (We are using "area" here since there is a question as to in what sense this area exists. Essentially, this take this argument as *defining* this area as zero. This same idea will form the basis behind the notion of the Lebesgue measure soon enough.) This function will be Lebesgue integrable, and its Lebesgue integral will be zero.

Riemann and pointwise convergence. For a second limitation of the Riemann integral, consider the sequence of functions $f_n : [0,1] \to \mathbb{R}$ defined by

$$f_n(x) = \begin{cases} 1 & x \in \{r_1, r_2, \ldots, r_n\} \\ 0 & \text{otherwise,} \end{cases}$$

where the r_i denote the same enumeration of the elements of $\mathbb{Q} \cap [0,1]$ as before. (So, f_n is the indicator function of $\{r_1, r_2, \ldots, r_n\}$.) Since each r_n is nonzero are only a finite number of points, each f_n is Riemann integral with integral 0. However, the f_n converge pointwise to the indicator function f of $\mathbb{Q} \cap [0,1]$ from the previous example, which is not Riemann integrable. This shows that Riemann integrability is not preserved under pointwise convergence.

The situation is even worse: even if a pointwise limit is Riemann integrable, the value of the integrals themselves are not necessarily preserved when taking pointwise limits. Consider now the sequence of functions $f_n : [0,1] \to \mathbb{R}$ defined by

$$f_n(x) = \begin{cases} n & 0 < x < \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Each of these is Riemann integrable with Riemann integral

$$\int_0^1 f_n(x)\, dx = \int_0^{1/n} n\, dx = n(\tfrac{1}{n}) = 1.$$

The pointwise limit of these functions, however, is the constant zero function (since for any $x > 0$, we eventually have $f_n(x) = 0$ once n is large enough so that $\frac{1}{n} < x$). This limit *is* integrable, but with integral 0 so that

$$\int_0^1 f_n(x)\, dx \not\to \int_0^1 0\, dx.$$

Now, the Lebesgue integral will get around this flaw in some sense, but not in absolute sense: $f_n \to f$ pointwise with all functions Lebesgue integrable does not imply that the Lebesgue integrals of the f_n converges to the Lebesuge integral of f. But, we will see that in fact there are some *mild* assumptions we can impose on the f_n so that the values of the Lebesgue integrals do behave nicely when taking pointwise limits. This is a whole lot better than what is the case with the Riemann integral, where you need full blown uniform convergence to guarantee such a conclusion.

Riemann and completeness of L^2. Finally, last quarter you saw a definition of the space L^2 of *square-integrable* functions, something like

$$L^2([a,b]) = \left\{ f : [a,b] \to \mathbb{R} \;\middle|\; \int_a^b f(x)^2\, dx \text{ exists and is finite} \right\}.$$

(Note that the integral used here is the Riemann integral, so by "exists" we mean that f^2 is Riemann integrable.) On this space we have the so-called L^2-metric defined by

$$d(f,g) = \left(\int_a^b (f(x) - g(x))^2\, dx \right)^{1/2},$$

and this metric space provides the natural setting which underlies the theory of Fourier series. (Actually, some care here is needed since as written d is technically not a metric, since $d(f,g) = 0$

does not imply $f = g$ because there exist *many* non-negative functions with the property that they have integral zero and yet are nonzero themselves. We'll see how to fix this later.)

But this metric space is in fact not complete, in that there exist Cauchy sequences which do not converge. The issue is that these Cauchy sequences "converge" to a function which is not in this space, namely one which is not Riemann integrable. To get true completeness we must instead consider the Lebesgue-analog of this space, which, as we see, will be complete. We will revisit this space and the theory of Fourier series later, and see the *correct* context behind it all.

Plan for the course. Our goal then is to develop first the theory of the Lebesgue measure (and measure theory more generally), then the Lebesgue integral, and finally the structure of L^2. After this we have some leeway. My hope is to *also* consider some other well-known spaces besides L^2, such as the L^p spaces which are the standard examples of what are called *Banach spaces*, and in so doing see a glimpse of functional analysis. Time permitting, we might also very briefly touch upon the subject of *ergodic theory*.

But before all of this, we will actually come back to some MATH 321-2 material, and finish the discussion of differentiability you began last quarter. Specifically, we will study the Inverse and Implicit Function Theorems in more detail than what you saw last quarter already. These theorems will not play a big role once we move to measure theory, although the Inverse Function Theorem will at least show up briefly when considering the change of variables formula. Nonetheless, these theorems are crucial pieces in any analyst's toolbox, so they are worth being familiar with. If nothing else, they will give us a chance to understand the role that *contractions* and *fixed-point* problems play in analysis in general, which is certainly a worthy goal.

Lecture 2: Contractions

Warm-Up. Suppose that X is a metric space and that (p_n) is a sequence in X such that

$$d(p_n, p_{n+1}) \leq c^n \text{ for all } n$$

where $0 < K < 1$ is fixed. We show that (p_n) is Cauchy. Note that then if X is complete, (p_n) in fact converges.

Let $\epsilon > 0$. Since $\sum c^n$ is a convergent geometric series, by the Cauchy criterion for series convergence there exists N such that

$$|c^n + c^{n+1} + \cdots + c^{n+k}| < \epsilon \text{ for } n \geq N, k > 0.$$

Now, for $k > 0$ we have

$$\begin{aligned} d(p_n, p_{n+k}) &\leq d(p_n, p_{n+1}) + d(p_{n+1}, p_{n+k}) \\ &\leq d(p_n, p_{n+1}) + d(p_{n+1}, p_{n+2}) + d(p_{n+2}, p_{n+k}) \\ &\vdots \\ &\leq d(p_n, p_{n+1}) + d(p_{n+1}, p_{n+2}) + \cdots + d(p_{n+k-1}, p_{n+k}) \end{aligned}$$

after repeated applications of the triangle inequality. Thus if $n \geq N$ and $k > 0$, we have

$$\begin{aligned} d(p_n, p_{n+k}) &\leq d(p_n, p_{n+1}) + d(p_{n+1}, p_{n+2}) + \cdots + d(p_{n+k-1}, p_{n+k}) \\ &\leq c^n + c^{n+1} + \cdots + c^{n+k-1} \\ &= |c^n + c^{n+1} + \cdots + c^{n+k-1}| \end{aligned}$$

$$< \epsilon,$$

which shows that (p_n) is Cauchy.

Banach Contraction Principle. The proof we will give of the Inverse Function Theorem uses heavily the notion of a contraction. A *contraction* on a metric space X is a function $f : X \to X$ for which there exists $0 < C < 1$ such that

$$d(f(p), f(q)) \leq Cd(p, q) \text{ for all } p, q \in X.$$

Thus, f "contracts" distances, and the "contraction factor" K is the same regardless of the points used. The key fact we will need is the *Banach Contraction Principle*, which states that any contraction on a *complete* metric space has a unique fixed point, where a fixed point is a point $p \in X$ such that $f(p) = p$.

Here is the proof. The fact that the fixed point, once we know it exists, is unique is a simple consequence of the contraction property: if $f(p) = p$ and $f(q) = q$, so that p and q are both fixed points of f, then

$$d(f(p), f(q)) \leq Cd(p, q) \text{ comes } d(p, q) \leq Cd(p, q).$$

But $0 < C < 1$, so this inequality above holds forces $d(p, q) = 0$, so that $p = q$ and f has at most one fixed point.

To show that a fixed point exists, pick any $q \in X$. Consider the sequence of iterates

$$q, f(q), f(f(q)), \ldots, f^n(q), \ldots,$$

where $f^n(q)$ denotes the point obtained by applying f repeatedly n times. Then the contraction condition implies that

$$d(f^n(q), f^{n+1}(q)) \leq Cd(f^{n-1}(q), f^n(q)) \leq C^2 d(f^{n-2}(q), f^{n-1}(q)),$$

and inductively we get

$$d(f^n(q), f^{n+1}(q)) \leq C^n d(q, f(q)) \text{ for all } n.$$

If $d(q, f(q)) = 0$, then $q = f(q)$ so q is the fixed point we want. Otherwise, we can divide by $d(q, f(q))$ to get

$$\frac{d(f^n(q), f^{n+1}(q)}{d(q, f(q)} \leq C^n,$$

and the argument in the Warm-Up implies that the sequence of iterates $f^n(q)$ is Cauchy. (The additional factor of $d(q, f(q))$ in the denominator does not affect this argument since it is just a constant.) Since X is complete, these iterates converge to some $p \in X$.

Now, the contraction property implies that f is continuous, so since $f^n(q) \to p$ we have $f(f^n(q)) \to f(p)$. But $f(f^n(q))$ looks like

$$f(q), f(f(q)), f(f(f(q))), \ldots$$

and hence is a subsequence of $(f^n(q))$ itself and thus also converges to p. Since limits of a convergent sequence are unique, we get $f(p) = p$, and thus p is a fixed point of f.

Application to ODEs. Before moving on to the Inverse Function Theorem next time, we give a first application of this contraction principle in the theory of differential equations. Truth be told, this is one of my favorite applications of all time, since it illustrates the core of what mathematics is all about: take a problem that seems intractable at first, rephrase it in a way which opens up a whole new way of approaching it, and then use some known result in this new perspective to immediately solve your original problem.

The problem we consider is that of showing that the following differential equation with initial condition has a unique solution:

$$y'(x) = 3|y(x)| - \log(e^{\sin(\cos x)} + 1), y(1) = 1.$$

A solution in this case is a differentiable function y of x satisfying both the first condition on y' and the condition $y(1) = 1$ as well. If we had some simpler differential equation such as $y' = y$ we would know that any function of the form $y = ce^x$ was a solution, and specifying an initial condition would single out what the value of the constant c should be. However, here we have no hope of writing down an explicit solution of this differential equation due to its complicated nature, so we need a more clever approach to show that a function satisfying the above conditions exists.

The point is that we can rephrase the problem of solving this differential equation as a fixed point problem, and then figure out how to apply the contraction principle. After integrating both sides, we see that a function y satisfies

$$y'(x) = 3|y(x)| - \log(e^{\sin(\cos x)} + 1)$$

if and only if it satisfies

$$y(x) = c + \int_1^x [3|y(t)| - \log(e^{\sin(\cos t)} + 1)]\, dt$$

for some constant c. Indeed, if y is continuous, the Fundamental Theorem of calculus implies that the integral expression on the right is differentiable and that its derivative is the integrand evaluated at $t = x$; thus taking derivatives of both sides indeed gives our original differential equation. (Note that if we assume only that y is continuous, it might not be clear that the derivative of the left side y even exists, but the point is that it will as a consequence of the fact that this left side equals the differentiable expression given on the right side.) The constant c is determined by the initial condition $y(1) = 1$: since an integral from 1 to 1 is always zero, we need $c = 1$ in order to have $y(1) = 1$. Thus, the upshot is that a function f satisfies

$$y'(x) = 3|y(x)| - \log(e^{\sin(\cos x)} + 1) \text{ with initial condition } f(1) = 1$$

if and only if it satisfies the integral equation:

$$y(x) = 1 + \int_1^x [3|y(t)| - \log(e^{\sin(\cos t)} + 1)]\, dt.$$

So our goal is now to show that there is a function satisfying this integral equation. Indeed, consider the metric space $C[a, b]$ of continuous functions $[a, b] \to \mathbb{R}$, for some to-be-determined interval $[a, b] \ni 1$, equipped with the sup metric. Define the map $T : C[a, b] \to C[a, b]$ by setting, for each $f \in C[a, b]$, Tf to be the function on $[a, b]$ whose value at $x \in [a, b]$ is:

$$(Tf)(x) = 1 + \int_1^x [3|f(t)| - \log(e^{\sin(\cos t)} + 1)]\, dt.$$

Note here that this function Tf is continuous since the function defined by the integral on the right is continuous. Then, saying that y satisfies

$$y(x) = 1 + \int_1^x [3|y(t)| - \log(e^{\sin(\cos t)} + 1)]\, dt$$

is the same as saying that the function Ty equals y itself! Thus, showing that this integral equation has a solution for y is the same as showing that this map T has a fixed point!!!

To show that T has a fixed point, we show that we can arrange for it to be a contraction by choosing $[a,b]$ appropriately. (Recall that $C[a,b]$ is complete with respect to the sup metric, so that the Banach contraction principle will indeed be applicable.) For $f,g \in C[a,b]$ and $x \in [a,b]$, we have that $|(Tf)(x) - (Tg)(x)|$ equals

$$\left|\left(1 + \int_1^x [3|f(t)| - \log(e^{\sin(\cos t)} + 1)]\,dt\right) - \left(1 + \int_1^x [3|g(t)| - \log(e^{\sin(\cos t)} + 1)]\,dt\right)\right|,$$

which simplifies to

$$\left|\int_1^x 3(|f(t)| - |g(t)|)\,dt\right|.$$

Then

$$\begin{aligned}
|(Tf)(x) - (Tg)(x)| &= \left|\int_1^x 3(|f(t)| - |g(t)|)\,dt\right| \\
&\le \int_{\min\{1,x\}}^{\max\{1,x\}} 3||f(t)| - |g(t)||\,dt \\
&\le \int_{\min\{1,x\}}^{\max\{1,x\}} 3|f(t) - g(t)|\,dt
\end{aligned}$$

where we use $||p| - |q|| \le |p - q|$ at the end. (Note that we must use the min and max of 1 and x in the integral bounds since we allow for both $x < 1$ and $x > 1$ as possibilities: $x \in [a,b]$ and $[a,b]$ contains 1, but we cannot be sure about the relation between 1 and x.) But for each $t \in [a,b]$, $|f(t) - g(t)| \le d(f,g)$ since $d(f,g)$ is the supremum of such expressions as t varies throughout $[a,b]$, and thus:

$$\begin{aligned}
|(Tf)(x) - (Tg)(x)| &\le \int_{\min\{1,x\}}^{\max\{1,x\}} 3|f(t) - g(t)|\,dt \\
&\le \int_{\min\{1,x\}}^{\max\{1,x\}} 3d(f,g)\,dt \\
&= 3|1 - x|d(f,g) \\
&\le 3(b-a)d(f,g),
\end{aligned}$$

where we use in the last step that the fact that $1, x \in [a,b]$.

Hence the number $3(b-a)d(f,g)$ is an upper bound for all expressions $|(Tf)(x) - (Tg)(x)|$ as x varies in $[a,b]$, so $3(b-a)d(f,g)$ is larger than or equal to the supremum of such expresions, which is the sup distance $d(Tf,Tg)$:

$$d(Tf,Tg) \le 3(b-a)d(f,g).$$

Thus, as long as $a < b$ satisfy $3(b-a) < 1$, the map T as defined above is a contraction from $C[a,b]$ to itself. Since $C[a,b]$ is complete, this gives a unique fixed point y of T, which, as discussed before, gives a unique solution of our original differential equation. Note that this solution is in the end only guaranteed to exist on the interval $[a,b] \ni 1$ since it is only on this interval that we got the necessary contraction property, so this should actually be included in the original statement of our differential equation problem: the given differential equation with initial condition has a unique solution defined on *some* interval containing the initial point 1.

Picard iteration. The work above only shows that the given differential equation has a solution, but we can still ask whether it is possible to determine what the solution concretely is. Although we cannot determine the solution explicitly, it turns out we can in fact approximate the solution however well we want. The key is in the proof of the Banach Contraction Principle: the fixed point is obtained as the limit of the sequence defined by starting with any point and repeatedly applying contraction. Thus, in our case, if we start with any continuous function $g : [a, b] \to \mathbb{R}$ (say a constant one), the functions obtained by repeatedly applying T:

$$g,\ T(g),\ T(T(g)),\ \ldots,$$

will eventually provide better and better approximations of the fixed point of T, and hence of the solution of the differential equation we're interested in. This approximation method is known as *Picard iteration*, and is a basic technique for approximating solutions of differential equations. (The general version of the "existence and uniqueness theorem" for differential equations is known as the Picard-Lindelöf theorem, and you will prove it on the homework.)

To see this all in action in a simpler example, consider the following initial value problem:

$$y' = y,\ y(0) = 1.$$

Of course, we know that the solution should be $y = e^x$. But, let us see what happens using the iterates described above. A function y satisfying this initial value problem is the same as one satisfying

$$y(x) = 1 + \int_0^x y(t)\,dt.$$

Take the right side to define a function Ty, where T is then some mapping from continuous functions to continuous functions. Take the constant function 1, and consider the sequence of iterates

$$1,\ T1,\ T^2 1,\ T^3 1,\ \ldots.$$

The function $T1$ is

$$(T1)(x) = 1 + \int_0^x 1\,dt = 1 + x.$$

The next iterate is obtained by substituting $(T1)(x) = 1 + x$ in place of y in $1 + \int_0^x y(t)\,dt$, so it is

$$(T^2 1)(x) = 1 + \int_0^x (1 + t)\,dt = 1 + t + \frac{x^2}{2}.$$

The next iterate is

$$(T^3 1)(x) = 1 + \int_0^x (1 + t + \tfrac{t^2}{2})\,dt = 1 + x + \tfrac{x^2}{2} + \tfrac{x^3}{3!}.$$

And so on, you might recognize that the the iterates being produced are precisely the Taylor polynomials of e^x centered at 0, which do indeed converge to e^x!

Thus, the proof of the Banach Contraction Principle gives the expected answer in this case: the iterates $1, T1, T^2 1, \ldots$ converge to the solution of $y' = y, y(0) = 1$. Most amazingly, we could have chosen any function to start with apart from 1, and this would still be true; for examples, the iterates

$$\sin x,\ T(\sin x),\ T^2(\sin x),\ T^3(\sin x), \ldots$$

will *also* converge to e^x, and these are a lot more complicated than the Taylor polynomials above!

Lecture 3: Inverse Functions

TO BE FINISHED

Lecture 4: Implicit Functions

TO BE FINISHED

Lecture 5: Step Functions

TO BE FINISHED

Lecture 6: Regulated Integrals

TO BE FINISHED

Lecture 7: Outer Measure

Warm-Up. We show that uniform limits of regulated functions are regulated. ***TO BE FINISHED***

Measures. We now seek to define a "better" of notion of integration. As we've discussed before, this better notion will come from a general notion of "length" of subsets of $\mathbb{R}$, so that's where we will start. We want to define a *measure* μ, which at first we will consider to be a function

$$\mu : 2^{\mathbb{R}} \to [0, \infty].$$

Here $2^{\mathbb{R}}$ denotes the *power set of* $\mathbb{R}$, which is the set of all subsets of $\mathbb{R}$. (If you've never seen this notation before, no big deal, just treat it literally as denoting the set of all subsets of $\mathbb{R}$. Ask me elsewhere if you'd like to see why this is indeed the correct notation to use for this set of subsets!) We will interpret the value $\mu(A)$ of this function on $A \subseteq \mathbb{R}$ as the "length" of A. Note that we allow for this value to be infinite, as will be the case with $\mu(\mathbb{R}) = \infty$ for example.

Now, if we want this to match our intuition as to what length should be, we will ask that μ satisfy the following properties:

- For any bounded interval I, $\mu(I) = \text{len}(I)$. (So, μ gives the answer we expect for intervals.)
- For any $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$, $\mu(A + x) = \mu(A)$, where $A + x$ is the set of translates of elements of A by x. (We say that μ is *translation invariant*, so that "moving" a set from one location to another shouldn't change its "length".)
- If $A \subseteq B$, then $\mu(A) \le \mu(B)$. (We refer to this property as *monotonicity*, so that the "length" of a subset is no larger than the length of the larger set.)
- If $A_1, A_2, A_3, \ldots$ is a countable collection of subsets of $\mathbb{R}$, then

  $$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) \le \sum_{k=1}^{\infty} \mu(A_k).$$

  We refer to this property as *countable subadditivity.* Moreover, if the A_k are pairwise disjoint, then the inequality above should actually be an equality, which we call *countable additivity.*

And there you have it. Note that all of these properties are indeed true if we consider only intervals as the subsets of $\mathbb{R}$ we care about, so it is reasonable to expect that they hold more generally. Alas, we will soon argue that there is no such function μ defined on the entirety of $2^{\mathbb{R}}$, and that if we want to demand that these properties hold, we will have to restrict the domain of μ to only certain subsets. In terms of a theory of integration, however, this restriction will be a fairly mild one and won't cause any real issues.

Null/zero sets. To get a sense of what such a "measure" might look like, we first give a name to those sets that will have "measure zero". We say that $A \subseteq \mathbb{R}$ is a *null set* (also called a *zero set*) if for every $\epsilon > 0$, there exists a countable covering $\{I_k\}$ of A by open intervals such that

$$\sum_n \operatorname{len}(I_k) < \epsilon.$$

(To be clear, when we say "countable covering" we mean a covering by countably many sets, not one so that the union of all sets in that covering is a countable set itself.) The intuition is that whatever $\mu(A)$ should be, it should satisfy

$$\mu(A) \leq \mu(\bigcup_k I_k) \leq \sum_k \mu(I_k),$$

where the first inequality is monotonicity and the second is countable subadditivity. But μ gives length on intervals, so the final sum is $\sum_k \operatorname{len}(I_k) < \epsilon$, so we should have $\mu(A) < \epsilon$. Since this will hold for all $\epsilon > 0$, $\mu(A)$—if it is defined—will be zero.

Here are some examples. First, any finite set is a null set. Indeed, if $A = \{x_1, \ldots, x_n\}$ has n elements, for any $\epsilon > 0$ pick an open interval I_k of length $\frac{\epsilon}{2n}$ around x_k. Then the I_k cover A and

$$\sum_{k=1}^n \operatorname{len}(I_k) = \sum_{k=1}^n \frac{\epsilon}{2n} = \frac{\epsilon}{2},$$

which is less than ϵ. Hence A is a null set.

More generally, any countable set is a null set. (So, for instance, $\mathbb{Q}$ is a null set. This is actually an argument we gave back on the first day of class when motivating the desire to have a more general theory of integration.) Suppose A is countable and enumerate its elements as

$$r_1,\ r_2,\ r_3,\ \ldots.$$

Fix $\epsilon > 0$ and for each k pick an open interval I_k around r_k of length less than $\frac{\epsilon}{2^k}$. Then

$$\sum_{k=1}^\infty \operatorname{len}(I_k) < \sum_{k=1}^\infty \frac{\epsilon}{2^k} = \epsilon,$$

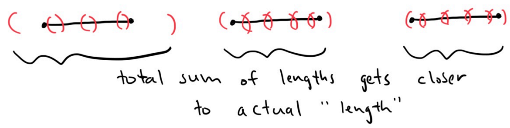
which shows that A is a null set. (Any series such that $\sum_k a_k = 1$ could have been used here instead of $\sum_k \frac{1}{2^k} = 1$, but this particular geometric series is a fairly standard choice. Take note of this "$\frac{\epsilon}{2^k}$" argument since it tends to show up quite often in measure theory, analogously to how "$\frac{\epsilon}{2}$" arguments show up often when discussing sequences, limits, and other things.) A similar argument to this shows that countable unions of null sets are always null sets, where again by "countable union" we mean a union of countably many null sets, not countability of the resulting union.

At this point one might ask whether there are any null sets which are not countable? Sure: the Cantor set is a null set, as we will show next time.

Outer measure. Finally we come to main definition of interest, that we've been building up towards. We define the *Lebesgue outer measure* of $A \subseteq \mathbb{R}$ to be

$$\mu^*(A) := \inf\left\{ \sum_k \text{len}(I_k) \;\middle|\; \{I_k\} \text{ is a countable covering of } A \text{ by open intervals} \right\}.$$

That is, we cover A by arbitrarily many countable intervals, and take the total sum of their lengths as an "overestimate" of the "length" of A. As we vary the covering we use, and can get this overestimate to be smaller and smaller, and we then take the infimum of such values as a measure of the *actual* "length" of A.



This is an "outer" measure since we are "approximating" from the "outside", by using open intervals which cover A. (There is a related notion of the Lebesgue *inner measure* $\mu_*(A)$ where we "approximate" A from the "inside" that we will mention later, as it takes some care to define correctly.) Note that $\mu^*(A)$ always exists for any $A \subseteq \mathbb{R}$ if we allow ∞ as a valid value.

So, this outer measure $\mu^*(A)$ is the "measure" we want, and intuitively does seem to capture the "length" of A by approximating it using lengths of intervals. We can verify that outer measure does indeed have at least some of the properties we laid out earlier that we want "measure" to satisfy. First, simply by its definition as an infimum, outer measure satisfies monotonicity

$$A \subseteq B \implies \mu^*(A) \leq \mu^*(B)$$

because $\mu^*(B)$ is the infimum of a subset of the things of which $\mu^*(A)$ is the infimum, since any open cover of B is also an open cover of A. Also, μ^* is translation invariant: the intervals in an open cover of A can be translated to get ones in an open cover of $A + x$, and vice versa, and translating intervals for sure does not alter their lengths. Thus, $\mu^*(A + x)$ and $\mu^*(A)$ are literally the infimum of the *same* set. (To be clear, we mean that the set of total sums of lengths $\sum_k \text{len}(I_k)$ for both are the same, not that the intervals I_k used to produce these sums are the same.)

Vitali sets. So, μ^* is monotone and translation invariant, and we will see soon enough that it gives the correct value on intervals and that it is countably subadditive. But, and this is a big but, it is not additive on all possible countable collections of pairwise disjoint sets. Here is the standard example that shows what can go wrong.

Consider the collection $\{\mathbb{Q}+x \mid x \in \mathbb{R}\}$ of sets of *rational* translates of real numbers. (For those of you who have had a course in group theory before, we are considering the *cosets* of the subgroup $\mathbb{Q}$ of $\mathbb{R}$.) These sets of rational translates partition $\mathbb{R}$ into disjoint pieces: the real number $x \in \mathbb{R}$ belongs to $\mathbb{Q} + x$ if nothing else, and two sets of translates $\mathbb{Q} + x$ and $\mathbb{Q} + y$ have some nonempty overlap only if they are completely the same. Indeed, if $z \in \mathbb{Q} + x$ and $z \in \mathbb{Q} + y$, then $z = r_1 + x$ and $z = r_2 + y$ for some $r_1, r_2 \in \mathbb{Q}$, but then any $r + x \in \mathbb{Q} + x$ (with $r \in \mathbb{Q}$) can be written as

$$r + x = r + (z - r_1) = r + (r_2 + y - r_1) = (r + r_2 - r_1) + y$$

where $r + r_2 - r_1 \in \mathbb{Q}$, showing that $r + x \in \mathbb{Q} + y$ as well. Hence $\mathbb{Q}+x \subseteq \mathbb{Q}+y$ if $(\mathbb{Q}+x) \cap (\mathbb{Q}+y)$ is nonempty, and the same argument shows that the opposite containment holds as well. To get a feel for what this partition is doing, consider the decimal expansion of π:

$$\pi = 3.14159...$$

Any number which has a rational difference with π will belong to the same partition class as π, so for example

$$\pi - 3 = 0.14159..., \ \pi - 3.1 = 0.04159..., \ \pi - 3.14 = 0.00159...$$

all belong to the same partition class as π. In a sense, this partition class keeps track of only the "tail" of the decimal expansion of π, and more generally the "tail" of any rational difference of π. This partition says to essentially ignore the effect of all rational numbers when adding or subtracting, and only consider real numbers "up to" such rational differences.

Now, by subtracting the "integer part" of a real number, any class $\mathbb{Q}+x$ will intersect $[0,1]$. For each partition class, pick such an element of $[0,1]$, and define the *Vitali set* to be the set $V \subseteq [0,1]$ of these chosen elements. (There are some set-theoretic issues here which should be noted, since it is not clear how to "pick such an element" for the partition classes all at once. The reason why we can do this is called the *Axiom of Choice*, but this is not something we will delve into more in this course. Ask in office hours if you'd like to learn more about this!) By construction, the sets of rational translates of two elements in the Vitali set are completely disjoint (since these two elements were meant to come from different partition classes), and every real number is the rational translate of an element (uniquely!) of the Vitali set since such sets of translates are meant to give all partition classes above. To get a sense for what V might consist of, perhaps for

$$\pi = 3.14159...$$

we will pick $\pi - 3 = 0.14159... \in [0,1]$ to be the representative of the partition class to which π belongs, while for

$$e = 2.71828...$$

perhaps we pick $e - 2.7 = 0.01828...$ as the element in the Vitali set.

The Vitali set is a strange set to wrap your head around, and indeed it is so strange that we claim it does *not* have a well-defined notion of "length"! To be clear, the Vitali set certainly has an outer measure $\mu^*(V)$, as all subsets of $\mathbb{R}$ do, but the point is that this particular outer measure value does not behave in the way you would ordinarily expect of "length" in general. The issue is that V is, in a sense, too "weirdly distributed" throughout $[0,1]$ that it makes any attempt to measure its "length" by approximating intervals fail. (The Vitali set is a 1-dimensional analog of a "fractal" in the plane, which is also an object for which standard geometric measurements—like area—don't quite make sense.) To see this, enumerate the rationals in $[-1,1]$ as

$$r_1, \ r_2, \ r_3, \ \dots$$

and consider the translates $V + r_k$. These translates are all pairwise disjoint since elements of the Vitali set were chosen to give distinct partition classes above, and we have that the union of these translates contains all of $[0,1]$ since each element of $[0,1]$ belongs to the same partition class as an element of V, meaning that each element of $[0,1]$ is the translate of an element of V by a rational between -1 and 1. (The restriction to rationals between -1 and 1 comes from the fact that V is itself a subset of $[0,1]$, so that the difference between an element of V and an element of V falls in $[-1,1]$.) Since $r_k \in [-1,1]$, we have $V + r_k \subseteq [0,1] + r_k \subseteq [-1,2]$, so altogether we get

$$[0,1] \subseteq \bigcup_{k=1}^{\infty} (V + r_k) \subseteq [-1,2].$$

By monotonicity of outer measure, this gives

$$\mu^*([0,1]) \leq \mu^* \left(\bigcup_{k=1}^{\infty} (V + r_k) \right) \leq \mu^*([-1,2]).$$

If countable additivity held in this case, the term in the middle would split up as the sum of the individual $\mu^*(V+r_k)$ since the $V+r_k$ are pairwise disjoint, so we would have

$$\mu^*([0,1]) \leq \sum_{k=1}^{\infty} \mu^*(V+r_k) \leq \mu^*([-1,2]).$$

By translation invariance, $\mu^*(V+r_k) = \mu^*(V)$ for all k, and the outer measure of an interval is its length, so the inequalities above become

$$1 \leq \sum_{k=1}^{\infty} \mu^*(V) \leq 3.$$

But now this is a problem: if $\mu^*(V) = 0$, then the sum in the middle is zero and does not fall between 1 and 3, while if $\mu^*(V) > 0$, the sum in the middle is infinite, and so again is not between 1 and 3. The conclusion is that countable additivity fails in this case, so that $\mu^*(V) = \mu^*(V+r_k)$ does not behave as a usual "length".

The upshot is that if we want outer measure μ^* to indeed give a general notion of "length" satisfying all properties we expect—including additivity—we will have to restrict the types of sets of which we can take the "length". The Vitali set is the standard example of a *nonmeasurable* set, which is a concept we will define next time and spend a few good days digesting.

Lecture 8: Measurability

Warm-Up 1. We show that the Cantor set is a null set.

Warm-Up 2. We show that the outer measure of a bounded interval is its length.

Sigma algebras.

Measurable sets.

Clean intersections.

Lecture 9: More on Measurability

Warm-Up. We show that A is measurable if and only if

$$\mu^*(A \cap X) + \mu^*(A^c \cap X) = \mu^*(X)$$

for all X.

Subaddtivity.

Null sets are measurable.

Rays are measurable.

Lecture 10: Sigma Algebras

Lecture 11: Regularity

Lecture 12: Simple Functions

Lecture 13: Measurable Functions

Lecture 14: Lebesgue Integration

Lecture 15: Bounded Convergence

Lecture 16: Unbounded Functions

Lecture 17: Absolute Continuity

Lecture 18: Dominated Convergence

Lecture 19: General Functions

Lecture 20: L^p Spaces

Warm-Up.

Towards functional analysis. We now move to the final portion of our course, where we will focus on studying *L^p spaces*. L^p spaces are spaces of functions with metrics defined by integrals, and are the prototypical examples of *Banach spaces*, which we will define later. The study of Banach spaces in general make up a large portion of the subject of *functional analysis*, which we will see glimpses of. Of particular interest will be the space L^2, which is an example of a *Hilbert space*, and is where the theory of Fourier series (which we will revisit as needed) finds its natural home.

We should note that our textbook *only* covers the structure of L^2 and does not mention more general L^p spaces. However, we will see that many of the results the book describes for L^2 have natural counterparts for L^p, with comparable proofs. Thus, we will take the perspective that we should prove what we can for L^p in general, and only specialize to the L^2 case when it is absolutely necessary. Indeed, to get a real sense of what "functional analysis" is all about, the general L^p is a much more illustrative example than L^2 alone. The fact that the functional analysis of L^2 is much simpler than that for L^p in general *is* the reason why the theory of Fourier series is even possible, but of course this statement would make no sense if we didn't know what L^p was in general.

Functions on finite sets. As a first basic introduction to what we aim to do, let us consider the finite-dimensional version of L^p. Take $X = \{1, 2, \ldots, n\}$ equipped with the counting measure, where the measure of a subset of X is the number of elements within it. (We consider here the σ-algebra of *all* subsets of X.) We define $L^2(X)$ to be the set of all integrable functions $X \to \mathbb{R}$, which, due to the nature of the counting measure (which implies that all functions $X \to \mathbb{R}$ are measurable) and the fact that X is finite (which implies the integral of every function with respect to counting measure is finite), is just the set of *all* functions $X \to \mathbb{R}$.

But, what is a function from $X = \{1, 2, \ldots, n\}$ to $\mathbb{R}$? Such a function f is completely determined by the values of $f(1), f(2), \ldots, f(n)$, and so is fully characterized by the element $(f(1), f(2), \ldots, f(n))$ of $\mathbb{R}^n$. Thus, as a set, we have that

$$L^2(\{1, 2, \ldots, n\}) = \mathbb{R}^n.$$

(This is why $L^2(\{1,2,\ldots,n\})$ is a finite-dimensional version of the other L^p spaces we will soon introduce, which should be viewed as infinite-dimensional analogs of $\mathbb{R}^n$.)

Now, the 2 in L^2 refers to a certain norm, namely the L^2-*norm*. We define this as

$$\|f\|_2 := \left(\int_{\{1,2,\ldots,n\}} |f|^2\,d\mu\right)^{1/2}.$$

To be clear, this is an integral taken with respect to counting measure (for now), which are just ordinary sums: evaluate the integrand at each point in the domain of integration and add them all together. Thus, in this particular case, we take the value of $|f(1)|^2$, plus the value of $|f(2)|^2$, plus the value of $|f(3)|^2$, and so on, taking the square root of the result:

$$\|f\|_2 = (|f(1)|^2 + \cdots + |f(n)|^2)^{1/2}.$$

Thus, this is nothing but a fancy way of talking about the usual norm operation on $\mathbb{R}^n$, and we then get a metric—the L^2-*metric*—by

$$d_2(f,g) := \|f-g\|_2 = \sqrt{|f(1)-g(1)|^2 + \cdots + |f(n)-g(n)|^2},$$

which is the standard Euclidean metric on $L^2(\{1,2,\ldots,n\}) = \mathbb{R}^n$.

But there are other norms and hence metrics we can consider on this same set of functions $\{1,2,\ldots,n\} \to \mathbb{R}$. The L^1-norm is defined by

$$\|f\|_1 := \int_{\{1,2,\ldots,n\}} |f|\,d\mu = |f(1)| + \cdots + |f(n)|,$$

and the L^1-metric is then

$$d_1(f,g) := \|f-g\|_1 = |f(1)-g(1)| + \cdots + |f(n)-g(n)|.$$

(When considering this particular norm we use the notation $L^1(\{1,2,\ldots,n\})$ for our space.) This is also known as the *taxicab metric*, since it measures the distance from one point in $\mathbb{R}^n$ to another when moving only horizontally and vertically, just as a taxicab on the streets would. A basic fact you can check (which my MATH 321-1 class in the fall did!) is that d_2 and d_1 actually determine precisely the same open sets, and the same notion of convergence. More generally, for $p \geq 1$ we can consider the L^p-metric on $\{1,2,\ldots,n\} \to \mathbb{R}$ defined by

$$d_p(f,g) = \left(\int_{\{1,2,\ldots,n\}} |f-g|^p\,d\mu\right)^{1/p} = (|f(1)-g(1)|^p + \cdots + |f(n)-g(n)|^p)^{1/p},$$

which *also* determines the same notion of open/convergence as d_1 or d_2, and hence the same version of *analysis*. (The fact that the triangle inequality for the L^p-metric on $\mathbb{R}^n$ holds is far from obvious, and takes some real effort to justify. For this it is important that $p \geq 1$, as the triangle inequality does not hold when $0 < p < 1$.)

p-norms. General L^p spaces are defined by the exact same setup as above, only by replacing X by some other domain and replacing μ by some other measure. For the most part, we will only care about $X = [a,b]$ and Lebesgue measure μ, but the theory works in pretty much the same way in other settings. For X an interval equipped with Lebesgue measure and $p \geq 1$, we thus define

$L^p(X)$ to be the set of functions $f : X \to \mathbb{R}$ such that $|f|^p$ is integrable, which means that $|f|^p$ is measurable and

$$\int_X |f|^p \, d\mu < \infty.$$

For example, $L^1(X)$ is then just the set of Lebesgue integrable functions in the usual sense (recall that f is integrable as a function of both negative and positive values if and only if $|f|$ is integrable as a nonnegative function), and $L^2(X)$ is the set of so-called *square-integrable* functions, for which $\int f^2 \, d\mu < \infty$. (Note that $\frac{1}{\sqrt{x}} \in L^1([0,1])$ and $\frac{1}{\sqrt{x}} \notin L^2([0,1])$, for example.)

For $f \in L^p(X)$, we define its L^p*-norm*, or simply p*-norm* for short, by

$$\|f\|_p := \left(\int_X |f|^p \, d\mu \right)^{1/p}.$$

Note that this definition makes sense since the integral used is a finite nonnegative number. This definition then mimics and generalizes the definition of the Euclidean metric on $\mathbb{R}^n$ (i.e. $L^2(\{1, 2, \ldots, n\})$ with counting measure, or other p-norms in finite dimensions. One big difference is that, although different p-norms on $\mathbb{R}^n$ gave the same versions of "analysis", this will not be true in infinite-dimensions, and different choices of p can lead to different behaviors.

Taking equivalence classes. We want to use the norm above to get a metric on $L^p(X)$, where we will define the distance between f and g to be $\|f - g\|_p$. But there is one subtle issue here which arises from the fact we are using the Lebesgue integral in defining L^p. Namely, we want it to be true that

$$\|f - g\|_p = 0 \iff f = g$$

as expected for a metric, but in fact we know that

$$\int_X |f - g|^p \, d\mu = 0 \iff f - g = 0 \text{ only } \textit{almost everywhere}.$$

That is, f and g can have zero "distance" between them without being the same, as long as they are the same off a set of measure zero since this is all the integral detects. For example, a function which is zero except at a finite number of points will still have norm 0, and hence distance 0 from the constant zero function.

To get around this, we simply declare by fiat that "almost everywhere" equality is *literal* equality. That is, we will treat f and g as being the *same* element of L^p if they agree almost everywhere. To be precise, the property of being the "same almost everywhere" defines an equivalence relation on sets of functions, and we are taking the equivalence classes. Thus, officially, $L^p(X)$ is defined to be the set of equivalence classes of functions satisfying $\int_X |f|^p \, d\mu < \infty$, where f and g are equivalent if $f = g$ almost everywhere. With this modification, we then do have

$$\|f\|_p = 0 \iff f = 0,$$

simply because the notation "$f = 0$" in this context literally means "$f = 0$ almost everywhere". On this space of equivalence classes, $\|f - g\|_p$ does give an honest metric, as we'll see.

Defining $L^p(X)$ to be made up of equivalence classes might seem like a jump in abstraction at first, but actually we will usually push these details aside and simply think of elements of $L^p(X)$ as if they were functions in the normal sense. Technically, though, evaluating an element at a point doesn't quite make sense, since for example the value of $f(0)$ depends on which f we pick from the respective equivalence class; we can always change the specific value of $f(0)$ to be whatever we want without changing the element of $L^p(X)$ we are considering. What *does* make sense is integration, since the value $\int_X f\,d\mu$ is independent of which f we use in the sense that if $f = g$ almost everywhere, so that they give the same element of $L^p(X)$, then $\int_X f\,d\mu = \int_X g\,d\mu$. Thus, the fact that elements of $L^p(X)$ are not quite honest functions won't play a big role in our discussion. (We will see later that, nonetheless, we *can* in a sense evaluate elements of $L^p(X)$ at certain points, essentially off a set of measure zero. This will come up when showing that L^p is complete.)

Lecture 21: Hölder and Minkowski

Warm-Up. Define $L^\infty(X)$ to be the set of *essentially bounded* functions on X, which are functions that are bounded almost everywhere. (As with the other L^p spaces, we really take equivalence classes of essentially bounded functions, so that $f = g$ almost everywhere means that f and g are the same.) Define the L^∞-*norm* of $f \in L^\infty(X)$ to be the infimum of all essential bounds on M:

$$\|f\|_\infty := \inf\{M \mid M \text{ is an essential bound on } f\}.$$

(This is usually called the *essential supremum* of f, and gives the supremum of f away from a set of measure zero. In other words, it's the supremum of the portion of f which *really* matters.)

We show that $\|f\|_\infty$ is itself an essential bound on f. This is the L^∞ analog of the fact that the infimum of upper bounds of a bounded function—namely its supremum—is itself an upper bound. In this latter case, the fact that this infimum is an upper bound is shown by saying that $|f(x)| \leq M$ for any x and upper bound M of f, so that each $|f(x)|$ is a lower bound on the set of upper bounds, and hence must be smaller than or equal to the infimum of upper bounds. Such an argument doesn't work in the "essential" setting, however, since we cannot guarantee that $|f(x)| \leq M$ for any one specific x, which might lie in the null set on which M does not bound f.

So, we argue as follows. For each $n \in \mathbb{N}$, by definition of infimum there exists an essential bound M of f such that $M < \|f\|_\infty + \frac{1}{n}$, so that

$$|f(x)| \leq M < \|f\|_\infty + \frac{1}{n}$$

off a set E_n of measure zero. Thus for $x \in \bigcap_n E_n^c = (\bigcup_n E_n)^c$, we have

$$|f(x)| \leq \|f\|_\infty + \frac{1}{n}$$

for all n, which implies that $|f(x)| \leq \|f\|_\infty$. Since each E_n is a null set, $\bigcup_n E_n$ is also a null set, so this shows that $\|f\|_\infty$ is an essential bound on f.

Conjugate pairs. As we develop the theory of L^p spaces, we will see that certain such spaces are naturally related to others. To make this (eventual) idea precise, let us introduce now the notion of a *conjugate* pair (sometimes called a *Hölder conjugate* pair) of numbers. For $p > 1$, we define its conjugate to be the number $q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Moreover, we define the conjugate of 1 to be ∞, and the conjugate of ∞ to be 1, which is motivated by the idea that $\frac{1}{\infty}$ should, in a sense, be "zero", so that $1 + \frac{1}{\infty} = 1$ is true. (Of course, the real reason why we take ∞ to be conjugate to 1 has to do with the relations between L^1 and L^∞ we will see later. The "motivation" in terms of $\frac{1}{\infty}$ is only to highlight the relation between p and q in the case of finite numbers.)

At this point it is not clear at all why the number q defined by $\frac{1}{p}+\frac{1}{q}=1$ should be of any special interest, but we will very soon see why. Note that 2 is its own conjugate (since $\frac{1}{2}+\frac{1}{2}=1$), a fact which underlies much the special structure of L^2 we will explore later.

Hölder's inequality. The inequality we will now state and prove, *Hölder's inequality*, is arguably the most important inequality in the entire theory of L^p spaces. Indeed, for now at least it will be crucial to proving the triangle inequality for the L^p-metric, which is defined as

$$d_p(f,g) := \|f-g\|_p = \left(\int_X |f-g|^p \, d\mu\right)^{1/p}.$$

The point is that the triangle inequality for this soon-to-be metric is not at all obvious, since it amounts to saying that

$$\left(\int_X |f-g|^p \, d\mu\right)^{1/p} \leq \left(\int_X |f-h|^p \, d\mu\right)^{1/p} + \left(\int_X |f-h|^p \, d\mu\right)^{1/p}$$

for any $f,g,h \in L^p$. There is no simple algebraic manipulation of integrals and exponents that immediately gives rise to this for all $p \geq 1$, without the use of Hölder's inequality.

Here's *Hölder*: If $f \in L^p$ and $g \in L^q$, with $p,q \geq 1$ conjugates, then $fg \in L^1$ and

$$\|fg\|_1 \leq \|f\|_p \|g\|_q .$$

Thus, the product of a p-power integrable function and a q-power integrable function is integrable in the usual L^1 sense, and the L^1-norm is bounded by the product of the p- and q-norms. If you write what this says in terms of integrals, it looks like

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

which again is highly nonobvious since there is no straightforward way of manipulating the exponents and roots. Ultimately, Hölder's inequality comes from monotonicity of the Lebesgue integral together with some non-obvious inequalities.

We will take the following inequality for granted: If $A,B>0$ and $\theta>0$, then

$$A^\theta B^{1-\theta} \leq \theta A + (1-\theta)B.$$

Note that when $\theta = \frac{1}{2}$, this becomes the standard *arithmetic-geometric mean inequality*:

$$\sqrt{AB} \leq \frac{1}{2}(A+B),$$

so $A^\theta B^{1-\theta} \leq \theta A + (1-\theta)B$ can be viewed as a generalization of this to other exponents. This general version can be proved using convexity of the log function, or by optimizing some appropriate single-variable function using a derivative. You can easily find the details elsewhere if interested.

First, if $\|f\|_p = 0$ or $\|g\|_q = 0$, then f or g are zero almost everywhere, and hence so is fg, so that fg is integrable and has 1-norm zero. Thus Hölder's inequality holds in this case. Suppose now that $\|f\|_p = 1 = \|g\|_q$. Set $A = |f(x)|^p$, $B = |g(x)|^q$, and $\theta = 1/p$ (so that $1-\theta = 1/q$) in this inequality to get

$$|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q \text{ for all } x.$$

(Incidentally, this is called *Young's inequality.* It seems that every useful inequality in analysis has its own name!) Integrating both sides (monotonicity for the win!) gives

$$\int_X |fg|\,d\mu \le \frac{1}{p}\int_X |f|^p\,d\mu + \frac{1}{q}\int_X |g|^q\,d\mu.$$

But the integrals on the right are equal to $\|f\|_p = 1$ and $\|g\|_q = 1$ respectively, since we are assuming these norms equal 1:

$$1 = \|f\|_p = \left(\int_X |f|^p\,d\mu\right)^{1/p} \iff 1 = \|f\|_p^p = \int_X |f|^p\,d\mu.$$

Thus, we have

$$\int_X |fg|\,d\mu \le \frac{1}{p}\int_X |f|^p\,d\mu + \frac{1}{q}\int_X |g|^q\,d\mu = \frac{1}{p}(1) + \frac{1}{q}(1) = \frac{1}{p} + \frac{1}{q} = 1.$$

Hence $\|fg\|_1 \le 1 = \|f\|_p\,\|g\|_q$, so Hölder's inequality holds in this case.

Finally, for the general case with $\|f\|_p\,, \|g\|_q > 0$, consider $\frac{f}{\|f\|_p} \in L^p$ and $\frac{g}{\|g\|_q} \in L^q$. Each of these have respective p- or q-norm 1, so by the case above we get

$$\left\|\frac{f}{\|f\|_p}\frac{g}{\|g\|_q}\right\|_1 \le \left\|\frac{f}{\|f\|_p}\right\|_p \left\|\frac{g}{\|g\|_q}\right\|_q = 1.$$

Multiplying through by the positive scalar $\|f\|_p\,\|g\|_q$ then gives Hölder's inequality $\|fg\|_1 \le \|f\|_p\,\|g\|_q$.

L^p is a vector space. In the argument above we used the fact that scaling an element of L^p still gives an element of L^P when saying that, for example, $f/\|f\|_p$ is in L^p. This comes from the fact that

$$\int_X |f|^p\,d\mu < \infty \iff \int_X |cf|^p\,d\mu = |c|^p\int_X |f|^p\,d\mu < \infty.$$

Thus, L^p is closed under scalar multiplication.

In fact, it is also closed under addition, so that L^p is actually a *vector space.* To see this, let $f, g \in L^p$. (The same p now!) For any x, we have

$$|f(x) + g(x)|^p \le (|f(x)| + |g(x)|)^p$$

by the usual triangle inequality for the absolute value. Now, if we let $h(x)$ denote whichever of $f(x)$ or $h(x)$ is larger, then

$$(|f(x)| + |g(x)|)^p \le (2|h(x)|)^p = 2^p|h(x)|^p \le 2^p(|f(x)|^p + |g(x)|^p),$$

where in the last step we replace $|h(x)|^p$ by $|f(x)|^p + |g(x)|^p$, which is for sure no smaller since $|h(x)|^p$ is simply one of the two terms being added together. Integrating now gives

$$\int_X |f+g|^p\,d\mu \le 2^p\int_X |f|^p\,d\mu + 2^p\int_X |g|^p\,d\mu.$$

The two terms on the right are finite since $f, g \in L^p$, so the term on the left is finite as well and hence $f + g \in L^p$.

Minkowski's inequality. We needed to know that L^p was closed under addition so that we can make sense of $\|f+g\|_p$, given $f, g \in L^p$. *Minkowski's inequality* is the inequality which says that this norm behaves in the way we expect of a "norm", in the sense that

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p .$$

This is an analog of the triangle inequality for absolute values, and is what gives the triangle inequality for the L^p-metric:

$$\|f-g\|_p = \|(f-h)+(h-g)\|_p \leq \|f-h\|_p + \|h-g\|_p .$$

Note that this L^p metric is clearly symmetric (switching f and g doesn't change the value), and $\|f-g\|_p = 0 \iff f = g$ by the way in which we technically define L^p using equivalence classes. Thus, the triangle inequality is only remaining property we need to verify to guarantee we have a metric. (As mentioned at some point a while back, it is crucial here that $p \geq 1$; the triangle inequality does not hold for the analogous notions when $0 < p < 1$.)

The proof of Minkowsi's inequality comes down to a clever use of Hölder's inequality. For fixed x, we have

$$\begin{aligned}
|f(x)+g(x)|^p &= |f(x)+g(x)||f(x)+g(x)|^{p-1} \\
&\leq (|f(x)|+|g(x)|)|f(x)+g(x)|^{p-1} \\
&= |f(x)||(f(x)+g(x))^{p-1}| + |g(x)||(f(x)+g(x))^{p-1}|.
\end{aligned}$$

Integrating then gives

$$\int |f+g|^p \, d\mu \leq \int |f(f+g)^{p-1}| \, d\mu + \int |g(f+g)^{p-1}| \, d\mu.$$

Note that the left side is $\|f+g\|_p^p$. Since $\frac{1}{p}+\frac{1}{q}=1$, $q(p-q)=p$, so

$$\int |(f+g)^{p-1}|^q \, d\mu = \int |f+g|^{q(p-1)} \, d\mu = \int |f+g|^p \, d\mu < \infty$$

since $f+g \in L^p$. But this thus says that the function $(f+g)^{p-1}$ is in L^q (!), since its q-th power is integrable. Thus, $f(f+g)^{p-1}$ is the product of a function in L^p and a function in L^q, so Hölder's inequality gives

$$\int |f(f+g)^{p-1}| \, d\mu = \left\|f(f+g)^{p-1}\right\|_1 \leq \|f\|_p \left\|(f+g)^{p-1}\right\|_q .$$

The same applies to the integral $\int |g(f+g)^{p-1}| \, d\mu$ above, so altogether we get

$$\|f+g\|_p^p \leq \|f\|_p \left\|(f+g)^{p-1}\right\|_q + \|g\|_p \left\|(f+g)^{p-1}\right\|_q = (\|f\|_p + \|g\|_p) \left\|(f+g)^{p-1}\right\|_q .$$

Now,

$$\left\|(f+g)^{p-1}\right\|_q^q = \int (|f+g|^{p-1})^q \, d\mu = \int (|f+g|^{q(p-1)}) \, d\mu = \int |f+g|^p \, d\mu = \|f+g\|_p^p ,$$

where we again use $\frac{1}{p}+\frac{1}{q}=1$. Taking q-th roots of both sides gives

$$\left\|(f+g)^{p-1}\right\|_q = \|f+g\|_p^{p/q} .$$

(Note the shift in q-norm to p-norm!) But $\frac{p}{q} = p - 1$, so

$$\|f+g\|_p^{p/q} = \|f+g\|_p^{p-1}.$$

Thus we get

$$\|f+g\|_p^p \leq (\|f\|_p + \|g\|_p) \left\|(f+g)^{p-1}\right\|_q = (\|f\|_p + \|g\|_p) \|f+g\|_p^{p-1},$$

which after dividing by $\|f+g\|_p^{p-1}$ gives precisely Minkowski's inequality. (Whew! No doubt this is a lot to keep track of, with non-obvious manipulations along the way, but all aimed at manipulating norms and exponents to move between different p-norms.)

Lecture 22: L^p Convergence

Warm-Up. We show that the Hölder and Minkowski inequalities hold for $p = \infty$ as well, where for Hölder's inequality we take $1 = \frac{1}{\infty}$ as the conjugate. (Recall that $L^\infty(X)$ is the space of essentially bounded functions on X, and $\|f\|_\infty$ is the essential supremum of f, which is the infimum of the set of all essential bounds of f.) We need the fact from the Warm-Up last time that $\|f\|_\infty$ is itself an essential bound of f.

First Hölder. Let $f \in L^1$ and $g \in L^\infty$. Then $|g(x)| \leq \|g\|_\infty$ almost everywhere, so

$$|f(x)g(x)| \leq |f(x)| \, \|g\|_\infty \text{ almost everywhere.}$$

Thus integrating gives

$$\int |fg| \, d\mu \leq \int |f| \, \|g\|_\infty \, d\mu = \|g\|_\infty \int |f| \, d\mu = \|f\|_1 \|g\| \infty.$$

The right side is finite, so the left side is as well and this gives $fg \in L^1$ with $\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$, which is Hölder's inequality.

For Minkowski, let $f, g \in L^\infty$. Since $|f(x)| \leq \|f\|_\infty$ almost everywhere and $|g(x)| \leq normg_\infty$ almost everywhere, we have

$$|f(x)+g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty \text{ almost everywhere.}$$

(To be clear, the null sets A and B off of which f and g, respectively, are bounded by their essential supremums could be different, so this bound on $f+g$ above only holds off of $A \cup B$, but the point is that this is still a null set.) This shows that $\|f\|_\infty + \|g\|_\infty$ is an essential bound of $f+g$, so

$$\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$$

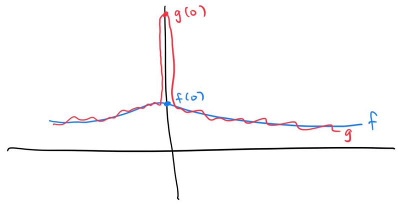
because the left side is the infimum of essential bounds.

The fact that Hölder's and Minkowski's inequality holds for L^∞ is one reason why we treat it in the same vein as other L^p spaces, even though the definition of this space is different because L^∞ is not defined via integrals. The reason why we use the symbol "∞" when describing this space or its norm has to do with the relation between $\|\|_\infty$ and $\|\|_p$, where you will show on the homework that the ∞-norm is in some sense a limit of p-norms as $p \to \infty$.

L^p convergence. We now seek to understand convergence in the space L^p. First we should understand what it means for functions to be "close" with respect to the L^p metric. The point is that

$$\|f-g\|_p = \left(\int |f-g|^p \, d\mu \right)^{1/p}$$

measures, not really the *values* of $f - g$, but instead essentially the "area" between the graphs of f and g. This area can be small even if the values of f and g can differ by a large amount, at least over certain portions of the domain. For example, the functions



are "close" in the L^p sense, although $f(0)$ and $g(0)$ can be quite far apart. (Of course, as we said earlier when introducing equivalence classes, it doesn't quite make sense to evaluate an element of L^p at a point anyway, so take the statement above with a grain of salt.) We can in fact make $f(0)$ arbitrarily far away from $g(0)$, as long as we compensate by making the graphs "narrower" in order to leave the value of the integral unchanged. (So, large differences between values can only take place over regions of "small" measures if f are meant to be "close" in L^p.)

The discussion above is meant to highlight the difference between convergence in L^p and pointwise convergence specifically, where the latter type of convergence is indeed a statement about the *values* of a function. But, we really only care about the values "almost everywhere", so perhaps there is a more direct relation between L^p convergence and almost everywhere pointwise convergence. In fact, you already saw an instance of this on the homework, where you showed (under appropriate assumptions) that if $f_n \to f$ pointwise almost everywhre, then $f_n \to f$ *in the mean.* Convergence "in the mean" was defined as

$$\lim_{n\to\infty} \int |f_n - f| \, d\mu = 0,$$

which we can now recognize is precisely convergence in L^1, since the integral above is the L^1-norm of $f_n - f$. Thus, the result is that almost everywhere pointwise convergence implies L^1-convergence, again under appropriate hypotheses.

The converse is not true, as a different problem on the homework showed: convergence in the mean (i.e. L^1 convergence) does not imply almost everywhere pointwise convergence, precisely because integrals being "close" does not mean values are "close", even almost everywhere. However, there was one final part to this homework problem, which showed that we do in fact get almost everywhere pointwise convergence from L^1 convergence, at least for a *subsequence* of our original sequence. This is a fact we will generalize next time, and is the key step in showing that L^p is complete. This will be the ultimate "gleam information about *values* from information about areas" statement we can hope to derive.

Riemann L^2 not complete. Back on the first day of class we gave as a motivation for introducing the Lebesgue integral the fact that the "Riemann version" of L^2 is in fact not complete, and we can now make this statement precise. Completeness is a good property to have since it guarantees that we can take limits and remain within our space, so the fact that the full Lebesgue version of L^p is complete is crucial in applications.

Let us denote by L^2_R the set of Riemann integrable functions $f : [0,1] \to \mathbb{R}$ for which

$$\int_0^1 f(x)^2 \, dx < \infty,$$

where the integral here is just the Riemann integral. We then consider the metric

$$\|f-g\|_2 = \sqrt{\int_0^1 |f(x)-g(x)|^2\,dx}$$

on L^2_R. (As in the general L^2 setup, we actually need to take equivalence classes of functions when defining L^2_R, since it is possible for a nonnegative Riemann integrable function to have Riemann integral zero while not being identically zero, only zero almost everywhere.) For each n, set

$$f_n(x) = x^{-1/4}1_{[1/n,1]} = \begin{cases} x^{-1/4} & \frac{1}{n} \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Each of these are Riemann integrable since they are bounded and piecewise continuous. We claim that the sequence (f_n) is Cauchy in L^2_R. Indeed, we have

$$\begin{aligned} \|f_{n+k}-f_n\|_2^2 &= \int_0^1 |f_{n+k}(x)-f_n(x)|^2\,dx \\ &= \int_0^1 x^{-1/2}1_{[1/(n+k),1/n]}\,dx \\ &= \int_{1/(n+k)}^{1/n} x^{-1/2}\,dx \\ &= \frac{2}{\sqrt{n}} - \frac{2}{\sqrt{n+k}}. \end{aligned}$$

Since the sequence $(\frac{2}{\sqrt{n}})$ converges in $\mathbb{R}$, it is Cauchy in $\mathbb{R}$, so the computation above implies that (f_n) is Cauchy in L^2_R as claimed.

But, we now claim that (f_n) does not converge in L^2_R, which will show that L^2_R is not complete. For each $0 < a < 1$, set $g_n = f_n 1_{(a,1]}$. Note that for $n > \frac{1}{a}$, we have $(a,1] \subseteq [\frac{1}{n},1]$, so

$$g_n = f_n 1_{(a,1]} = x^{-1/4}1_{(a,1]} \text{ for } n > \frac{1}{a}.$$

The point is that the sequence g_n is eventually constant (!), so that it converges in L^2_R (!!!) to the "eventual constant" $x^{-1/4}1_{(a,1]}$. (Note that this is statement about convergence in L^2_R, *not* pointwise convergence, which is not the type of convergence we want to consider. In any metric space, a sequence which is eventually constant will always converge to the constant point it eventually equals. This is why we introduced g_n in the first place, to work with eventually constant sequences.) Since the L^2_R-convergence above holds for all $0 < a < 1$, it also holds on $(0,1]$, so we get that

$$f_n \to x^{-1/4} \text{ with respect to } L^2 \text{ on } (0,1].$$

Thus if $x^{-1/4}$ were an element of L^2_R, it would have to be what (f_n) converged to. (It should be clear that $x^{-1/4}$ is the pointwise limit of (f_n), but here we are saying it would have to be the L^2 limit as well.)

However, $x^{-1/4}$ is not Riemann integrable on $[0,1]$ since it is unbounded, and Riemann integrable functions are always bounded. So, the only possible limit of (f_n) is not actually in L^2_R, so (f_n) does not converge in L^R_2. Technically, we need a bit more care here, since, due ot the "equivalence relation" construction of L^2_R, there would potentially be some other element of L^2_R which *did* serve as a valid limit for (f_n). However, any such element would have to differ from $x^{-1/4}$ on at most a set of measure zero, but modifying $x^{-1/4}$ on a set of measure zero will not change the fact that it is unbounded, so such a modification would still fail to belong to L^2_R. (You might object by saying that $x^{-1/4}$ is in fact *improperly Riemann integrable* on $[0,1]$, a notion which allows for integrating unbounded functions. Still the results stands, however, that we have a Cauchy sequence in L^2_R as defined using the (proper) Riemann integral that does not converge. But even if we allow improper Riemann integrals, the space L^2_R is not complete: you will see on the homework an example of this same phenomenon where all functions, including the candidate limit function, are bounded. The failure of the Riemann version of L^2 to be complete is not really about bounded vs unbounded, but is a reflection of a more serious issue with the Riemann integral itself.)

Strategy for completeness. We will show that L^p (full Lebesgue version) is complete next time, but let us outline strategy we will use now to set the stage. Given a Cauchy sequence (f_n) in L^p, the goal is to construct $f \in L^p$ such that $f_n \to f$ with respect to the L^p-metric. The first step will be to extract from (f_n) a subsequence which converges pointwise almost everywhere. This is the idea we've alluded to before, that we can extract some information about "values" from information about "areas", albeit only for a nicely-controlled subsequence and only almost everywhere. The pointwise limit of this subsequence is then the limiting function we're looking for, and we will show in the process of its construction that it is in L^p.

Second, we show that the pointwise convergence of the subsequence to the limit constructed above is actually L^p convergence. This is the generalization of the problem from the homework, that almost everywhere pointwise convergence implies convergence in the mean under mild hypotheses. (So, "values" being close implies "areas" being close.) Finally, we show that since (f_n) is Cauchy in L^p and it has a convergent subsequence in L^p, the full sequence converges in L^p as well, which gives completeness. This final step actually holds for metric space in general: if (p_n) is Cauchy in a metric space X and there exists a subsequence converging to $p \in X$, then $p_n \to p$ as well. (You probably saw this in the fall!) This last step is "easy", it is the first two steps that are truly new.

And just what is it about the Lebesgue integral vs the Riemann integral that allows for this all to work out. The answer, of course, is the Dominated Convergence Theorem! We will actually use this twice, in the first step (in the form of the Monotone Convergence Theorem) when constructing the subsequence we want, and then again in the second step when promoting almost everywhere pointwise convergence to L^p convergence. As we said back when first discussing the convergence theorems for the Lebesgue integral, the fact that they hold is really the key reason why the Lebesuge integral is of such importance.

Lecture 23: Completeness

Warm-Up. Suppose $f_n, f \in L^p$ and $f_n \to f$ pointwise almost everywhere. We show that if there exists $g \in L^p$ such that $|f_n| \leq g$ almost everywhere, then $f_n \to f$ in L^p, meaning with respect to the L^p-metric. (This is the L^p analog of the "convergence in the mean" problem from the homework.) Note first that if $|f_n| \leq g$ and $f_n \to f$ pointwise almost everywhere, then $|f| \leq g$ almost everywhere.

We have

$$|f_n - f|^p \leq (|f_n| + |f|)^p \leq (2g)^p = 2^p g^p.$$

Since $f_n \to f$ pointwise almost everywhere, $|f_n - f| \to 0$ and then $|f_n - f|^p \to 0$ almost everywhere. Thus $|f_n - f|^p$ is a sequence of functions dominated by an integrable function (since $g \in L^p$) converging pointwise to 0, so the dominated convergence theorem applies and gives

$$\lim_{n\to\infty} \int |f_n - f|^p \, d\mu = \int 0 \, d\mu = 0.$$

This says that $\|f_n - f\|_p^p \to 0$, so $\|f_n - f\|_p \to 0$ as well, which is convergence of (f_n) to f in L^p.

L^p is complete. We now prove that L^p (for $1 \leq p < \infty$) is complete. Let us recall the strategy we outlined last time: take a Cauchy sequence, extract a subsequence that converges pointwise almost everywhere, show that this convergence is actually L^p convergence, and then show that the full sequence converges to the same limit as this subsequence. Note the use of two of our convergence theorems in the proof.

Suppose (f_n) is a Cauchy sequence in L^p. First we pick a subsequence (f_{n_k}) such that

$$\left\|f_{n_{k+1}} - f_{n_k}\right\|_p < \frac{1}{2^k} \text{ for all } k.$$

To do this, we first pick f_{n_1} from among the terms for which $\|f_m - f_n\|_p < \frac{1}{2}$ holds. Then we pick f_{n_2} from among the terms for which $\|f_m - f_n\|_p < \frac{1}{2^2}$ holds, increasing n_2 if need be to ensure that $n_2 > n_1$. Note that this f_{n_2} is still among the terms for which the previous inequality held, so that $\|f_{n_2} - f_{n_1}\|_p < \frac{1}{2}$. Then we pick f_{n_3} from the among the terms for which $\|f_m - f_n\|_p < \frac{1}{2^3}$, increasing n_3 to guarantee that $n_3 > n_2$. This term is still among those satisfying the previous inequality, so $\|f_{n_3} - f_{n-2}\| < \frac{1}{2^2}$. And so on we continue in this manner, picking at the k-th stage a term f_{n_k} (with $n_k > n_{k-1}$) from the terms satisfying $\|f_m - f_n\|_p < \frac{1}{2^k}$, and then (f_{n_k}) is our desired subsequence. This is the subsequence we claim will converge pointwise almost everywhere.

Set

$$f(x) = f_{n_1}(x) + \sum_{k=1}^{\infty}(f_{n_{k+1}}(x) - f_{n_k}(x))$$

and

$$g(x) = |f_{n_1}(x)| + \sum_{k=1}^{\infty}|f_{n_{k+1}}(x) - f_{n_k}(x)|.$$

Note that if we allow ∞ as a valid value of a function, then the series defining g "converges" everywhere since it is just a sum of nonnegative numbers, so that g is well-defined everywhere, only with $g(x) = \infty$ being a possibility. We claim actually that g is finite almost everywhere. To see this, denote the partial sums of g as

$$S_M g = |f_{n_1}| + \sum_{k=1}^{M}|f_{n_{k+1}} - f_{n_k}|,$$

so that $S_M g \to g$ as $M \to \infty$ at any x, again allowing ∞ to be a valid limiting value. Then also $(S_M g)^p \to g^p$ at any x.

Now, we have

$$\|S_M g\|_p \leq \|f_{n_1}\|_p + \sum_{k=1}^{M}\left\|f_{n_{k+1}} - f_{n_k}\right\|_p < \|f_{n_1}\|_p + \sum_{k=1}^{M}\frac{1}{2^k},$$

where the final inequality follows from the choice of the f_{n_k}. But as $M \to \infty$ the sum on the right converges to a finite value due to its geometric nature, so we get that $\|S_M g\|_p$ converges to a finite value as $M \to \infty$ as well. This means that $\|S_M g\|_p^p$ does too, so

$$\lim_{M\to\infty}\int (S_M g)^p \, d\mu < \infty.$$

But the $(S_M g)^p$ form an increasing sequence of measurable functions (as M increases we add more nonnegative forms, so $S_M g < S_{M+1} g$) converging pointwise to g^p, so the fact that the limit above is finite implies by the monotone convergence theorem that g^p is integrable:

$$\int g^p \, d\mu < \infty.$$

Thus $g \in L^p$, so f, which is dominated by g in the sense that $|f| \leq g$, is also in L^p. In particular, this means that f^p, and hence f, can only take on the value ∞ on a set of measure zero, so we can say that the partial sums of the series defining f converge to f pointwise almost everywhere, no longer needing to allow ∞ as a value.

Note that if $S_M f$ denote the partial sums of f, then

$$S_M f = f_{n_1} + \sum_{k=1}^{M} (f_{n_{k+1}} - f_{n_k}) = f_{n_{k+1}}$$

due to the telescoping nature of the sum. These partial sums converge pointwise to f almost everywhere, so we have achieved the first aim of our proof: extract a subsequence (f_{n_k}) of (f_n) converging pointwise almost everywhere to a function f in L^p. For the second part, we use the result of the Warm-Up, which depended on the dominated convergence theorem. Namely, we have that

$$|S_M f| = \left| f_{n_1} + \sum_{k=1}^{M} (f_{n_{k+1}} - f_{n_k}) \right| \leq |f_{n_1}| + \sum_{k=1}^{M} |f_{n_{k+1}} - f_{n_k}| = S_M g \leq g$$

and $|f| \leq g$, so since $S_M f \to f$ pointwise almost everywhere, the Warm-Up directly implies that $S_M f \to f$ in L^p. But $S_M f = f_{n_{k+1}}$, so the subsequence (f_{n_k}) converges to f in L^p.

Finally, we use the general metric space fact that if a Cauchy sequence has a convergent subsequence, then it converges. In this particular case, this comes from

$$\|f_n - f\|_p \leq \|f_n - f_{n_k}\|_p + \|f_{n_k} - f\|_p \, .$$

Given ϵ, we can make the first term on the right smaller than $\epsilon/2$ once n, n_k are large enough by the Cauchy assumption, and we can make the second term smaller than $\epsilon/2$ by the convergence of (f_{n_k}) to f. Putting these together gives that $\|f_n - f\|_p < \epsilon$ for large enough n, so $f_n \to f$ in L^p and hence L^p is complete!

This concludes our proof, which can take a while to digest. The real subtle part is the first one, where we use the monotone convergence theorem to ensure that a series of nonnegative terms actually has finite value at almost all points. After this, the rest is a bit more direct. Note that the book gives essentially the same proof, although only in the case of L^2.

Pointwise for subsequence. Let us record the following consequence of the proof, which is the best we can do in general in terms of relating L^p convergence (i.e., convergence of areas) to pointwise convergence (i.e., convergence of values): If $f_n \to f$ in L^p, then there exists a subsequence (f_{n_k}) which converges pointwise to f almost everywhere. This comes from treating (f_n) as the Cauchy sequence in the proof of completeness, and then extracting the desired subsequence via the proof.

We will also point out that you saw in instance of this on the homework. A recent problem gave an example of a sequence of integrable functions (i.e., functions in L^1) which converged in the mean to 0, but not pointwise almost everywhere to 0. Nevertheless, the final part of that problem asked to construct a subsequence of that sequence which *did* converge pointwise almost everywhere to 0, which is precisely what the claim above says should happen!

Banach spaces. Vector spaces which are complete with respect to a norm defined on it are known as *Banach spaces*, of which L^p for $1 \leq p < \infty$ is the prototypical example. L^∞ is also an example, as shown on a recent problem from discussion section. In the finite-dimensional setting, $\mathbb{R}^n$ is also a Banach space, although such examples are not as interesting as infinite-dimensional ones. We will say a bit more next time about why Banach spaces are useful.

The spaces $L^p(\mathbb{N})$ or $L^p(\mathbb{Z})$, with $\mathbb{N}$ or $\mathbb{Z}$ equipped with counting measure are also complete, and are hence Banach. These space are usually denoted by $\ell^p(\mathbb{N})$ or $\ell^p(\mathbb{Z})$, and are called *sequence spaces* since their elements are sequences. For example, $\ell^1(\mathbb{Z})$ is the space of sequences such that

$$\sum_{n=-\infty}^{\infty} |c_n| < \infty,$$

and $\ell^\infty(\mathbb{Z})$ is the space of bounded sequences. (There is no difference between "bounded" and "essentially bounded" when considering counting measure.)

In certain cases, the norm we are considering actually arises from an *inner product* via

$$\|v\| = \sqrt{\langle u, u\rangle}.$$

An inner product is a generalization of the usual dot product on $\mathbb{R}^n$, and we will give a precise definition later. Banach spaces whose norms come from inner products are called *Hilbert spaces*, and L^2 is the key example of this. The fact that the norm on L^2 comes from an inner product, and that it is complete, is something we will explore in the final week.

Lecture 24: Dual Spaces

Warm-Up. Suppose V is a *normed* vector space, meaning a vector space equipped with a norm. (To be sure, a *norm* on V is a function $\|\cdot\|$ that assigns a nonnegative real number to each element of v such that: $\|v\| = 0$ if and only if $v = 0$, $\|cv\| = |c|\,\|v\|$ for any scalar c, and $\|u + v\| \leq \|u\| + \|v\|$. Any norm gives a metric via $\|u - v\|$.) We show that V is a Banach space (i.e., it is complete) if and only if

$$\text{convergence of } \sum_{n=1}^{\infty} \|v_n\| \text{ implies convergence of } \sum_{n=1}^{\infty} v_n.$$

The point is that these convergences are happening in different spaces: $\sum \|v_n\|$ is a series in $\mathbb{R}$, so here we mean usual convergence with respect to the absolute value metric, whereas $\sum v_n$ is a series in V, so for this we mean convergence with respect to the given norm. In general, convergence of a series of real numbers is simpler to think about than convergence of a series of vectors, so this characterization of completeness says that for *series*, convergence in $\mathbb{R}$ of norms is indeed enough to guarantee convergence in V.

Suppose first that V is complete and let $\sum v_n$ be a series in V for which $\sum \|v_n\|$ converges in $\mathbb{R}$. Denote by $S_M = \sum_{n=1}^M v_n$ the partial sums of $\sum v_n$. Then for $k \geq 0$ we have

$$\|S_{M+k} - S_M\| = \left\| \sum_{n=M+1}^{M+k} v_n \right\| \leq \sum_{n=M+1}^{M+k} \|v_n\|.$$

The right side is a difference of partial sums for $\sum \|v_n\|$, so since this series converges in $\mathbb{R}$, we can make right side smaller than any ϵ for M large enough. But this then makes the left side smaller than ϵ as well for such M, so (S_M) is a Cauchy sequence in V. Since V is complete, this sequences of partial sums converges V, which is what it means to say that the series $\sum v_n$ converges in V.

The proof of the converse is actually the same as the argument we gave last time when showing that L^p is complete, only that here we use the assumption of convergence of $\sum \|v_n\|$ implying convergence of $\sum v_n$ as a replacement for the use of any integral convergence theorems specific to the setting of L^p. To be clear, suppose V has the aforementioned property and let (v_n) be a Cauchy sequence in V. Define a subsequence of (v_n) by picking, for each k, terms such that

$$\left\|v_{n_{k+1}} - v_{n_k}\right\| < \frac{1}{2^k}.$$

Then set

$$v = v_{n_1} + \sum_{k=1}^{\infty}(v_{n_{k+1}} - v_{n_k}).$$

(Note that at this point we do not yet know that this series converges, so v is not yet a well-defined element of V.) The corresponding series of norms in $\mathbb{R}$ satisfies

$$\|v_{n_1}\| + \sum_{k=1}^{\infty}\left\|v_{n_{k+1}} - v_{n_k}\right\| \leq \|v_{n_1}\| + \sum_{k=1}^{\infty}\frac{1}{2^k},$$

so since the series on the right converges, so does the series of (nonnegative) norms on the left. By our assumption, this means that the series defining v converges as well, in V. (This is the conclusion of the first part of the L^p completeness proof.)

But the partial sums of the series defining v are

$$v_{n_1} + \sum_{k=1}^{M}(v_{n_{k+1}} - v_{n_k}) = v_{n_{M+1}},$$

so we have that the subsequence v_{n_k} converges to v in V. (This is the conclusion of the second part of the L^p completeness proof.) Since (v_n) is a Cauchy sequence in V with a subsequence converging to $v \in V$, (v_n) converges to v as well, so V is complete and hence a Banach space.

Why Banach? The result above gives the reason why we care about Banach spaces: they are the vector spaces in which we have a clear notion of taking infinite sums of vectors. Of course, not all such infinite sums will converge, but Banach spaces are the spaces for which we have a direct way of relating convergence of such infinite sums in convergence of a series of numbers instead, which are much better behaved.

Why should we care about infinite sums of vectors? Because they allow us to talk about *infinite linear combinations* like

$$\sum_{n=1}^{\infty} c_n v_n = c_1 v_1 + c_2 v_2 + c_3 v_3 + \cdots.$$

With infinite linear combinations we can think about a certain notion of an infinite-dimensional "basis" $\{v_1, v_2, v_3, \ldots\}$ where we try to express arbitrary $v \in V$ as expressions such as above. (We will be more precise about what "basis" means in this context when we talk about L^2 and Hilbert spaces.) As a contrast, if we consider the space of all real sequences

$$\mathbf{x} = (x_1, x_2, x_3, \ldots)$$

as an "infinite" version of $\mathbb{R}^n$, we find that the "standard basis elements"

$$\mathbf{e}_1 = (1, 0, 0, \ldots),\ \mathbf{e}_2 = (0, 1, 0, \ldots),\ \ldots,$$

where $\mathbf{e}_i$ has a 1 in the i-th location and 0 elsewhere, is not actually a basis since an expression like

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 + \dots$$

has no meaning because this space of sequences has no nice norm defined on all of it. (There are subspaces that have nice norms, such as the ℓ^p spaces we mentioned last time, but no norm on the entire space at once.) If we want to make sense of infinite-dimensional bases in a way that mimics our expectations coming from considering finite bases, then we need a norm, and better yet a Banach space so that we can ensure infinite linear combinations exist.

Once we can talk about infinite linear combinations, we can also make sense of infinite-dimensional linearity of linear transformations. To be linear means that

$$T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2),$$

and it follows from induction that this extends to any finite number of terms:

$$T(c_1v_1 + \dots + c_nv_n) = c_1T(v_1) + \dots + c_nT(v_n).$$

But if we want to extend this to *infinitely* many terms like

$$T\left(\sum_{k=1}^{\infty} c_kv_k\right) = \sum_{k=1}^{\infty} c_kT(v_k),$$

then we need to work with infinite linear combinations, so again there is no way of avoiding the eventual use of Banach spaces. (Even then, being to exchange the application of T with infinite sums as above requires more of T, namely that it be continuous. We will discuss continuous linear maps next.)

To give one final reason why completeness of L^p specifically is important, let us go back to the second day of class where we discussed contractions and their use in the study of differential equations. Now we want to consider what are called *partial differential equations*, which are differential equations involving multivariable functions and their partial derivatives. It turns out that such functions often naturally end up being elements of some (or multiple!) L^p, so if we want to use some contraction property to show that a given partial differential equation might have a solution, completeness of L^p is essential. Take a graduate level course in partial differential equations to learn more. (Actually, you need to consider more than just L^p spaces to make this all accurate—you need to consider what are called *Sobolev spaces*, which are built from L^p by adding more conditions. But we'll leave that to a proper course in analysis of partial differential equations.)

Dual spaces. The final topic dealing with general L^p we will consider before specializing to L^2 is that of *duality*. The notion of a "dual space" depends on the notion of a linear functional, so we will start here. A *linear functional* on L^p is nothing but a linear map (in the sense of linear algebra) from L^p to $\mathbb{R}$. So, a linear functional $\ell : L^p \to \mathbb{R}$ satisfies

$$\ell(f+g) = \ell(f) + \ell(g) \text{ and } \ell(cf) = c\ell(f).$$

But, this notion so far is a purely algebraic one and makes no use of the additional structure we have on L^p coming from the p-norm. So, what we really want to consider are those linear functionals that are *continuous* as maps from the metric space L^p to the metric space $\mathbb{R}$. We define the *dual space* $(L^p)^*$ (sometimes the adjective "continuous" is thrown in front of the word "dual") of L^p to be the set of continuous linear functionals on L^p:

$$(L^p)^* := \{\ell : L^p \to \mathbb{R} \mid \ell \text{ is linear and continuous}\}.$$

Note that this is itself a vector, since we can add linear maps together and multiply them by scalars, and such sums and scalar multiples will also be continuous.

The reasons for why one should care about dual spaces of vector spaces in general are not ones we will get into here, but suffice it to say that they play an important role in studying "coordinate-independent" linear algebra where we avoid picking a basis as much as possible. (Essentially, dual spaces keep track of all possible coordinates with respect to all possible bases.) For us, we consider dual spaces because of the special things that happen in the setting of L^p, and later L^2 specifically.

Bounded linear functionals. It turns out that for *linear* maps, there is an easy way to characterize continuity. The fact is that a linear functional $\ell : L^p \to \mathbb{R}$ is continuous if and only if it is bounded. (This is true for linear maps between any normed vector spaces, not just L^p and $\mathbb{R}$.) But, we have to be careful here about what we mean by saying that a linear map is "bounded". If we ask for the usual notion of boundedness, namely

$$|\ell(f)| \leq M \text{ for some } M > 0 \text{ and } \textit{all } f,$$

then it is only the zero map that can possibly satisfy such a definition. Indeed, if ℓ is nonzero, and specifically $\ell(f) \neq 0$, then taking $|c| \to \infty$ (c a scalar) in

$$|\ell(cf)| = |c\ell(f)| = |c||\ell(f)|$$

gives $|\ell(cf)| \to \infty$, so no finite bound M as that above can exist.

The way around this is to ask for boundedness only among vectors of norm 1, so that a linear functional $\ell : L^p \to \mathbb{R}$ is *bounded* if there exists $M > 0$ such that

$$|\ell(f)| \leq M \text{ whenever } \|f\|_p = 1.$$

(The same definition applies for linear maps between any normed spaces, and is something you already saw last quarter when discussing the norm of a matrix in the context of differentiability in $\mathbb{R}^n$.) For a nonzero vector f not of norm 1, we have that $\frac{f}{\|f\|_p}$ *is* of norm 1, so

$$\left|\ell\left(\frac{f}{\|f\|_p}\right)\right| \leq M, \text{ and hence } |\ell(f)| \leq M \|f\|_p,$$

and hence we can interpret "bounded" as saying that there is a restriction on the amount by which ℓ can scale the norm of an input. The fact that "continuous $\iff$ bounded" is true for linear maps comes from using linearity to get

$$|\ell(f) - \ell(g)| = |\ell(f - g)|,$$

so that making the left side small as required by continuity will be equivalent to being able to control the amount by which ℓ can scale norms. We'll leave the precise details for you to work out.

So, $(L^p)^*$ is equivalently the space of bounded linear functionals on L^p. For $\ell \in (L^p)^*$, the supremum of $|\ell(f)|$ as f ranges among all elements in L^p of norm 1 thus exists, and we call this the *functional norm* (or *operator norm* as the same definition applies to any linear map between normed spaces) of ℓ:

$$\|\ell\| := \sup_{\|f\|_p} |\ell(f)|.$$

Another way of saying this is that $\|\ell\|$ is the infimum of all possible $M > 0$ satisfying $|\ell(f)| \leq M \|f\|_p$, and hence in particular we get $|\ell(f)| \leq \|\ell\| \|f\|_p$. (Again, you likely saw this type of inequality last quarter in the context of matrix norms.) The functional "norm" $\|\ell\|$ defined in this way is indeed a norm in the sense of linear algebra, and it turns out that $(L^p)^*$ is complete with respect to this norm, so that it too is a Banach space. We'll leave the details of verying all this to elsewhere.

The dual of L^p. The dual $(L^p)^*$ of L^p is not such a bad space to study since its elements have a concrete description as bounded linear maps $L^p \to \mathbb{R}$, but actually we can do much better. It turns out that we can more explicitly describe *all* such bounded linear functionals in terms of the elements of L^q (!!!), where q is the conjugate of p. *This* is the ultimate reason why conjugate pairs are introduced in the first place, namely because of their use in describing dual spaces of L^p spaces. (Conjugate pairs are also often called *dual pairs*, exactly for this reason.)

To see this, we first define a map $L^q \to (L^p)^*$. Now, what should such a thing do? It will take as input a function $f \in L^q$, and output a *functional* ℓ_f, meaning a map $L^p \to \mathbb{R}$. That is, $L^q \to (L^p)^*$ should send $f \in L^q$ to something that sends $g \in L^p$ to a real number. But if we want to take $f \in L^q$ and $g \in L^p$ and produce a real number out of them, I claim that we have already seen the tool needed to do so... Hölder's inequality! Hölder's inequality guarantees that $\int fg\,d\mu$ is a bona fide real number, since

$$\left|\int fg\,d\mu\right| \leq \int |fg|\,d\mu = \|fg\|_1 \leq \|f\|_q \|g\|_p < \infty.$$

Thus, for $f \in L^q$, we set $\ell_f : L^p \to \mathbb{R}$ to be the functional defined by

$$\ell_f(g) = \int fg\,d\mu \text{ for } g \in L^p.$$

The map $L^q \to (L^p)^*$ sending $f \mapsto \ell_f$ thus, in all its glory, looks like

$$f \mapsto \left(g \mapsto \int fg\,d\mu\right).$$

Note that the ℓ_f thus defined is linear since integration is linear (i.e., $\int f(g_1+g_2)\,d\mu = \int fg_1\,d\mu + \int fg_2\,d\mu$ and $\int f(cg)\,d\mu = c\int fg\,d\mu$), and it is bounded since

$$|\ell_f(g)| \leq \int |fg|\,d\mu \leq \|f\|_q \|g\|_p = \|f\|_q \text{ for any } \|g\|_p = 1.$$

This in particular implies that the functional norm of ℓ_f is bounded by the q-norm of f:

$$\|\ell_f\| \leq \|f\|_q.$$

In fact, the reverse inequality is also true, as we will leave to a problem on the homework. (It comes down to finding a concrete $g \in L^p$ for which $|\ell_f(g)|$ in fact *equals* $\|f\|_q$.) Thus we get

$$\|\ell_f\| = \|f\|_q.$$

This means that the map $L^q \to (L^p)^*$ we've defined is an *isometry*, meaning that it preserves norms, and hence preserves distance. But *this* in particular guarantees that this map is injective (!!): if $\ell_{f_1} = \ell_{f_2}$ for some $f_1, f_2 \in L^q$, then

$$\|\ell_{f_1} - \ell_{f_2}\| = 0, \text{ so that } \|f_1 - f_2\|_q = 0$$

(this uses the fact that $f \mapsto \ell_f$ is norm-preserving), and hence $f_1 = f_2$. (More generally, isometries between metric spaces are always injective.) Thus, we can think of L^q as being a *subspace* of $(L^p)^*$.

Functionals as measures. The amazing fact is that the map $L^q \to (L^p)^*$ we've constructed above is (now we have to be careful and specify that $1 < p < \infty$) is actually surjective (!!!), so that *every* bounded linear functional $L^p \to \mathbb{R}$ arises from an element of L^q via integration. That is, the main result in this story is that L^q *is* literally (or more precisely, "isometrically isomorphic" to) the dual of L^p. (The same is true for $p = 1$, so that L^∞ is the dual of L^1, but surprisingly L^1 is not the dual of L^∞: the dual of L^∞ does contain L^1 as a subspace, but is in fact *larger*. You'll explore this a bit on the homework.)

To prove that L^q is indeed the *full* dual of L^p for $1 < p < \infty$, we need to know that given a bounded linear functional $\ell : L^p \to \infty$ there exists $f \in L^q$ such that $\ell = \ell_f$, which amounts to saying that

$$\ell(g) = \int fg\,d\mu \text{ for all } g \in L^p.$$

This is a very difficult thing to do, since we have to produce such an f that works for all g at once seemingly out of nowhere. Indeed, we will not prove this in full in this course, as it does require more advanced measure theory than what we've covered. (Side note: you should all take graduate real analysis MATH 410, which covers measure theory in the fall and functional analysis in the winter. *Functional analysis* is—surprise surprise—the study of functionals on Banach spaces, and is precisely what we are seeing a glimpse of now.)

But, even though we will not give a proof of this result, we will say something about the context which underlies the proof, where we will see a direct connection to something we spoke about previously. If we are to have an f satisfying

$$\ell(g) = \int fg\,d\mu$$

for all $g \in L^q$, then in particular if we take $g = 1_A$ for some measurable A we get

$$\ell(1_A) = \int f1_A\,d\mu = \int_A f\,d\mu.$$

(Indicator functions are always in L^q, as we will show next time as part of our Warm-Up.) Lo and behold, we have seen expressions such as $\int_A f\,d\mu$ before, where f is *fixed* and it is A that varies, when discussing *absolute continuity* for Lebesgue measure. Indeed, we spoke about how we could then view $\int_A f\,d\mu$ as defining a *new* measure of A, and that is precisely what is going on here: ℓ is after all giving a way to assign to a measurable set A a number $\ell(1_A)$, which is what a measure should do. The "measure" obtained in this way

$$\nu(A) := \ell(1_A)$$

is not quite a measure in the sense we've thought of before, specifically because it can take on negative values in addition to positive ones. This is an example of what's called a *signed* measure, where $\nu(A) < 0$ is allowed. (Think about how when the graph of f is below the x-axis, we interpret the integral of f as the "signed" area of the region between its graph and the x-axis.) The important part is that all the other requirements of being a "measure"—such as countable additivity—do indeed hold, as we will verify next time as a Warm-Up.

So, here is the context we need. We have a bounded linear functional $\ell : L^q \to \mathbb{R}$, and we want to find $f \in L^p$ for which $\ell = \ell_f$. We use ℓ to define a signed measure

$$\nu(A) := \ell(1_A).$$

We then show that this measure is "absolutely continuous" with respect to Lebesgue measure μ, in the sense that

$$\mu(A) = 0 \implies \nu(A) = 0.$$

(This is something we previously spoke about briefly when discussing absolutely continuity as a statement about making $\int_A f\,d\mu$ small when $\mu(A)$ is small. We showed as a Warm-Up back then that this property is equivalent to saying that any null set with respect to μ is still a null set with respect to ν. Well actually, we showed it for nonnegative measures, but the same is true for signed measures as well.) Expressions like $\int_A f\,d\mu$ give examples of absolutely continuous signed measures with respect to μ, and a result known as the *Radon-Nikodym theorem*—which we also very briefly mentioned a while back—guarantees that *any* measure which is absolutely continuous with respect to Lebesgue measure is indeed of this form. (The function f for which $\nu(A) = \int_A f\,d\mu$ is called the *Radon-Nikodym* derivative of ν with respect to μ, and, as we briefly alluded to earlier, does behave in a sense like a "derivative".)

Distribution theory. So, for our $\nu(A) = \ell(1_A)$, which is absolutely continuous with respect to Lebesgue measure μ, we thus get a function f such that

$$\ell(1_A) = \int_A f\,d\mu.$$

By building up from indicator functions to simple functions and then to measurable functions, one can then show that

$$\ell(g) = \int_A fg\,d\mu \text{ for all } g \in L^q,$$

which, after verifying that $f \in L^q$, gives the amazing result $\ell = \ell_f$, and hence surjectivity of $L^q \to (L^p)^*$, we wanted. Simply amazing mind-boggling stuff.

The idea of viewing a linear functional $L^q \to \mathbb{R}$ as a type of measure is the foundation of the subject known as *distribution theory*. Indeed, the overarching idea is that there really is no difference between "measures" and "functionals" when everything is interpreted in the right way. *Distributions* are functionals (or just measures!) defined on spaces of functions, and can be thought of as generalizations of functions . (Note that a function always gives rise to a distribution, precisely in the way we constructed ℓ_f above: f gives the distribution which assigns to g the number $\int fg\,d\mu$. The so-called *Dirac delta function*, which you might have heard of, is not actually a function as the name suggests, but is in fact a distribution.) Distributions are crucial tools in various areas such as mathematical physics and what's called *microlocal analysis*, and their study is intimately connected with the study of L^p spaces and their duals. We'll leave it at that in this course, but explore the topic of analysis of partial differential equations (which is essentially what microlocal analysis is) to learn more.

Lecture 25: Hilbert Spaces

Warm-Up. Suppose $\ell : L^p \to \mathbb{R}$ is a bounded linear functional, and define the "measure" ν by $\nu(A) = \ell(1_A)$. We show that ν is countably additive, which is one of the properties required of a (signed) "measure." As a quick remark, note that the requirement $\nu(\emptyset) = 0$ of a measure is just the fact that linear maps send zero to zero: the indicator function of the empty set is the zero function, and $\nu(\emptyset) = \ell(1_\emptyset) = \ell(0) = 0$. Countable additivity for ν amounts to the fact that linearity for ℓ extends to infinite sums, not just finite ones.

Let $\{A_n\}_n$ be a pairwise disjoint countable collection of measurable sets, and set

$$E_M = \bigcup_{n=M+1}^{\infty} A_n.$$

Then $\bigcup_{n=1}^{\infty} A_n = E_M \cup A_1 \cup A_2 \cup \cdots \cup A_M$, so the indicator function of the set on the left is the sum of indicator functions of the sets on the right. (This uses that the A_n are pairwise disjoint.) Thus by linearity of ℓ, which gives finite additivity for ν, we have

$$\nu(\bigcup_{n=1}^{\infty} A_n) = \nu(E_M) + \sum_{n=1}^{M} \nu(A_n).$$

The goal is to show that the first term on the right vanishes as $M \to \infty$, which will result in $\nu(\bigcup_n A_n) = \sum_n \nu(A_n)$, as required by additivity.

Now, we have

$$|\nu(E_M)| = |\ell(1_{E_M})| \leq \|\ell\| \, \|1_{E_M}\|_p \,,$$

where $\|\ell\|$ is the functional norm of ℓ. (To be clear, for any nonzero f, $|\ell(f/\|f\|_p)|$ is bounded by $\|\ell\|$ since $f/\|f\|_p$ has p-norm 1, and multiplying through by $\|f\|_p$ gives $|\ell(f)| \leq \|\ell\| \, \|f\|_p$. This is also an inequality we highlighted last time.) We can compute $\|1_{E_M}\|_p$ explicitly as

$$\left(\int 1_{E_M}^p \, d\mu\right)^{1/p} = \left(\int 1_{E_M} \, d\mu\right)^{1/p} = \mu(E_M)^{1/p}.$$

(So, in particular, the indicator function of any measurable set of finite measure is in L^p for any p.) Thus we get

$$|\nu(E_M)| \leq \|\ell\| \, \mu(E_M)^{1/p}.$$

Since the E_M form a decreasing sequence of measurable sets (i.e., $E_M \subseteq E_{M+1}$), we have that

$$\lim_{M\to\infty} \mu(E_M) = \mu(\bigcap_{M=1}^{\infty} E_M).$$

But the intersection of all E_M is empty since, by construction, each E_M excludes an additional A_n; more precisely, all E_M are contained in E_1, but $x \in E_1$ belongs to some A_n for $n \geq 2$, which is then excluded from E_n for this n, and hence x is not in the intersection of all the E_M. Thus $\mu(E_M) \to 0$ as $M \to \infty$, so $\nu(E_M) \to 0$ as well since the $|\nu(E_M)|$ are bounded by a finite number $\|\ell\|$ times $\mu(E_M)^{1/p} \to 0$. Hence taking limits in

$$\nu(\bigcup_{n=1}^{\infty} A_n) = \nu(E_M) + \sum_{n=1}^{M} \nu(A_n)$$

gives

$$\nu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \nu(A_n)$$

as desired.

This all shows that $\nu(A) := \ell(1_A)$ does indeed define a (signed) measure. Note that since $|\nu(A)| \leq \|\ell\| \, \mu(A)^{1/p}$ (same reasoning for general A as for E_M above), we also see that if A is a null set with respect to μ, then it will still be a null set with respect to ν. This is what is means to say that ν is *absolutely continuous* with respect to μ, which then guarantees that the "Radon-Nikodym theorem" kicks in to say that ν is of the form $\ell(1_A) = \nu(A) = \int_A f\,d\mu$ for some $f \in L^q$, which is then the function f that will satisfy $\ell(g) = \int fg\,d\mu$ for all $g \in L^p$. Whew!

Hilbert spaces. We bow specialize to the case of L^2. To set the stage, let us first properly define the notion of a Hilbert space. First, an *inner product* on a vector space H is an assignment of a real number $\langle u, v\rangle \in \mathbb{R}$ to each pair $u, v \in H$ such that

- $\langle u, u\rangle \geq 0$ for all $u \in H$ and $\langle u, u\rangle = 0$ if and only if $u = 0$;
- $\langle u, v\rangle = \langle v, u\rangle$ for all $u, v \in H$; and
- $\langle c_1u_1 + c_2u_2, v\rangle = c_1 \langle u_1, v\rangle + c_2 \langle u_2, v\rangle$ for all $u_1, u_2, v \in H$ and $c_1, c_2 \in \mathbb{R}$.

These are nothing but the properties one expects of the usual dot product in $\mathbb{R}^n$. Note the final property says that for fixed $v \in H$, the map $H \to \mathbb{R}$ defined by $u \mapsto \langle u, v\rangle$ is a linear functional.

These properties allow us to define a norm on H via $\|u\| := \sqrt{\langle u, u\rangle}$. Indeed, $\langle u, u\rangle$ is never negative, so that it makes sense to take its square root, and $\|u\| = 0$ if and only if $u = 0$. Next, $\|cu\| = \sqrt{\langle cu, cu\rangle} = \sqrt{c^2 \langle u, u\rangle} = |c|\,\|u\|$. And finally, the triangle inequality for $\|u\|$ follows from the so-called *Cauchy-Schwarz inequality* for $\langle u, v\rangle$, of which Hölder's inequality for L^p is a generalization and which we will come back to next time. The norm $\|u\|$ then gives a metric, and to say that H is a *Hilbert space* is then to say that H is complete with respect to this metric. Thus, Hilbert spaces are just Banach spaces whose norms come from inner products.

The standard examples of Hilbert spaces are $\mathbb{R}^n$ with the usual dot product, and L^2 with the inner product

$$\langle f, g\rangle = \int fg\,d\mu.$$

That this is an inner product on L^2 can be verified directly using linearity and monotonicity of the integral, and the fact that $\langle f, f\rangle = \int f^2\,d\mu = 0$ if and only if f^2, and hence f, is zero almost everywhere. (Recall that L^2 is technically defined by taking equivalence classes of functions.) In fact, the $\mathbb{R}^n$ example is a special case of the more general L^2 example of $L^2(\{1, 2, \ldots, n\})$ where $\{1, 2, \ldots, n\}$ is equipped with counting measure. The point is that you should view $\langle f, g\rangle$ above for L^2 on some interval as doing literally the same thing as the usual dot product: if we view a function as an "uncountable" vector indexed by $x \in [a, b]$ whose "x-th component" is $f(x)$, then in $\langle f, g\rangle$ we multiply corresponding components $f(x)$ and $g(x)$ together, and then "add" these via the integral as x ranges over $[a, b]$. That L^2 is complete with respect to this inner product is just the $p = 2$ case of our general L^p completeness result.

Self-duality.

Orthogonal projections.