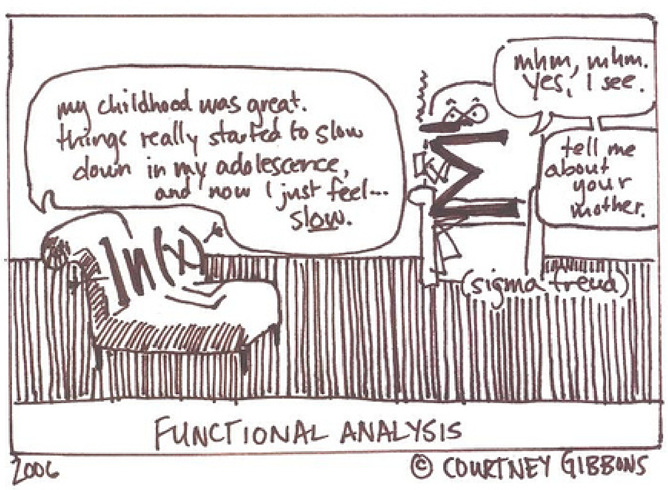
M383C: Methods of Applied Mathematics

Arun Debray

December 12, 2015



Source: http://brownsharpie.courtneygibbons.org/?pa=78.

These notes were taken in UT Austin's M383C class in Fall 2015, taught by Todd Arbogast. I live-TEXed them using vim, and as such there may be typos. Please send questions, comments, complaints, and corrections to a.debray@math.utexas.edu.

Contents

Chapter 1. Normed Linear Spaces and Banach Spaces 3
1.1 General Remarks: 8/26/15 3
1.2 Banach Spaces: 8/28/15 5
1.3 Bounded Linear Operators: 8/31/15 7
1.4 ℓ^p-norms: 9/2/15 10
1.5 ℓ^p and L^p-spaces: 9/4/15 12
1.6 $L^p(\Omega)$ is Banach: 9/9/15 15
1.7 The Hahn-Banach Theorem: 9/11/15 18
1.8 The Hahn-Banach Theorem, II: 9/14/15 19
1.9 Separability: 9/16/15 21
1.10 The Minkowski Functional and the Baire Category Theorem: 9/18/15 23
1.11 The Open Mapping Theorem: 9/21/15 25
1.12 The Uniform Boundedness Principle: 9/23/15 27
1.13 Weak and Weak-$*$ Convergence: 9/25/15 29
1.14 The Banach-Alaoglu Theorem: 9/28/15 31
1.15 The Generalized Heine-Borel Theorems: 9/30/15 33
1.16 The Dual to an Operator: 10/2/15 35

Chapter 2. Inner Product Spaces and Hilbert Spaces 38
2.1 Orthogonality: 10/5/15 38
2.2 Projections: 10/7/15 40
2.3 Orthonormal Bases: 10/9/15 42
2.4 Midterm Breakdown: 10/12/15 45
2.5 Classification of Hilbert Spaces: 10/14/15 47
2.6 Fourier Series and Weak Convergence in Hilbert Spaces: 10/16/15 50

Chapter 3. Spectral theory 53
3.1 Basic Spectral Theory in Banach Spaces: 10/19/15 53
3.2 Compact Operators: 10/21/15 55
3.3 Spectra of Compact Operators: 10/23/15 57
3.4 The Spectral Theorem for Compact Operators: 10/26/15 59
3.5 The Spectral Theorem for Self-Adjoint Operators: 10/28/15 61
3.6 The Spectral Theorem for Self-Adjoint Operators, II: 10/30/15 62
3.7 Positive Operators: 11/2/15 63
3.8 Compact, Self-Adjoint Operators and the Ascoli-Arzelà Theorem: 11/4/15 66
3.9 Sturm-Liouville Theory: 11/6/15 68
3.10 Solving Sturm-Liouville Problems With Green's Functions: 11/9/15 70
3.11 Applying Spectral Theorems to Sturm-Liouville Problems: 11/11/15 73

Chapter 4. Distributions 77
4.1 The Space of Test Functions: 11/16/15 77
4.2 Distributions: 11/18/15 79
4.3 Examples of and Operations on Distributions: 11/20/15 81

4.4 Differentiation of Distributions: 11/23/15 84
4.5 Midterm 2 Review: 11/25/15 86
4.6 Convolution of Distributions: 11/30/15 88
4.7 Applications of Distributions to Linear Differential Equations: 12/2/15 90
4.8 Linear Differential Operators with Constant Coefficients: 12/4/15 93

CHAPTER 1

Normed Linear Spaces and Banach Spaces

Lecture 1: 8/26/15

General Remarks.

Though the course name is "Methods of Applied Mathematics," this is a misnomer; the course is really about functional analysis.

The course will use the Canvas website (http://canvas.utexas.edu/), and office hours will be after class (modulo lunch), Mondays and Wednesdays from 12:30 to 1:50. Under UT Direct, there's also a CLIPS page, but that's less central to the course.

The textbook is a set of course notes; it hasn't changed much since 2013, so if you have that version, you'll be fine. They'll be ready at the copy center by Friday or Monday.

Homework will be due every week, assigned one Friday, and due the next. The first assignment will be due in a little over a week. We're encouraged to work in groups, but must write up our own individual proofs. Midterms will be weeks 7 and 12, probably, and will be topical; the final, at the end of the semester, will be comprehensive.

In this course, we'll cover chapters 2 – 5 of the lecture notes. Some elementary topology and Lesbegue integration (the first chapter) will be assumed.

Now, for some math. The professor is an applied mathematician, doing numerical analysis, and more specifically, approximation of differential equations. Functional analysis is useful for that, but also plenty of other fields, even including abstract algebra! Nonetheless, the course will be presented from an applied perspective.

The background is that we're trying to solve a problem of the form $T(u) = f$. Here, T is a model or differential equation; it's some kind of operator. f is the data that we're given, and we want to find the solution u. We use the framework of functional analysis to understand the nature of the functions u and f: their properties and what classes of functions they live in. We also want to know the nature of the operator T. In particular, we'll focus on cases where T is linear, since anything nonlinear can usually be locally approximated with a linear one. Thus, we should start with the linear case.

The set of all functions is a vector space, of course, so we're led to study vector spaces. At the undergraduate level, one studies finite-dimensional spaces, but here we'll use infinite-dimensional ones. Vector spaces also give us the required linearity. But since we also have questions of convergence, we'll introduce topology, so this course combines algebra and topology.

In this class, $\mathbb{F}$ will denote a field, either $\mathbb{R}$ or $\mathbb{C}$ (a lot of the time, the stuff we're doing won't depend on which).

Definition. Let X be a vector space over $\mathbb{F}$. Then, X is a *normed linear space* (henceforth NLS) if it has a *norm*, a function $\|\cdot\| : X \to \mathbb{R}^+ = [0, \infty)$ such that for every $x, y \in X$ and $\lambda \in \mathbb{F}$,

- $\|\lambda x\| = |\lambda|\|x\|$,
- $\|x\| = 0$ iff $x = 0$, and
- $\|x + y\| \le \|x\| + \|y\|$.

The last stipulation is called the *triangle inequality*.

These conditions on the norm mean it's a measure of size: stretching a vector stretches the norm, the only thing with size 0 is the origin, and the triangle inequality corresponds to the familiar geometric one. It turns out these are the only properties we need to measure size.

Example 1.1.1.

(1) d-dimensional *Euclidean space* $\mathbb{F}^d$ comes with a familiar norm: if $x = (x_1, \dots, x_n)$ for $x_j \in \mathbb{F}$, then

$$\|x\| = \sqrt{\sum_{j=1}^{d} |x_j|^2}.$$

Sometimes, this is simply denoted $|x|$. Thus, whenever we talk about $\mathbb{F}^d$, we really mean $(\mathbb{F}^d, \|\cdot\|)$, the normed linear space.

(2) If $a < b$, where $a, b \in [-\infty, \infty]$, let $C([a, b])$ denote the space of continuous functions $f : [a, b] \to \mathbb{F}$ such that $\sup_{x\in[a,b]}|f(x)|$ is finite.[1] This is indeed a vector space; then, it turns to a normed linear space with the norm

$$\|f\| = \sup_{x\in[a,b]} |f(x)|.$$

Notice that the norm must be finite, which is satisfied here. The first two properties are clearly satisfied, and because the absolute value is a norm on $\mathbb{R}$, then the triangle equality is also satisfied.

(3) We can pair $C([a, b])$ with a different norm $\|\cdot\|_{L^1}$, defined by

$$\|f\|_{L^1} = \int_a^b |f(x)|\,\mathrm{d}x.$$

The integral certainly exists, since f is continuous, but it might be infinite; thus, we assume that a and b are finite, so $[a, b]$ is compact, and

$$\int_a^b |f(x)|\,\mathrm{d}x \le (b - a) \sup_{x\in[a,b]} |f(x)|,$$

so we're bounded. It's also not that hard to show that $\|\cdot\|_{L^1}$ is a norm, as the integral is linear.

We now have two norms on $C([a, b])$; are they "the same?" Though the underlying vector spaces are the same, the measures of size are different, so as normed linear spaces they are not the same.

We can find more examples sitting inside other NLSes.

Proposition 1.1.2. *Let $(X, \|\cdot\|)$ be an NLS and $V \subseteq X$ be a linear subspace. Then, $(V, \|\cdot\|)$ is an NLS.*

It's easy to check that the three requirements are still met.

We can measure size, so since we're in a vector space, we can measure distance. In general, we have a metric. Specifically, if $(X, \|\cdot\|)$ is an NLS, define $d : X \times X \to \mathbb{R}^+$ by $d(x, y) = \|x - y\|$. Why is this a metric? It has to satisfy the following three properties for all $x, y, z \in X$.

(1) $d(x, y) = 0$ iff $x = y$.
(2) $d(x, y) = d(y, x)$.
(3) $d(x, y) + d(y, z) \ge d(x, z)$.

It's easy to check that the d induced from the norm is indeed a metric; each metric property follows from one of the norm properties.

And now that we can measure distance, we have a topology; specifically a metric topology, the simplest of all topologies. That is, a normed linear space is a metric space. To be specific, define the *ball of radius r about x*, where $r > 0$ and $x \in X$, as

$$B_r(x) = \{y \in X \mid d(x, y) < r\}.$$

This is an open ball, so the distance must be strictly less than r.

The topology is defined by setting $U \subseteq X$ to be *open* if for every $x \in U$, there exists an $r > 0$ such that $B_r(x) \subseteq U$. In other words, an open set doesn't contain its boundary. A set $F \subseteq X$ is *closed* if its complement $F^c = X \setminus F$ is open.

Definition. A subset F of a metric space X is *sequentially closed* if whenever $\{x_n\}_{n=1}^\infty$ is a sequence in F converging to an $x \in X$ (in the sense of the metric, i.e. $d(x_n, x) \to 0$), then $x \in F$.

[1] Recall that the *supremum* of a set is its least upper bound: for example, $\sup(0, 1) = 1$, even though 1 isn't part of the set. This distinguishes the supremum from the maximum.

In a metric space (this is *not* true in general!), F is closed iff F is sequentially closed.

Now, we have algebra (the vector space), the metric (giving us convergence, compactness, etc.), and the norm. How are they related?

Proposition 1.1.3. *In an NLS X, addition, scalar multiplication, and the norm are all continuous functions.*

PROOF. We'll prove this for addition and the norm; scalar multiplication is analogous to addition.

Addition is a function $+ : X \times X \to X$. Let $\{x_n\} \subseteq X$ with $x_n \to x$ and $\{y_n\} \subseteq X$ with $y_n \to y$. Continuity is equivalent to $x_n + y_n \to x + y$ for all such sequences. That is, I need $d(x_n + y_n, x + y) \to 0$, but that's equivalent to $\|(x_n + y_n) - (x + y)\| \to 0$.

Since $x_n \to x$ and $y_n \to y$, then $\|x_n - x\| \to 0$ and $\|y_n - y\| \to 0$. It looks like we should use the triangle inequality.

$$\begin{aligned}\|(x_n + y_n) - (x + y)\| &= \|(x_n - x) + (y_n - y)\| \\ &\le \|x_n - x\| + \|y_n - y\| \to 0.\end{aligned}$$

The norm is a little different. Suppose $x_n \to x$, which means we need to show that $\|x_n\| \to \|x\|$. Well,

$$\begin{aligned}\|x\| &= \|x - x_n + x_n\| \\ &\le \|x - x_n\| + \|x_n\| \\ &\le 2\|x - x_n\| + \|x\|.\end{aligned}$$

Since we've sandwiched $\|x - x_n\|$, then $\lim\|x_n\| = \|x\|$.[2] ⊠

Lecture 2: 8/28/15

Banach Spaces.

Recall that if $(X, \|\cdot\|)$ is an NLS, we have a metric $d(x, y) = \|x - y\|$ and a topology. More generally, if (X, d) is a metric space, $x_n \to x$ is the same as $d(x_n, x) \to 0$. In our case, this means that $\|x_n - x\| \to 0$.

Definition. A sequence $\{x_n\}_{n=1}^{\infty}$ is a *Cauchy sequence* if $\lim\limits_{n,m\to\infty} d(x_n, x_m) = 0$.

Here, n and m go to infinity independently, which might be confusing; an alternate way to phrase this is that $\{x_n\}$ is Cauchy if for all $\varepsilon > 0$, there exists an $N = N_\varepsilon > 0$ such that $d(x_n, x_m) < \varepsilon$ whenever $m, n \ge N$.

In a Cauchy sequence, the terms get closer and closer together, but do they converge? Consider $(0, \infty)$ and $x_n = 1/n$. This is Cauchy, but would converge to 0, which isn't part of our set; in a sense, it's a "hole" in our set. This is annoying.

Definition.

- A metric space X is *complete* if every Cauchy sequence on X converges in X.
- A complete NLS is called a *Banach space*.

We'll also give some properties of subspaces of NLSes.

Definition. Let X be an NLS. A set $M \subseteq X$ is *bounded* if there exists an $R > 0$ such that $M \subseteq \overline{B_R(0)} = \{x : \|x\| \le R\}$.

Equivalently, M is bounded if there's a finite R such that $\|x\| \le R$ for all $x \in M$.

Proposition 1.2.1. *Every Cauchy sequence in an NLS is bounded.*

PROOF. The idea is that all but a finite number of points in a sequence are within distance 1 of each other.

Let $\{x_n\}_{n=1}^{\infty}$ be a Cauchy sequence in an NLS X. By definition (using $\varepsilon = 1$), there's an $N > 0$ such that $\|x_n - X_N\| \le 1$ for all $n \ge N$. Using the triangle inequality, $\|x_n\| \le \|x_N\| + 1$ for all $n \ge N$.

Now, let $M = \max\{\|x_1\|, \dots, \|x_{N-1}\|\}$ and $R = \max\{\|x_N\| + 1, M\}$; both of these are finite sets, and therefore have maxima. Thus, $\|x_n\| \le R$ for all n. ⊠

Even if the limit isn't there, the sequence is still bounded, which is nice. Also, notice how we used the norm; boundedness in metric spaces maybe isn't so interesting.

[2] This was all that the professor said about the proof that the norm is continuous. Here's an alternate proof in case you, like me, didn't get it: since $x_n \to x$, then for any $n \in \mathbb{N}$, there's an N_n such that if $m \ge N_n$, then $x_m - x \in B_{1/n}(0)$. But that means that $\|x_m - x\| < 1/n$. Since $1/n \to 0$, then $\|x_n - x\| \to 0$ as well.

Example 1.2.2. Let's give some examples of Banach spaces.

(1) $\mathbb{R}^d$ and $\mathbb{C}^d$, as we learned in elementary real analysis.
(2) $C([a,b])$ with $\|f\| = \sup_{x\in[a,b]}|f(x)|$ is Banach, because a sequence $\{f_n\}$ is Cauchy iff it converges uniformly, and we know the uniform limit of continuous functions is continuous.

$C([a,b])$ with norm
$$\|f\|_{L^1} = \int_a^b |f(x)|\,\mathrm{d}x$$
is *not* complete, and therefore not Banach! This will verify the statement we made last lecture, that these spaces aren't the same. This is interesting behavior, because it doesn't happen in finite dimensions, and is an example of the subtle differences in behavior between finite-dimensional and infinite-dimensional vector spaces.

We'll let $a = -1$ and $b = 1$, though by suitable rescaling or translation this works for any $[a,b]$ with a and b finite.

Let $f_n(x)$ be 1 on $[-1,0]$, then decrease linearly on $[0,1/n]$, and then be 0 on $[1/n,1]$. Then,
$$\begin{aligned}\|f_n - f_m\|_{L^1} &= \int_{-1}^1 |f_n(x) - f_m(x)|\,\mathrm{d}x\\ &= \int_0^1 |f_n(x) - f_m(x)|\,\mathrm{d}x\\ &\le \int_0^1 (|f_n(x)| + |f_m(x)|)\,\mathrm{d}x\\ &= \frac{1}{2n} + \frac{1}{2m}.\end{aligned}$$
This goes to 0, so $\{f_n\}$ is Cauchy. But it converges to the step function
$$f(x) = \begin{cases} 1, & x < 0\\ 0, & x > 0.\end{cases}$$
This is because
$$\begin{aligned}\|f_n - f\|_{L^1} &= \int_{-1}^1 |f_n(x) - f(x)|\,\mathrm{d}x\\ &= \int_0^1 |f_n(x)|\,\mathrm{d}x = \frac{1}{2n},\end{aligned}$$
which goes to 0, so $f_n \to f$ after all.

This means that when we talk about $C([a,b])$, unless otherwise specified, we'll use the other norm, which makes it into a Banach space.

This situation, where the same vector space has two norms with different topological properties, is actually fairly common.

Definition. Let X be a vector space and $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on X. One says that the two norms are *equivalent* if there exist $c,d > 0$ such that for all $x \in X$, $c\|x\|_1 \le \|x\|_2 \le d\|x\|_1$.

This means that, though they might not agree precisely, the vague notions of "small" and "large" are the same in both norms.

We'll see eventually that all norms on a finite-dimensional space are equivalent, even though we already know that $\|\cdot\|$ and $\|\cdot\|_{L^1}$ are inequivalent on $C([a,b])$. We do know, however, that for $f \in C([0,1])$, $\|f\|_{L^1} \le \|f\|$,[3] but the other bound fails: there is no constant C such that $\|f\| \le C\|f\|_{L^1}$. We'll see this using the sequence $\{f_n\}$, where f_n increases linearly from 0 to n on $[0,1/n]$, decreases on $[1/n,2/n]$, and is 0 elsewhere. This sweeps out a triangle, so $\|f_n\| = n$, but $\|f_n\|_{L^1} = 1$ for all n, and thus no such C exists.

Proposition 1.2.3. *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two equivalent norms on X. Then, their induced topologies are the same.*

To be precise, the collections of open sets $\mathscr{U}_1$ and $\mathscr{U}_2$ induced from $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively, are identical.

[3] More generally, on $C([a,b])$, $\|f\|_{L^1} \le (b-a)\|f\|$.

PROOF. We'll let $B_r^1(x)$ denote the ball of radius r around x in $\|\cdot\|_1$, and define $B_r^2(x)$ similarly.

Since $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent, there exist c and d such that for any x and r, $B_{r/d}^1(x) \subseteq B_r^2(x) \subseteq B_{r/c}^1(x)$. Thus, if U_2 is any open set in $\mathscr{U}_2$, then for any $x \in U_2$, there's an r such that $B_r^2(x) \subseteq U_2$, and therefore $B_{r/d}^1(x) \subseteq U_2$, and so U_2 is open in $\mathscr{U}_1$, and the argument in the other direction is similar. ⊠

Convexity. Convexity is an important notion because it allows us to talk about the line joining two points.

Definition. Let X be a vector space over $\mathbb{F}$. Then, a set $C \subseteq X$ is *convex* if whenever $x, y \in C$, the line $\{tx + (1-t)y : 0 \le t \le 1\}$ is contained in C.

Proposition 1.2.4. *In any NLS, $B_r(x)$ is convex.*

PROOF. Let $y, z \in B_r(x)$ and $t \in [0,1]$. We want to show that $ty + (1-t)z \in B_r(x)$. We'll have to write x as $x + tx - tx$ and then use the triangle inequality. Specifically,

$$\begin{aligned} \|ty + (1-t)z - x\| &= \|t(y-x) + (1-t)(z-x)\| \\ &\le t\|y-x\| + (1-t)\|z-x\| \\ &< tr + (1-t)r = r. \end{aligned}$$ ⊠

This is more interesting than it looks, because in some spaces that are otherwise similar to NLSes, there exist balls that are non-convex.

Even in finite dimensions, balls aren't necessarily round; they can even be square! But that doesn't make much of a difference.

Linear Operators. We'll talk about linear operators in order to manipulate and transform functions.

Definition. A *linear operator* is a function $T : X \to Y$ of vector spaces X and Y such that

(1) $T(x+y) = T(x) + T(y)$, and
(2) $T(\lambda x) = \lambda T(x)$.

The idea is that scalar multiplication and addition in X and Y (which are *a priori* very different) are considered the same by T, which commutes with them.

Definition. A linear operator $T : X \to Y$, where X and Y are NLSes, is *bounded* if it takes bounded sets to bounded sets.

That is, if $C \subseteq X$ is bounded, then $T(C) = \{y : y = T(x) \text{ for some } x \in C\}$.

The definition is nice, but everybody thinks of bounded operators by the following characterization.

Proposition 1.2.5. *Let X and Y be normed linear spaces and $T : X \to Y$ be linear. Then, T is bounded iff there exists an $C > 0$ such that $\|Tx\|_Y \le C\|x\|_X$ for all $x \in X$.*

PROOF. First, suppose T is bounded. Then, the image of $B_1(0)$ (in X) is some bounded set, and therefore contained in a ball $B_R(0)$ for some R. In particular, if $y \in B_1(0)$, then $\|Ty\|_Y \le R$.

Given $x \in X$, if $x = 0$ then $Tx = 0$, so we're good. If $x \ne 0$, let $y = (1/2\|x\|_X) \cdot x$, so that $\|y\| = 1/2$, and therefore $y \in B_1(0)$, and therefore $\|Ty\| \le R$. That is,

$$\left\| T\left(\frac{1}{2\|x\|}\|x\|\right) \right\| = \frac{1}{2\|x\|}\|Tx\| \le R,$$

and therefore $\|Tx\| \le 2R\|x\|$, so with $C = 2R$ we're done.

Conversely, suppose there exists a $C > 0$ such that $\|Tx\| \le C\|x\|$ for all $x \in X$. Let $M \subseteq X$ be bounded; then, $M \subseteq B_R(0)$ for some R. For an $x \in M$, $\|Tx\| \le C\|x\| \le CR$, so $T(X) \subseteq B_{CR}(0)$ in Y, and thus T is bounded. ⊠

Lecture 3: 8/31/15

Bounded Linear Operators.

Let X and Y be normed linear spaces; the maps between them that we'll consider are linear operators $T : X \to Y$, as in the previous lecture.

If T is one-to-one and onto, then we should have an inverse $T^{-1} : Y \to X$. It's easy to check that T^{-1} is linear; you probably checked this as an undergraduate. In this situation, we have structure preservation: it doesn't matter whether you check addition in X or in Y, or scalar multiplication. Thus, in the sense of linear algebra, X and Y look the same; they have the same addition and scalar multiplication. In this case, we say that X and Y are *isomorphic*; they may be unequal as sets (e.g. sequences or functions), but identical from the perspective of linear algebra.

For vector spaces, these maps are pretty cool, but for topology, we care about continuous maps $f : X \to Y$. Thus, as you might guess, when studying normed linear spaces, we care about maps $X \to Y$ that are both linear and continuous.

Definition. If X and Y are NLSes, then $B(X, Y)$ denotes the set of functions $f : X \to Y$ that are both linear and continuous.

Continuity means that for all $\varepsilon > 0$ there exists a $\delta > 0$ depending on x and ε such that when $d(x, y) < \delta$, then $d(f(x), f(y)) < \varepsilon$. But since there's a norm defining the metric, this is equivalent to stating that when $\|x-y\| < \delta$, then $\|f(x)-f(y)\| < \varepsilon$. And if $f = T$ is a linear operator, then $\|T(x)-T(y)\| < \varepsilon$ is equivalent to requiring $\|T(x-y)\| < \varepsilon$. In other words, this doesn't depend on x at all: letting $z = x - y$, continuity of a linear $T : X \to Y$ means that when $\|z\| < \delta$, then $\|Tz\| < \varepsilon$.

In other words, if you know what a linear map does around 0, you know what it looks like everywhere.

Proposition 1.3.1. *Let X and Y be NLSes and $T : X \to Y$ be linear. Then, the following are equivalent:*

(1) *T is continuous.*
(2) *T is continuous at some $x_0 \in X$.*
(3) *T is bounded.*

This is why we used the notation $B(X, Y)$: it stands for "bounded." And we can now talk about bounded linear maps, with continuity understood.

Proof. Clearly, (1) $\implies$ (2). For (2) $\implies$ (3), suppose T is continuous at some $x_0 \in X$. With $\varepsilon = 1$, this means there's a $\delta > 0$ such that $\|x - x_0\| \le \delta$ implies $\|Tx - Tx_0\| \le 1$, i.e. $\|T(x - x_0)\| \le 1$. In other words, with $z = x - x_0$, when $\|z\| \le \delta$, we have $\|Tz\| \le 1$.

For $x = 0$ boundedness is clear, but if $x \ne 0$, then

$$\begin{aligned} \|Tx\|_Y &= \left\| \frac{\|x\|}{\delta} T\left(\frac{\delta x}{\|x\|}\right) \right\|_Y \\ &= \frac{\|x\|}{\delta} \left\| T\left(\frac{\delta x}{\|x\|}\right) \right\| \le \frac{1}{\delta}\|x\|_X, \end{aligned}$$

so with $C = 1/\delta$, T is a bounded operator.

For (3) $\implies$ (1), we know $\|Tx\|_Y \le C\|x\|_X$ for some fixed C and all $x \in X$. Let $\varepsilon > 0$ and pick any $x_0 \in X$. Then, if $\delta = \varepsilon/C$ and $\|x - x_0\| \le \delta$, then

$$\|T(x - x_0)\| \le C\|x - x_0\| \le C\delta = \varepsilon,$$

so T is continuous at x_0 and therefore everywhere. ⊠

It turns out $B(X, Y)$ is a vector space itself, with $(f + g)(x) = f(x) + g(x)$ and $(\lambda \cdot f)(x) = \lambda \cdot (f(x))$, which is little surprise. But we do have to check that if $f = T$ and $g = S$ are linear, $f + g$ and λf are also linear, i.e. $(T + S)(x + y) = (T + S)(x) + (T + S)(y)$, and similarly for scalar multiplication.

What makes this more interesting is that $B(X, Y)$ is an NLS itself. What's the norm, you ask? Excellent question. The norm is

$$\|T\| = \|T\|_{B(X,Y)} = \sup_{x \in B_1(0)} \|Tx\|_Y.$$

Since T is continuous and bounded, $T(B_1(0))$ is a bounded set. Then, the norm of T is the radius of the smallest ball that contains $T(B_1(0))$, which is the supremum of the amount that T scales any point in the unit ball. Since T is bounded, the norm is a finite, nonnegative number.

Note that, even though we called this a norm, we still have to check that it's a norm!

Proposition 1.3.2. *Let X and Y be NLSes. Then, $\|\cdot\|_{B(X,Y)}$ is a norm on $B(X, Y)$. Moreover, if $T \in B(X, Y)$,*

$$\|T\| = \sup_{\|x\|_X \le 1} \|Tx\|_Y = \sup_{\|x\|_X = 1} \|Tx\|_Y = \sup_{x \ne 0} \frac{\|Tx\|_X}{\|x\|_X}.$$

Furthermore, if Y is Banach, then $B(X,Y)$ is too.

This last point is quite interesting: completeness follows when the range is complete, but the domain doesn't matter.

PROOF. First, that $\|\cdot\|$ is a norm: we have three properties to show.

- We need $\|T\| = 0$ iff $T = 0$. Clearly, if $T = 0$ (i.e. $T(x) = 0$ for all x), then $\|T\| = \sup_{x\in B_1(0)}\|Tx\| = \|0\| = 0$. Conversely, if we assume $\|T\| = 0$, then for any $x \in B_1(0)$, $\|Tx\| = 0$, so $Tx = 0$. Thus, $T|_{B_1(0)} = 0$. For general x, we'll scale $x = 2\|x\|(x/2\|x\|)$, so
$$Tx = 2\|x\|T\left(\frac{x}{2\|x\|}\right) = 2\|x\| \cdot 0 = 0,$$
since $x/2\|x\| \in B_1(0)$. Thus, $T = 0$.
- For linearity of the norm,
$$\|\lambda T\| = \sup_{x\in B_1(0)} \|\lambda Tx\| = \sup_{x\in B_1(0)} |\lambda|\|Tx\| = |\lambda| \sup_{x\in B_1(0)} \|Tx\| = |\lambda|\|T\|.$$

Exercise. Finish the proof that this is a norm by addressing the triangle inequality, which isn't too complicated.

Next, we have the different ways of calculating the norm. The idea is that since T is continuous, the supremum shouldn't depend on whether the boundary is present or not. One interesting corollary of the formulas for calculating $\|T\|$ is that for any $x \in X$, $\|Tx\| \le \|T\|\|x\|$.

The last part does require care. Let $\{T_n\}_{n=1}^{\infty}$ be a Cauchy sequence. That is, given an $\varepsilon > 0$, there's an $N > 0$ such that if $m, n \ge N$, then $\|T_n - T_m\|_{B(X,Y)} < \varepsilon$. Thus, given an $x \in X$, $\|T_n x - T_m x\|_Y \le \|T_n - T_m\|\|x\|_X$. The right-hand side goes to 0 as a Cauchy sequence in m and n, and therefore the left-hand side does too. That is, $\{T_n x\}_{n-1}^{\infty} \subset Y$ is a Cauchy sequence. Since Y is Banach, this means there's a limit $\lim_{n\to\infty} T_n x = T(x) \in Y$. This defines a map $T : X \to Y$; we need to prove that it's bounded linear and that $T_n \to T$.

First, let's look at linearity.
$$T(x+y) = \lim_{n\to\infty} T_n(x+y) = \lim_{n\to\infty} (T_n x + T_n y).$$
Since addition is continuous, we can break this up as
$$= \lim_{n\to\infty} T_n x + \lim_{n\to\infty} T_n y = Tx + Ty.$$
Similarly, since scalar multiplication is continuous,
$$T(\lambda x) = \lim_{n\to\infty} T_n(\lambda x) = \lambda T(x).$$
Next, let's check that T is bounded. Since the norm is continuous,
$$\begin{aligned}\|Tx\|_Y &= \left\|\lim_{n\to\infty} T_n x\right\|_Y \\ &= \lim_{n\to\infty} \|T_n x\|_Y.\end{aligned}$$
However, this limit *a priori* might not exist, so we have to use the lim sup.
$$\begin{aligned}&\le \limsup_{n\to\infty}\|T_n\|\|x\|_X \\ &= M\|x\|_X.\end{aligned}$$
Here, M is an upper bound on $\|T_n\|$, because $\{T_n\}$ is Cauchy and therefore bounded. Thus, we know $T \in B(X,Y)$.

Finally, to show $T_n \to T$, we need to be careful: limits depend on the topology that we're using, and so we should be careful that we're using the topology defined by $\|\cdot\|_{B(X,Y)}$.

Let $x \in B_1(0)$. Then,
$$\begin{aligned}\|Tx - Ty\|_Y &= \lim_{m\to\infty} \|T_m x - T_n x\| \\ &= \lim_{m\to\infty} \|(T_m - T_n)x\| \\ &\le \limsup_{m\to\infty}\|T_m - T_n\|\|x\|.\end{aligned}$$
Since $\{T_n\}$ is Cauchy, then for any $\varepsilon > 0$, $\|T_m - T_n\| < \varepsilon$ when m, n are sufficiently large, and therefore the lim sup goes to 0 as $n \to \infty$, and so $T_n \to T$. ⊠

There's one particularly important case, in which $Y = \mathbb{F}$.

Definition. The *dual space* of an NLS X is $X^* = B(X, \mathbb{F})$.

By Proposition 1.3.2, X^* is always a Banach space.

Though $B(X, Y)$ can be complicated for general Y, one can often understand it more easily using X^*.

Example 1.3.3. We can connect this with finite-dimensional linear algebra that we're more familiar with, and see that it's actually quite special.

Let X be a d-dimensional vector space over $\mathbb{F}$ with basis $\{e_n\}_{n=1}^d$. Thus,

$$\begin{aligned} X &= \operatorname{span}\{e_1, \dots, e_d\} \\ &= \{\alpha_1 e_1 + \cdots + \alpha_d e_d \mid \alpha_i \in \mathbb{F}\}, \end{aligned}$$

and we can write $x = x_1 e_1 + \cdots + x_d e_d \in X$. The map $T : X \to \mathbb{F}^d$ sending $x \mapsto (x_1, \dots, x_d)$ is one-to-one, onto, and linear, so all finite-dimensional vector spaces over a specified field are isomorphic. Moreover, we will show in Proposition 1.4.4 that all norms over a finite-dimensional vector space are equivalent, so as NLSes, they're all isomorphic too! There are many norms, which may still be interesting, but there's only one topology.

Lecture 4: 9/2/15

ℓ^p-norms.

Recall that we were looking at examples of Banach spaces, and that the first examples we saw (Example 1.3.3) were finite-dimensional vector spaces. If $d = \dim X$ is finite, so that $X = \operatorname{span}\{e_1, \dots, e_n\}$ (so that $\{e_1, \dots, e_n\}$ is a basis for X), then the map $T : X \to \mathbb{F}^d$ sending $(x_1 e_1 + \cdots + x_d e_d) \mapsto (x_1, \dots, x_d)$ is an isomorphism of vector spaces, and the claim is that these maps define the same topology as well.

But first, let's define some norms on $\mathbb{F}^d$. Let $1 \le p \le \infty$, and define

$$\|x\|_{\ell^p} = \begin{cases} \left(\displaystyle\sum_{n=1}^d |x_n|^p \right)^{1/p}, & p < \infty \\ \max_n |x_n|, & p = \infty. \end{cases}$$

Sometimes, these are denoted $\|x\|_{\ell_p}$. Also, the case $p = 2$ is our familiar Euclidean norm $\|x\|_{\ell^2} = |x|$.

We do have to show that these are norms. When $p = 1$ or $p = \infty$, it's a straightforward check, and when $1 < p < \infty$, the first two properties are pretty simple, but the triangle inequality is harder.

Lemma 1.4.1 (Young's inequality[4]). *Let $1 < p < \infty$ and q be the* conjugate exponent *defined such that $1/p + 1/q = 1$. If $a, b \ge 0$, then $ab \le a^p/p + b^q/q$, with equality iff $a^p = b^q$. Moreover, for all $\varepsilon > 0$, there exists a C depending on p and ε such that $ab < \varepsilon a^p + C b^q$.*

Proof. The proof is easy once you know the trick, to look at the right function. Let $u : [0, \infty) \to \mathbb{R}$ send

$$u(t) = \frac{t^p}{p} + \frac{1}{q} - t.$$

Its derivative is well-defined: $u'(t) = t^{p-1} - 1$, so $u'(0) = 1$. In particular, $u(0) = 1/q$, and $u(1) = 0$ is a strict minimum.

We'll apply this to $t = ab^{-q/p}$:

$$\begin{aligned} 0 \le u(ab^{-q/p}) &= \frac{a^p}{pb^q} + \frac{1}{q} - \frac{a}{b^{q/p}} \\ &= \frac{1}{b^q}\left(\frac{a^p}{p} + \frac{b^q}{q} - \frac{ab^q}{b^{q/p}} \right), \end{aligned}$$

but $b^q / b^{q/p} = b$, since $q - q/p = q(1 - 1/p) = 1$. Thus, $0 \le a^p/p + b^q/q - ab$, and equality holds iff $t = ab^{-q/p} = 1$, where $u(t)$ is equal to 0.

For the second part, we can write

$$ab = \big((\varepsilon p)^{1/p} a\big)\big((\varepsilon p)^{-1/p} b\big) \le \frac{\varepsilon p a^p}{p} + \frac{(\varepsilon p)^{-q/p}}{q} b^q. \qquad \boxtimes$$

[4]Young's inequality technically refers to a more general statement; this could be called "Young's inequality for products."

For conjugate exponents, we have the convention that the conjugate of 1 is ∞, and vice versa.

Theorem 1.4.2 (Hölder's inequality)**.** *Let $1 \le p \le \infty$ and q be its conjugate exponent. If $x, y \in \mathbb{F}^d$, then*

$$\sum_n |x_n y_n| \le \|x\|_{\ell^p} \|y\|_{\ell^q}.$$

When $p = 2$, this is also known as the *Cauchy-Schwarz inequality*.

Proof. The cases $p = 1, \infty$ are trivial; expand their definitions out. Similarly, if $x = 0$ or $y = 0$, there's not a lot to say. Thus, we're left with $1 < p < \infty$, so we can use Lemma 1.4.1.

Let $a = |x_n|/\|x\|_{\ell^p}$ and $b = |y_n|/\|y\|_{\ell^q}$. Then, by Lemma 1.4.1,

$$\frac{|x_n|}{\|x\|_{\ell^p}} \frac{|y_n|}{\|y\|_{\ell^q}} \le \frac{|x_n|^p}{p\|x\|_{\ell^p}^p} + \frac{|y_n|^q}{q\|y\|_{\ell^q}^q},$$

so summing all n of those,

$$\begin{aligned}
\frac{\sum_n |x_n y_n|}{\|x\|_{\ell^p}\|y\|_{\ell^q}} &\le \frac{\sum |x_n|^p}{p\|x\|_{\ell^p}^p} + \frac{\sum |y_n|^q}{q\|y\|_{\ell^q}^q} \\
&= \frac{\|x\|_{\ell^p}^p}{p\|x\|_{\ell^p}^p} + \frac{\|y\|_{\ell^q}^q}{q\|x\|_{\ell^q}^q} \\
&= \frac{1}{p} + \frac{1}{q} = 1.
\end{aligned}$$

⊠

Now, we can use this to prove the triangle inequality for $\|\cdot\|_{\ell^p}$. We'll need two things for the Hölder inequality, so just take one term out of the p^{th} power:

$$\begin{aligned}
\|x+y\|_{\ell^p}^p &= \sum_{n=1}^d |x_n + y_n|^p \\
&\le \sum_{n=1}^d |x_n + y_n|^{p-1}(|x_n| + |y_n|) \\
&\le \left(\sum_{n=1}^d |x_n + y_n|^{(p-1)q}\right)^{1/q} (\|x\|_{\ell^p} + \|y\|_{\ell^q}).
\end{aligned}$$

Since p and q are conjugate, $p = (p-1)q$, so the first term is $\|x - y\|_{\ell^p}^{p/q}$. Thus,

$$\|x+y\|_{\ell^p}^{p-p/q} \le \|x\|_{\ell^p} + \|y\|_{\ell^p},$$

and $p - p/q = 1$, so we're done.

Moreover, all these norms are equivalent.

Proposition 1.4.3. *Let $1 \le p \le \infty$. Then, for all $x \in \mathbb{F}^d$,*

$$\|x\|_{\ell^\infty} \le \|x\|_{\ell^p} \le d^{1/p}\|x\|_{\ell^\infty}.$$

These estimates are sharp, the first at $x = (1, 0, 0 \dots, 0)$, and the second at $x = (1, 1, \dots, 1)$.

Proof. Let m be an index for which $|x_m| = \max_n |x_n|$. Since $f(x) = x^{1/p}$ is an increasing function,

$$\|x\|_{\ell^\infty} = |x_m| = (|x_m|^p)^{1/p} \le \left(\sum_{n=1}^d |x_n|^p\right)^{1/p} = \|x\|_{\ell^p},$$

and

$$\|x\|_{\ell^p} = \left(\sum_{n=1}^{d}|x_n|^p\right)^{1/p} \le \left(\sum_{1}^{d}|x_m|^p\right)^{1/p} = (d|x_m|^p)^{1/p} = d^{1/p}\|x\|_{\ell^\infty}. \qquad \boxtimes$$

Notice that some of these proof methods fail horribly in infinite dimensions.

It turns out that on all finite-dimensional vector spaces, all norms are equivalent.

Proposition 1.4.4. *All norms on a finite-dimensional NLS are equivalent. Moreover, a $K \subset X$ is compact iff it is closed and bounded.*

That means there's only one topology.

Proof. Let $d = \dim X$ and $\{e_n\}_{n=1}^d$ be a basis. Then, let $T : X \to \mathbb{F}^d$ be the coordinate map defined above. Let $\cong$ denote an isomorphism of NLSes.

We'll define a norm $\|\cdot\|_1$ on x by $\|x\|_1 = \|Tx\|_{\ell^1}$: of the three properties, the last two are trivial (since T is linear), so we just need to prove that $\|x\|_1 = 0$ iff $x = 0$. But T is one-to-one and onto, so this follows, and $\|\cdot\|_1$ is in fact a norm.[5]

Thus, $(X, \|\cdot\|_1) \cong (\mathbb{F}^d, \|\cdot\|_{\ell^1})$, so they really are the "same" space. This is because $T : X \to \mathbb{F}^d$ is a bounded map, with $C = 1$, and therefore continuous, and T^{-1} is also linear and continuous. Thus, T is an isomorphism of vector spaces and a homeomorphism of topological spaces, so we can take results in $\mathbb{F}^d$ and apply them to X.

The Heine-Borel theorem from undergraduate real analysis tells us that $K \subset \mathbb{F}^d$ is closed and bounded iff it's compact. But since X and $\mathbb{F}^d$ have the same topology, then this is also true in X. In particular, $S_1^1 = \{x \in X : \|x\|_1 = 1\}$ is also compact.

Now, for any norm $\|\cdot\|$ on X and $x \in X$,

$$\|x\| = \left\|\sum_{n=1}^{d} x_n e_n\right\| \le \sum_{n=1}^{d}|x_n|\|e_n\| \le C\|x\|_1,$$

where $C = \max_n\|e_n\|$. Notice that this step won't work in infinite dimensions. Our upper bound implies that $(Top)_{\|\cdot\|} \subseteq (Top)_{\|\cdot\|_1}$, so the former topology is said to be stronger. We'll prove the two are equal by providing a lower bound.

We have a continuous map $\|\cdot\| : (X, \|\cdot\|_1) \to \mathbb{R}$. It's also continuous as a map $\|\cdot\| : (X, \|\cdot\|) \to \mathbb{R}$. Let $a = \inf_{x\in S_1^1}\|x\|$; since S^1 is compact and the norm is continuous, the minimum is attained, and it must be positive (because $0 \notin S_1^1$).

Thus, for any $x \in X$, $\|x/\|x\|_1\| \ge a$, so $\|x\| \ge a\|x\|_1$, which is our desired lower bound. $\boxtimes$

Corollary 1.4.5. *If X is a d-dimensional NLS, then $X \cong \mathbb{F}^d$.*

Corollary 1.4.6. *If X and Y are NLSes and X is finite-dimensional, then every linear $T : X \to Y$ is bounded and $X^* = \mathbb{F}^d$, given by $T(x) = y \cdot x$.*

Lecture 5: 9/4/15

ℓ^p and L^p-spaces.

"There are different sizes of infinity, and this one is the best."

Last time we showed that if $(X, \|\cdot\|)$ is a finite-dimensional NLS, then it's isomorphic and homeomorphic to $(\mathbb{F}^d, \|\cdot\|_{\ell^2})$, where $d = \dim X$. Moreover, X is Banach, and $(\mathbb{F}^d)^* \cong \mathbb{F}^d$. Finite dimensions aren't very interesting, but they're a good place to gain intuition.

A lot of this nice stuff goes away for infinite-dimensional spaces, and some are nicer than others.

[5]A great way to create a new norm is to map from one space to another (or the same one) and pull the norm back.

Example 1.5.1. Let $1 \le p \le \infty$. We'll define a space ℓ^p which behaves sort of like an "$\mathbb{F}^\infty$." Specifically,

$$\ell^p = \left\{ x = \{x_n\}_{n=1}^\infty : x_n \in \mathbb{F}, \|x\|_{\ell^p} < \infty \right\},$$

where

$$\|x\|_{\ell^p} = \begin{cases} \left(\sum_{n=1}^\infty |x_n|^p \right)^{1/p}, & p \text{ finite} \\ \sup_n |x_n|, & p = \infty. \end{cases}$$

The same proofs for the ℓ^p-norms in finite-dimensional spaces apply, and show that ℓ^p is an NLS.

Theorem 1.5.2 (Hölder's inequality in ℓ^p). *If $1 \le p \le \infty$, $1/p + 1/q = 1$, and $x \in \ell^p$ and $y \in \ell^q$, then*

$$\sum_{n=1}^\infty |x_n y_n| \le \|x\|_{\ell^p} \|y\|_{\ell^q}.$$

Again, the proof is identical to the one for the finite-dimensional ℓ^p-norm.

Note that ℓ^∞ can be a bit weird relative to the rest of the ℓ^p spaces.

If p is finite, then ℓ^p has countably infinite dimension, i.e. it has a basis that's countable. This is subtle: the span of a basis is the set of *finite* linear combinations; in the infinite case, we would have to worry about convergence. Anyways, set

$$e^{i_n} = \begin{cases} 1, & i = n \\ 0, & i \ne n. \end{cases}$$

Then, a basis for ℓ^p, called the *Schauder basis*, is $\mathscr{B} = \{e^i\}_{i=1}^\infty$, and its span is

$$\operatorname{span}(\mathscr{B}) = \left\{ \alpha_{i_1} e^{i_1} + \alpha_{i_2} e^{i_2} + \cdots + \alpha_{i_n} e^{i_n} : n \in \mathbb{N}, \alpha_{i_j} \in \mathbb{F} \right\}.$$

Note that this is *not* a basis in the linear-algebraic sense (which would have to be uncountable); rather, this means that ℓ^p is the closure of $\operatorname{span}(\mathscr{B})$. That is, for all $x \in \ell^p$, there's a unique representation $x = \sum_{j=1}^\infty x_j e^j$, meaning that if x_N denotes the N^{th} partial sum, then $x_N \in \operatorname{span}(\mathscr{B})$ for all N, and

$$\|x - x_N\|_{\ell^p} = \left(\sum_{n=N+1}^\infty |x_n|^p \right)^{1/p} \longrightarrow 0.$$

This is a little weird, but the point is that, since you can't take infinite sums in a basis, things can get a little strange. But everything comes from the finite case.

ℓ^∞ does *not* have a countable basis. As a result, we sometimes consider subspaces with a countable basis. Define

$$c_0 = \{x \in \ell^\infty : \lim_{n\to\infty} x_n = 0\} \text{ and}$$
$$f_0 = \{x \in \ell^\infty : x_n = 0 \text{ for all but finitely many } n\}.$$

For example, $(1, 1, 1\dots) \in \ell^\infty$, but it's not in c_0 or f_0, and $(1, 1/2, 1/3, \dots)$ is in c_0 but not f_0. f_0 and c_0 inherit the ℓ^∞-norm and become NLSes in their own right.

If $1 \le p \le q < \infty$, then we have the following chain of inclusions:

$$f_0 \subsetneq \ell^p \subseteq \ell^q \subsetneq c_0 \subsetneq \ell^\infty.$$

If you're looking for examples (or, sometimes, counterexamples), c_0 and f_0 are often useful. For example, on f_0, we have a function $T : f_0 \to \mathbb{F}$ defined by

$$T(x) = \sum_{n=1}^\infty n x_n.$$

Since each $\alpha \in f_0$ is a finite sequence, then this is well-defined, and it's linear, but it's not bounded, since $T(e^i) = i$ but $\|e^i\|_{\ell^\infty} = 1$ for all i. Thus, we have a linear map which is not continuous.

Exercise. If $1 \le p \le \infty$, show that ℓ^p is Banach.

This is conceptually easy but a bit of work, coming down to calculus, and so we know that limits of Cauchy sequences exist. However, since ℓ^1 is a subspace of ℓ^∞, we can consider the NLS $(\ell^1, \|\cdot\|_{\ell^\infty})$; this space is not Banach.

Lemma 1.5.3. *Let $0 < p < 1$ and define ℓ^p in the same way as above. In this case, however, ℓ^p is not an NLS, because $\|\cdot\|_{\ell^p}$ isn't a norm.*

PROOF IDEA. We can look at $(\mathbb{F}^2, \|\cdot\|_{\ell^p})$ to see this: we proved that, given the triangle inequality, the unit ball is convex. However, the unit ball isn't convex when $p < 1$. The same proof works for ℓ^p, but with a less explicit picture. ⊠

The Hölder inequality allows us to create many continuous linear functionals $T : \ell^p \to \mathbb{F}$ when $1 \le p \le \infty$. Let q be the conjugate exponent (so $1/p + 1/q = 1$), and choose any $y \in \ell^q$. Then, we can produce a $T_y \in (\ell^p)^*$, i.e. $T_y : \ell^p \to \mathbb{F}$, defined by

$$T_y(x) = \sum_{n=1}^{\infty} x_n y_n.$$

Moreover, T_y is bounded, because $|T_y(x)| \le \|y\|_{\ell^q}\|x\|_{\ell^p}$.

This defines an inclusion $\ell^q \hookrightarrow (\ell^p)^*$.

Exercise. In fact, when p is finite, $\ell^q = (\ell^p)^*$. Moreover, $T : \ell^q \to (\ell^p)^*$ sending $T(y) \to T_y$ is a bounded operator, as $\|T_y\|_{(\ell^p)^*} = \|y\|_{\ell^q}$.

That is, the dual space is the conjugate space; to show this, figure out how to write $T(e^i)$ as y_i for some $y_i \in \ell^q$.

The above result is untrue for ℓ^∞; in fact, $(\ell^\infty)^* \supsetneq \ell^1$, but $c_0^* = \ell^1$.

That's all that we really need to say about ℓ^p for now; it's one step up from finite-dimensional spaces, and is a bit different, but not all that exotic. Right now, our examples are $\mathbb{F}^d$, which is finite-dimensional; ℓ^p when p is finite, which has a countable basis, and ℓ^∞, which has no countable basis.

Lesbegue spaces. Let $\Omega \subseteq \mathbb{R}^d$ be a measurable set with nonzero measure. We want to define a space of functions on Ω. However, when we talk about functions and measure, we really want to define two functions f and g as "the same" if $f(x) = g(x)$ except on a set of measure zero. If this is true, no integral can distinguish f and g.

FIGURE 1.1. An example of an LP space. Source: http://iloveaustin.tumblr.com/.

Definition. Let $1 \leq p < \infty$, and define $L^p(\Omega)$ be the set of measurable functions[6] $f : \Omega \to \mathbb{F}$ such that $\int_\Omega |f(x)|^p \, \mathrm{d}x$ is finite. $L^p(\Omega)$ becomes an NLS with the norm

$$\|f\|_p = \left(\int_\Omega |f(x)|^p \right)^{1/p},$$

though we'll have to show that.

Once again, we can define this for $p < 1$, but it won't end up being a norm.
When $p = \infty$, we'll do things a little differently, as usual.

Definition.

- A measurable $f : \Omega \to \mathbb{F}$ is *essentially bounded* by $K \in \mathbb{R}$ if $|f(x)| \leq K$ for almost every $x \in \Omega$ (i.e. the set where this is not true has measure zero).
- The *essential supremum* of f, denoted $\operatorname{ess\,sup}_{x\in\Omega}|f(x)|$, is the infimum of the K that essentially bound f.

Then, we can define $L^\infty(\Omega)$ as the set of (equivalence classes of) measurable functions whose essential suprema are finite, and $\|f\|_\infty = \operatorname{ess\,sup}_{x\in\Omega}|f(x)|$. This will also be an NLS, though we'll have to show that too.

Proposition 1.5.4. *If $0 < p \leq \infty$, then $L^p(\Omega)$ is a vector space, and $\|f\|_p = 0$ iff $f = 0$ almost everywhere on Ω.*

Proof. First, why is $L^p(\Omega)$ closed under addition? If p is finite, then

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)^p \leq 2^p(|f(x)|^p + |g(x)|^p),$$

so when one integrates, if $f, g \in L^p(\Omega)$, then the rightmost quantity is bounded and therefore the leftmost one is. Scalar multiplication (and the scaling property of the norm) is easy: just write down the definition.

For $p = \infty$, the maximum of the sum cannot be bigger than the sum of the maxima, so $\|f + g\|_\infty = \|f\|_\infty + \|g\|_\infty$. Scaling and scalar multiplication are also straightforward. ⊠

Thus, all we have left is the triangle inequality, which we'll show next class.

Lecture 6: 9/9/15

$L^p(\Omega)$ is Banach.

Recall that if $\Omega \subseteq \mathbb{R}^d$, then $L^p(\Omega)$ is the set of equivalence classes of measurable functions $\Omega \to \mathbb{F}$ with $\|f\|_p < \infty$, where $f \sim g$ if they differ on a set of measure zero. Then, the p-norm is

$$\|f\|_p = \begin{cases} \left(\int_\Omega |f(x)|^p \, \mathrm{d}x \right)^{1/p}, & p < \infty \\ \operatorname{ess\,sup}_{x\in\Omega}|f(x)|, & p = \infty. \end{cases}$$

Last time, we showed that $L^p(\Omega)$ is a vector space, and two of the properties of NLSes, the zero and scaling properties. Today we'll attack the triangle inequality; just as for ℓ^p, we'll need Hölder's inequality.

Proposition 1.6.1 (Hölder's inequality for L^p)**.** *Let $1 \leq p \leq \infty$ and $1/p + 1/q = 1$. If $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$, then $fg \in L^1(\Omega)$ and $\|fg\|_1 \leq \|f\|_p\|g\|_q$, with equality iff $|f(x)|^p$ is proportional to $|g(x)|^q$.*

Proof. If $p = 1$ or $p = \infty$, we already know that $\int_\Omega |f(x)g(x)| \, \mathrm{d}x \leq \|g\|_\infty \int_\Omega |f| \, \mathrm{d}x = \|f\|_1\|g\|_\infty$.

If $1 < p < \infty$, we know from Lemma 1.4.1 that $ab \leq a^p/p + b^q/q$, with equality when $a^p = b^q$. If $\|f\|_p = 0$ or $\|g\|_q = 0$, then we're done; otherwise,

$$\frac{|f(x)|}{\|f\|_p} \frac{|g(x)|}{\|g\|_q} \leq \frac{|f(x)|^p}{\|f\|_p^p p} + \frac{|g(x)|^q}{\|g\|_q^q q},$$

so integrating, we get

$$\frac{\int |fg|}{\|f\|_p\|g\|_q} \leq 1,$$

with equality when $|f(x)|^p/\|f\|_p^p = |g(x)|^q/\|g\|_q^q$, which gives us our proportionality. ⊠

[6] To be pedantic, the elements of $L^p(\Omega)$ are equivalence classes of functions that differ from f on a set of measure zero, since the integrals are the same.

Theorem 1.6.2 (Minkowski's inequality). *If $1 \leq p \leq \infty$, then $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.*

Proof. Notice that if f or g isn't in $L^p(\Omega)$, then its p-norm is infinite, so we're done. The result is also clear if $p = 1$ or $p = \infty$: the supremum of the sum is less than the sum of the suprema, and similarly with absolute value.

So we only have to worry about $1 < p < \infty$, and here we'll use a similar trick as for ℓ^p spaces, taking one copy of a p^{th} power.

$$\begin{aligned}\|f+g\|_p^p &= \int_\Omega |f(x)+g(x)|^p \,\mathrm{d}x \\ &\leq \int_\Omega |f(x)+g(x)|^{p-1}(|f(x)|+|g(x)|)\,\mathrm{d}x.\end{aligned}$$

Using Hölder's inequality,

$$\begin{aligned}&\leq \left(\int_\Omega |f(x)+g(x)|^{(p-1)q}\right)^{1/q}\big(\|f\|_p + \|g\|_p\big) \\ &= \|f+g\|_p^{p-1}\big(\|f\|_p + \|g\|_p\big),\end{aligned}$$

so dividing by $\|f+g\|^{p-1}$, we're done. ⊠

L^p spaces are very important in analysis, and form an important set of examples for NLSes. A little later, we'll show that they're complete, but we should note that we're measuring the size of a function using varying p, which measure different things, between emphasizing large values at a point, or large values at infinity.

On $\mathbb{R}$, imagine a function that goes to ∞ as $x \to 0^+$ and 0 as $x \to \infty$. If p is large, we're emphasizing the large values of the function, so if it grows too quickly it might not be in $L^p(\mathbb{R})$. If p is small, then we're emphasizing the long tail as $x \to \infty$; if it dies too slowly, it might not be in $L^p(\mathbb{R})$. An instructive example is x^p, which is in some L^q spaces but not others.

An easier way to think about this is to bound Ω, so we don't have to worry about long tails.

Proposition 1.6.3. *Let μ denote the Lesbegue measure, and suppose $\mu(\Omega)$ is finite. Let $1 \leq p \leq q \leq \infty$.*

(1) If $f \in L^q(\Omega)$, then $f \in L^p(\Omega)$, and in fact $\|f\|_p = (\mu(\Omega))^{1/p-1/q}\|f\|_q$.
(2) If $f \in L^\infty(\Omega)$, then $f \in L^p(\Omega)$ for $1 \leq p \leq< \infty$, and $\lim_{p\to\infty}\|f\|_p = \|f\|_\infty$.
(3) If $f \in L^p(\Omega)$ for $1 \leq p < \infty$ and $\|f\|_p \leq K$ for all such p, then $f \in L^\infty(\Omega)$ and $\|f\|_\infty \leq K$.

These will be proven in the homework. Part (2) is the reason the L^∞-norm is named such. Note also that there exist f such that $f \in L^p(\Omega)$ for $1 \leq p < \infty$ but $f \notin L^\infty(\Omega)$, even when Ω has finite measure.

The general proof idea is to consider sets of bad points and see what happens.

Proposition 1.6.4. *For $1 \leq p \leq \infty$ and Ω measurable, $L^p(\Omega)$ is complete.*

Thus, we have another useful class of Banach spaces.

Proof. As usual, we'll start with a Cauchy sequence $\{f_n(x)\}_{n=1}^\infty$ in $L^p(\Omega)$. The idea will be to write

$$f_n(x) = f_1(x) + f_2(x) - f_1(x) + f_3(x) - f_2(x) + \cdots + f_n(x) - f_{n-1}(x),$$

so if we group the $f_i(x) - f_{i-1}(x)$, then these pieces should be small, and therefore we ought to converge to some function $f(x)$. There are technical problems, though, since we don't know how fast the f_n converge, so we need to try $f_i(x) - f_{i-k}(x)$ for $k > 1$. Moreover, we'll use absolute values. This is the idea; now, let's write it down carefully.

First, select a subsequence such that $\|f_{n_{j+1}} - f_{n_j}\| \leq 2^{-j}$ for all j; we can do this because if we have n_{j-1}, there's an n_j such that $\|f_{n_j} - f_m\| \leq 2^{-j}$ when $m \geq n_j \geq n_{j-1}$.

Let

$$F_m(x) = |f_{n_1}(x)| + \sum_{j=1}^m |f_{n_{j+1}}(x) - f_{n_j}(x)| \geq 0,$$

so that $\{F_m(x)\}$ is increasing in m pointwise, so there's a limit (which might be ∞, but that's OK). Let $F(x) = \lim_{m\to\infty} F_m(x) \in [0,\infty]$. Then,

$$\|F_m\|_p \leq \|f_{n_1}\|_p + \sum_{j=1}^n 2^{-j} \leq \|f_{n_1}\|_p + 1,$$

which is finite. But more interestingly, $F \in L^p(\Omega)$ too! We'll have to treat L^∞ as a special case again.

If p is finite, we'll use the monotone convergence theorem.

$$\begin{aligned}\int_\Omega |F(x)|^p \, dx &= \int_\Omega \lim_{m\to\infty} |F_m(x)|^p \, dx \\ &\le \lim_{m\to\infty} \int_\Omega |F_m(x)|^p \, dx \\ &\le \|f_{n_1}\|_p + 1,\end{aligned}$$

which is finite.

When $p = \infty$, then $|F_m(x)| \le \|F_m\|_\infty \le \|f_{n_1}\|_\infty + 1$ for all $x \notin A_m$, where $\mu(A_m) = 0$. Thus, if $A = \bigcup_{n=1}^\infty A_n$, then $\mu(A) = 0$ too. Thus, $|F(x)| = \lim_{m\to\infty} |F_m(x)| \le K$ for some K and all $m, x \notin A$, so $F \in L^\infty(\Omega)$.

Now,

$$f_{n_j+1}(x) = f_{n_1}(x) + (f_{n_2}(x) - f_{n_1}(x)) + \cdots + (f_{n_j+1}(x) - f_{n_j}(x)).$$

Thus, this converges absolutely *pointwise*[7] to some $f(x)$, so f is measurable. Now, $|f_{n_j}(x)| \le F(x)$, so $|f(x)| \le F(x)$, and therefore $f \in L^p(\Omega)$.

But we need that $\|f_{n_j} - f\|_p \to 0$, so let's think about that. Again, we have to argue differently when $p = \infty$.

When p is finite, we'll use the dominated convergence theorem on $|f_{n_j}(x) - f(x)| \le F(x) + |f(x)| \in L^p(\Omega)$:

$$\lim_{j\to\infty} \int_\Omega |f_{n_j}(x) - f(x)|^p \, dx \le \int_\Omega \lim_{j\to\infty} |f_{n_j}(x) - f(x)|^p \, dx \longrightarrow 0.$$

When p is infinite, for any j and k, there's a set B_{n_j,n_k} with measure zero such that on $\Omega \setminus B_{n_j,n_k}$, $|f_{n_j}(x) - f_{n_k}(x)| \le \|f_{n_j} - f_{n_k}\|_\infty$. Thus,

$$B = \bigcup_j \bigcup_k B_{n_j,n_k}$$

is a countable union, so $\mu(B) = 0$. Since $\{f_{n_j}\}$ is Cauchy, then for any $x \notin B$ and $\varepsilon > 0$, there's an $N > 0$ such that if $j, k \ge N$, then $|f_{n_j}(x) - f_{n_k}(x)| < \varepsilon$, so taking the pointwise limit $f_k(x) \to f(x)$, $|f_{n_j}(x) - f(x)| < \varepsilon$. Thus, since we're avoiding B, $\|f_{n_j} - f\|_\infty < \varepsilon$.

We're almost done: we have $f_{n_j} \to f$ in L^p, but we need $f_n \to f$ in L^p. If $\varepsilon > 0$, then there exists an $N > 0$ such that $\|f_n - f_{n_j}\|_p < \varepsilon/2$ for all $n, n_j \ge N$. Therefore $|f_{n_j} - f|_p < \varepsilon/2$ for all $n_j \ge N$, and therefore the triangle inequality tells us that

$$\|f_n - f\|_p \le \|f_n - f_{n_j}\|_p + \|f_{n_j} - f\|_p < \varepsilon. \qquad \boxtimes$$

If you examine the proof, we've also proven an interesting result.

Corollary 1.6.5. *If $1 \le p < \infty$ and $\{f_n\}_{n=1}^\infty$ is a sequence in $L^p(\Omega)$ converging to f in the L^p-norm, then there exists a subsequence $\{f_{n_j}\}_{j=1}^\infty$ such that $f_{n_j}(x) \to f(x)$ pointwise a.e.*

So convergence in L^p implies pointwise convergence of a subsequence almost everywhere. We'll use this later.

It turns out that the dual space to $L^p(\Omega)$ is $L^q(\Omega)$, where q is the conjugate exponent. Given a $g \in L^q(\Omega)$, define an operator $T_g : L^p(\Omega) \to \mathbb{F}$ by

$$T_g(f) = \int_\Omega f(x) g(x) \, dx,$$

which makes sense and is finite by Proposition 1.6.1. Thus, this is well-defined, and linear because the integral is. It's continuous, because it's bounded (by Hölder's inequality again): $T_g(f) \le \|g\|_q \|f\|_p$, so $\|t_g\| \le \|g\|_q$, and it's probably not a surprise that's actually an equality: choose something like $f(x) = |g(x)|^{q/p}/\|g\|_q$ (maybe with a power in the denominator), to see that the bound is sharp.

Thus, we've shown that $L^q(\Omega) \subseteq (L^p(\Omega))^*$ in some sense, for $1 \le p \le \infty$. However, if p is finite, then $L^q(\Omega) = (L^p(\Omega))^*$; there are no other continuous linear functionals. When $p = \infty$, there are more, so the dual space is the space of positive measures: $g(x)\,dx$ is a measure, but there are other measures that aren't of that form.

We won't prove this, but it follows from a deep theorem in analysis called the Radon-Nikodym theorem.

[7] We have multiple notions of convergence floating around; be careful to distinguish pointwise convergence, uniform convergence, and convergence in L^p.

Lecture 7: 9/11/15

The Hahn-Banach Theorem.

"Almost everything has three properties. Have you noticed that?"

Corollary 1.7.1. *Let X be an NLS, $Y \subset X$ be a linear subspace, and $f : Y \to \mathbb{F}$ be bounded. Then, there exists an $F \in X^*$ such that $F|_Y = f$ and $\|F\|_{X^*} = \|f\|_{Y^*}$.*

Though L^p functions can be complicated, all of them can be well-approximated by less complicated functions. Recall that a *simple* function is a Lesbegue-integrable function that takes on only finitely many values, and that a function is *compactly supported* if it is equal to 0 outside of a compact set.

Proposition 1.7.2. *For $1 \le p \le \infty$, the set $\mathscr{S}$ of all measurable simple functions with compact support is dense in $L^p(\Omega)$.*

This says that for any $f \in L^p(\Omega)$ and $\varepsilon > 0$, there's a $\varphi \in \mathscr{S}$ such that $\|f - \varphi\|_{L^p(\Omega)} < \varepsilon$. The proof comes from measure theory: the integral was defined by the limit of approximations by simple functions, and so these simple functions are successively better approximations.

Definition. Let $C_0^\infty(\Omega)$ denote the space of compactly supported, continuous functions.

Proposition 1.7.3. *If Ω is an open set and $1 \le p < \infty$, then $C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$.*

The proof follows from another measure-theoretic result called Lusin's theorem.

Now, we'll move into some deeper (and, well, harder) theorems and questions in functional analysis. We'll start with a question.

Let X be a finite-dimensional NLS and $Y \subset X$ be a subspace. Given a linear $f : Y \to \mathbb{R}$, can we extend f to X? The answer is yes. But what about the infinite-dimensional case? Here, we care about continuous (so bounded) linear operators.

Once again, the answer is that it's possible, but this is hard to prove, and it'll take us a while to prove that. We won't need all the properties of a norm to prove that, so we can weaken what we need in terms of the norm.

Definition. Let X be a vector space over $\mathbb{F}$. We say that $p : X \to [0, \infty)$ is *sublinear* if

(1) $p(\lambda x) = \lambda p(x)$ for all $\lambda \ge 0$ and $x \in X$, and
(2) $p(x + y) \le p(x) + p(y)$ for all $x, y \in X$.

If in addition p satisfies (1) for all $\lambda \in \mathbb{F}$, p is called a *seminorm*.

If a seminorm also satisfies $p(x) = 0$ implies $x = 0$, then p is a norm.

The Hahn-Banach theorem about extension of linear operators will apply perfectly well to sublinear operators. First, let's deal with the simplest version we can think of.

Lemma 1.7.4. *Let X be a vector space over $\mathbb{R}$ and $Y \subsetneq X$ be a linear subspace. Let p be sublinear on X and $f : Y \to \mathbb{R}$ be linear such that $f(y) \le p(y)$ for all $y \in Y$. For a given $x_0 \in X \setminus Y$, let $\widetilde{Y} = \operatorname{span}\{Y, x_0\} = Y + \mathbb{R}x_0 = \{y + \lambda x_0 : y \in Y, \lambda \in \mathbb{R}\}$; then, there exists a linear map $\widetilde{f} : \widetilde{Y} \to \mathbb{R}$ such that $\widetilde{f}|_Y = f$ and $-p(-x) \le \widetilde{f}(x) \le p(x)$ for all $x \in \widetilde{Y}$.*

The definitions of $\widetilde{Y}$ all show that it's "Y plus one more dimension."

Proof. If $\widetilde{f}(x) \le p(x)$, then $-\widetilde{f}(x) = \widetilde{f}(-x) \le p(-x)$, so $\widetilde{f}(x) \ge -p(-x)$, and so the lower bound comes for free.

We'll present the proof not as a cleaned-up proof, but how one would think of the proof when trying to prove it.

If we had such an $\widetilde{f}$, what would it look like? $\widetilde{y} \in \widetilde{Y}$ can be written $\widetilde{y} = y + \lambda x_0$ for some $y \in Y$ and $\lambda \in \mathbb{R}$, so $\widetilde{f}(\widetilde{y}) = \widetilde{f}(y + \lambda x_0) = \widetilde{f}(y) + \lambda \widetilde{f}(x_0) = f(y) + \lambda \widetilde{f}(x_0)$, since $\widetilde{f}|_Y = f$.

So if we had defined $\alpha \in \mathbb{R}$ to be $\widetilde{f}(x_0)$, then we get a function, and correspondingly, given $\widetilde{f}$, we get $\alpha = \widetilde{f}(x_0)$. Thus, $\widetilde{f}$ is characterized by α.

However, we need to be careful: is this really well-defined? We chose y; what if you choose a different one than I do? It turns out that you have to choose the same y: suppose $\widetilde{y} = y + \lambda x_0 = z + \mu x_0$ for $y, z \in Y$ and $\lambda, \mu \in \mathbb{R}$. Thus, $y - z = (\mu - \lambda)x_0$, but $y - z \in Y$, so since $x_0 \notin Y$, then $\mu - \lambda = 0$, and therefore $y = z$; thus, this choice of y is well-defined, so $\widetilde{f}$ really is characterized by α.

So now we need to find an α such that $\widetilde{f}(\widetilde{y}) = f(y) + \lambda\alpha \leq p(y + \lambda x_0)$. If $\lambda = 0$ this works, so let's focus on $\lambda \neq 0$. Rescale: let $y = -\lambda x$, so we want to show that $f(-\lambda x) + \lambda \cdot \alpha \leq p(\lambda(x_0 - x))$, or $\lambda(-f(x) + \alpha) \leq p(-\lambda(x - x_0))$.

If $\lambda < 0$, then divide by $-\lambda$: $f(x) - \alpha \leq p(x - x_0)$; when $\lambda > 0$, we get a change in sign: $-(f(x) - \alpha) \leq p(-(x - x_0))$. Together, this means $-p(-(x - x_0)) \leq f(x) - \alpha \leq p(x - x_0)$. Rearranging,

$$f(x) - p(x - x_0) \leq \alpha \leq f(x) + p(x_0 - x).$$

This is our requirement; that is, if there's an α that satisfies this for all $x \in Y$, then we have our desired linear functional.

So let $a = \sup_{x \in Y}(f(x) - p(x - x_0))$ and $b = \inf_{x \in Y}(f(x) + p(x_0 - x))$. Now we can ignore α and ask, is it true that $a \leq b$? If so, we're done.

Let $x, y \in Y$. Since p is sublinear, then

$$\begin{aligned} f(x) - f(y) &= f(x - y) \leq p(x - y) \\ &\leq p(x - x_0) + p(x_0 - y) \\ \implies f(x) - p(x - x_0) &\leq f(y) + p(x_0 - y). \end{aligned}$$

In the last equation, first take the infimum on the left, which is a, and the right side doesn't change; then, take the supremum on the right, which is b, and the left side doesn't change. Thus $a \leq b$. ⊠

This proof can be shortened: if you start with α, then suddenly magical things happen, but our proof helps it make more sense and feel more rigorous.

Transfinite Induction and Generalizing Lemma 1.7.4. Applying this inductively, we can extend a finite number of dimensions, and even a countable number of dimensions! However, standard induction doesn't allow us to extend by an uncountable number of dimensions. This will require a technique called transfinite induction, and therefore a brief vacation into set theory.

Definition. A *ordering* on a set $\mathscr{S}$ is a binary relation $\preceq$ such that for all $x, y, z \in \mathscr{S}$,

(1) $x \preceq x$,
(2) if $x \preceq y$ and $y \preceq x$, then $x = y$, and
(3) if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

Not every set can be ordered. However, some can be partially ordered; a *partial order* on a set is the same except that only some pairs $x \preceq y$ are defined, but the same order axioms are satisfied (in particular, $x \preceq x$ is always defined and true, and if $x \preceq y$ and $y \preceq z$, then $x \preceq z$ is defined and true). A *chain* in a partially ordered set $\mathscr{S}$ is a $\mathscr{C} \subset \mathscr{S}$ such that $\preceq|_{\mathscr{C}}$ is a total order: all pairs of elements of $\mathscr{C}$ can be compared.

Example 1.7.5. On $\mathbb{C}$, write $z = r_z e^{i\theta_z}$, with $\theta_z \in [0, 2\pi)$.

(1) An ordering on $\mathbb{C}$ can be given by $x \preceq y$ iff $r_x < r_y$ or $r_x = r_y$ and $\theta_x \leq \theta_y$.
(2) A partial ordering on $\mathbb{C}$ can be given by $x \preceq y$ iff $\theta_x = \theta_y$ and $r_x \leq r_y$ (and is undefined if $\theta_x \neq \theta_y$).

We'll need a more complicated order, which requires using Zorn's lemma. This comes from an axiom of set theory called the Axiom of Choice, which states that, given any collection of nonempty sets, it's possible to choose one element out of each set.

Zorn's lemma is equivalent to the Axiom of Choice, but it somehow seems harder to believe.

Lemma 1.7.6 (Zorn's lemma). *Let $\mathscr{S}$ be a nonempty, partially ordered set, and suppose every chain $\mathscr{C} \subseteq \mathscr{S}$ has an upper bound, i.e. for all $\mathscr{C}$, there's a $u \in \mathscr{C}$ such that $x \preceq U$ for all $x \in \mathscr{C}$. Then, $\mathscr{S}$ has at least one maximal element m, i.e. if $m \preceq x$ for some $x \in \mathscr{S}$, then $x = m$.*

Next time, we'll use this to extend by an uncountable number of dimensions; then, we'll remove the requirement that the base field is real.

Lecture 8: 9/14/15

The Hahn-Banach Theorem, II.

Recall that we're in the middle of proving the Hahn-Banach theorem, and therefore should remember the results we're going to need. We defined orders and partial orders and chains within partially ordered sets last lecture, and cited Zorn's lemma, Lemma 1.7.6, which gives conditions for when a partially ordered set has a maximal element. Finally, we have Corollary 1.7.1 in mind as a long-term goal.

Since we have a possibly countable number of dimensions, we have to use transfinite induction to prove the most general theorem, which is why Zorn's lemma shows up.

Theorem 1.8.1 (Hahn-Banach theorem for real vector spaces). *Let X be a vector space over $\mathbb{R}$, $Y \subset X$ be a subspace, and p be sublinear on X. If $f : Y \to \mathbb{R}$ is linear on Y and $f(x) \le p(x)$ for all $x \in Y$, then there exists a linear $F : X \to \mathbb{R}$ such that $F|_Y = f$ and $-p(-x) \le F(x) \le p(x)$ for all $x \in X$.*

Proof. Let $\mathscr{S}$ denote the set of all linear extensions g of f to a subspace $D(g) \subset X$ containing Y, and such that $g(x) \le p(x)$ for all $x \in D(g)$. Since $f \in \mathscr{S}$, then f is nonempty. We'll turn $\mathscr{S}$ into a partially ordered set by saying that $g \preceq h$ if h extends g, i.e. $D(g) \subseteq D(h)$ and $h|_{D(g)} = g$.

Let $\mathscr{C}$ be a chain in $\mathscr{S}$, and let

$$D = \bigcup_{g \in \mathscr{C}} D(g).$$

Since these $D(g)$ are nested (i.e. one of $D(g) \subset D(h)$ or $D(g) \supset D(h)$ for all $g, h \in \mathscr{C}$), then D is a vector space.[8] Then, we'll define $g_{\mathscr{C}}$ as follows: if $x \in D$, then $x \in D(g)$ for some $g \in \mathscr{C}$, so define $g_{\mathscr{C}}(x) = g(x)$. Is this well-defined? Yes, because if $x \in D(g) \cap D(h)$, then without loss of generality $g \preceq h$, and so $g(x) = h(x)$. Thus, we get a function $g_{\mathscr{C}} : D \to \mathbb{R}$, which is linear (which follows from its definition), and is bounded by p (specifically, $g(x) \le p(x)$ for all $x \in D$), since each $g \in \mathscr{C}$ is. Thus, $g_{\mathscr{C}} \in \mathscr{C}$, and it's an upper bound for $\mathscr{C}$.

Applying Zorn's lemma, we have a maximal element F for $\mathscr{S}$; since $F \in \mathscr{S}$, then it's a linear extension of f and is bounded by p. So the final question is, what's $D(F)$? Suppose $D(F) \subsetneq F$; then, there exists some $x_0 \in X \setminus D(F)$, so by Lemma 1.7.4 we can extend F to $\operatorname{span}\{D(F), x_0\}$. But this contradicts the fact that $D(F)$ is maximal. Thus, $D(F) = X$. ⊠

Awesome. Now, let's deal with complex vector spaces. Since we want scalar multiplication for all $\lambda \in \mathbb{C}$, we'll have to use a seminorm instead.

Theorem 1.8.2 (Hahn-Banach theorem for complex vector spaces). *Let X be a vector space over $\mathbb{F}$, $Y \subset X$ be a linear subspace, and p be a seminorm. If $f : Y \to \mathbb{F}$ is a linear functional such that $|f(x)| \le p(x)$ for all $x \in Y$, then there exists an extension $F : X \to \mathbb{F}$ such that $F|_Y = f$ and $|F(x)| \le p(x)$ for all $x \in X$.*

Proof. We'll assume $\mathbb{F} = \mathbb{C}$, since the real case comes from Theorem 1.8.1. Then, we can write $f(x) = g(x) + ih(x)$ for g, h real linear, since

$$\begin{aligned} f(x+g) &= g(x+y) + ih(x+y) \\ &= f(x) + f(y) = g(x) + g(y) + ih(x) + ih(y), \end{aligned}$$

and scalar multiplication is similar, though only for real scalars. Instead, $f(ix) = if(x) = -h(x) + ig(x)$, and this is also $g(ix) + ih(ix)$. Thus, $h(x) = -g(ix)$. That is, since f is linear, $f(x) = g(x) - ig(ix)$, which is a general fact.

Since g is real linear, then Theorem 1.8.1 yields a real extension G on X, because $|g(x)| \le |f(x)| \le p(x)$, and we have that $|G(x)| \le p(x)$.

Define $F(x) = G(x) - iG(ix)$, which is a function $F : X \to \mathbb{C}$ that commutes with addition and real scalar multiplication. Thus, we need to check complex scalar multiplication, and therefore that $F(ix) = iF(x)$. Let's check that:

$$F(ix) = G(ix) - iG(-x) = G(ix) + iG(x) = i(G(x) - iG(ix)).$$

Therefore F is $\mathbb{C}$-linear. Moreover, if $x \in Y$, then $F(x) = G(x) - iG(ix) = g(x) - ig(ix)$, and therefore $F|_Y = f$ as desired. Thus, the only thing left to check is the bound.

Let $x \in X$, and write $F(x) = re^{i\theta}$. Then,

$$r = |F(x)| = e^{-i\theta} F(x) = F\big(e^{-i\theta} x\big) = G\big(g^{-i\theta}(x)\big) - iG\big(-e^{-i\theta} x\big),$$

but the second term is imaginary, and therefore must be zero. Then,

$$\le p\big(e^{-i\theta}(x)\big) = |e^{-i\theta}| p(x) = p(x).$$ ⊠

[8]This is an important point; the union of subspaces isn't in general a vector subspace when they're not nested.

As a corollary, notice that $p(x) = \|f\|_{Y^*}\|x\|_X$.

The Hahn-Banach theorem has a great number of corollaries, which provide a lot of insight into NLSes and Banach spaces.

Corollary 1.8.3. *Let X be an NLS and $x_0 \in X \setminus 0$ be fixed. Then, there exists an $f \in X^*$ such that $\|f\|_{X^*} = 1$ and $f(x_0) = \|x_0\|$.*

The idea is to define f on a subspace where it's easy to define, and then extend.

Proof. Let $Z = \mathbb{F}x_0$, and define $h : Z \to \mathbb{F}$ by $h(\lambda x_0) = \lambda\|x_0\|$. Then, $|h(x_0)| = |\lambda|\|x_0\| = \|\lambda x_0\|$, so $|h(x)| \le \|x\|$ for all $x \in Z$ and $\|h\| = 1$. Then, we use Theorem 1.8.2 to extend h to the desired f. ⊠

Corollary 1.8.4. *For any $\alpha \in \mathbb{F}$, there exists an $f \in X^*$ such that $f(x_0) = \alpha\|x_0\|$ (and therefore $\|f\|_{X^*} = |\alpha|$).*

The proof is the same as for Corollary 1.8.3, but one defines $h(\lambda x_0) = \alpha\lambda\|x_0\|$ instead.

Here's a more interesting corllary.

Corollary 1.8.5. *Let X be an NLS and $x \in X$. Then,*

$$\|x\| = \sup_{\substack{f \in X^* \\ f \neq 0}} \frac{|f(x)|}{\|f\|_{X^*}} = \sup_{\substack{f \in X^* \\ \|f\|_{X^*}=1}} \frac{|f(x)|}{\|f\|_{X^*}}.$$

Often, when one knows the structure of the dual space better than that of the original space, this can be a useful way to calculate a norm.

Proof. For all $f \in X^*$ with $f \neq 0$, we know $|f(x)| \le \|f\|_{X^*} \le \|x\|$, so we know the supremum is still bounded by $\|x\|$. To get the other bound, we need the Hahn-Banach theorem, which says that there exists a $\widetilde{f} \in X^*$ such that $\widetilde{f}(x) = \|\widetilde{f}\|\|x\|$; then,

$$\sup_{\substack{f \in X^* \\ f \neq 0}} \frac{|f(x)|}{\|f\|_{X^*}} \ge \frac{\widetilde{f}(x)}{\|f\|} = \|x\|.$$ ⊠

The idea here is that we can look at $\|x\|$, which is a calculation involving an abstract vector, or $\{|f(x)|\}_{f \in X^*}$, which is a collection of numbers, which sometimes is nicer. This is a common theme in functional analysis. The following result is related, at least in ideas.

Proposition 1.8.6. *If $f(x) = f(y)$ for all $f \in X^*$, then $x = y$.*

We'll prove this next time.

Lecture 9: 9/16/15

Separability.

"Quis separabit? [Who will separate us?]" – an Irish motto

Recall that we're in the middle of exploring the consequences of the Hahn-Banach theorem, Theorems 1.8.1 and 1.8.2. For example, if X is an NLS and $x_0 \in X$, then there's an $f \in X^*$ such that $f(x_0) = \|x_0\|$ (Corollary 1.8.4), that you can calculate $\|x\|$ from the norms of $f \in X^*$ (Corollary 1.8.5), and more.

Proposition 1.9.1. *If X is an NLS, then X^** separates *points in X, i.e. for any $x, y \in X$, there exists an $f \in X^*$ such that $f(x) \neq f(y)$, and if $f(x) = f(y)$ for all $f \in X^*$, then $x = y$.*

The recurring theme is that if you know what all the linear functionals do to an element, you know what that element is.

Proof. Choose $x, y \in X^*$ such that $x \neq y$. Then, $x - y = x_0 \in X$ and $x_0 \neq 0$, and there exists an $f \in X^*$ such that $f(x_0) \neq 0$, so $0 \neq f(x-y) = f(x) - f(y)$. ⊠

Corollary 1.9.2. *If $f(x) = 0$ for all $x \in X^*$, then $x = 0$.*

Oftentimes, one creates simple functionals by doing something interesting on a finite-dimensional subspace and then extending à la the Hahn-Banach theorem.

Definition. In an NLS X, the *distance* between a subspace $Y \subset X$ and a $w \in X$ is $\operatorname{dist}(w, Y) = \inf_{y \in Y}\|w - y\|$.

This is nonnegative, and sometimes it's zero even when $w \notin Y$.

Lemma 1.9.3 (Mazur Separation Thm. I)**.** *Let X be an NLS, $Y \subset X$ be a subspace, and $w \in X \setminus Y$. Suppose $d = \operatorname{dist}(w, Y) > 0$. Then, there exists an $f \in X^*$ with*
- $\|f\| \le 1$,
- $f(w) = d$, *and*
- $f(y) = 0$ *for all* $y \in Y$.

Proof. Let $Z = Y + \mathbb{F}w$. Then, any $z \in Z$ has a unique representation as $z = y + \lambda w$ for exactly one choice of $y \in Y$ and $\lambda \in \mathbb{F}$ (which we discussed last time).

Then, define $g : Z \to \mathbb{F}$ by $g(y + \lambda w) = \lambda d$. g is clearly linear, but it's less clear why $\|g\| \le 1$.

$$\left| g\left(\frac{y + \lambda w}{\|y + \lambda w\|} \right) \right| = \frac{|\lambda| d}{\|y + \lambda w\|} = \frac{d}{\|(1/\lambda)y + w\|}.$$

Since $(1/\lambda)y \in Y$, then $\|(1/\lambda)y + w\| \ge d$, and therefore $d/\|(1/\lambda)y + w\| \le 1$. Then, we use the Hahn-Banach theorem to extend to X. ⊠

We'll introduce another notion, entirely topological, which will be useful.

Definition. A topological space X is *separable* if it contains a countable dense subset, i.e. a $\mathscr{D} \subset X$ such that $\overline{\mathscr{D}} = X$.

A space might be large and scary, but if it's separable, then everything is close, and therefore we can get a little control on it.

Example 1.9.4.

(1) $\mathbb{Q} \subset \mathbb{R}$. $\mathbb{Q}$ is countable and every real number can be arbitrarily well approximated by rational numbers, so $\mathbb{R}$ is separable.
(2) $\mathbb{Q}(i) = \mathbb{Q} + i\mathbb{Q} \subseteq \mathbb{C}$ is countable and dense, so $\mathbb{C}$ is separable.
(3) $\mathbb{F}^d$ is also separable, with the countable dense subset either $\mathbb{Q}^d$ or $\mathbb{Q}(i)^d$.
(4) If $1 \le p < \infty$, our Schauder basis for ℓ^p is uncountable, but we can take instead the $\mathbb{Q}(i)$-span (or the $\mathbb{Q}$-span if $\mathbb{F} = \mathbb{R}$) of $\{e_i\}_{i=1}^\infty$; this is a countable dense subset of ℓ^p, so ℓ^p is separable.
(5) If $1 \le p < \infty$, then $L^p(\Omega)$ is separable. This one is a little more surprising. The set S of simple functions (functions which are constant on a finite number of intervals) is dense in Ω, but uncountable, so we have to restrict it in two ways: first, restrict the allowed intervals to have rational coefficients, and then restrict the functions to take on values in $\mathbb{Q}(i)$ (or $\mathbb{Q}$; we'll assume that when we talk about $\mathbb{Q}(i)$, then we mean $\mathbb{Q}$ for $\mathbb{R}$). Thus restricted, we have our countable dense subset.

This argument doesn't work for $L^\infty(\Omega)$, since simple functions aren't dense in it, and in fact $L^\infty(\Omega)$ isn't separable.

Proposition 1.9.5. *Let X be an NLS. If X^* is separable, then X is separable.*

The converse isn't true, because $L^1(\Omega)$ is separable but $L^\infty(\Omega)$ isn't. So if you start with a separable space, your dual might be bigger.

Proof. Let $\{f_n\}_{n=1}^\infty$ be a countable, dense subset of X^*. We'll use this to construct a countable, dense subset of X. Since $\|f\| = \sup_{\|x\|=1} |f(x)|$, then we can choose for each n an x_n such that $\|x_n\| = 1$ and $|f_n(x_n)| \ge (1/2)\|f_n\|$, giving us a sequence $\{x_n\}_{n=1}^\infty$.

Then, let $\mathscr{D} = \operatorname{span}_{\mathbb{Q}(i)}\{x_n\}$, which is still countable, and we'll show that $\overline{\mathscr{D}} = X$. Suppose that it weren't: then, there exists a $w \in X \setminus \overline{\mathscr{D}}$. Let $d = \operatorname{dist}(w, \overline{\mathscr{D}}) = \inf_{x \in \overline{\mathscr{D}}} \|w - x\| > 0$. If we can show that $\|w - y_n\| \to 0$ for some sequence $\{y_n\}_{n=1}^\infty$, then since $\overline{\mathscr{D}}$ is closed, that would imply $w \in \overline{\mathscr{D}}$.

Since $\overline{\mathbb{Q}(i)} = \mathbb{C}$ (or, in the real case, $\overline{\mathbb{Q}} = \mathbb{R}$),[9] then $\overline{\mathscr{D}}$ is a linear subspace of X; thus, by Lemma 1.9.3, there exists an $f \in X^*$ such that $f|_{\overline{\mathscr{D}}} = 0$ and $f(w) = d > 0$. But there's a sequence $\{f_{n_k}\}_{k=1}^\infty$ such that $f_{n_k} \to f$. Thus,

$$\|f_{n_k} - f\| \ge |f(x_{n_k}) - f_{n_k}(x_{n_k})| = |f_{n_k}(x_{n_k})| \ge \frac{1}{2}\|f_{n_k}\|.$$

Since $f_{n_k} - f \to 0$, then this means $f_{n_k} \to 0$, and so $f = 0$. But this is a contradiction. ⊠

[9]Usually, $\overline{\mathbb{Q}}$ tends to denote algebraic closure; today, we're talking about topological closure.

So far, we've always looked at sets that are subspaces. Here's an example where we don't do that.

Definition. Let X be an NLS and $C \subseteq X$ be a sub*set* (not necessarily a subspace). Then, C is *balanced* if for any $\lambda \in \mathbb{F}$ with $|\lambda| \leq 1$ and any $x \in C$, we have $\lambda x \in C$.

For example, if $\mathbb{F} = \mathbb{C}$, then this implies that C is invariant under rotation, as well as contractions. Note that all subspaces are balanced.

Lemma 1.9.6 (Mazur Separation Thm. II)**.** *Let X be an NLS and $C \subseteq X$ be a closed, convex, and balanced set. Then, for any $w \in X \setminus C$, there exists an $f \in X^*$ such that $|f(x)| \leq 1$ for $x \in C$ and $f(w) > 1$.*

PROOF. Since C is closed and $w \notin C$, we can choose a ball $B + w$ about w (so B is a ball centered at the origin) such that $B \cap C = \emptyset$. Then, we can define the *Minkowski functional* $p : X \to [0, \infty)$ by

$$p(x) = \inf\left\{t > 0 : \frac{x}{t} \in C + B\right\}.$$

Here, $C + B$ is a slight fattening of our set C, but we can guarantee that $w \notin C + B$. Moreover, $0 \in C$, because C is balanced; therefore, $p(x)$ is always finite. We also know that $p(x) \leq 1$ if $x \in C$ and $p(w) > 1$.

Moreover, p is a seminorm: since C is balaced, $p(\lambda x) = p(|\lambda| x) = |\lambda| p(x)$. We also have the triangle inequality, which is left to the reader.

Now, we use Theorem 1.8.2: let $Y = \mathbb{F}w$, and if $f(\lambda w) = \lambda p(w)$, then $f(w) = p(w) > 1$, and $|f(\lambda w)| = |\lambda| p(w) = p(\lambda w)$, so we have a nice bound. Therefore, we can extend f to an F such that $F(w) > 1$ and $F(x) \leq 1$ if $x \in C \subset C + B$. ⊠

The key idea is the Minkowski functional; once you write that down, you're basically done.

Lecture 10: 9/18/15

The Minkowski Functional and the Baire Category Theorem.

Last time, we had to rush through the Minkowski functional, so today we'll talk a little more about it. This is *not* a linear functional, but it does map into $\mathbb{F}$, so it's called a functional.

Specifically, given a nonempty $A \subseteq X$, where X is an NLS, the Minkowski functional is defined as

$$p(x) = \inf\{t > 0 : x \in tA\},$$

which takes values in $[0, \infty]$. We then showed the following.

(1) If there's an open ball containing 0 and contained in A, then $p(x)$ is finite.
(2) p is *positively homogeneous*, i.e. if $\lambda \geq 0$, then $p(\lambda x) = \lambda p(x)$.
(3) If A is convex, then $p(x + y) \leq p(x) + p(y)$.
(4) If A is balanced, then p is a seminorm.

Well, we didn't actually show (3), so let's do that now. Suppose $x/r, y/s \in A$ (so that $r \geq p(x)$ and $s \geq p(y)$). By convexity,

$$\frac{x+y}{s+r} = \frac{r}{s+r}\frac{x}{r} + \frac{s}{s+r}\frac{y}{s} \in A,$$

and therefore $s + r \geq p(x + y)$. Since this is true for all such s and r, passing to their infimum replaces them with $p(x)$ and $p(y)$, so $p(x) + p(y) \geq p(x + y)$.

This was sufficient to prove Lemma 1.9.6, but we have one more separating theorem to prove. This time, we don't need sets to be balanced, but we will require convexity.

Lemma 1.10.1 (Separating hyperplane theorem)**.** *Let A and B be disjoint, nonempty, convex subsets of an NLS X.*

(1) If A is open, then there exists an $f \in X^*$ and a $\gamma \in \mathbb{R}$ such that $\mathrm{Re}(f(x)) \leq \gamma \leq \mathrm{Re}(f(y))$ for all $x \in A$ and $y \in B$.
(2) If A and B are open, the above inequality is strict.
(3) If A is compact and B is closed, then the above inequality is also strict.

PROOF. We'll prove part (1); the others are similar. Moreover, it suffices to prove it for real fields, because if $\mathbb{F} = \mathbb{C}$, then we can view X as a real vector space and get a real linear functional g that satisfies the lemma over $\mathbb{R}$. Then, $f(x) = g(x) - ig(ix)$ satisfies the lemma for $\mathbb{C}$.

All right, so $\mathbb{F} = \mathbb{R}$, and A is open and both are convex. We'll have to put the Minkowski functional into this proof somehow, so let's start by picking an $x \in A$ and a $y \in B$. Let $A - x = \{t - x : t \in A\}$, and define $B - y$ similarly.

Then, let $C = (A - x) - (B - y) = \{t - s - x + y : t \in A, s \in B\}$, and for convenience, let $w = y - x$. We'll want to construct a Minkowski functional on C.

C is open, since A is; convex, because A and B are; and contains 0 (because we've moved x and y to the origin). But $w \notin C$, since A and B are disjoint. Let $Y = \mathbb{R}w$ and $g(tw) = t$; then, our Minkowski functional is $p(x) = \inf\{t > 0 : x \in tC\}$, which is well-defined and sublinear, and satisfies $p(w) \geq 1 = g(w)$, so $g(w) \leq p(w)$, and therefore for any $y \in Y$, $g(y) \leq p(y)$: if $\lambda \geq 0$, this follows from the positive homogeneity of p and the linearity of g, and if $\lambda < 0$, $-\lambda g(w) \geq -\lambda p(w)$, and therefore $-g(-\lambda w) \leq p(-\lambda w)$.

Thus, we can extend g to X, and $g(x) \leq 1$ on C, since $g(x) \leq p(x)$ everywhere on X, and therefore $g(x) \geq -1$ for $x \in -C$, and so $|g(x)| \leq 1$ on $C \cap (-C)$. Since this contains a neighborhood of the origin, g is bounded, so $g \in X^*$.

If $a \in A$ and $b \in B$, then $a - b + w \in C$, and therefore $1 \geq g(a - b + w) = g(a) - g(b) + 1$, so $g(b) \geq g(a)$. Let $\gamma = \sup g(a)$ (or $\gamma = \inf g(b)$), and we're done. ⊠

This concludes our discussion of the Hahn-Banach theorem and its applications.

The Open Mapping Theorem. The open mapping theorem, which is the next major result for Banach spaces, helps us characterize what linear functionals can look like.

The following theorem is important in its own right, but we'll use it as an ingredient in the proof of the open mapping theorem.

Theorem 1.10.2 (Baire category theorem). *Let (X, d) be a complete metric space, and let $\{V_j\}_{j=1}^\infty$ be a sequence of open, dense subsets of X. Then, $V = \bigcap_{j=1}^\infty V_j$ is dense.*

In other words, open dense sets aren't exactly thin: they're actually surprisingly fat, so fat that a countable intersection of them is still fat, in a sense.

Proof. Let $W \subseteq X$ be any nonempty open set. Then, we have to show that $V \cap W \neq \emptyset$, since V being dense is equivalent to intersecting every nonempty open.

Since V_1 is dense, then $W \cap V_1 \neq \emptyset$, so there's an $x_1 \in W \cap V_1$, and since W is open, there's a $\overline{B_{r_1}(x_1)} \subseteq W \cap V_1$ (we have an open neighborhood, and can take the closure of a smaller ball); also, we can without loss of generality take $0 < r_1 < 1$, by shrinking r_1 if necessary.

In the same way, V_2 is open and dense and $B_{r_1}(x_1)$ is a nonempty open set, so there exists an $x_2 \in V_2 \cap B_{r_1}(x_1)$ and an $r_2 \in (0, 1/2)$, such that $\overline{B_{r_2}(x_2)} \subseteq B_{r_1}(x_1) \cap V_2$. Then, we can continue in this way, choosing for each n an x_n and an r_n such that $\overline{B_{r_n}(x_n)} \subseteq B_{r_{n-1}}(x_{n-1}) \cap V_n$ and $0 < r_n < 1/n$.

Now, consider the sequence $\{x_n\}_{n=1}^\infty$, which is Cauchy, because if $i, j \geq n$, then $x_i, x_j \in B_{r_n}(x_n)$ and therefore $d(x_i, x_j) < 2/n$. Since X is complete, this sequence converges to some $x \in X$. Since $x_i \in B_{r_n}(x_n)$ for all $i > n$, then $x \in \overline{B_{r_n}(x_n)}$, so $x \in V_n$ for all n. Since $x \in \overline{B_{r_1}(x_1)} \subseteq W$, then $x \in W \cap V$, so the intersection is nonempty, and thus V is dense. ⊠

Some of you may have been disappointed to see that no category theory appeared in the statement or proof; in this part of mathematics, "category" has a different definition.

Definition. Let (X, d) be a metric space.

- A is *nowhere dense* if it has empty interior: $(\overline{A})^0 = \emptyset$.
- A is *first category* if it can be written as a countable union of nowhere dense sets.
- If A isn't first category, then it is called *second category*.

Using these definitions, the Baire category theorem says that a complete metric space is second category.

Corollary 1.10.3. *A complete metric space is not the countable union of nowhere dense sets.*

In other words, a complete metric space is fatter than that.

Proof. Suppose X is such a union: $X = \bigcup_{j=1}^\infty M_j$ with each M_j nowhere dense. Without loss of generality, each M_j is closed (or just take their closures, which still cover X). Thus, by de Morgan's law, $\emptyset = \bigcap_{j=1}^\infty (M_j)^c$. Since M_j is closed and nowhere dense, then M_j^c is a dense open set, and therefore $\emptyset$ is the countable intersection of dense open sets, which contradicts Theorem 1.10.2. ⊠

Next time, we'll return to the world of Banach spaces.

Lecture 11: 9/21/15

The Open Mapping Theorem.

Last time, we learned about the Baire category theorem. Today, we'll use it to prove the open mapping theorem.

Definition. A continuous map $f : X \to Y$ is *open* if it maps open sets to open sets, i.e. if $U \subseteq X$ is open, then $f(U) \subseteq Y$ is open.

An arbitrary continuous map is not open; for example, $T : \mathbb{R}^2 \to \mathbb{R}^2$ sending $(x, y) \mapsto (x, 0)$ is perfectly continuous, but the image of $B_1(0)$ is $(0, 1) \times \{0\}$, which isn't open in $\mathbb{R}^2$. Surjective linear maps, however, are open.

In the infinite-dimensional case, things can become more interesting; for example, $T : \ell^2 \to \ell^2$ sending $e_n \to (1/n)e_n$ isn't open (the image of the unit ball isn't open), but is linear and surjective; the discrepancy is that this T isn't bounded.

Theorem 1.11.1 (Open mapping)**.** *Let X and Y be Banach and $T : X \to Y$ be a bounded, linear surjection. Then, T is an open map.*

A bounded linear map is typical in this class, so the key hypothesis in this theorem is that T is surjective. The example $(x, y) \mapsto (x, 0)$ shows that this is important.

This is a pretty fundamental theorem about Banach spaces.

Proof. It suffices to show that $T(B_1(0))$ contains a $B_r(0)$ for some $r > 0$: if $U \subset X$ is open, then to check that $T(U)$ is open, we can pick a $y \in T(U)$ and a preimage x (i.e. $T(x) = y$). Then, we can look at $U - x$, and since T is linear, then $T(U - x) = T(U) - y$. But since x and y are now sent to the origin, we just need to pick a neighborhood of x and make sure its image contains a neighborhood of y.

This is the proper way to think about the theorem: if you know what a bounded linear map looks like at the origin, you know what it looks like everywhere.

Since T is onto, then we can write

$$Y = \bigcup_{j=1}^{\infty} T(B_j(0)).$$

Since Y is a complete metric space, then the Baire category theorem tells us it's not the union of nowhere dense sets. Thus, there's some k such that $T(B_k(0))$ isn't nowhere dense, i.e. there's an open $W_1 \subset \overline{T(B_k(0))}$. Thus, we can scale: $(1/2k)W \subseteq (1/2k)\overline{T(B_k(0))} = \overline{T(B_{1/2}(0))}$.

Since W_1 is open, there's a $y_0 \in Y$ and an $r > 0$ such that $B_r(y_0) \subseteq W \subseteq \overline{T(B_{1/2}(0))}$. This is almost everything:

$$\begin{aligned} B_r(0) &= B_r(y_0) - y_0 \\ &\subseteq B_r(y_0) - B_r(y_0) \\ &\subseteq \overline{T(B_{1/2}(0))} + \overline{T(B_{1/2}(0))} \\ &\subseteq \overline{T(B_1(0))}, \end{aligned}$$

by the triangle inequality. We'd be done, except that we had to take the closure (which ultimately came from the Baire category theorem). Thus, we'll show that if $\varepsilon > 0$, then $T(B_{1+\varepsilon}(0)) \supseteq B_r(0)$, because then

$$T(B_1(0)) = \frac{1}{1+\varepsilon} T(B_{1+\varepsilon}(0)) \supseteq B_{r/(1+\varepsilon)}(0).$$

Then, we won't need the closure anymore. Note that this isn't obvious, even if it seems obvious in the finite-dimensional case.

Fix a $y \in B_r(0)$ and an $\varepsilon > 0$. We know that $T(B_1(0)) \cap B_r(0)$ is dense in $B_r(0)$ (since we showed already its closure contains $B_r(0)$), so we can pick an $x_1 \in B_1(0)$ such that $\|y - T x_1\| < \varepsilon/2$.

Inductively, when $n \ge 1$, suppose we've picked $x_1, \dots, x_n$ such that $\|x_1\| \le 1$ and $\|x_j\| \le 2^{-j+1}\varepsilon$ and $\|y - T(x_1 + \cdots + x_n)\| < 2^{-n}\varepsilon r$. Let $z = y - T(x_1 + \cdots + x_n)$, so that $z \in B_{2^{-n}\varepsilon r}(0)$. Since $T(B_1(0)) \cap B_r(0)$ is dense in $B_r(0)$, we can scale things: there's an $x_{n+1} \in B_{2^{-n}\varepsilon}(0)$ such that $\|z - T x_{n+1}\| \le 2^{-(n+1)}\varepsilon r$; thus, $\|y - T(x_1 + \cdots + x_{n+1})\| \le 2^{-(n+1)}\varepsilon r$.

Since the terms get smaller and smaller, $\sum_{j=1}^{n} x_j$ is a Cauchy sequence, so since X is complete, then this sum converges to a point $x \in X$, such that

$$\|x\| \leq 1 = \sum_{j=2}^{\infty} \|x_j\| < 1 + \sum_{n=2}^{\infty} 2^{-n+1}\varepsilon = 1 + \varepsilon.$$

Since T is continuous, then $Ts_n \to Tx = y$. ⊠

The first part, showing it's true for $\overline{T(B_1(0))}$, is pretty easy, but then getting just one more ε is surprisingly fussy.

Corollary 1.11.2. *If X and Y are Banach spaces and $T : X \to Y$ is a bounded, linear bijection (one-to-one and onto), then the inverse map exists and is a bounded linear functional.*

In other words, the inverse of a bounded linear functional is bounded linear. This is nice, and very useful.

Proof. It's easy to show T^{-1} is linear. T is open, so it takes open sets to open sets, and therefore for T^{-1}, the preimage of every open set is open, so T^{-1} is continuous. ⊠

We now know enough to make the following definition.

Definition. If X and Y are Banach spaces, we say that they're *isomorphic as Banach spaces* if there exists a linear, bounded bijection $T : X \to Y$. If in addition T preserves the norm, it's called a *isometry*.

This means that X and Y have the same vector space structure (since there's a bijective linear map) and same topology (there's a homeomorphism). If T isn't an isometry, then the norms may be different, but they'll be equivalent, so X and Y are basically the same.

There's a closely related result about graphs of maps; we could have proven this first and used it to derive the open mapping theorem, though we'll go about it in the other direction.

Definition. Let X and Y be topological spaces, $D \subseteq X$ and $f : D \to Y$. Then, the *graph* of f is $\mathrm{graph}(f) = \{(x, f(x)) : x \in D\} \subseteq X \times Y$.

$X \times Y$ is where we're used to drawing graphs (such as $X, Y = \mathbb{R}$); we chose D because the function might not be defined everywhere.

Proposition 1.11.3. *Let X be a topological space, Y be a Hausdorff space, and $f : X \to Y$ be continuous. Then, $\mathrm{graph}(f)$ is closed in $X \times Y$.*

In the case of graphs we're most familiar with, this makes sense, as it's how we're used to thinking of continuity intuitively.

Proof. Let $U = X \times Y \setminus \mathrm{graph}(f)$, and let $(x_0, y_0) \in U$, so $f(x_0) \leq y_0$.

Recall that since Y is Hausdorff, we can choose open neighborhoods V and W in Y around y_0 and $f(x_0)$, respectively, that don't intersect. Then, we're going to consider $f^{-1}(W) \subset X$; specifically, $f^{-1}(W) \times V$ doesn't intersect $\mathrm{graph}(f)$, and is an open neighborhood of (x_0, y_0). ⊠

If that proof didn't make sense, drawing a picture will likely help.

Definition. Let X and Y be NLSes, $D \subseteq X$ be a linear subspace, and $T : D \to Y$ be linear. Then, we say that T is a *closed* operator if $\mathrm{graph}(T)$ is closed in $X \times Y$.

An open map takes open sets to open sets; a bounded map takes bounded sets to bounded sets; but a closed operator *doesn't* take closed sets to closed sets. This can be confusing.

If X and Y are metric spaces and $f : D \to Y$ is continuous, then $\mathrm{graph}(f)$ being closed means that if $\{x_n\}_{n=1}^{\infty} \subset D$ is such that $x_n \to x$ and $Tx_n \to y$, then $x \in D$ and $y = Tx$.

Theorem 1.11.4 (Closed graph theorem)**.** *Let X and Y be Banach spaces, and $T : X \to Y$ be linear. Then, T is bounded iff it's closed.*

PROOF. The forward direction is true in general (continuous implies closed).

In the other direction, suppose graph(T) is closed, and therefore is a closed linear subspace of $X \times Y$. This is a very important point: since X and Y are Banach, then $X \times Y$ is Banach, and since graph(T) is closed in it, then graph(T) is also a Banach space, with the *graph norm*[10]

$$\|(x, Tx)\| = \|x\|_X + \|Tx\|_Y.$$

Define two projection operators $\pi_1 : (x, y) \mapsto x$ and $\pi_2 : (x, y) \mapsto y$, so that we have maps

$$\begin{tikzcd} & \mathrm{graph}(T) \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ X & & Y. \end{tikzcd}$$

π_1 is a linear bijection between Banach spaces, and is bounded (by the triangle inequality, $\|x\|_X \le \|(x, Tx)\| + \|Tx\|_Y$), so by Corollary 1.11.2, its inverse π_1^{-1} is a bounded linear functional. Moreover, π_2 is bounded linear for the same reasons, so $T = \pi_2 \circ \pi_1^{-1}$ is bounded as well. ⊠

Lecture 12: 9/23/15

The Uniform Boundedness Principle.

Recall that last time, we proved the closed graph theorem, Theorem 1.11.4, which states that if X and Y are Banach spaces and $T : X \to Y$ is linear, then T has a closed graph iff it's bounded.

Corollary 1.12.1. *Let X and Y be Banach spaces, $D \subset X$ be a subspace, and $T : D \to Y$ be closed. Then, T is bounded iff D is closed.*

PROOF. In the reverse direction, if D is closed, then it's Banach, so we can apply Theorem 1.11.4.

Conversely, suppose T is bounded, and let $\{x_n\}_{n=1}^\infty \subset D$ be such that $x_n \to x$ in X. Since T is bounded and linear, then $\{Tx_n\}_{n=1}^\infty$ is Cauchy.[11]

Since Y is complete, then $Tx_n \to y$ for some $y \in Y$. And since T is a closed operator, then graph(T) $= \{(x, Tx)\}$ is closed. Since $x_n \to x$ and $Tx_n \to y$, then $(x, y) \in$ graph(T), so $y = Tx$ and thus $x \in D$. Therefore D contains its limit points, and so is closed. ⊠

Hopefully this illustrates some uses of the closed graph theorem.

Example 1.12.2. Though continuous implies closed, the converse isn't true. Here's an example. Let $X = C([0, 1])$, the continuous functions with the L^∞ norm, and let $D = C^1([0, 1]) \subset X$, the C^1 functions. Let $T : D \to X$ be the derivative operator, so $T(f) = f'$.

This is perfectly well defined: if f is C^1, then f' is continuous. Then, $D \ne X$ (e.g. $f(x) = |x - 1/2|$), but $\overline{D} = X$, so D is *not* closed in X (intuitively, any continuous but not differentiable function can be well approximated by a C^1 function).

- First, we'll see that T isn't continuous (equiv. bounded). $T(x^n) = nx^{n-1}$, but $\|x_n\| = 1$ and $\|Tx_n\| = n \to \infty$.
- However, it is closed. Let $\{f_n\}_{n=1}^\infty \subseteq D$ have a limit $f_n \to f$ in X, and such that $f_n' \to g$ in X. We want to show that $g = f'$. This follows from the fundamental theorem of calculus: for each n,

$$f_n(t) = f_0(t) + \int_0^t f_n'(\tau)\,\mathrm{d}\tau.$$

Since convergence in L^∞ implies pointwise convergence. Then, by the dominated convergence theorem, these integrals also converge (recall that continuous functions on compact sets are bounded), so

$$f(t) = f(0) + \int_0^t g(\tau)\,\mathrm{d}\tau = f(0) + \int_0^\tau f'(\tau)\,\mathrm{d}\tau.$$

Thus, $f'(t)$ exists and $g = f'$, so $f \in C^1([0, 1])$.

[10] Though this was defined in a way mirroring the ℓ^1 norm, you can use the analogous definition with any ℓ^p including ℓ^∞, since, as we proved on a problem set, they're all equivalent.

[11] This is because $\|Tx_n - Tx_m\| = \|T(x_n - x_m)\| = \|T\|\|x_n - x_m\| \to 0$.

Continuous is better than closed, but closed is part of the way there, in some sense.

We've talked about two of the three important theorems about NLSes: the Hahn-Banach theorem and the open mapping theorem. Here's the third.

Theorem 1.12.3 (Uniform boundedness principle). *Let X be Banach, Y be an NLS, and $\{T_\alpha\}_{\alpha\in\mathcal{I}} \subset B(X,Y)$. Then, one of the following holds.*

(1) *The collection is uniformly bounded: there's an M such that $\sup_{\alpha\in\mathcal{I}}\|T_\alpha\|_{B(X,Y)} \le M$.*
(2) *There's a point where they're not: there's an $x \in X$ such that $\sup_{\alpha\in\mathcal{I}}\|T_\alpha X\| = \infty$.*

The point is, if these functions aren't uniformly bounded, then they all blow up at some given point, when *a priori* they could do so in different places. This is true no matter how large the collection $\mathcal{I}$ is; it could very well be uncountable.

The proof in the notes is nice, but a little fussy to prove, and uses the Baire category theorem. We'll give a proof based on the following lemma, which is a little nicer.

Lemma 1.12.4. *Let X and Y be NLSes and $T : X \to Y$ be a bounded linear map. For any $x \in X$ and $r > 0$,* $\sup_{y\in B_r(x)}\|Ty\| \ge r\|T\|$.

The idea is if $x = 0$, we have equality, but even if we don't, this is still a one-sided bound.

Proof. For a visualization, it may help to think about the case $X = Y = \mathbb{R}$, where graph(T) is a line with slope $\|T\|$ through the origin. Here, $\sup_{y\in B_r(x)}\|Ty\| = \|T\|(\|x\| + r) \ge \|T\|r$.

More generally, we'll think of the "larger" and "smaller" parts of $B_r(x)$. The triangle inequality tells us that since $z = (1/2)(x+z) - (x-z)$, then

$$\begin{aligned}\|Tz\| &\le \frac{1}{2}(\|T(x+z)\| + \|T(x-z)\|)\\ &\le \max\{\|T(x+z)\|, \|T(x-z)\|\}.\end{aligned}$$

If we take the supremum over $z \in B_r(0)$, we know that

$$r\|T\| \le \sup_{z\in B_r(0)}\|T(x+z)\| = \sup_{y\in B_r(x)}\|Ty\|. \qquad \boxtimes$$

This is a nice geometric result, and relatively easy to prove. Then, we'll use it to attack the uniform boundedness principle.

Proof of Theorem 1.12.3. Since we want to show one of two things is true, let's assume $\sup_\alpha\|T_\alpha\| = \infty$. Choose a countable subcollection $\{T_n\}_{n=1}^\infty$ such that $\|T_n\| \ge 4^n$. Then, choose $x_0 = 0 \in X$ and choose $x_n \in X$ such that $\|x_n - x_{n-1}\| \le 3^{-n}$ (so that $x_n \in B_{3^{-n}}(x_{n-1})$), so that by Lemma 1.12.4, $\|T_n x_n\| \ge (2/3)3^{-n}\|T_n\|$.

Since $\{x_n\}$ is Cauchy, then it converges to some $x \in X$, and $\|x - x_n\| \le (1/2)3^{-n}$, since

$$\begin{aligned}\|x - x_n\| &= \lim_{m\to\infty}\|x_m - x_n\|\\ &\le \lim_{m\to\infty}\sum_{j=m}^{n+1}\|x_j - x_{j-1}\|\\ &\le \lim_{m\to\infty}\left(3^{-m} + 3^{-(m-1)} + \cdots + 3^{-(n+1)}\right)\\ &\le \lim_{m\to\infty} 3^{-n}\sum_{k=1}^{\infty}\left(\frac{1}{3}\right)^k = \frac{1}{2}3^{-n},\end{aligned}$$

since it's a nice old geometric series.

Finally, we're going to look at $T_n x$.

$$\|T_n x\| \le \|T_n(x - x_n)\| + \|T_n x_n\| \le \|T_n\|\|x - x_n\| + \|T_n x\|.$$

Thus, we know that

$$\frac{2}{3}3^{-n}\|T_n\| \le \|T_n\|\frac{1}{2}3^{-n} + \|T_n x\|,$$

and therefore

$$\|T_n x\| \ge \frac{1}{6}3^{-n}\|T_n\| \ge \frac{1}{3}\left(\frac{4}{3}\right)^n,$$

which goes to infinity. $\boxtimes$

The Double-Dual. If X is an NLS, then $X^* = B(X, \mathbb{F})$ is a Banach space, and we can form the *double-dual* $X^{**} = B(X^*, \mathbb{F})$, which is also a Banach space. It's possible to interpret X as sitting inside X^{**}.

For any $x \in X$, define the evaluation map $E_x : X^* \to \mathbb{F}$ by $E_x(f) = f(x)$: that is, we evaluate f at x. Since $f : X \to \mathbb{F}$, then E_x is well-defined, and it's linear: if $f, g \in X^*$ and $\lambda \in \mathbb{F}$, then

$$\begin{aligned} E_x(f+g) &= (f+g)(x) = f(x) + g(x) = E_x(f) + E_x(g) \\ E_x(\lambda f) &= (\lambda f)(x) = \lambda(f(x)) = \lambda E_x(f). \end{aligned}$$

E_x is also bounded, which is more interesting. If you think about what the norm means, then

$$\|E_x\|_{X^{**}} = \sup_{\substack{f \in X^* \\ f \neq 0}} \frac{|E_x(f)|}{\|f_{X^*}\|} = \sup_{\substack{f \in X^* \\ f \neq 0}} \frac{f(x)}{\|f\|_{X^*}} = \|x\|_X,$$

where the last equality is due to Corollary 1.8.5.

Definition. Let (M, d) and (N, ρ) be metric spaces. Then, $f : M \to N$ is an *isometry* if $\rho(f(x), f(y)) = d(x, y)$ for all $x, y \in M$. If f is surjective, M and N are said to be *isometric*.

Note that f is always injective, because if $f(x) = f(y)$, then $\rho(f(x), f(y)) = 0 = d(x, y)$, so $x = y$. Thus, isometric spaces are given by a bijection f.

Anyways, that's what's going on here: not only is the metric the same, but the norm is the same. We have a map $E : X \to X^{**}$ sending $x \mapsto E_x$. E is a bounded linear map, and an isometry. Therefore, $\widetilde{X} = \{E_x \in X^{**} : x \in X\} \subset X^{**}$ is isomorphic and isometric to X. It might not be all of X^{**}, but we've embedded X into its double-dual.

Definition. Sometimes, $X = X^{**}$ (i.e. $\widetilde{X} = X^{**}$). If this is true, X is said to be *reflexive*.

In general, $X \subseteq X^{**} \subseteq X^{****} \subseteq \cdots$; if $X = X^{**}$, then this entire chain collapses to equalities. Similarly, we could have started with X^* and X^{***}, and so on.

Theorem 1.12.5. *If X is reflexive, then X^* is reflexive.*

This is left to the exercises, but isn't hard to prove. Notice, however, that the converse isn't true.

Example 1.12.6. If $1 \le p < \infty$, we know that $(\ell^p)^* = \ell^q$, so ℓ^p is reflexive for $1 < p < \infty$. ℓ^1 and ℓ^∞ aren't as nice, so we don't have reflexivity. (The duality follows from something on the homework this week.)

Since $(L^p(\Omega))^* = L^q(\Omega)$ for $1 < p < \infty$, then $L^p(\Omega)$ is reflexive (which follows from the Radon-Nikodym theorem). Similarly, $L^1(\Omega)$ and $L^\infty(\Omega)$ are more complicated.

Lecture 13: 9/25/15

Weak and Weak-$*$ Convergence.

In a finite-dimensional NLS, we know that we're basically looking at $\mathbb{F}^d$, where we have the Heine-Borel theorem: a set is closed and bounded iff it's compact. But, of course, infinite dimensions are weirder, and in fact the unit ball isn't compact in an infinite-dimensional vector space. This was the reasoning behind one of the less intuitive HW problems, about embedding infinitely many disjoint balls of a fixed radius into the unit ball.

Theorem 1.13.1. *Let $Y \subseteq X$ be a closed subspace, and Z be another subspace containing Y. If $Z \neq Y$ and $\theta \in (0,1)$, then there exists a $z \in Z$ such that $\|z\| = 1$ and $\operatorname{dist}(z, Y) = \theta$.*

The intuition in the finite-dimensional case is that z should be "orthogonal" to Y, but not all infinite-dimensional spaces have an inner product structure (which is what gives us angles), so we have to be careful.

PROOF. Pick a $z_0 \in Z \setminus Y$ and let

$$d = \operatorname{dist}(z_0, Y) = \inf_{y \in Y} \|y - z_0\|.$$

Since Y is closed, $d > 0$, so we can choose a $y_0 \in Y$ such that $d/\theta \geq \|z_0 - y_0\| \geq d$. Then, let $z = (z_0 - y_0)/\|z_0 - y_0\| \in Z$. For all $y \in Y$,

$$\begin{aligned}\|z - y\| &= \frac{\|z_0 - \overbrace{y_0 - y\|z_0 - y_0\|}^{y_1 \in Y}\|}{\|z_0 - y_0\|} \\ &= \frac{\|z_0 - y_1\|}{\|z_0 - y_0\|} \\ &\geq \|z_0 - y_1\| \frac{\theta}{d} \geq \theta.\end{aligned}$$

⊠

Corollary 1.13.2. *If X is an infinite-dimensional NLS and $M \subset X$ is a closed, bounded set with nonempty interior, then M is* not *compact.*

The Heine-Borel theorem is false in infinite dimensions. Alas.

Proof. It's sufficient to show it for the (closed) unit ball, because then translations and scalings cover all such closed, bounded sets.

Let $x_1 \in X$ with $\|x_1\| = 1$, and we'll inductively assume we've chosen $x_1, \dots, x_n$ such that $\|x_i\| = 1$ and $\|x_i - x_j\| \geq 1/2$ for all $i \neq j$. Let $Y = \operatorname{span}\{x_1, \dots, x_n\}$, so since Y is finite-dimensional and X is infinite-dimensional, we can choose an $x \in X \setminus Y$ and let $Z = \operatorname{span}\{x, Y\}$. With $\theta = 1/2$, Theorem 1.13.1 gives us an $x_{n+1} \in Z$ with $\|x_{n+1}\| = 1$ and $\operatorname{dist}(x_{n+1}, Y) \geq 1/2$.

Thus, we can pick an infinite sequence of points on the unit sphere such that all of them are at least distance $1/2$ from each other. Thus, $\{x_n\}$ is bounded, but has no convergent subsequence, so $\overline{B_1(0)}$ cannot be compact.[12] ⊠

A modification of this proof shows that the volume is arbitrarily large too, not just the surface area.

What we'll do about this is to define a new topology, which is weaker (has fewer open sets); since compactness is a condition on open covers, this makes things more likely to be compact. We'll start with a new notion of convergence, and then extract the topology afterwards.

Definition. Let X be an NLS and $\{x_n\}_{n=1}^\infty \subset X$.

- We say that x_n *converges weakly* to x, written $x_n \rightharpoonup x$ or $x_n \xrightarrow{w} x$, if for any $f \in X^*$, $f(x_n) \to f(x)$.
- If $\{f_n\}_{n=1}^\infty$, then f_n *converges weak-∗* (said "weak-star") to f, written $f_n \xrightarrow{w^*} f$, if for all $x \in X$, $f(x_n) \to f(x)$.

As a contrast, $x_n \to x$ given by $\|x_n - x\| \to 0$ is sometimes called *strong convergence*.

So in a space X^* that's the dual of some X, we have three notions of convergence floating around: the strong topology $\|f_n - f\|_{X^*} \to 0$, the weak topology $F(f_n) \to F(f)$ for all $F \in X^{**}$, and weak-∗, where we only consider evaluation maps in X^{**}.

These notions of convergence represent new topologies. Let's establish some propositions before we continue.

Proposition 1.13.3. *If $x_n \to x$ as $n \to \infty$ (i.e. it converges strongly), then $x_n \rightharpoonup x$.*

This is obvious: $\|x_n - x\| \to 0$, and any $f \in X^*$ is continuous in the strong topology, so $\|f(x - x_n)\| \to 0$ too.

Example 1.13.4. The converse is not true: weak convergence does not imply strong convergence. Let $1 < p < \infty$ and consider ℓ^p, with our Schauder basis $\{e^n\}_{n=1}^\infty$. This sequence does not converge strongly, because

$$\|e^n - e^m\|_{\ell^p} = \left(\sum_{i=1}^\infty |e_i^n - e_i^m|^p\right)^{1/p} = 2^{1/p},$$

which does not go to zero. But $e_n \rightharpoonup 0$! Take any $f \in (\ell^p)^* = \ell^q$; then, $f(e^n) = \sum f_i e_i^n = f_n$, and since $\left(\sum |f_n|^q\right)^{1/q}$ is finite, then $|f_n| \to 0$.

This is clearly a very different kind of convergence than strong convergence, but we have more things converging, which is nice if you like convergence. It's in some sense a sampling notion.

[12] Here, we're using the theorem that compactness is equivalent to sequential compactness in metric spaces, and we should also remark that no subsequence of $\{x_n\}$ converges, which is the criterion we need.

Proposition 1.13.5. *Let X be an NLS, $\{x_n\} \subset X$ and $\{f_n\} \subset X^*$. Then:*

- *If $\{x_n\}$ converges weakly, then the limit is unique and $\|x_n\|$ is bounded.*
- *If $\{f_n\}$ converges weak-$*$, then the limit is unique. If in addition X is Banach, then $\|f_n\|$ is bounded.*

PROOF. Suppose $x_n \rightharpoonup x$ and $x_n \rightharpoonup y$. Therefore for any $f \in X^*$, $f(x_n) \to f(x)$ and $f(x_n) \to f(y)$. But since limits are unique in $\mathbb{F}$, then $f(x) = f(y)$, so since this is true for all $f \in X^*$, then $x = y$.

Fix an $f \in X^*$, so that $\{f(x_n)\}_{n=1}^\infty$ converges in $\mathbb{F}$, and therefore is bounded. Thus, $|E_x(f)| = |f(x_n)| \le C_f$ for all n (i.e. some constant C that depends only on f). In particular, $\{E_{x_n}\}_{n=1}^\infty \subseteq X^{**}$ is pointwise bounded, so by the uniform boundedness principle, $\|x_n\| = \|E_{x_n}\| \le C$.

The second part is left as an exercise; but since uniform boundedness requires the space to be Banach, we'll have to assume that of X. ⊠

Proposition 1.13.6. *Let X be an NLS and $x_n \rightharpoonup x$. Then, $\|x\| \le \liminf_{n\to\infty} \|x_n\|$.*

This one was left as an exercise.

Though most of the time one only cares about convergence, the topologies are worth knowing about. Recall that a topology is a collection of open sets, so one topology being "smaller" than another means that it's a subset of the other collection of open sets.

Definition. The *weak topology* on X is the smallest topology such that each $f \in X^*$ is continuous. The *weak-$*$ topology* on X^* is the smallest topology such that each E_x (the image of x in the canonical map $X \to X^{**}$) is continuous.

Since we know these are continuous in the strong topology, these are smaller topologies; there are fewer open sets, and convergence is nicer.

What this means is that if $U \subset \mathbb{F}$ is open and $f \in X^*$, then $f^{-1}(U)$ is open in X. Since these f are linear, then we really only need to talk about open sets containing 0, and therefore contain an open ball $B_\varepsilon(0)$, which will give us all the open sets in the weak topology (by translation).

We know that if $f_i \in X^*$, then $f_i^{-1}(B_{\varepsilon_i}(0))$ is open, and we can take arbitrary unions and finite intersections. Specifically, our *basic open sets* are translates of sets of the form

$$\bigcap_{i=1}^n f_i^{-1}\big(B_{\varepsilon_i}(0)\big).$$

All open sets in the weak topology are translations of unions of these sets, which can be rewritten as

$$U = \{x \in X : |f_i(x)| < \varepsilon_i, i = 1, \dots, n\}.$$

For the weak-$*$ topology, the basic opens are the translates of

$$\begin{aligned} V &= \{f \in X^* : |f(x_i)| < \varepsilon_i, i = 1, \dots, n\} \\ &= \bigcap_{i=1}^n E_{x_i}^{-1}(B_{\varepsilon_i}(0)). \end{aligned}$$

The next thing we need to do is show that these topologies imply the notions of convergence we defined at the start of lecture.

Lecture 14: 9/28/15

The Banach-Alaoglu Theorem.

"Just take these epsilons; don't worry about it..."

Recall that we've defined weak convergence $x_n \rightharpoonup x$ if $f(x_n) \to f(x)$ for all $f \in X^*$, and an even weaker notion called weak-$*$ convergence where $f_n \xrightarrow{w^*} f$ if $f_n(x) \to f(x)$ for all $x \in X$. Then, we defined the weak and weak-$*$ topologies: the smallest topologies on X (resp. X^*) such that every $f \in X^*$ (resp. $E_x \in X^{**}$) is continuous.

We saw that the basic open sets in the weak topology are (the translates of) finite intersections of sets of the form $f_i^{-1}(B_{\varepsilon_i}(0))$ for $f_i \in X^*$; in other words, a set of points where $f_i(x) < \varepsilon_i$ for some finite number of $f_i \in X^*$ and $\varepsilon_i > 0$. General open sets are unions of translations of these sets. We also had a similar notion for the weak$*$ topology.

Proposition 1.14.1. *Let X be an NLS and $\{x_n\}_{n=1}^\infty \subseteq X$. Then, $x_n \to x$ in the weak topology iff $x_n \rightharpoonup x$, and if $\{f_n\}_{n=1}^\infty \subseteq X^*$, then $f_n \to f$ in the weak$*$ topology iff $f_n \xrightarrow{w^*} f$.*

PROOF. For the first part and in the forward direction, suppose $f \in X^*$, so that f is continuous in the weak topology. Thus, $\lim_{n\to\infty} f(x_n) = f(x)$, so $x_n \rightharpoonup x$.

Conversely, suppose $x_n \rightharpoonup x$, and let U be a basic open set about x, so that

$$U = x + \{y \in X : |f_i(y)| < \varepsilon_i, i = 1, \dots, n\}$$

for some $f_i \in X^*$ and $\varepsilon_i > 0$. Since $f(x_n) \to f(x)$ for all f, then there's an $N > 0$ such that $|f_i(x_n - x)| = |f_i(x_n) - f_i(x)| < \varepsilon_i$ for all $i = 1, 2, \dots, n$, and therefore $x_n = (x_n - x) + x \in U$.

The proof of the second part is similar. ⊠

Remark.

- The Hahn-Banach theorem implies that the weak topology is Hausdorff, since it implies there's a linear functional that separates any two points.
- If τ_w denotes the (collection of open sets of the) weak topology, and τ denotes the strong topology, then $\tau_w \subset \tau$ (i.e. it actually is weaker): since each f_i is continuous in the strong topology, then inverse images of open sets under it remain open. A similar result holds for the weak-$*$ topology.

Theorem 1.14.2 (Banach-Alaoglu[13]). *Let X be an NLS and $B_1^* = \{f \in X^* : \|f\|_{X^*} \le 1\}$ be the closed unit ball in X^*. Then, B_1^* is compact in the weak-$*$ topology.*

PROOF. Let $B_x = \{\lambda \in \mathbb{F} : |\lambda| \le |x|\}$. Then, each B_x is a closed and bounded subset of $\mathbb{F}$, and therefore compact, and let $C = \prod_{x\in X} B_x$. By Tychonoff's theorem, which is an amazing theorem, any product of compact sets is still compact, so C is compact, even though it's (in some sense) huge![14] A function $g : X \to \mathbb{F}$ such that $|g(x)| \le \|x\|$, whether it's linear or not, can be viewed as an element of C by sending $g \mapsto (g(x))_{x\in X}$. Then, the coordinate map $\pi_x : C \to B_x$ sends $g \mapsto g(x)$, so it's just the evaluation map, and we know by the definition of the product topology that this map is continuous. In particular, this defines a continuous inclusion of $B_1^* \subseteq C$, where the former has the weak-$*$ topology. Since C is compact, then B_1^* is compact if it's closed in C.

We probably won't get around to showing it, but the weak and weak-$*$ topologies aren't metrizable, so the nice proof techniques from metric spaces don't work, and we'll have to use more basic topological methods. Specifically, we'll have to show that all accumulation points of B_1^* are in B_1^*.

Let g be an accumulation point of B_1^*, and fix $x, y \in X$ and a $\lambda \in \mathbb{F}$; then, let

$$U = g + \{h \in C : |h(x_i)| < \varepsilon, i = 1, \dots, m\}.$$

Specifically, we'll take $m = 4$ and the points $x_i = x$, y, $x + y$, and λx. If $f = g + h \in B_1^*$, we'll choose the epsilons $|h(x)| < \varepsilon/3\max(1, |\lambda|)$, $|h(y)\} < \varepsilon/3$, $|h(x+y)| < \varepsilon/3$, and $|h(\lambda x)| < 2\varepsilon/3$, for some arbitrary $\varepsilon > 0$. Then,

$$\begin{aligned} |g(x+y) - g(x) - g(y)| &= |h(x+y) - h(x) - h(y)| \\ &\le |h(x+y)| + |h(x)| + |h(y)| < \varepsilon. \\ |g(\lambda x) - \lambda g(x)| &= h(\lambda x) - \lambda h(x)| < \varepsilon. \end{aligned}$$

This is why we chose the strange epsilons; as $\varepsilon \to 0$, this forces g to be linear! Moreover,

$$|g(x)| = |f(x) - h(x)| \le \|f\| + |h(x)| \le 1 + \frac{\varepsilon}{3},$$

so in particular $\|g\| \le 1$, so $g \in B_1^*$, and so B_1^* has to be compact. ⊠

Now, we can begin to unpack the applications of this theorem. Even though these topologies aren't metrizable, we can get some nice topological results for suitably nice spaces.

Theorem 1.14.3. *Let X be a separable Banach space and $K \subseteq X^*$ be weak-$*$ compact. Then, K with the weak-$*$ topology is metrizable.*

In particular, on this set, sequential compactness is equivalent to compactness, which can be useful!

[13] Pronounced approximately "ala-glue."

[14] Note that this compactness is in the product topology, not the box topology. If $\{X_\alpha\}_{\alpha\in I}$ is a set of topological spaces and $X = \prod_{\alpha\in I} X_\alpha$, then for each α there's a *coordinate map* $\pi_\alpha : X \to X_\alpha$ sending each element to its coordinate in that index. Then, the *product topology* is defined to be the smallest topology in which each of these coordinate maps is continuous.

Proof. Let $D = \{x_n\}_{n=1}^\infty \subseteq X$ be a countable, dense subset of X. Then, the evaluation maps $E_n = E_{x_n}$ are weak-$*$ continuous; since D is dense, then these E_n separate points. That is, if $E_n(f^*) = E_n(g^*)$ for all n (i.e. $f^*(x_n) = g^*(x_n)$ for all n), then $f^* = g^*$.

Let $c_n = \sup_{f^* \in K} |E_n(f^*)|$, which is a continuous, finite function on a compact set and therefore has a maximum, and let

$$f_n = \begin{cases} \dfrac{E_n}{c_n}, & c_n \neq 0 \\ 0, & c_n = 0. \end{cases}$$

Now that we've scaled suitably, we may define our metric: let

$$d(f^*, g^*) = \sum_{n=1}^\infty 2^{-n} |f_n(f^*) - f_n(g^*)|.$$

This is a metric: it's clearly symmetric, and it's 0 iff each term is, which is true iff $f^* = g^*$, as observed above. Moreover, since the triangle inequality holds termwise, then it holds here.

However, we still need to prove that the topology τ_d induced by this metric agrees with the weak-$*$ topology τ on K. First, we'll show that $\tau_d \subseteq \tau$: for $N \geq 1$, let

$$d_N(f^*, g^*) = \sum_{n=1}^N 2^{-n} |f_n(f^*) - f_n(g^*)|,$$

and consider $d_N(\cdot, g^*) : X^* \to [0, \infty)$ sending $f^* \mapsto d_N(f^*, g^*)$. This is weak-$*$ continuous, and $d_N(\cdot, g^*) \to d(\cdot, g^*)$ converges uniformly, so the limit $d(\cdot, g^*)$ is continuous. Therefore $B_r(g^*) = \{f^* \in K : d(f^*, g^*) < r\}$ is the inverse image of $(-\infty, r)$ (which is open in $\mathbb{R}$) under $d(\cdot, g^*)$, which is continuous, so it's open, and thus $\tau_d \subseteq \tau$ (since these balls generate all opens).

Conversely, let $A \in \tau$, so that $A^c \subset K$ is τ-closed, and therefore τ-compact. Since we already showed that $\tau_d \subseteq \tau$, then A^c must be τ_d-compact (as there are fewer open sets), and in particular τ_d-closed. Thus, A is open in τ_d. ⊠

Lecture 15: 9/30/15

The Generalized Heine-Borel Theorems.

Recall that we proved the Banach-Alaoglu theorem, that the unit ball in X^* is weak-$*$ compact. This isn't super useful, but we found that if X is a separable Banach space, then B_1^* is a metric space. Relatedly, since any weak-$*$ compact subspace of X^* is metrizable, then if $\{f_n\} \subset X^*$ and $\|f_n\| \leq C$, then there exists a subsequence f_{n_k} which weak-$*$ converges to an $f \in X^*$.

In other words, we know the following.

Theorem 1.15.1 (Generalized Heine-Borel I). *Let X be a separable Banach space and $K \subseteq X^*$. Then, the following are equivalent.*

(1) K is weak-$*$ compact.
(2) K is weak-$*$ closed and bounded.
(3) K is weak-$*$ sequentially compact.[15]

We've already seen that (1) $\implies$ (3), and the discussion above is (2) $\implies$ (1) for the unit ball (which generalizes by scaling to any ball, and therefore to any bounded set, since it can be contained in a large ball). So there's only one step left in the proof.

Proof of Theorem 1.15.1, (3) $\implies$ (2). Let K be weak-$*$ sequentially compact. Then, K must be bounded: if not, then there's a sequence $\{f_n\} \subseteq K$ such that $\|f_n\|_{X^*} \geq n$, but then there can be no convergent subsequence, since all weak-$*$ convergent sequences are bounded.

We also want to show that K is weak-$*$ closed, but we don't have a metric topology yet, so we must be careful. Let $\{x_n\}_{n=1}^\infty \subset X$ be a countable dense subset and f be an accumulation point of K, so that there's a sequence $f_n \to f$. In particular, $f_n(x_j) \to f(x_j)$. Let $U_n = f + \{g \in X^* : |g(x_j)| \leq 1/n, j = 1, 2, \ldots, n\}$, which is weak-$*$ open neighborhood of f; thus, there exists an $f_m \in U_n \cap K$, so that $f = f_m + g$.

[15] *Sequential compactness* of a set K means that for every sequence $\{x_n\} \subseteq K$, there's a convergent subsequence x_{n_j}.

If $x \in X$, then there's a subsequence $x_{n_j} \to x$; we want to know whether $f_n(x) \overset{?}{\to} f(x)$.

$$|f_n(x) - f(x)| \le |f_n(x) - f_m(x_j)| + |f_m(x_j) - f(x_j)| + |f(x_j) - f(x)|.$$

Each of the three terms on the right goes to zero, but differently (in m or in j). The first one is the only problem, but since it's bounded by $\|f_m\|\|x - x_m\|$ and $\|f_m\| \le C$, then we may take a large j to make this small, and then the whole expression goes to 0 as $m \to \infty$, so $f_m \overset{w^*}{\to} f$, and therefore, since K is weak-$*$ sequentially compact, then $f \in K$. ⊠

We want to take statements about X^* and turn them into statements about x. If X is reflexive (i.e. $X = X^{**}$), then everything we've talked about applies.

Theorem 1.15.2 (Generalized Heine-Borel II). *Let X be a separable, reflexive Banach space and $K \subseteq X$. Then, the following are equivalent.*

(1) K is compact in the weak topology.
(2) K is closed in the weak topology and bounded.
(3) K is sequentially compact in the weak topology.

There's a harder theorem which offers a sort of converse (the Heine-Borel result implies reflexivity), and therefore we can remove the separability hypothesis. However, in applications, one's Banach are almost always separable, and therefore it's not really a huge deal.

Example 1.15.3. Let $1 < p < \infty$ and $\Omega \subseteq \mathbb{R}^d$ be measurable. Then, if $\{f_n\}_{n=1}^\infty \subseteq L^p(\Omega)$ is a sequence such that $\|f_n\|_{L^p} \le C$, then we know there's a subsequence $f_{n_j} \rightharpoonup f$; that is, for all $g \in L^q(\Omega)$ (where q is the conjugate exponent),

$$\int_\Omega f_{n_j}(x)g(x)\,\mathrm{d}x \longrightarrow \int_\Omega f(x)g(x)\,\mathrm{d}x.$$

This is used all the time in analysis, and we'll use it later.

We're not done with compactness yet; this next theorem provides a nice connection between weak and strong convergence.

Theorem 1.15.4 (Banach-Saks). *Let X be an NLS and $x_n \rightharpoonup x$. Then, for all n, there exist $\alpha_j^{(n)}$ for $j = 1, \dots, n$ with $\alpha_j^{(n)} \ge 0$ and $\sum_{j=1}^b \alpha_j^{(n)} = 1$ such that*

$$y_n = \sum_{j=1}^n \alpha_j^{(n)} x_j \longrightarrow x.$$

In other words, there is a convex combination of a weakly convergent sequence that strongly converges. Think about this in the case $e^n \rightharpoonup 0$.

Proof. We'll start by considering all such convex combinations. To wit, let

$$M = \left\{ \sum_{j=1}^n \alpha_j^{(n)} x_j : n \ge 1, \alpha_j^{(n)} \ge 0, \sum_{j=1}^n \alpha_j^{(n)} = 1 \right\}.$$

We want to show that $x \in \overline{M}$ (here denoting the strong closure). Notice that M is the convex hull of $\{x_j\}$, and therefore it (and its strong closure) are convex sets.

Let's assume $x \notin \overline{M}$ and use the separating hyperplane theorem (Lemma 1.10.1). We know that $\overline{M}$ is a closed, convex set and $\{x\}$, being a single point, is compact, so there's an $f \in X^*$ and a $\gamma \in \mathbb{R}$ such that $\operatorname{Re} f(x_n) \ge \gamma$, but $\operatorname{Re} f(x) < \gamma$.

Taking the liminf, we see that $f(x_n) \not\to f(x)$, and therefore $x_n \not\rightharpoonup x$, which is a contradiction. ⊠

Perhaps more interesting is an immediate corollary.

Corollary 1.15.5. *Let X be an NLS and $S \subseteq X$ be convex. Then, the weak and strong closures of S agree.*

Proof. Let $\overline{S}^w$ denote the weak closure of S, and $\overline{S}$ denote the strong closure. Theorem 1.15.4 tells us that $\overline{S}^w \subseteq \overline{S}$: if $x_n \rightharpoonup x$, there's a sequence $y_n \to x$ with $y_n \in S$. Moreover, we already know that $\overline{S} \subseteq \overline{S}^w$, since we already knew that if $x_n \to x$, then $x_n \rightharpoonup x$. ⊠

Dual of an Operator. The dual of an operator goes by many names: dual may be the most common, but it is also known as the *transpose*, *conjugate transpose*, *adjoint*, and so forth.

Definition. Let X and Y be NLSes and $T \in B(X,Y)$. Then, define the *dual* to T to be the map $T^* \in B(Y^*, X^*)$ by $(T^*g)(x) = g(Tx)$.

This makes sense because if $g \in Y^*$ and $x \in X$, then Tx in y, so $g(Tx) \in \mathbb{F}$. Thus, $T^*g : X \to \mathbb{F}$. We can write $T^*g = g \circ T$, which writes it as the composition of two continuous (so that T^*g is continuous) and linear (so that T^*g is linear) functions. Thus, $T^*g \in X^*$.

We also have to check that $T^* : Y^* \to X^*$ is bounded — we've got a perfectly good definition already, but the boundedness of T^* was part of the definition. Suppose $g \in Y^*$ and $x \in X$; then,

$$\begin{aligned} |Tg^*(x)| &= |g(Tx)| \le \|g\|_{Y^*} \|Tx\|_Y \\ &\le \big(\|g\|_{Y^*} \|T\|_{B(X,Y)}\big) \|x\|_X. \end{aligned}$$

Thus, T^* is bounded, and moreover $\|T^*g\| \le \|g\|_{Y^*} \|T\|_{B(X,Y)}$. Finally, why is T^* linear? By definition,

$$\begin{aligned} T^*(g+h)(x) &= (g+h)(Tx) = g(Tx) + h(Tx) \\ &= (T^*g)(x) + (T^*h)(x) = (T^*g + T^*h)(x) \\ T^*(\lambda g)(x) &= (\lambda g)(Tx) = \lambda g(Tx) = \lambda (T^*g)(x). \end{aligned}$$

We already know that $\|T^*\|_{B(Y^*,\, X^*)} \le \|T\|_{B(X,Y)}$; the two are actually equal, and we'll prove this next lecture.

Lecture 16: 10/2/15

The Dual to an Operator.

Recall that last time, we defined the dual to an operator $T \in B(X,Y)$: the dual is $T^* : Y^* \to X^*$ given by $T^*(g) = g \circ T : X \to \mathbb{F}$. We showed this is linear and bounded, and in fact $\|T^*g\|_{X^*} \le \|T\|_{B(X,Y)} \|g\|_{Y^*}$, so taking the supremum, $\|T^*\| \le \|T\|_{B(X,Y)}$.

Claim. In fact, more is true: $\|T^*\| = \|T\|$.

PROOF. By the Hahn-Banach theorem,

$$\begin{aligned} \|T\|_{B(X,Y)} &= \sup_{\substack{x \in X \\ x \ne 0}} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\substack{x \in X \\ x \ne 0}} \sup_{\substack{g \in Y^* \\ g \ne 0}} \frac{|g(Tx)|}{\|x\|\|g\|} \\ &\le \sup_x \sup_g \frac{\|T^*g\|\|x\|}{\|x\|\|g\|} \\ &\le \sup_g \frac{\|T^*\|\|g\|}{\|g\|} = \|T^*\|. \end{aligned}$$ ⊠

We have another operator around, $* : B(X,Y) \to B(Y^*, X^*)$ sending $T \mapsto T^*$. One can show that it's linear and bounded, though this is left as an exercise.

Proposition 1.16.1.

(1) $*$ *is an isometry:* $\|T\| = \|T^*\|$.
(2) $*$ *is linear:* $(\lambda T + \mu S)^* = \lambda T^* + \mu S^*$.
(3) $*$ *is contravariant: if* $S : X \to Y$ *and* $R : Y \to Z$*, then* $(R \circ S)^* = S^* \circ R^*$ *as maps* $Z^* \to Y^* \to X^*$.
(4) $(\mathrm{id}_X)^* = \mathrm{id}_{X^*}$.

This is also left as an exercise, though maybe it reminds you of something: transposes of matrices.[16] For example, if $T : \mathbb{R}^m \to \mathbb{R}^n$, then T is represented by an $(n \times m)$-matrix A ($T(x) = Ax$). Then, $T^* = A^{\mathrm{T}}$; showing this is a useful exercise.

[16] What it reminds you of definitely depends on who you are; I see a contravariant functor!

Example 1.16.2. Let $1 < p < \infty$ and $f \in L^p(0,1)$. Define an *integral operator* $T : L^p(0,1) \to L^p(0,1)$ by

$$Tf(x) = \int_0^1 K(x,y) f(y)\,\mathrm{d}y,$$

where $K \in L^\infty((0,1)\times(0,1))$. We know that the dual space is $(L^p)^* = L^q$, where $1/p + 1/q = 1$, in the sense that $g \in L^q$ iff the map $\Lambda_g \in (L^p)^*$, where

$$\Lambda_g(f) = \int_0^1 g(x) f(x)\,\mathrm{d}x.$$

Let's find out what T^* is: this should be a map $(L^p)^* \to (L^p)^*$. We know

$$\begin{aligned}(T^*\Lambda_g)(f) = \Lambda_g(Tf) &= \int_0^1 g(x) Tf(x)\,\mathrm{d}x \\ &= \int_0^1 g(x)\,\mathrm{d}x \int_0^1 K(x,y) f(y)\,\mathrm{d}y\,\mathrm{d}x.\end{aligned}$$

Using the Fubini theorem, we may change the order of integration, because we want to recast this as an operator in terms of f.

$$\begin{aligned}&= \int_0^1 f(y) \int_0^1 K(x,y) g(x)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^1 f(x) \underbrace{\int_0^1 K(y,x) g(y)\,\mathrm{d}y}_{T^*\Lambda_g}\,\mathrm{d}x.\end{aligned}$$

That is, if $h(x) = \int_0^1 K(y,x) g(y)\,\mathrm{d}y$, then $T^*\Lambda_g = \Lambda_h$. Once again, $K(x,y)$ being sent to $K(y,x)$ is a sort of transposition.

Whenever we can do something once, we like to do it twice.

Lemma 1.16.3. *Let X and Y be NLSes and $T \in B(X,Y)$. Then, $T^{**} \in B(X^{**}, Y^{**})$ is a bounded* extension *of T, i.e. under the canonical inclusion $X \hookrightarrow X^{**}$, $T^{**}|_X = T$.*

If X is reflexive, this means that $T^{**} = T$. For example, in finite-dimensional vector spaces, applying the transpose twice gets you back where you started.

PROOF. Let $x \in X$ and $g \in Y^*$. Then, $T^{**}(E_x)(g) = E_x(T^*g) = (T^*g)(x) = g(Tx) = E_{Tx}(g)$, i.e. $T^{**}E_x = E_{Tx}$. ⊠

The proof amounts to unwinding definitions.

Back in the world of matrices, if the matrix is square we can sometimes take the inverse; remember that this commutes with the transpose. Let's generalize this.

Lemma 1.16.4. *Let X be a Banach space, Y be an NLS, and $T \in B(X,Y)$. Then, T has a bounded inverse on Y iff T^* has a bounded inverse on all of X^*; in this case, $(T^*)^{-1} = (T^{-1})^*$.*

PROOF. First, the forward direction: suppose $S = T^{-1} \in B(Y,X)$. Then, $S^*T^* = (TS)^* = (I_Y)^* = I_{Y^*}$ (where I_Y is the identity on Y), and therefore T^* is one-to-one. If we go the other way, $T^*S^* = (ST)^* = (I_X)^* = I_{X^*}$, so T^* is onto. Thus, $(T^*)^{-1}$ exists (as a linear map), and these calculations showed us that it's S^*, so in particular, $(T^*)^{-1}$ is bounded.

In the other direction, we know $(T^*)^{-1}$ exists, and therefore $(T^{**})^{-1}$ exists, and in particular is bijective, and $T^{**}|_X = T$, so in particular T is one-to-one (since T^{**} is), so we need to show that it maps onto Y, because then the open mapping theorem will imply it has a bounded inverse.

We know T^{**} maps onto and is an open map, so it takes open sets to open sets, and therefore closed sets to closed sets. Since X is Banach and therefore closed in X^{**}, then $T^{**}(X) = T(X)$ is closed in Y^{**}, and therefore $T(X)$ is closed in Y. Suppose T isn't onto Y, so that there exists a $y \in Y \setminus T(X)$, and the Hahn-Banach theorem allows us to create a $g \in Y^*$ such that $g|_{T(X)} = 0$ but $g(y) = \|y\| \neq 0$. But then we see that if $x \in X$, $(T^*g)(x) = g(Tx) = 0$, but since T^* is invertible, this means $g = 0$, and therefore $\|y\| = 0$, so $y = 0$. But $0 = T(0)$, so this means T is onto. ⊠

So ends Chapter 2 of the book. Chapter 3 is easier: many introductions to functional analysis start with Hilbert spaces, which are easier, and then ramp it up to where we were. And some abstract introductions start with a very general notion of a locally convex vector space!

A Hilbert space is a Banach space, but with more structure. They solve the problem that, though we have a nice notion of size, we don't have a notion of angle, like in finite-dimensional vector spaces. Hilbert spaces have a solution to this.

Definition. An *inner product* on a vector space H, denoted $(\cdot,\cdot)$, $(\cdot,\cdot)_H$, $\langle\cdot,\cdot\rangle$, or $\langle\cdot,\cdot\rangle_H$ is a map $H \times H \to \mathbb{F}$ such that:

(1) $(\cdot,\cdot)$ is linear in its first argument.
(2) $(\cdot,\cdot)$ is *conjugate symmetric*, i.e. $(x,y) = \overline{(y,x)}$: reversing the arguments produces the complex conjugate. If $\mathbb{F}$ is real, then this means that $(\cdot,\cdot)$ is symmetric.
(3) For any $x \in H$, $(x,x) \geq 0$, and $(x,x) = 0$ iff $x = 0$.

H along with this inner product is called an *inner product space*, which we'll abbreviate IPS.

If you combine properties (1) and (2), one sees that $(\cdot,\cdot)$ is *conjugate linear* in its second argument:

$$\begin{aligned}(x, \alpha y + \beta z) &= \overline{(\alpha y + \beta z, x)} \\ &= \overline{\alpha}\overline{(y,x)} + \overline{\beta}\overline{(z,x)} \\ &= \overline{\alpha}(x,y) + \overline{\beta}(x,z).\end{aligned}$$

This property, also known as *sesquilinearity*, means that it commutes with addition, but scalar multiplication in the second argument gets replaced with its conjugate.

Example 1.16.5.

(1) $\mathbb{F}^d$ is an IPS with the *complex dot product*

$$(x,y) = x \cdot \overline{y} = \sum_{i=1}^{d} x_i \overline{y}_i.$$

It's not hard to show that this satisfies the three defining properties.

(2) ℓ^p is *not* an inner product space, unless $p = 2$. The inner product on ℓ^2 is

$$(x,y) = x \cdot \overline{y} = \sum_{i=1}^{\infty} x_i \overline{y}_i.$$

By the Hölder inequality, this is bounded by $\|x\|_{\ell^2}\|\overline{y}\|_{\ell^2} = \|x\|_{\ell^2}\|y\|_{\ell^2}$, so the inner product is finite, and has the desired properties.

(3) Similarly, $L^p(\Omega)$ isn't an inner product space unless $p = 2$; an inner product on $L^2(\Omega)$ is given by

$$(f,g) = \int_\Omega f(x)\overline{g(x)}\,\mathrm{d}x.$$

The idea here is that 2 is its own conjugate exponent, because $1/2 + 1/2 = 1$. Hölder's inequality once again guarantees that this is finite.

CHAPTER 2

Inner Product Spaces and Hilbert Spaces

"Gentlemen: there's lots of room left in Hilbert space." – Saunders Mac Lane

Lecture 17: 10/5/15

Orthogonality.

Recall that last time, we defined an inner product $(\cdot,\cdot) : X \times X \to \mathbb{F}$ on a vector space X, which is linear in the first argument, satisfies $(x, y) = \overline{(y, x)}$, and is positive definite: $(x, x) \geq 0$, and is equal to 0 iff $x = 0$. Together with such an inner product, X is called an inner product space (IPS).

An inner product buys us a lot of other structure, too.

Definition. The *induced norm* on an IPS $(H, (\cdot,\cdot))$ is the function $\|\cdot\| : H \to [0, \infty)$ defined by $\|x\| = (x, x)^{1/2}$.

This is, unsurprisingly, a norm: first, $\|x\| \geq 0$ and is 0 iff $x = 0$ by positive definite-ness. Since the inner product is linear, then $\|\lambda x\| = |\lambda|\|x\|$, because

$$\|\lambda x\|^2 = (\lambda x, \lambda x) = \lambda\overline{\lambda}(x, x) = |\lambda|^2\|x\|^2.$$

Using the norm squared sometimes allows us to save writing some square roots.

For the triangle inequality we need an intermediate result, analogous to Hölder's inequality with $p = 2$.

Lemma 2.1.1 (Cauchy-Schwarz inequality)**.** *If H is an IPS and $x, y \in H$, $|(x, y)| \leq \|x\|\|y\|$.*

PROOF. We're done if $y = 0$, so assume $y \neq 0$ and $\lambda \in \mathbb{F}$. Then,

$$\begin{aligned} 0 \leq \|x - \lambda y\|^2 &= (x - \lambda y, x - \lambda y) \\ &= (x, x - \lambda y) - \lambda(x, x - \lambda y) \\ &= (x, x) - \overline{\lambda}(x, y) - \lambda(y, x) + |\lambda|^2(y, y) \\ &= \|x\|^2 - 2\operatorname{Re}(\lambda(y, x)) + |\lambda|^2\|y\|^2. \end{aligned} \tag{2.1}$$

This is a positive, quadratic function in λ, so by taking the derivative with respect to λ, the minimum is $\lambda = (x, y)/\|y\|^2$. Choose this λ, so as to obtain the maximum amount of information. Thus, in this case, (2.1) becomes

$$\begin{aligned} 0 \leq \|x\|^2 - 2\frac{|(x, y)|^2}{\|y\|^2} + \frac{|(x, y)|^2}{\|y\|^4}\|y\|^2 \\ = \|x\|^2 - \frac{|(x, y)|^2}{\|y\|^2}. \end{aligned}$$

⊠

The proof reminds us of Hölder's inequality, too. However, since the inner product is sesquilinear, it's important to keep track of complex conjugates to avoid adding errors.

Now, we can prove the triangle inequality for the induced norm.

$$\begin{aligned} \|x + y\|^2 &= (x + y, x + y) \\ &= \|x\|^2 + \|y\|^2 + 2\operatorname{Re}(x, y) \\ &\leq \|x\|^2 + \|y\|^2 + 2|(x, y)| \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \\ &= (\|x\| + \|y\|)^2. \end{aligned}$$

In other words, *inner product spaces are normed spaces too*.

The Cauchy-Schwarz inequality also suggests to us that we have a well-defined notion of angle.

Definition.

- If H is an inner product space over $\mathbb{F} = \mathbb{R}$, we can define the *angle* θ between two points $x, y \in H$ as the solution to
$$\cos\theta = \frac{(x,y)}{\|x\|\|y\|} \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$
- If H is a complex IPS, we define the *angle* between $x, y \in H$ to be the θ such that
$$\cos\theta = \frac{|(x,y)|}{\|x\|\|y\|} \in \left[0, \frac{\pi}{2}\right].$$
- If $(x, y) = 0$, so that the angle between them is $\pm\pi/2$, then x and y are said to be *orthogonal*, written $x \perp y$.

Though we only have a reduced notion of angle in complex vector spaces, it's OK, because we mostly care about orthogonality.

We also have a nice formula for addition: x, y, and 0 are three vertices of a parallelogram, and $x + y$ is the fourth vertex.

Proposition 2.1.2 (Parallelogram law)**.** $\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$.

Note that this is not true for all norms; in particular, if you can find one that doesn't satisfy the parallelogram law, then you've shown that not every NLS is an IPS.

We can also talk about topology.

Lemma 2.1.3. $(\cdot,\cdot) : X \times X \to \mathbb{F}$ *is continuous.*

Corollary 2.1.4. *Suppose $\lambda_n \to \lambda$ and $\mu_n \to \mu$ in $\mathbb{F}$, and that $x_n \to x$ and $y_n \to y$ in H. Then, $(\lambda_n x_n, \mu_n y_n) \to (\lambda x, \mu y)$.*

Proof of Lemma 2.1.3. To avoid confusion between the inner product and elements of the product space, we'll use $\langle\cdot,\cdot\rangle$ to denote the inner product on H in this proof.

Since H (and thus also $H \times H$) and $\mathbb{F}$ are metric spaces, continuity is equivalent to sequential continuity, so suppose $(x_n, y_n) \to (x, y)$ in $H \times H$, i.e. $x_n \to x$ and $y_n \to y$ in H. Thus,
$$\begin{aligned}|\langle x_n, y_n\rangle - \langle x, y\rangle| &\le |\langle x_n, y_n\rangle - \langle x_n, y\rangle| + |\langle x_n, y\rangle - \langle x, y\rangle|\\ &= |\langle x_n, y_n - y\rangle| + |\langle x_n - x, y\rangle|\\ &\le \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\|.\end{aligned}$$
Since $\{x_n\}$ and $\{y_n\}$ are convergent sequences, then their norms are bounded, so we're done. ⊠

Then, Corollary 2.1.4 follows directly: we know that scalar multiplication is continuous in any vector space. Lemma 2.1.3 shouldn't come as a surprise: we've defined a continuous norm, after all.

Best Approximation and Orthogonal Projection.

Definition. Just as a Banach space is a complete normed space, a *Hilbert space* is a complete IPS.

Theorem 2.1.5 (Best approximation)**.** *Let $(H, (\cdot,\cdot))$ be an IPS and $M \subseteq H$ be a nonempty, convex, and complete*[1] *subset. If $x \in H$, then there exists a unique $y \in M$ such that*
$$\operatorname{dist}(x, M) = \inf_{z\in M}\|x - z\| = \|x - y\|.$$

In other words, y is the best approximation (minimizing distance) to x that's in M.

Proof. Let $\delta = \inf_{z\in M}\|x - z\|$. If $\delta = 0$, then $y = x$, and moreover $x \in M$, because M is complete and therefore closed. This tells us that the best approximation to a point in the space is itself, which is perhaps unsurprising.

[1] If H is a Hilbert space, then completeness is the same as being closed.

If $\delta > 0$, then $x \neq M$, so there exists a sequence $\{y_n\}_{n=1}^{\infty}$ such that $\delta_n = \|x - y_n\| \to \delta$. It's easy to see that $\{y_n\}$ is Cauchy: by Lemma 2.1.2,

$$\begin{aligned}\|y_n - y_m\|^2 &= \|(y_n - x) + (x - y_m)\| \\ &= 2\big(\|y_n - x\|^2 + \|x - y_n\|^2\big) - \|y_n + y_m - 2x\| \\ &= 2\big(\delta_n^2 + \delta_m^2\big) - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2.\end{aligned}$$

Since M is convex, $(y_n + y_m)/2 \in M$, and therefore

$$\le 2\big(\delta_n^2 + \delta_m^2\big) - 4\delta^2 \longrightarrow 0.$$

Since M is complete and $\{y_n\}$ is Cauchy, then $y_n \to y$, which therefore attains the infimum and is our best approximation. Now, we have to prove uniqueness: if $\|x - z\| = \delta$, then

$$\|y - z\|^2 = 4\delta^2 - 4\left\|\frac{y + z}{2} - x\right\|^2 \le 4\delta^2 - 4\delta^2 = 0. \qquad \boxtimes$$

Corollary 2.1.6. *If $M \subset H$ is a complete linear subspace, $x \in H$, and $y \in M$ is its best approximation, then $x - y \perp M$.*[2]

Proof. Let $m \in M$ be nonzero. Then, for any $\lambda \in \mathbb{F}$,

$$\|x - y\|^2 \le \|x - y + \lambda m\|^2 = \|x - y\|^2 + |\lambda|^2\|m\|^2 + 2\operatorname{Re}\overline{\lambda}(x - y, m).$$

Taking $\lambda = -(x - y, m)/\|m\|^2$, we conclude that

$$\begin{aligned}0 &\le |\lambda|^2\|m\|^2 - \frac{2|(x - y, m)|}{\|m\|^4}\|m\|^2 \\ &= |\lambda|^2\|m\|^2 - 2|\lambda|^2\|m\|^2 = -\lambda^2\|m\|^2,\end{aligned}$$

but this only makes sense when $\lambda = 0$, which means that $(x - y, m) = 0$. $\boxtimes$

This is nothing more than projection from a vector space down to a subspace (though we do require completeness in this case). It'll be useful to consider all of the points which project down.

Definition. The *orthogonal complement* of any set $M \subseteq H$ is the set $M^\perp = \{x \in H : x \perp M\}$ (i.e. the set of x such that $(x, m) = 0$ for all $m \in M$).

For example, $\mathbb{R}^2$ is a Hilbert space with its usual inner product; if M is the x-axis, then $M^\perp$ is the y-axis. As in that case, we'd more generally want to write $H = M + M^\perp$, but we'll do that next time.

Lecture 18: 10/7/15

Projections.

Today's the exam, from 7 to 9 pm in UTC 1.104; it will cover §2.1 – 2.7 from the textbook.

We were talking about best approximation: if H is an IPS, $M \subseteq H$ is a nonempty, convex, and complete subset, and $x \in H$, then there exists a best approximation, i.e. a unique $y \in H$ such that $\operatorname{dist}(x, M) = \|x - y\|$. As a corollary, if M is a complete linear subspace (which is automatically nonempty and convex), we know $x - y \perp M$, so if $\langle \cdot, \cdot \rangle$ denotes the inner product, then $\langle x - y, m\rangle = 0$ for all $m \in M$. This motivated the definition of the orthogonal complement $M^\perp$, the set of all vectors orthogonal to M.

Proposition 2.2.1. *If H is an IPS and $M \subseteq H$, then $M^\perp$ is a closed linear subspace of H; furthermore, $M \perp M^\perp$ and $M \cap M^\perp = \{0\}$ or is empty.*

This will be left as an exercise, because it's not very hard. That $M^\perp$ is closed ultimately follows from the continuity of the inner product.

The best approximation is a projection, a concept you might recall from finite-dimensional linear algebra.

Definition. If X is an NLS and $P : X \to X$ satisfies $P^2 = P$, then P is called a *projection*.

[2] $x \perp V$ if for all $v \in V$, $x \perp v$, just as in linear algebra.

For example, on $\mathbb{R}^2$, we could project something down to the x-axis; if you do this twice, you're still on the x-axis, and nothing more changes. Also, if P is a projection and $M = \mathrm{Im}(P)$, then $P|_M = \mathrm{id}$: if $m \in M$, then $m = P(x)$ for some x, so $P(m) = P^2(x) = P(x) = m$.

Proposition 2.2.2. *Let X be an NLS and $P : X \to X$ be a projection mapping onto $M \subseteq X$. If $Q = I - P$, then Q is a projection, and $QP = PQ = 0$.*

Proof. This is just algebra:

$$Q^2 = (I-P)(I-P) = I^2 - IP - PI + P^2 = I - 2P + P = I - P = Q,$$

since P is a projection, so $P^2 = P$. Then, $QP = (I-P)P = P - P^2 = 0$, and $PQ = P(I-P) = P - P^2 = 0$. ⊠

We're particularly interested in projections that are also linear operators.

Proposition 2.2.3. *Let P and Q be as in Proposition 2.2.2, where P projects onto M and Q projects onto N. If P is linear, then so is Q, and $X = M \oplus N$.*[3] *If P is bounded and $M \neq \{0\}$, then $\|P\| \geq 1$.*

Proof. Since $Q = I - P$ is a difference of linear functions, then it's also linear.

To show that $X = M \oplus N$, we must show that $X = M + N$ and $M \cap N = \{0\}$. For the first claim, any $x \in X$ can be written as $x = x - Px + Px = (I-P)x + Px = Qx + Px$, and $Qx \in N$ and $Px \in M$; for the second, if $x \in M \cap N$, then $Px = x$ and $Qx = x$, so $PQx = x$, but $PQ = 0$, so $x = 0$.

Finally, for any $m \in M \setminus 0$,

$$\|P\| = \sup_{x \in X \setminus 0} \frac{\|Px\|}{\|x\|} \geq \frac{\|Pm\|}{\|m\|} = \frac{\|m\|}{\|m\|} = 1. \quad ⊠$$

The last part of this proposition is interesting: there are linear projections where the norm could increase; this is a little counterintuitive.

In an inner product space, we also have a notion of orthogonal projection.

Definition. Let H be an IPS and $M \subseteq H$ be a complete linear subspace. Then, define $P = P_M : X \to M$ as sending X to its best approximation in M, and define $P^\perp = P_M^\perp$ to be $I - P$.

By Theorem 2.1.5, Px (and therefore also $P^\perp x$) is uniquely defined for every $x \in X$.

Lemma 2.2.4. *P is a projection onto M and $P^\perp$ is a projection onto $M^\perp$.*

Proof. We proved that the best approximation in M of an $m \in M$ is just m again, so for any $x \in M$, since $Px \in M$, its best approximation $P(Px)$ is just Px again. Thus, $P^2 = P$. By Proposition 2.2.2, $P^\perp$ is a projection as well, and by Corollary 2.1.6, for any $x \in X$, $(I-P)x \in M^\perp$. Then, if $x \in M^\perp$, then its best approximation $Px \perp M$, but $Px \in M$ as well, so $Px = 0$, and therefore $x = P^\perp x$. Thus, $P^\perp$ is onto $M^\perp$. ⊠

By Proposition 2.2.2, this means $PP^\perp = P^\perp P = 0$ and $P^2 = (P^\perp)^2 = 0$; Lemma 2.2.4 also implies that $P_M^\perp = P_{M^\perp}$. Additionally, $x \in M$ iff $x = Px$ iff $P^\perp x = 0$, and thus also $x \in M^\perp$ iff $x = P^\perp x$ iff $Px = 0$.

Theorem 2.2.5. *P and $P^\perp$ are bounded linear operators.*

Proof. Let $x, y \in X$ and $\alpha, \beta \in \mathbb{F}$; then,

$$\begin{aligned} \alpha x + \beta y &= P(\alpha x + \beta y) + P^\perp(\alpha x + \beta y) \\ &= \alpha(Px + P^\perp x) + \beta(Py + P^\perp y), \end{aligned}$$

and therefore

$$\alpha Px + \beta Py - P(\alpha x + \beta y) = P^\perp(\alpha x + \beta y) - \alpha P^\perp x - \beta P^\perp y, \tag{2.2}$$

so (2.2) lies in both M (on the left) and $M^\perp$ (on the right), so it's equal to 0. Thus, we can conclude that $P(\alpha x + \beta y) = \alpha P(x) + \beta P(y)$, and similarly for $P^\perp$ in place of P, so they're both linear.

It suffices to show that P is bounded, because then it follows that $P^\perp$ is too. Since $P + P^\perp = I$, then

$$\begin{aligned} \|x\|^2 = \|Px + P^\perp x\|^2 &= \langle Px + P^\perp x, Px + P^\perp x\rangle \\ &= \|Px\|^2 + \underbrace{\langle Px, P^\perp x\rangle}_{0} + \underbrace{\langle P^\perp x, Px\rangle}_{0} + \|P^\perp x\|^2. \end{aligned} \tag{2.3}$$

That is, $\|Px\|^2 = \|x\|^2 - \|P^\perp x\|^2 \leq \|x\|^2$, meaning P is bounded and $\|P\| \leq 1$. ⊠

[3] If two subspaces $M, N \subseteq X$ are such that $X = M + N$ and $M \cap N = \{0\}$, then one says that $X = M \oplus N$, the *direct sum* of M and N.

(2.3) has a familiar-looking corollary.

Corollary 2.2.6 (Pythagorean theorem)**.** *For any $x \in X$, $\|x\|^2 = \|Px\|^2 + \|P^\perp x\|^2$.*

It also leads us to the the next result.

Corollary 2.2.7. *If $M \neq 0$, then $\|P\| = 1$, and if $M \neq X$, then $\|P^\perp\| = 1$.*

Proof. The proof of Theorem 2.2.5 shows $\|P\| \leq 1$, and Proposition 2.2.3 says that if $M \neq 0$, then $\|P\| \geq 1$, so in this case $\|P\| = 1$. Since $M^\perp \neq 0$ iff $M \neq X$, then in this case $\|P^\perp\| = 1$ too. ⊠

These results (particularly that $P_M^\perp = P_{M^\perp}$) are why the best approximation is often called the *orthogonal projection*, and we will adopt this term. The following result, which is a useful characterization of Px, is sometimes taken as its definition.

Proposition 2.2.8. *Let X, M, and P be as above and $x, y \in X$; then, $y = Px$ iff $y \in M$ and $\langle x - y, m\rangle = 0$ for all $m \in M$.*

Proof. The forward direction is Corollary 2.1.6, so suppose in the other direction that $y \in M$ and $\langle x - y, m\rangle = 0$ for all $m \in M$. Then, $\|x - (y + m)\|^2 = \|x - y\|^2 + \|m\|^2$ is minimal over $m \in M$ for the best approximation $y + m = Px$, but is also minimal if $m = 0$ (i.e. for y), so $Px = y$. ⊠

Dual Spaces. Let H be a Hilbert space; then, we'll define some elements of the dual space. For example, if $y \in H$, then let $L_y : H \to \mathbb{F}$ be given by $L_y(x) = \langle x, y\rangle$. This is linear and bounded: by the Cauchy-Schwarz inequality, $|L_y(x)| = |\langle x, y\rangle| \leq \|x\|\|y\|$, so $\|L_y\|_{H^*} \leq \|y\|$. But since

$$L_y\left(\frac{y}{\|y\|}\right) = \left\langle \frac{y}{\|y\|}, y\right\rangle = \|y\|,$$

then $\|L_y\|_{H^*} = 1$.

It turns out this is everything.

Theorem 2.2.9 (Riesz representation theorem)**.** *If $(H, \langle\cdot,\cdot\rangle)$ is a Hilbert space and $L \in H^*$, then there's a unique y such that $L = L_y$ (i.e. $Lx = \langle x, y\rangle$), and so $\|L\|_{H^*} = \|y\|_H$.*

Proof. The proof is very clever, and uses a trick.

First, let's use uniqueness: suppose $Lx = \langle x, y_1\rangle = \langle x, y_2\rangle$. Then, $\langle x, y_1 - y_2\rangle = 0$ for every $x \in X$, but this means that $y_1 - y_2 = 0$ (test $x = y_1 - y_2$; we know $\|y_1 - y_2\|^2 = 0$ iff $y_1 = y_2$).

For existence, first note that if $L = 0$, we can choose $y = 0$. Thus, suppose $L \neq 0$, and let $M = \ker(L)$, i.e. $M = \{x \in H : Lx = 0\}$. Since $L \neq 0$, then $M \subsetneq H$. Since $M = L^{-1}(0)$ and $\{0\}$ is closed, then M is closed.

Now for the weird part: choose any $z \in M^\perp$ such that $\|z\| = 1$, and consider $u = (Lx)z - (Lz)x$, so that $Lu = 0$, and thus $u \in M$. In particular, $u \perp z$, so

$$0 = \langle u, z\rangle = \langle (Lx)z - (Lz)x, z\rangle = (Lx)\|z\|^2 - Lz(x, z)$$

and so $Lx = (x, (Lz)z)$, so let $y = (Lz)z$. ⊠

Lecture 19: 10/9/15

Orthonormal Bases.

"Looks like a thinner class after the exam..."

In a finite-dimensional space, you can produce a basis of orthogonal vectors. It turns out you can do this in Hilbert spaces as well; our next goal is to prove this.

Definition.

- Let X be an IPS and $\mathcal{I}$ be an index set. Then, a set $A = \{x_\alpha\}_{\alpha\in\mathcal{I}} \subset X$ is *orthogonal* if $x_\alpha \neq 0$ for all $\alpha \in \mathcal{I}$ and $x_\alpha \perp x_\beta$ for all $\alpha, \beta \in \mathcal{I}$ such that $\alpha \neq \beta$ (they're pairwise orthogonal). If in addition $\|x_\alpha\| = 1$ for all α, A is said to be *orthonormal* (sometimes abbreviated ON).
- Let X be an NLS (no inner product needed here) and $A \subseteq X$. Then, A is *linearly independent* if every finite subset of A is linearly independent, i.e. if $\{x_i\}_{i=1}^n \subseteq A$ and $c_1x_1 + \cdots + c_nx_n = 0$, then $c_1 = \cdots = c_n = 0$.

For the rest of this lecture, H will denote a Hilbert space.

Proposition 2.3.1. *Let X be an inner product space and $A \subseteq X$ be an orthogonal subset. Then, A is linearly independent.*

PROOF. This is essentially the same proof as one does in finite dimensions: let $\{x_i\}_{i=1}^n$ be any finite subset of A and $c_i \in \mathbb{F}$ be such that $\sum_{i=1}^n c_i x_i = 0$. For any j, $(\cdot, x_j)$ is linear, so

$$0 = \sum_{i=1}^n c_i (x_i, x_j) = c_j \|x_j\|^2.$$

Since $x_j \neq 0$, then $c_j = 0$. ⊠

In order to talk about bases, we'll need to talk about projections again. Suppose that $\{x_1, \dots, x_n\}$ is a linearly independent subset of H, so that $M = \operatorname{span}\{x_1, \dots, x_n\}$ is a closed subspace of H. For any $x \in X$, $P_M x \in M$, so there exist $c_1, \dots, c_n \in \mathbb{F}$ such that

$$P_M x = \sum_{j=1}^n c_j x_j. \tag{2.4}$$

Let's calculate these c_j: since $P_M x - x \perp M$ by Proposition 2.2.8, then $(P_M x, x_i) = (x, x_i)$ for all i. Let $\mathbf{c} = (c_1, \dots, c_n)$, A be the matrix whose entries are $a_{ij} = (x_j, x_i)$, and $\mathbf{b} = (b_1, \dots, b_n)$, where $b_i = (x, x_i)$. Thus, (2.4) means that $A\mathbf{x} = \mathbf{b}$, so we can recover the c_j coefficients by $\mathbf{c} = A^{-1}\mathbf{b}$, assuming A is invertible.

This may be harder to compute in general, but orthogonality comes to our assistance: $A = I$, so $\mathbf{c} = \mathbf{b}$. In other words, we've proven (2.5a) in the following theorem.

Theorem 2.3.2. *Let $\{u_1, \dots, u_n\} \subseteq H$ be an orthonormal set. Let $M = \operatorname{span}\{u_1, \dots, u_n\}$ and $x \in H$; then,*

$$P_M x = \sum_{i=1}^n (x, u_i) u_i, \tag{2.5a}$$

and

$$\sum_{i=1}^n |(x, u_i)|^2 \leq \|x\|^2. \tag{2.5b}$$

PROOF. We've seen the proof of (2.5a), so let's look at (2.5b). Since P_M is an orthogonal projection, then $\|P_M\| = 1$, so

$$\sum_{i=1}^n |(x, u_i)|^2 = \left(\sum_i (x, u_i) u_i, \sum_j (x, u_j), u_j \right) = \|P_M x\|^2 \leq \|x\|^2.$$ ⊠

This proof once again looks a lot like what we did in finite-dimensional linear algebra. So let's do something that doesn't.

Definition. Let $\mathcal{I}$ be any index set, possibly uncountable, and choose $\{x_\alpha\}_{\alpha \in \mathcal{I}}$ with $x_\alpha \in [0, \infty)$. Then, define their sum to be

$$\sum_{x \in \mathcal{I}} x_\alpha = \sup_{\substack{J \subseteq \mathcal{I} \\ J \text{ finite}}} \sum_{\beta \in J} x_\beta.$$

For example, if $\mathcal{I} = \mathbb{N}$, then this looks familiar:

$$\sum_{\alpha=0}^\infty x_\alpha = \lim_{n \to \infty} \sum_{\alpha=0}^n x_\alpha.$$

However, adding uncountability doesn't get us very much: it turns out that if $\sum_{\alpha \in \mathcal{I}} x_\alpha$ is finite, then at most countably many x_α are nonzero! This is worth thinking about, but it makes sense: only finitely many can be greater than ε for any $\varepsilon > 0$.

For the rest of this lecture, $\{u_\alpha\}_{\alpha \in \mathcal{I}}$ will denote an orthonormal set in H.

Theorem 2.3.3 (Bessel's inequality). *For any $x \in H$,*

$$\sum_{\alpha \in \mathcal{I}} |(x, u_\alpha)|^2 \leq \|x\|^2.$$

Proof. If $J \subseteq \mathcal{I}$ is a finite set, then Theorem 2.3.2 tells us that
$$\sum_{\beta \in J} |(x, u_\beta)|^2 \le \|x\|^2,$$
so this is still true when we take the supremum. ⊠

Corollary 2.3.4. *At most finitely many (x, u_α) are nonzero.*

This might be a little strange: no matter how large this Hilbert space is, every vector can only project down to finitely many vectors in an orthonormal set.

We're working up to having an orthonormal basis for a Hilbert space, so let's consider some examples. $\mathbb{F}^d$ is a Hilbert space with $\|\cdot\|_{\ell^2}$ induced by the usual inner product, so we can take an indexing set $\mathcal{I} = \{1, \dots, d\}$. ℓ^2 with the usual norm and inner product (this time given by an infinite sum) can take the indexing set $\mathcal{I} = \mathbb{N}$ (starting from 1, not 0), which produces an orthonormal basis. We want to generalize this to possibly uncountable index sets $\mathcal{I}$, producing larger Hilbert spaces which are in some sense indexed in $\mathcal{I}$; then, the inner product should be similar, but with the sum over $\mathcal{I}$. Corollary 2.3.4 implies that such a large sum still makes sense.

Let's make this formal.

Definition. Let $\mathcal{I}$ be an index set, and let
$$\ell^2(\mathcal{I}) = \left\{ f : \mathcal{I} \to \mathbb{F} : \sum_{\alpha \in \mathcal{I}} |f(\alpha)|^2 \text{ is finite} \right\}.$$
Often, we'll let f_α denote $f(\alpha)$.

We'll end up proving that *all* Hilbert spaces are isomorphic to some $\ell^2(\mathcal{I})$! All you really need to know about a Hilbert space is how big it is.

Theorem 2.3.5 (Riesz-Fisher)**.** *Define $F : H \to \ell^2(\mathcal{I})$ by $F(x) = f_x$, where $f_x(\alpha) = x_\alpha = (x, u_\alpha)$. Then, F is a bounded linear surjection.*

We'd like to find an orthonormal set for which F is also one-to-one, but this will be useful for us nonetheless.

Proof. First, why is F linear? Take $\lambda \in \mathbb{F}$ and $x, y \in H$. Then,
$$\begin{aligned} F(\lambda x + y) &= \{(\lambda x + y)_\alpha\}_{\alpha \in \mathcal{I}} \\ &= \{(\lambda x + y, u_\alpha)\}_{\alpha \in \mathcal{I}}. \end{aligned}$$
By the definition of addition and scalar multiplication of functions,
$$\begin{aligned} &= \lambda \{(x, u_\alpha)\}_{\alpha \in \mathcal{I}} + \{(y, u_\alpha)\}_{\alpha \in \mathcal{I}} \\ &= \lambda F(x) + F(y). \end{aligned}$$
Then, Bessel's inequality tells us that
$$\|F(x)\|^2_{\ell^2(\mathcal{I})} = \sum_{\alpha \in \mathcal{I}} |x_\alpha|^2 \le \|x\|^2_H,$$
so $\|F\| \le 1$, and in particular F is bounded.

All the interesting content in the proof is the surjectivity: let $f \in \ell^2(\mathcal{I})$ and $n \in \mathbb{N} \setminus \{0\}$. We want to make things finite, where we can get a handle on them, so let $\mathcal{I}_n = \{\alpha \in \mathcal{I} : |f(\alpha)| \ge 1/n\}$. Then, using $|\cdot|$ to denote cardinality of a set,
$$|\mathcal{I}_n| = \sum_{\alpha \in \mathcal{I}_n} 1 < \sum_{\alpha \in \mathcal{I}_n} (n|f(\alpha)|)^2 \le n^2 \|f\|^2_{\ell^2(\mathcal{I})},$$
and the rightmost quantity is finite, so each $\mathcal{I}_n$ is a finite set. Thus, $\mathcal{J} = \bigcup_{i=1}^n \mathcal{I}_n$ is countable, and if $\beta \notin \mathcal{J}$, then $f(\beta) = 0$. Define an $x_n \in H$ by
$$x_n = \sum_{\alpha \in \mathcal{I}_n} f(\alpha) u_\alpha.$$
$\{x_n\}_{n=1}^\infty$ is a Cauchy sequence, because
$$\|x_n - x_m\|^2 = \left\| \sum_{\alpha \in \mathcal{I}_n \setminus \mathcal{I}_m} f(\alpha) u_\alpha \right\|^2 = \sum_{\alpha \in \mathcal{I}_n \setminus \mathcal{I}_m} |f(\alpha)|^2 \le \sum_{\alpha \in \mathcal{I} \setminus \mathcal{I}_m} |f(\alpha)|^2.$$

Since $\|f\|^2_{\ell^2(\mathcal{I})}$ is finite, then the tail (summing over $\alpha \in \mathcal{I} \setminus \mathcal{I}_n$) goes to 0. Since H is Hilbert, then there's an $x \in H$ such that $x_n \to x$, and therefore that $F(x_n) \to F(x)$. Now, we just have to show that $F(x) = f$: for any $\alpha \in I$,

$$\begin{aligned} F(x)(\alpha) = (x, u_\alpha) &= \lim_{n\to\infty} (x_n, u_\alpha) \\ &= \lim_{n\to\infty} \sum_{\beta \in \mathcal{I}_n} f(\beta)(u_\beta, u_\alpha) = f(\alpha). \end{aligned} \qquad \boxtimes$$

Next time, we'll take the maximal orthonormal set, and therefore get a basis, making F one-to-one as well as onto.

Lecture 20: 10/12/15

Midterm Breakdown.

First, we went over the midterm. Questions 1a and 1b were just stating definitions; for 1c, we want to show that in an NLS X,

$$\|x\| = \sup_{\substack{f \in X^* \\ f \neq 0}} \frac{|f(x)|}{\|f\|_{X^*}}.$$

Clearly this is true when $x = 0$, so assume $x \neq 0$ and let $Y = \operatorname{span}\{x\}$. Define $g(\lambda x) = \lambda\|x\|$, so that g is linear on Y and

$$\|g\| = \sup_{\|\lambda x\|=1} |g(\lambda x)| = \sup|\lambda|\|x\| = 1,$$

so by the Hahn-Banach theorem, $g \in Y^*$ extends to X^*. Thus,

$$\sup_{\substack{f \in X^* \\ f \neq 0}} \frac{|f(x)|}{\|f\|} \geq \frac{|g(x)|}{\|g\|} = \|x\|,$$

but $\|x\| \geq |f(x)|/\|f\|$ for all nonzero f, so $\|x\| \geq \sup|f(x)|/\|f\|$, and thus $\|x\|$ realizes the supremum. We proved this in class as Corollary 1.8.5. Then, part d follows directly from c, since $(L^3)^* = L^{3/2}$, so we get that

$$\|f\|_{L^3} = \sup_{\substack{g \in L^{3/2} \\ g \neq 0}} \frac{\left|\int fg\right|}{\|g\|_{L^{3/2}}}.$$

For question 2, we have a larger setup: let $a : [0, \infty) \to [0, \infty)$ be a continuous bijection such that $a(0) = 0$, and let $b = a^{-1}$. Then, define

$$A(t) = \int_0^t a(s)\,\mathrm{d}s \qquad \text{and} \qquad B(t) = \int_0^t b(s)\,\mathrm{d}s,$$

and suppose $A(st) \leq k(s)A(t)$ for all $s, t \geq 0$, where $k(s)$ is a continuous function such that $k(s) \to 0$ as $s \to 0$. Then, we can define

$$L_A = \left\{ u \mid \int_{\mathbb{R}} A(|u(x)|)\,\mathrm{d}x < \infty \right\}.$$

This integral doesn't scale nicely, so we modify it to get a norm

$$\|u\|_A = \inf\left\{ r > 0 \mid \int_{\mathbb{R}} A\left(\frac{|u(x)|}{r}\right) \mathrm{d}x \leq 1 \right\}.$$

Notice that if $a(t) = t^{p-1}/(p-1)$, then $A(t) = t^p$, and $L_A(\mathbb{R})$ is $L^p(\mathbb{R})$; maybe this provides some intuition for why we like conjugate exponents.

For part a, to show that L_A is a vector space, since it's a subset of the space of all functions, we just need to show it's a subspace, i.e. that it's closed under addition and scalar multiplication, and that it's nonempty. Since $0 \in L_A$, then the last property is true. For scalar multiplication, we know

$$\int A(|\lambda u|)\,\mathrm{d}x \leq k(\lambda) \int A|u(x)|\,\mathrm{d}x < \infty.$$

Addition relies on the fact that A is a convex function, which means that

$$\begin{aligned}\int A(|u(x)+v(x)|)\,\mathrm{d}x &\leq \int A(|u(x)|+|v(x)|)\,\mathrm{d}x\\ &= \int A\left(\frac{1}{2}\cdot 2|u|+\frac{1}{2}\cdot 2|v|\right)\mathrm{d}x\\ &\leq \int \frac{1}{2}A(2|u|)+\frac{1}{2}A(2|v|)\,\mathrm{d}x,\end{aligned}$$

and this last value is finite, so L_A is a vector space.

Part b asks us to show that $\|\cdot\|_A$ is a norm. First, why is it even finite? We know that $\int A(|u|)=R$ is finite, so

$$\int A\left(\frac{|u(x)|}{r}\right)\mathrm{d}x \leq \int k\left(\frac{1}{r}\right)A(|u|)\,\mathrm{d}x = k\left(\frac{1}{r}\right)R,$$

and since $k(1/r)\to 0$ as $r\to\infty$, then this is bounded by 1 for some finite r. Then, for scalar multiplication, $|\lambda u|/r = |u|/(r/|\lambda|)$, so if $S=r/|\lambda|$, then

$$\begin{aligned}\|u\|_A &= \inf\left\{r>0 \mid \int A\left(\frac{|\lambda u(x)|}{r}\right)\mathrm{d}x \leq 1\right\}\\ &= \left\{|\lambda|s>0 \mid \int A\left(\frac{|u|}{s}\right)\mathrm{d}x\leq 1\right\}\\ &= |\lambda|\|u\|_A.\end{aligned}$$

Clearly, $\|0\|=0$, but the other direction is more interesting: suppose $\|u\|_A=0$, so that $\int A(|u|/r)\,\mathrm{d}x\leq 1$ for all $r>0$. If $|u|\neq 0$, then there must exist a set $S\subset\mathbb{R}$ with nonzero measure on which $|u|\geq\varepsilon>0$, and so $\int A(|u|/r)\,\mathrm{d}x \geq \int_S A(\varepsilon/r)\,\mathrm{d}x\leq 1$, but since $A(t)\to\infty$ as $t\to\infty$, this is a contradiction.

Finally, we have to show tha triangle inequality. Since the norm is the infimum, then when $\varepsilon>0$ is small, then

$$\int A\left(\frac{|u(x)|}{\|u\|+\varepsilon}\right)\mathrm{d}x\leq 1.$$

Then, using the monotone convergence theorem, we can remove the ε. In any case, convexity allows us to do the following.

$$\begin{aligned}\int A\left(\frac{|u|+|v|}{\|u\|+\|v\|+2\varepsilon}\right)\mathrm{d}x &\leq \int A\left(\frac{\|u\|+\varepsilon}{\|u\|+\|v\|+2\varepsilon}\frac{|u|}{\|u\|+\varepsilon}+\frac{\|v\|+\varepsilon}{\|u\|+\|v\|+2\varepsilon}\frac{|v|}{\|v\|+\varepsilon}\right)\mathrm{d}x\\ &\leq \frac{\|u\|+\varepsilon}{\|u\|+\|v\|+2\varepsilon}\int\left(A\left(\frac{|u|}{\|u\|+\varepsilon}\right)+A\left(\frac{|v|}{\|v\|+\varepsilon}\right)\right)\mathrm{d}x\\ &\leq 1,\end{aligned}$$

so letting $\varepsilon\to 0$,

$$\|u+v\| = \inf_{\varepsilon\to 0} r\leq \|u\|+\|v\|+2\varepsilon.$$

Part c is akin to Hölder's inequality. Just by the definitions of a and b, we know that for any $s,t\geq 0$,

$$st\leq\int_{\mathbb{R}} a(s)+\int_{\mathbb{R}} b(t).$$

(If this doesn't make sense, draw a picture.) Then, let $s=|u|/\|u\|$ and $t=|v|/\|v\|$, so that

$$\int\frac{|u||v|}{\|u\|_A\|u\|_B}\,\mathrm{d}x\leq A\left(\frac{|u|}{\|u\|_A}\right)+B\left(\frac{|v|}{\|v\|_B}\right)\leq 2.$$

This one had more real analysis than one might have expected, but this is typical of examples.

For question 3, let X and Y be Banach spaces and $\{x_i\}_{i=1}^\infty$ be a dense subset of X. Let $\{T_n\}\subset B(X,Y)$ such that $\max_n\|T_nx\|$ is finite for all $x\in X$, and suppose $\{T_nx_i\}_{n=1}^\infty$ is Cauchy for all i; then, we want to show $T_n\to T$ for some bounded linear $T:X\to Y$.

Immediately, the uniform boundedness principle, Theorem 1.12.3, tells us that there's an M such that $\|T_n\|\leq M$ for all n. We can also deduce that for each i, there's a unique y_i such that $T_nx_i\to y_i$ as $n\to\infty$. We'll let $Tx_i=y_i$.

For any $x \in X$, there's a subsequence $x_{i_j} \to x$ as $j \to \infty$. Then, $\{T_n x\}$ is Cauchy, because

$$\begin{aligned}\|T_n x - T_m x\| &\le \|T_n x - T_n x_{i_j}\| + \|T_n x_{i_j} - T_m x_{i_j}\| + \|T_m x_{i_j} - T_m x\| \\ &\le 2M\|x - x_{i_j}\| + \|(T_n - T_m)x_{i_j}\|.\end{aligned}$$

When j is large, the first term is small, and so we can then take n and m to be large, which makes the second term go to zero. Then, we can define $T : X \to Y$ by $Tx = \lim_{n\to\infty} T_n x$. Clearly, T is linear, and it's bounded because

$$\|Tx\| = \lim_{n\to\infty} \|T_n x\| \le M\|x\|.$$

Thus, $T \in B(X,Y)$. This is where the problem should have stopped; instead, it asked that $T_n \to T$ in $B(X,Y)$. Nobody showed that; moreover, it may be false! Thus, points were awarded for realizing there was something more to say. The trick is that the Cauchy convergence of $T_n x_i$ may not be uniform.

Back to Hilbert Spaces. Recall that the Riesz-Fischer theorem, Theorem 2.3.5, allows us to surject onto $\ell^2(\mathcal{I})$ if there's an orthonormal set in a Hilbert space indexed by $\mathcal{I}$. We want to make this map an isomorphism, but we might not have picked the largest orthonormal set.

Theorem 2.4.1. *Let H be a Hilbert space and $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ be an orthonormal set. Then, the following are equivalent.*

(1) *$\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is a maximal orthonormal set, i.e. adding any nonzero vector to it breaks orthogonality.*
(2) *$\operatorname{span}\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is dense in H.*
(3) *$\{u_\alpha\}_{\alpha\in\mathcal{I}}$ captures the norm: for all $x \in H$,*

$$\|x\|_H^2 = \sum_{\alpha\in\mathcal{I}} |(x,u_\alpha)|^2.$$

(4) *$\{u_\alpha\}_{\alpha\in\mathcal{I}}$ captures the inner product: for all $x, y \in H$,*

$$(x,y) = \sum_{\alpha\in\mathcal{I}} (x,u_\alpha)\overline{(y,u_\alpha)}.$$

Proof that (1) $\Longrightarrow$ (2). Suppose $\operatorname{span}\{u_\alpha\}_{\alpha\in\mathcal{I}}$ isn't dense in H; then, $M = \overline{\operatorname{span}\{u_\alpha\}}$ is a closed subspace of H that isn't all of H. Since M is closed, then $H = M \oplus M^\perp$, and since $M \subsetneq H$, then $M^\perp$ is nontrivial. Thus, we may pick a $v \in M^\perp$ such that $\|v\| = 1$, but $v \perp u_\alpha$ for each $\alpha \in \mathcal{I}$, because $u_\alpha \in M$, which contradicts the assumption that $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is a maximal orthonormal set. ⊠

Lecture 21: 10/14/15

Classification of Hilbert Spaces.

Recall that we were in the middle of proving Theorem 2.4.1; we proved that (1) $\Longrightarrow$ (2) last lecture.

Proof of Theorem 2.4.1 (Continuation). For (2) $\Longrightarrow$ (3), we have $M = \overline{\operatorname{span}\{u_\alpha\}} = H$. Let $x \in H$ and for every $\alpha \in \mathcal{I}$, let $x_\alpha = (x, u_\alpha)$. Then,

$$\|x\|^2 \ge \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2,$$

and if $\varepsilon > 0$ there exist $c_i \in \mathbb{F}$, $\alpha_i \in \mathcal{I}$, and an $N \in \mathbb{N}$ such that

$$\left\| x - \sum_{i=1}^N c_i u_{\alpha_i} \right\| < \varepsilon.$$

But the x_{α_i} are the coefficients in a best approximation, so

$$\left\| x - \sum_{i=1}^N x_{\alpha_i} u_{\alpha_i} \right\|^2 \le \left\| x - \sum_{i=1}^N c_i u_{\alpha_i} \right\| < \varepsilon^2.$$

The term on the left is equal to

$$\|x\|^2 - \sum_{i=1}^N |x_{\alpha_i}|^2,$$

so

$$\|x\|^2 \le \sum_{i=1}^N |x_{\alpha_i}|^2 + \varepsilon^2 \le \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 + \varepsilon^2.$$

Then, Bessel's inequality provides the bound in the other direction.

For (3) $\implies$ (4), we want to relate the inner product and the norm. Since the definition of the coefficients x_α is linear in x and

$$\|x+y\|^2 = \|x\|^2 + (x,y) + (y,x) + \|y\|^2,$$

then

$$\sum_{\alpha\in\mathcal{I}} |x_\alpha + y_\alpha|^2 = \sum_{\alpha\in\mathcal{I}} \left(|x_\alpha|^2 + x_\alpha \overline{y}_\alpha + \overline{x}_\alpha y_\alpha + |y_\alpha|^2\right). \tag{2.6}$$

Thus,

$$\|x+iy\|^2 = \sum_\alpha |x_\alpha + iy_\alpha| = \sum_\alpha \left(|x_\alpha|^2 - ix_\alpha\overline{y}_\alpha + i\overline{x}_\alpha y_\alpha + |y_\alpha|^2\right). \tag{2.7}$$

Combining (2.6) and (2.7), we have that

$$\begin{aligned}(x,y) + \overline{(x,y)} &= \sum_\alpha \left(x_\alpha\overline{y}_\alpha + \overline{x_\alpha\overline{y}_\alpha}\right)\\ -(x,y) + \overline{(x,y)} &= \sum_\alpha \left(-x_\alpha\overline{y}_\alpha + \overline{x_\alpha\overline{y}_\alpha}\right).\end{aligned}$$

Subtracting these two,

$$2(x,y) = 2\sum_{\alpha\in\mathcal{I}} x_\alpha \overline{y}_\alpha,$$

so given the norm and a maximal orthonormal subset, we can reconstruct the inner product (once we finish proving the theorem).

For (4) $\implies$ (1), suppose $\{u_\alpha\}$ isn't a maximal orthonormal set. Then, there's some $u\in H$ such that $u \perp u_\alpha$ for all $\alpha\in\mathcal{I}$, and $\|u\| = 1$. However, then

$$1 = \|u\|^2 = \sum_{\alpha\in\mathcal{I}} |(u,u_\alpha)|^2 = 0. \qquad \boxtimes$$

This result has a number of corollaries.

For the rest of this lecture, H will denote a Hilbert space, and $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ will denote a maximal orthonormal basis for H.

Corollary 2.5.1. *Suppose H is infinite-dimensional. For any $x\in H$, then there exist $\alpha_i \in \mathcal{I}$, with $i = 1,2,\dots$, such that*

$$x = \sum_{i=1}^\infty (x, u_{\alpha_i}) u_{\alpha_i} = \sum_{\alpha\in\mathcal{I}} (x,u_\alpha)u_\alpha.$$

That is, every element only sees countably many elements of any orthonormal subset, no matter how large our space is.

And now, the moment we've been waiting for.

Corollary 2.5.2. *The Riesz-Fischer map $F : H \to \ell^2(\mathcal{I})$ is a Hilbert space isomorphism.*

Proof. We already know F is linear and surjective, but it's injective: if $F(x) = 0$, then we've just seen that $x = 0$. Theorem 2.4.1 also tells us that the inner product structures are exactly the same, so H and $\ell^2(\mathcal{I})$ are abstractly isomorphic. ⊠

We're missing one thing, though: what if there's a Hilbert space without a maximal orthonormal basis?

Theorem 2.5.3. *Let $\{v_\alpha\}_{\alpha\in\mathcal{I}} \subset H$ be an orthonormal set. Then, there exists a maximal orthonormal set $\{v_\beta\}_{\beta\in\mathcal{J}}$ containing $\{v_\alpha\}_{\alpha\in\mathcal{I}}$.*

This theorem is the final step in proving the following corollary.

Corollary 2.5.4. *Every Hilbert space H is isomorphic to $\ell^2(\mathcal{I})$ for some $\mathcal{I}$. If H is separable and infinite-dimensional, then $H \cong \ell^2(\mathbb{N})$.*

This is kind of impressive: up to isomorphism, there is exactly one separable, infinite-dimensional Hilbert space.

PROOF OF THEOREM 2.5.3. The general result uses Zorn's lemma. This is a bit mysterious (setting up chains and maximal elements and stuff), so we won't do it; instead, we'll provide an explicit construction in the separable case.

Let $\{\widetilde{x}_j\}_{j=1}^{\infty}$ be a dense subset of a separable Hilbert space H and $M = \overline{\operatorname{span}\{v_\alpha\}}$. Let $\hat{x}_j = \widetilde{x}_j - P_M \widetilde{x}_j$, so that $\hat{x}_j \perp M$. Thus, $\operatorname{span}\{v_\alpha\} \cup \{\hat{x}_j\}$ is dense, but the $\hat{x}_j$ might not be orthogonal to each other. Thus, we use the Gram-Schmidt process.

Define $x_1 = \hat{x}_1$, and for $j \in \mathbb{N}$, we'll do induction. Let $N_j = \overline{\operatorname{span}\{x_1, \dots, x_j\}}$ and define $x_{j+1} = \hat{x}_{j+1} - P_{N_j}\hat{x}_{j+1}$, so that $x_{j+1} \perp N_j$ and $x_{j+1} \perp M$ as before. Then, we can consider the set $\operatorname{span}\{v_\alpha\} \cap \{x_j\}_{j=1}^{\infty}$, which is an orthogonal, dense set. Then, throw out the elements that are 0 and normalize, and we have an orthonormal set, so since (1) and (2) in Theorem 2.4.1 are equivalent, we're done. ⊠

Example 2.5.5. Let's talk about Fourier series. You likely saw this in undergrad, but probably not rigorously.

Consider the functions $f : \mathbb{R} \to \mathbb{C}$ that are periodic of period T. Then, $g(x) = f(\lambda x)$ has period T/λ, so we can rescale to get any period we like. We'll thus restrict to a particularly convenient case, $T = 2\pi$. So that means we're looking at the space $L^2_{\text{per}}(-\pi, \pi)$, the set of $f : \mathbb{R} \to \mathbb{C}$ such that $f \in L^2([-\pi, \pi])$ and $f(x + 2n\pi) = f(x)$ for all $n \in \mathbb{Z}$ and for almost all $x \in [-\pi, \pi]$.

It's not a huge surprise that $L^2_{\text{per}}(-\pi, \pi)$ is a Hilbert space, with the inner product

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,\mathrm{d}x.$$

There are a few things to check, but this is not difficult; it's essentially the same proof as for L^2.

Now, why do Fourier series work? The claim is that $\{e^{inx} : n \in \mathbb{Z}\} \subset L^2_{\text{per}}$ is orthonormal.[4] This is because

$$\int_{-\pi}^{\pi} e^{inx}e^{-imx}\,\mathrm{d}x = \left.\frac{e^{i(n-m)x}}{i(n-m)}\right|_{-\pi}^{\pi} = \begin{cases} 2\pi, & \text{if } m = n \\ 0, & \text{if } m \neq n. \end{cases}$$

Theorem 2.5.6. $\operatorname{span}\{e^{inx} : n \in \mathbb{Z}\}$ *is dense in* $L^2_{\text{per}}[-\pi, \pi]$*, so it's an orthonormal basis.*

PROOF. First off, since $C_0(-\pi, \pi)$ is dense in $L^2_{\text{per}}[-\pi, \pi]$, then $C_{\text{per}}[-\pi, \pi]$ is dense in $L^2_{\text{per}}[-\pi, \pi]$ (just extend the function so that it's periodic). Thus, we can reduce to showing the theorem for continuous functions.

For any $m \geq 0$, then let $k_m : [-\pi, \pi] \to \mathbb{C}$ be defined by

$$k_m(x) = c_m\left(\frac{1 + \cos x}{2}\right)^m,$$

where c_m is defined so that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} k_m(x)\,\mathrm{d}x = 1.$$ [5]

In particular, since we can write

$$k_m(x) = c_m\left(\frac{2 + e^{ix} + e^{-ix}}{4}\right)^m,$$

which is in $\operatorname{span}\{e^{inx} : -m \leq n \leq m\}$, so there exist coefficients λ_n such that

$$k_m(x) = \sum_{n=-m}^{m} \lambda_n e^{inx}.$$

[4] You may have seen this in the alternate form $e^{inx} = \cos nx + i \sin nx$.

[5] This might seem a little arbitrary or magical, but when we talk about distributions, specifically in Theorem 4.6.8 and Corollary 4.6.10, this will be more motivated. Specifically, as $m \to \infty$, k_m converges to the δ-"function" which is 0 when $x \neq 0$ and is infinite at $x = 0$.

We're going to take a convolution (which we'll learn more about later, in §4.6): for any $f \in C_{\text{per}}[-\pi,\pi]$, let

$$\begin{aligned} f_m(x) &= \int_1^{2\pi} k_m(x-y)f(y)\,\mathrm{d}y \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{n=-m}^{m}\lambda_n e^{in(x-y)}f(y)\,\mathrm{d}y \\ &= \sum_{n=-m}^{m}\frac{\lambda_n}{2\pi}\left(\int_{-\pi}^{\pi}e^{-iny}f(y)\,\mathrm{d}y\right)e^{inx}, \end{aligned}$$

which is also in $\operatorname{span}\{e^{inx}\}_{n=-m}^{m}$.

The remainder of the proof, which we'll do next time (since we've run out of time today), involves showing that $f_m \to f$ uniformly, i.e. in L^∞. Thus, since we're on a finite interval, this implies convergence in L^2.

Lecture 22: 10/16/15

Fourier Series and Weak Convergence in Hilbert Spaces.

"Fourier series have a sound foundation."

Recall that we were in the midst of proving the validity of Fourier series for functions in $L^2_{\text{per}}(-\pi,\pi)$, the functions $f:\mathbb{R}\to\mathbb{C}$ that are L^2 on $(-\pi,\pi)$ and 2π-periodic. This involved showing that the continuous periodic functions on $[-\pi,\pi]$ are dense in L^2_{per}, and that $\{e^{inx}\}$ is an orthonormal basis in the inner product

$$(f,g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,\mathrm{d}x.$$

CONTINUATION OF PROOF OF THEOREM 2.5.6. We had defined

$$f_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} k_m(x-y)f(y)\,\mathrm{d}y,$$

which is contained in $\operatorname{span}\{e^{inx}\}_{n=-m}^{n=m}$. So we want to show that $f_n \to f$ uniformly, and since $\|f_n - f\|_{L^\infty} \ge \|f_n - f\|_{L^2}$, that's sufficient to prove the theorem. First, as (eventually) implied by the Cauchy-Schwarz theorem,

$$f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} k_m(y)f(x)\,\mathrm{d}y,$$

and so

$$\begin{aligned} |f_m(x) - f(x)| &\le \frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x-y)-f(x)|k_m(y)\,\mathrm{d}y \\ &= \underbrace{\int_{|y|\le\delta}|f(x-y)-f(x)|k_m(y)\,\mathrm{d}y}_{I_1} + \underbrace{\int_{\delta\le|y|\le\pi}|f(x-y)-f(x)|k_m(y)\,\mathrm{d}y}_{I_2}, \end{aligned}$$

for any $\delta \in (0,\pi)$. Since f is continuous on $[-\pi,\pi]$, which is compact, then it's uniformly continuous, so for any $\varepsilon > 0$, there's a $\delta > 0$ such that $|f(x-y)-f(x)| < \varepsilon/2$ for all y with $|y| \le \delta$, and thus

$$I_1 \le \frac{1}{2\pi}\int_{|y|\le\delta}\frac{\varepsilon}{2}k_m(y)\,\mathrm{d}y = \frac{\varepsilon}{2}.$$

For I_2, we'll need to make a more careful estimate. If $\delta \le |y| \le \pi$, then $k_m(y) \le c_m((1+\cos\delta)/2)^m$, and therefore

$$\begin{aligned} 1 &= \frac{c_m}{\pi}\int_0^\pi \left(\frac{1+\cos x}{2}\right)^m \mathrm{d}x \\ &\ge \frac{c_m}{\pi}\int_0^\pi \left(\frac{1+\cos x}{2}\right)\sin x\,\mathrm{d}x \\ &= -\frac{2c_m}{\pi}\frac{1}{m+1}\left(\frac{1+\cos x}{2}\right)^{m+1}\Bigg|_0^\pi \\ &= \frac{2c_m}{\pi}\left(\frac{1}{m+1}\right), \end{aligned}$$

so $c_m \le (\pi/2)(m+1)$. The point here is that we showed that $c_m = O(m)$. In particular, for m sufficiently large, $k_m(y) < \varepsilon/4M$ and therefore

$$I_2 \le \frac{1}{2\pi}2M\int_{\delta\le|y|\le\pi}\frac{\varepsilon}{4M}\,\mathrm{d}x = \frac{\varepsilon}{2}. \qquad ⊠$$

Thus, we actually have Fourier series: if $f \in L^2_{\mathrm{per}}(-\pi,\pi)$, then

$$\begin{aligned} f(x) &= \sum_{n=-\infty}^{\infty}\left\langle f, e^{-in(\cdot)}\right\rangle e^{-inx} \\ &= \sum_{n=-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)e^{-iny}\,\mathrm{d}y\right)e^{-inx}. \end{aligned}$$

This is also the basis[6] for other common techniques, such as separation of variables.

Weak Convergence. Let H be a Hilbert space. In contrast to the Banach case, we have more control on what our linear functionals are: specifically, $x_n \rightharpoonup x$ is equivalent to $(x_n, y) \to (x, y)$ for all $y \in H$.

It would be really nice if we only had to check on an orthonormal basis. This is almost true.

Lemma 2.6.1. *Let $\{e_\alpha\}_{\alpha\in\mathcal{I}}$ be an orthonormal basis of H. Then, $x_n \rightharpoonup x$ iff $(x_n, e_\alpha) \to (x, e_\alpha)$ for all $\alpha \in \mathcal{I}$* and $\|x_n\|$ is bounded.

Proof. The forward direction is true pretty much by definition, so let's prove the converse.

Let $y \in H$ and $\varepsilon > 0$, so that there exists a $z \in \operatorname{span}\{e_\alpha\}_{\alpha\in\mathcal{I}}$ such that $\|y - z\| < \varepsilon$. Thus, $(x_n, z) \to (x, z)$, and therefore

$$\begin{aligned} \limsup_{n\to\infty}|(x_n - x, y)| &= \limsup_{n\to\infty}|(x_n - x, y - z)| \\ &\le \left(\sup_n \|x_n\| + \|x\|\right)\|y - z\| \\ &\le C\|y - z\| \le C\varepsilon. \qquad ⊠ \end{aligned}$$

Since Hilbert spaces are reflexive, then the Banach-Alaoglu theorem (Theorem 1.14.2) automatically applies, and in fact this is the primary use of this theorem.

Lemma 2.6.2. *If H is separable and $\{x_n\} \subset X$ is a bounded sequence, then there exists a subsequence $\{x_{n_j}\}$ and an $x \in X$ such that $x_n \rightharpoonup x$.*

This is the end of this chapter; next week, we'll begin talking about spectral theory. We'll do just a little bit of it today.

Recall that if $A : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, one looks for its eigenvalues, the $\lambda \in \mathbb{R}$ such that $Ax = \lambda x$. The way to do this is to find the kernel of $(A - \lambda I)$, or equivalently check whether it's invertible.

In this chapter, the field will always be complex (it's nice to have an algebraically closed field), so suppose X is a complex NLS and $T : X \to X$ is a bounded linear operator. These are kind of rare, so let's generalize: we'll let $D = D(T) \subseteq X$ be a subspace (the *domain* of T), so that we can consider $T : D \to X$. For example, differentiation is a map $C^1 \to C^0$. In other words, we're considering partially defined functions. Normally, though, we take the domain to be dense inside X, as in the example.

[6] No pun intended.

Definition. Let T be a linear operator.

- The *range* $R(T) = \{x \in X : Ty = x \text{ for some } y \in Y\}$ (the set of points hit by T).
- The *kernel* or *null space* of T, written $\ker(T)$ or $N(T)$, is $\{x \in D : Tx = 0\}$.
- If $\lambda \in \mathbb{C}$, T_λ will denote $T - \lambda I$, and $R_\lambda = R(T_\lambda)$.

So our question can be recast as: does $T_\lambda : D \to R_\lambda$ have an inverse? That's the set of less interesting λ, so to speak. But we need to check that it's one-to-one (since it's onto its range by definition). If D is dense in X, then R_λ ought to be dense in X too, so we want that to be true too. Finally, we need T (and T_λ^{-1}, if it exists) to be bounded. These questions are different from the finite-dimensional case, where we only need to check injectivity; this is what makes spectral theory more complicated in Banach spaces.

Definition. The *resolvent* $\rho(T)$ is the set of $\lambda \in \mathbb{C}$ such that T_λ is injective and maps onto a dense subset of X, and T_λ^{-1} is bounded. Sometimes, T_λ^{-1} is called the *resolvent operator for T at λ*.

The resolvent is the case where everything is nice… and boring. We won't worry about these λ very often.

Definition. The *spectrum* of T is $\sigma(T) = \mathbb{C} \setminus \rho(T)$. We divide it as follows.

(1) The *point spectrum* $\sigma_p(T)$, the set of $\mu \in \mathbb{C}$ such that T_μ is not one-to-one. These are the classical eigenvalues: there is no inverse.
(2) The *continuous spectrum* $\sigma_c(T)$, the set of $\mu \in \mathbb{C}$ where T_μ is one-to-one and $R_\lambda \subseteq X$ is dense, but T_μ^{-1} isn't bounded.[7]
(3) The *residual spectrum* $\sigma_r(T)$, the remaining $\mu \in \mathbb{C}$: these are the most pathological examples, where T_μ is injective, but $R(T_\mu)$ is not dense in X.

The next proposition follows directly from the definition.

Proposition 2.6.3. *$\sigma_p(T)$, $\sigma_c(T)$, and $\sigma_r(T)$ are disjoint, and their union is $\sigma(T)$.*

Notice that in the finite-dimensional case, we only have the point spectrum, so we will have to look to infinite-dimensional cases for examples.

If $\lambda \in \sigma_p(T)$, then $N(T_\lambda) \neq 0$, and so there exist nonzero $x \in X$ such that $Tx = \lambda x$. In this case, λx is called an *eigenvalue* and x is called an *eigenvector*, or more commonly an *eigenfunction*.

Example 2.6.4. Let $T : \ell^2 \to \ell^2$ be the shift operator: $T(x_1, x_2, x_3, \dots) = (0, x_1, x_2, \dots)$. Then, T^{-1} exists, but $R(T)$ isn't dense in ℓ^2, so $0 \in \sigma_r(T)$.

This is a good example to have in your pocket, because it really helps illustrate the difference between the finite-dimensional and infinite-dimensional cases.

[7]Yes, it is a little odd that the continuous spectrum is where the inverse is not continuous.

CHAPTER 3

Spectral theory

Lecture 23: 10/19/15

Basic Spectral Theory in Banach Spaces.

Recall the setup from last Friday: we have a linear operator $T : D(T) \to R(T)$, where $D(T)$, the domain, is dense in our space X. We'll take $\lambda \in \mathbb{C}$ and consider $T_\lambda = T - \lambda I$. T_λ has the same domain, but its range $R_\lambda = R(T_\lambda)$ may be different. We want to know whether T_λ is one-to-one, whether R_λ is dense, and whether T_λ^{-1} is bounded.

The spectrum is where these notions fail, and (the failure to satisfy) each of these three conditions gives rise to the point spectrum, the continuous spectrum, and the residual spectrum, respectively. We also defined the resolvent as the space where all of these properties are satisfied.

Example 3.1.1. The canonical kind of operator we want to look at is differentiation $D : C^1(\mathbb{R}) \to C^0(\mathbb{R})$, as $C^1(\mathbb{R})$ is dense in $C^0(\mathbb{R})$. In this case, $\sigma = \sigma_p = \mathbb{C}$, and the resolvent is $\rho = \emptyset$, because if $D_\lambda = D - \lambda I$, suppose $D_\lambda u = 0$ but $u \neq 0$. This is equivalent to the differential equation $u' = \lambda u$, so that $u(t) = Ce^{\lambda t}$ is an eigenfunction, and λ is an eigenvalue.

Now, we'll assume $D(T) = X$, so $T : X \to X$ is a (usually bounded) linear functional.

Lemma 3.1.2. *If X is a Banach space, $T \in B(X,X)$, and $\lambda \in \rho(T)$, then T_λ is surjective.*

This might not be a surprise, but we do need to prove it.

Proof. Since $\lambda \in \rho(T)$, then R_λ must be dense in X. Suppose $R_\lambda \neq X$; then, $S = T_\lambda^{-1} : R_\lambda \to X$ is a bounded linear functional, so we can extend it to all of X, producing a bounded linear operator $\widetilde{S} : X \to X$.

Since R_λ is dense in X, then for any $y \in X$, there's a sequence $y_n \in R_\lambda$ with $y_n \to y$. And since S is bounded, then $\{Sy_n\}$ is still Cauchy. Since X is Banach, we can take the limit, and let $\widetilde{S}(y) = \lim_{n\to\infty} Sy_n$. This was a choice, so we have to check that it's well-defined; what if we chose a different sequence $z_n \to y$, where the $z_n \in R_\lambda$? Then,

$$\lim_{n\to\infty} \|Sz_n - \widetilde{S}(y)\| = \lim_{n\to\infty} \lim_{m\to\infty} \|Sz_n - Sy_m\| \leq \lim_{n\to\infty} \lim_{m\to\infty} \|S\| \|z_n - y_m\| = 0.$$

Thus, $\widetilde{S}$ is well-defined, and by its definition, it's linear, and $\widetilde{S}y = Sy$ if $y \in R_\lambda$. Moreover, $\widetilde{S}$ is sequentially continuous by definition, and so it's bounded.

Now, suppose $y \in Y$, so there's a sequence $y_n \in R_\lambda$ for which $y_n \to y$. Let $x_n = Sy_n = T_\lambda^{-1} y_n$, so that $x_n \to \widetilde{S}y$. Let $x = \widetilde{S}y$; then, $y_n = T_\lambda x_n \to T_\lambda x$, and $y_n \to y$, so $T_\lambda x = y$, and so $R_\lambda = X$. ⊠

Corollary 3.1.3. *If X is a Banach space and $T \in B(X,X)$, then $\lambda \in \rho(T)$ iff T_λ is invertible on all of X.*

We've clearly proven the first direction; the converse follows because, by the open mapping theorem, an invertible, bounded linear map has a continuous inverse. The takeaway is that if you can work with a fully defined function, things can be a little nicer.

Recall the geometric series: if $|r| < 1$, then

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

The easiest way to prove this is to show that $(1-r)(1+r+r^2+\cdots) = 1$, and prove that the terms get smaller. We'll prove a suspiciously similar-looking result in the same way.

Lemma 3.1.4. *Let X be Banach and $V \in B(X,X)$ such that $\|V\| < 1$. Then,*[1]

$$(I-V)^{-1} = \sum_{n=0}^{\infty} V^n. \tag{3.1}$$

(3.1) is called the *Neumann series* for $(I-V)^{-1}$.

PROOF. Let's take partial sums: take $N \in \mathbb{N}$ and $S_N = I + V + \cdots + V^N$, so that $S_N \in B(X,X)$. Then, $\{S_N\}_{N=1}^{\infty}$ turns out to be Cauchy in $B(X,X)$: if $M > N$, then

$$\|S_M - S_N\|_{B(X,X)} = \left\| \sum_{n=N+1}^{M} V^n \right\|.$$

We showed that $\|AB\| \le \|A\|\|B\|$, so

$$\le \sum_{n=N+1}^{M} \|V\|^n,$$

but since $\|V\| < 1$, this can be made as small as you like for M and N sufficiently large.

Since X is Banach, then so is $B(X,X)$, and therefore there exists an $S \in B(X,X)$ such that $S_N \to S$. To show that $S_N = (I-V)^{-1}$, notice that

$$(I-V)S_N = I - V^{N+1} = S_N(I-V),$$

but $\|V^{N+1}\| \le \|V\|^{N+1} \to 0$, so as $N \to \infty$, $(I-V)S_N$ and $S_N(I-V)$ both converge to the identity, and in particular $(I-V)S = S(I-V) = I$. ⊠

There's nothing necessarily magical about 1 in this proof.

Corollary 3.1.5. *Suppose $\lambda \in \mathbb{C}$ and $\|T\|_{B(X,X)} < |\lambda|$. Then, $\lambda \in \rho(T)$ and*

$$T_\lambda^{-1} = -\frac{1}{\lambda} \sum_{n=0}^{\infty} \left(\frac{1}{\lambda} T \right)^n.$$

PROOF. Let $V = T_\lambda = -\lambda(I - (1/\lambda)T)$, so that $\|V\| < 1$; then, apply Lemma 3.1.4. ⊠

We haven't related this to spectral theory yet, and the next corollary won't either, but it's still very important.

Definition. If X is a Banach space, we define the *general linear group* $\mathrm{GL}(X) \subset B(X,X)$ to be the set of bounded linear invertible operators $X \to X$.

This does in fact have a group structure under composition.

Corollary 3.1.6. $\mathrm{GL}(X)$ *is open in* $B(X,X)$*.*

Intuitively, anything sufficiently close to an invertible operator is still invertible.

PROOF. Let $A \in \mathrm{GL}(X)$, so that A and A^{-1} are both in $B(X,X)$. Choose an $\varepsilon > 0$ such that $\varepsilon \le 1/\|A^{-1}\|$. Choose a $B \in B(X,X)$ such that $\|B\| < \varepsilon$; then, we want to show that $A+B$ is invertible (this shows that $B_\varepsilon(A) \subset \mathrm{GL}(X)$, so this is sufficient).

Then, $A + B = A(I + A^{-1}B)$, but $\|A^{-1}B\| \le \|A^{-1}\|\|B\| < \varepsilon\|A^{-1}\| < 1$. In particular, it's small enough that we can use Neumann series to get that $I + A^{-1}B$ has an inverse; thus, $A+B$ is the product of two invertible operators (A and $(I + A^{-1}B)$), so it's invertible too, and in fact, $(A+B)^{-1} = (I + A^{-1}B)^{-1}A^{-1} = A^{-1} + B^{-1}$. ⊠

Now let's say something about spectral theory.

Corollary 3.1.7. *If X is a Banach space and $T \in B(X,X)$, then $\rho(T) \subset \mathbb{C}$ is open, and $\sigma(T)$ is compact; specifically, if $\lambda \in \sigma(T)$, then $|\lambda| \le \|T\|$.*

PROOF. First, if $\lambda \in \rho(T)$, then $T - \lambda I \in \mathrm{GL}(X)$, but since $\mathrm{GL}(X)$ is open, then $T - \lambda I + B$ is still invertible, if $\|B\|$ is small, e.g. $B = -\mu I$ for $|\mu|$ sufficiently small. Thus, $T - (\lambda + \mu)I$ is still invertible, so $\lambda + \mu \in \rho(T)$, and thus $\rho(T)$ is open.

By Corollary 3.1.5, we know that if $\lambda \in \sigma(T)$, then $|\lambda| \le \|T\|$; then, since $\rho(T)$ is open, $\sigma(T)$ is closed, and we just saw that it's bounded, so it's compact. ⊠

[1]Note that $I - V \in B(X,X)$ and is bijective by the open mapping theorem.

Spectral theory is useful for lots of things, but it's particularly useful for some nicely behaved operators, called compact operators.

Definition. Let X and Y be NLSes and $T : X \to Y$. Then, T is a *compact* linear operator, sometimes called a *completely continuous* linear operator, if T is linear and if whenever $M \subset X$ is bounded, then $\overline{T(M)} \subset Y$ is compact.

Intuitively, T must send bounded sets to precompact sets, much like a bounded operator takes bounded sets to bounded sets. And we like compact operators because we can control their range.

Proposition 3.1.8. *Let $T : X \to Y$ be a compact linear operator; then, T is bounded.*

This is the reason behind the alternate name "completely continuous."

Lecture 24: 10/21/15

Compact Operators.

If X and Y are NLSes and $T : X \to Y$ is linear, recall that a bounded operator sends bounded sets to bounded sets, and a compact operator sends bounded sets to precompact sets (sets with compact closure). Thus, compact operators are bounded, because (pre)compact sets have to be bounded.

Definition. Let $C(X,Y) \subset B(X,Y)$ denote the set of compact operators $T : X \to Y$. It's quick to check that sums and scalar multiples of compact operators are compact, and thus $C(X,Y)$ is a subspace.

First, we'll need the following lemma from general topology.

Lemma 3.2.1. *Let (X,d) be a metric space; then, X is compact iff every sequence in X has a convergent subsequence.*[2]

It allows us to prove a useful criterion for compact operators.

Lemma 3.2.2. *A linear operator $T : X \to Y$ is compact iff T maps every bounded sequence to a sequence with a convergent subsequence.*

Proof. The forward direction is trivial: if T is compact, it maps bounded sets (e.g. $\{x_n\}$) to precompact ones (meaning $\{Tx_n\}$ has a convergent subsequence, using Lemma 3.2.1).

The other direction requires more work. Let $B \subset X$ be bounded and consider $\overline{T(B)}$; let $\{y_n\}_{n=1}^\infty \subseteq \overline{T(B)}$. The interesting part is when these points aren't in $T(B)$, so if $y_n \in \partial T(B) \setminus T(B)$, then choose a sequence $\{y_{n,m}\}_{m=1}^\infty \subset T(B)$ such that $\|y_{n,m} - y_n\| \le 1/m$ (and therefore $y_{n,m} \to y_n$ as $m \to \infty$). If instead $y_n \in T(B)$, let $y_{n,m} = y_n$ for all m.

Now, take the diagonal subsequence $\{y_{n,n}\} \subset T(B)$, so that there exist $x_n \in B$ such that $Tx_n = y_{n,n}$, and in particular $\{x_n\}$ is bounded. Thus, by hypothesis, $\{y_{n,n}\}$ has a convergent subsequence: there's a sequence $\{n_k\}$ so that y_{n_k,n_k} converges to some $y \in \overline{T(B)}$. But $y_{n_k} \to y$ as well, because

$$\begin{aligned}\|y_{n_k} - y\| &\le \|y_{n_k} - y_{n_k,n_k}\| + \|y_{n_k,n_k} - y\| \\ &\le \frac{1}{n_k} + \|y_{n_k,n_k} - y\| \longrightarrow 0.\end{aligned}$$

Since $\{y_n\}$ was arbitrary, this means every bounded sequence in $\overline{T(B)}$ has a convergent subsequence, and therefore $\overline{T(B)}$ is compact. ⊠

Now let's look at examples.

Proposition 3.2.3. *Let $T : X \to Y$ be a linear operator.*

(1) If X is finite-dimensional, then T is compact.
(2) If T is bounded and Y is finite-dimensional, then T is compact.
(3) If X is infinite-dimensional, then the identity $I : X \to X$ is not *compact.*

[2] This property, called *sequential compactness*, is not equivalent to compactness in more general topological spaces.

PROOF. For (1), the range of T is finite-dimensional, so by the Bolzano-Weierstrass theorem, any closed and bounded set is compact, and in particular, any bounded set is precompact. $T : X \to R(T)$ is a linear map of finite-dimensional spaces, so it must be bounded. Thus, the image of any bounded set is bounded in $R(T)$, and therefore precompact in $R(T)$, and therefore precompact in Y, and so T is a compact operator.

(2) is true because the range is necessarily finite-dimensional, so once again Bolzano-Weierstrass tells us that every bounded set is precompact. Thus, since T is bounded, the image of any bounded set is bounded, and therefore precompact, so T is a compact operator.

For (3), the image of the unit ball (which is bounded) is the unit ball, which we've seen is not compact. ⊠

So there are nontrivial examples (and nonexamples) of compact operators, which is nice, I guess.

Theorem 3.2.4. *If Y is Banach, $C(X,Y) \subset B(X,Y)$ is a closed subspace.*

PROOF. We want to show that if $\{T_n\}_{n=1}^\infty \subseteq C(X,Y)$ converges to some $T \in B(X,Y)$, then T is actually compact. We'll have to use a diagonalization argument again.

Let $\{x_n\}_{n=1}^\infty \subseteq X$ be bounded, so that, since T_1 is compact, there is a subsequence $\{x_{1,n}\} \subseteq \{x_n\}$ such that $\{T_1 x_{1,n}\}_{n=1}^\infty$ converges. Then, we can play the same game with T_2 and $\{x_{1,n}\}$, producing a subsequence $\{x_{2,n}\} \subseteq \{x_{1,n}\}$. Doing this again and again, we obtain a convergent sequence $\{T_j x_{j,n}\}$ such that $\{x_{j,n}\} \subseteq \{x_{j-1,n}\}$ for all j.

Let $\widetilde{x}_n = x_{n,n}$. Then, for all $n \ge 1$, $\{T_n \widetilde{x}_m\}_{m=1}^\infty$ converges (because we know it does when $m \ge n$).

We want to show that $\{T\widetilde{x}_n\}_{n=1}^\infty$ is Cauchy, which suffices to prove the theorem (since Y is Banach, it converges to something, and then Lemma 3.2.2 finishes the proof). Let $\varepsilon > 0$, so that there's an $N_0 \in \mathbb{N}$ for which $\|T_N - T\| < \varepsilon$ when $N \ge N_0$. Since $\{x_n\}$ is bounded, let M be an upper bound for it, so that

$$\begin{aligned} \|T\widetilde{x}_n - T\widetilde{x}_m\| &\le \|T\widetilde{x}_n - T_N\widetilde{x}_n\| + \|T_N(\widetilde{x}_n - \widetilde{x}_m)\| + \|T_N\widetilde{x}_m - T\widetilde{x}_m\| \\ &\le \|T - T_N\|(\|\widetilde{x}_n\| + \|\widetilde{x}_m\|) + \|T_n(\widetilde{x}_n - \widetilde{x}_m)\| \\ &\le 2M\varepsilon + \|T_n(\widetilde{x}_n - \widetilde{x}_m)\|, \end{aligned}$$

and since T_n is compact, this goes to zero. ⊠

This is surprisingly useful; one great way to prove an operator is compact is to show it's a limit of some other compact operators.

Example 3.2.5. Let $X = \ell^2$ and $T(x_1, x_2, \dots) = (x_1, x_2/2, x_3/3, \dots)$. Thus, if

$$T_n(x_1, x_2, \dots) = \left(x_1, \frac{x_2}{2}, \dots, \frac{x_n}{n}, 0, 0, \dots\right),$$

then T_n has finite-dimensional image and is bounded, so T_n is compact. Then,

$$\begin{aligned} \|T_n - T\|^2 &= \sup_{\|x\|=1} \|T_n x - Tx\|^2 \\ &= \sup_{\|x\|=1} \sum_{j=n+1}^\infty \frac{1}{j^2}|x_j|^2 \\ &\le \sup_{\|x\|=1} \frac{1}{(n+1)^2} \sum_{j=n+1}^\infty |x_j|^2 = \frac{1}{(n+1)^2} \to 0. \end{aligned}$$

Thus, by Theorem 3.2.4, T is compact, and it's a nice, nontrivial example.

Notice again that compact operators are "small" in some sense.

Theorem 3.2.6. *Let X and Y be NLSes and $T \in C(X,Y)$. If $x_n \rightharpoonup x \in X$, then $Tx_n \to Tx$.*

In other words, compact operators convert weak convergence into strong convergence!

PROOF. Let $y_n = Tx_n$ and $y = Tx$; we'll show first that, $y_n \rightharpoonup y$. For any $g \in Y^*$, let $f : X \to \mathbb{F}$ be given by $f = g \circ T$, so $f \in X^*$ and $f(x_n) \to f(x)$ (since $x_n \rightharpoonup x$), but $f(x_n) = g(y_n)$ and $f(x) = g(y)$, so $y_n \rightharpoonup y$.

Now, we know that $\{y_n\}_{n=1}^\infty$ is bounded, so since T is compact, then there's a subsequence $\{y_{n_k}\}$ such that Ty_{n_k} converges to some $\widetilde{y} \in Y$, and so $Ty_{n_k} \rightharpoonup \widetilde{y}$ as well. We also know that $Tx_n \rightharpoonup Tx$, and since the weak topology is Hausdorff, then limits are unique, so $Tx = \widetilde{y}$, and thus $Tx_{n_k} \to Tx$.

Okay, but what about the whole sequence? if $Tx_n \not\to Tx$, then there must be some $\varepsilon > 0$ and a subsequence $y_{n_j} = Tx_{n_j}$ such that $\|y_{n_j} - y\| \geq \varepsilon$, so we can run the whole argument again with $\{x_{n_j}\}$, which is a weakly convergent sequence, and therefore has a strongly convergent subsequence, which is a contradiction. Thus, no such subsequence x_{n_j} exists. ⊠

This is an example of a nice general principle about the weak and strong topologies: if $x_n \rightharpoonup x$ and $x_n \to y$, then $x = y$.

Now, we would like to relate this back to spectral theory. The takeaway is that the spectrum of a compact operator is particularly simple.

Proposition 3.2.7. *Let X be an NLS and $T \in C(X,X)$. Then, $\sigma_p(T)$ is countable, and if infinite, it accumulates at 0 and only at 0. If X is infinite-dimensional, then $0 \in \sigma(T)$.*

Corollary 3.2.8. *If X is an infinite-dimensional space, the eigenvalues (i.e. point spectrum) of a compact operator $T \in C(X,X)$ can be ordered by absolute value $|\lambda_1| \geq |\lambda_2| \geq \cdots$, and $\lambda_n \to 0$.*

Partial proof of Proposition 3.2.7. Suppose X is infinite-dimensional and T is compact, but $0 \in \rho(T)$. Thus, T is boundedly invertible on a dense subset of X. Let $B = R(T) \cap B_1(0)$; since $R(T)$ is dense in X, then B is dense in $B_1(0)$. Since T^{-1} is bounded, then $T^{-1}(B)$ is a bounded set, so since T is compact, then $T(T^{-1}(B)) = B$ has compact closure. But this is a contradiction, since $\overline{B} = \overline{B_1(0)}$ is noncompact, since X is infinite-dimensional, so $0 \in \sigma(T)$.[3]

For the other half of the theorem, it suffices to prove that $\sigma_p(T) \cap \{\lambda : |\lambda| \geq r\}$ is finite, which we'll show next time.

Lecture 25: 10/23/15

Spectra of Compact Operators.

Recall that we're talking about compact operators, building up to the spectral theorem for compact operators. To be precise, we're in the middle of proving Proposition 3.2.7, addressing the accumulation point (it'll be unique) of the spectrum of a compact operator.

Continuation of Proof of Proposition 3.2.7. We showed that it suffices to show that $\sigma_p(T) \cap \{\lambda : |\lambda| \geq r\}$ is finite for any $r > 0$. We'll argue by contradiction, eventually showing that our operator isn't compact if this isn't true.

Suppose there is an $r > 0$ and a sequence $\{\lambda_n\}_{n=1}^{\infty}$ of distinct eigenvalues of T with $|\lambda_n| > r$, and let x_i be an eigenvector corresponding to λ_i.

In finite dimensions, it would be obvious that $\{x_i\}$ is linearly independent, but, Toto, I've a feeling we're not in finite dimensions anymore. So suppose they're linearly dependent; then, there exists an $N > 0$ and some α_j for $j = 1, \ldots, N$ not all 0 such that

$$\sum_{j=1}^{N} \alpha_j x_j = 0. \tag{3.2}$$

Take N minimal with this property; now, the standard proof in finite dimensions applies (oh, there's no place like home!).

$$0 = T_{\lambda_N}\left(\sum_{j=1}^{N} \alpha_j x_j\right) = \sum_{j=1}^{N} \alpha_j(\lambda_j - \lambda_N)x_j,$$

but we know $\lambda_j - \lambda_N \neq 0$ if $j < N$, so since N is the minimal number for which (3.2) is true, then $\alpha_j = 0$ for all $j \leq N-1$. Thus, $\alpha_N \neq 0$, but $\alpha_N x_N = 0$, which is a contradiction.

Thus, eigenvectors of distinct eigenvalues are linearly independent. That seems useful outside of just this proof. Anyways, define $M_n = \operatorname{span}\{x_1, \ldots, x_n\}$, so if $x \in M_n$, we can write $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$. Thus,

$$Tx = \sum_{j=1}^{n} (\alpha_j \lambda_j) x_j,$$

[3] The professor said this part of the proof was clear. Is there a simpler argument that I am missing?

so $T : M_n \to M_n$. Moreover, our resolvent operator T_{λ_n} satisfies

$$T_{\lambda_n} x = \sum_{j=1}^{n-1} \alpha_j(\lambda_j - \lambda_n)x_j,$$

so $T_{\lambda_n} : M_n \to M_{n-1}$.

Let $z_1 = x_1/\|x_1\|$ and for $n > 1$, let $y \in M_n \setminus M_{n-1}$, which is nonempty because $\{x_1, \dots, x_n\}$ is linearly independent; since M_{n-1} is finite-dimensional and therefore closed, then $d = \operatorname{dist}(y, M_{n-1}) > 0$. Thus, there exists a $y_0 \in M_{n-1}$ such that $d \le \|y - y_0\| \le 2d$, so if $z_n = (y - y_0)/\|y - y_0\|$, then $\|z_n\| = 1$.

If $w \in M_{n-1}$, then

$$\begin{aligned} \|z_n - w\| &= \frac{1}{\|y - y_0\|}\|y - \underbrace{y_0 - \|y - y_0\|w}_{\in M_{n-1}}\| \\ &\ge \frac{1}{\|y - y_0\|} d \ge \frac{1}{2}. \end{aligned}$$

Thus, we have a sequence $\{z_n\}_{n=1}^\infty$ such that $\|z_n - w\| \ge 1/2$ for any $w \in M_{n-1}$, $z_n \in M_n$, and $\|z_n\| = 1$.

If $n > m$, define $\widetilde{x} = Tz_n - \lambda_n z_n - Tz_m = T_{\lambda_n} z_n - Tz_m$, so

$$Tz_n - Tz_m = \lambda_n z_n + Tz_n - \lambda_n z_n - Tz_m = \lambda_n z_n - \widetilde{x},$$

and therefore $\widetilde{x} \in M_{n-1}$. Then,

$$\|Tz_n - Tz_m\| = |\lambda_n| \left\| z_n - \frac{\widetilde{x}}{|\lambda_n|} \right\| \ge \frac{r}{2} > 0,$$

so $\{Tz_n\}$ has no convergent subsequence, meaning T isn't compact. Oh dear; we've reached a contradiction. ⊠

Example 3.3.1. Let's look again at Example 3.2.5 again: $X = \ell^2$ and $T(x_1, x_2, \dots) = (x_1, x_2/2, x_3/3, \dots)$, which is compact, as we showed last lecture, and its spectrum is $\sigma_p(T) = \{1/n : n \in \mathbb{N}\}$.

What about 0? T is injective, and has dense range: let $y \in \ell^2$, so that for any $\varepsilon > 0$ there's an N such that

$$\sum_{j=N+1}^\infty |y_i|^2 < \varepsilon.$$

It's easy to write down an inverse operator $T^{-1}(y_1, y_2, \dots) = (y_1, 2y_2, 3y_3, \dots)$, If $x = (y_1, 2y_2, \dots, Ny_N, 0, 0, \dots)$, then $\|Tx - y\| < \varepsilon$; thus, the image is dense.

However, T^{-1} isn't bounded, and so $0 \in \sigma_c(T)$ rather than the point spectrum.

The next proposition tells us a little more about compact operators.

Proposition 3.3.2. *Let X be an NLS and $T \in C(X, X)$. If $\lambda \ne 0$, then $N(T_\lambda)$ is finite-dimensional.*

Proof. If $\lambda \ne \sigma_p(T)$, then $\dim N(T_\lambda) = 0$ and we're done, so suppose that $\lambda \in \sigma_p(T)$.

Let $B = \overline{B_1(0)} \cap N(T_\lambda)$. Recall that the closed unit ball is compact iff X is finite-dimensional; we will end up using this fact. If $\{x_n\}_{n=1}^\infty \subseteq B$, then since T is compact, then there's a subsequence $\{x_{n_k}\}_{k=1}^\infty$ converging to some $z \in X$. But x_{n_k} is in the λ-eigenspace for T, so $Tx_{n_k} = \lambda x_{n_k}$, and thus $x_{n_k} \to z/\lambda$. Thus, B is compact.[4] But since the unit ball in the normed space $N(T_\lambda)$ is compact iff $N(T_\lambda)$ is finite-dimensional, then $N(T_\lambda)$ must be finite-dimensional. ⊠

Our next goal is to prove that all spectral values of compact operators are eigenvalues (except possibly 0, as Example 3.3.1 just showed us). To do this, we need an ancillary result which is more useful than it looks.

Proposition 3.3.3. *Let X be a Banach space and $T \in C(X, X)$. If $\lambda \ne 0$, then $R(T_\lambda)$ is closed, and so is $R(T_\lambda^n)$.*

Partial proof. It takes a while to set things up for this proof, but once all the actors are onstage, the proof is relatively simple.

Intuitively, we want to remove $N = N(T_\lambda)$, which is finite-dimensional by Proposition 3.3.2. Thus, N is closed, and so there's a subspace $M \subseteq X$ such that $X = M \oplus N$. Note that this is *not* an orthogonal complement, since we don't have an inner product.

[4] To be precise, it's sequentially compact; the notions are identical in metric spaces, cf. Lemma 3.2.1.

Let's do this more precisely. Since N is finite-dimensional, choose a basis $\{e_1, \dots, e_n\}$ for N, so for any $x \in N$, we can write

$$x = \sum_{j=1}^{n} \alpha_j(x) e_j.$$

But it's easy to see that these $\alpha_j : N \to \mathbb{F}$ must be linear functions, and since N is finite-dimensional, they're continuous. So by the Hahn-Banach theorem, we can extend them to X. The intersection of finitely many vector spaces is still a vector space, and the intersection of finitely many closed sets is closed, so

$$M = \bigcap_{j=1}^{n} N(\alpha_j)$$

is a closed vector space![5]

Now, we will show that $X = M \oplus N$. First, if $x \in M \cap N$, then $\alpha_j(x) = 0$ for all j, and therefore $x = 0$ (since $e_1, \dots, e_n$ are linearly independent). Next, given an $x \in X$, let

$$y = \sum_{j=1}^{n} \alpha_j(x) e_j,$$

so $y \in N$, and let $z = x - y$. It suffices to show that $z \in M$, but for any $j = 1, \dots, n$, $\alpha_j(x - y) = \alpha_j(x) - \alpha_j(x) = 0$, so $z \in M$, so $X = M \oplus N$ as desired.

Now let's actually prove the result. Once again, we look at the resolvent operator $T_\lambda : M \to X$. We'll first show that T_λ is "bounded below," i.e. there exists a $\gamma > 0$ such that $\gamma\|g\| \le \|T_\lambda x\|$ when $x \neq 0$, which implies T_λ is one-to-one. Well, if this isn't true, then there is some sequence $\{x_n\}_{n=1}^{\infty} \subseteq M$ with $\|x_n\| = 1$ and $T_\lambda x_n \to 0$; then, because T is compact, there's a subsequence $\{x_{n_k}\}$ such that $T x_{n_k} \to x$ for some $x \in X$, and since $T_\lambda x_{n_k} \to 0$, then $\lambda x_{n_k} \to x$, so that $x_{n_k} \to x/\lambda$, and thus $x \in M$. This means $T_\lambda x_{n_k} \to (1/\lambda) T_\lambda x$ and $T_\lambda x_{n_k} \to 0$, so $T_\lambda x = 0$.

We'll show that T_λ is one-to-one so $x = 0$, and that will give us our contradiction, because $\|x_{n_k}\| = 1$ for each k. Then, we'll finish the proof next lecture.

Lecture 26: 10/26/15

The Spectral Theorem for Compact Operators.

Last time, we were mired in the proof of Proposition 3.3.3, which is perhaps a small step for us, but a great leap on the way to the spectral theorem.

Continuation of Proof of Proposition 3.3.3. We had set up $N = N(T_\lambda)$, which is finite-dimensional, spanned by $\{e_1, \dots, e_n\}$. If $x \in N$, then $x = \sum \alpha_j(x) e_j$, giving us functions $\alpha_j : N \to \mathbb{C}$ that the Hahn-Banach theorem allows us to extend to functions $X \to \mathbb{C}$. Thus, if M is the intersections of the kernels of these α_j, then $X = M \oplus N$ and $T_\lambda : M \to X$ is one-to-one. Lastly, we proved that there's a $\gamma \in \mathbb{R}$ such that $\gamma\|x\| \le \|T_\lambda x\|$ for all nonzero $x \in M$.

Let us suppose that $y_n \in R(T_\lambda)$ and $y_n \to y \in X$. Then, since T_λ is injective, there exist $x_n \in M$ such that $T_\lambda x_n = y_n$. Since $\{y_n\}$ is Cauchy, so is $\{x_n\}$, and therefore $x_n \to x \in M$. Thus, $T_\lambda x_n$ converges to both y and $T_\lambda x$, so $y = T_\lambda x$ and therefore $y \in R(T_\lambda)$, so the range is closed.

For the last part, we use the binomial theorem:[6]

$$T_\lambda^n = \sum_{k=1}^{n} \frac{n!}{k!(n-k)!} (-\lambda I)^{n-k} T^k + (-\lambda)^n I = \underbrace{\sum_{k=1}^{n} \frac{n!}{k!(n-k)!} (-\lambda)^{n-k} T^k}_{S} + (-\lambda)^n I.$$

Since compact sets are bounded, then if B is a bounded set, then $T(B)$ is a compact set, which is bounded, so $T^2(B)$ is compact, and therefore bounded, so $T^3(B)$ is compact, and thus bounded, and so… thus, T^k maps bounded sets to compact sets for each $k > 0$, so $T^k \in C(X, X)$. In particular, this means S is a linear combination of compact operators, and since $C(X, X)$ is a vector space, then S is also compact. In particular, $T_\lambda^n = S_{(-\lambda)^n}$, and $(-\lambda)^n \neq 0$, and by the part of this theorem that we already proved, $S_{(-\lambda)^n}$ has closed image, so T_λ^n does too. ⊠

[5] Alternatively, each $N(\alpha_j)$ is the preimage of 0 under a continuous map, and then finite intersections of closed sets are closed.

[6] About which we are teeming with a lot o' news, of course.

The proof is predicated on the decomposition $X = M \oplus N$: the null space of T_λ doesn't help us, and in fact restricts convergence, so we threw it out. This is a common approach.

Theorem 3.4.1. *Let X be a Banach space and $T \in C(X,X)$. If $\lambda \in \sigma(T)$ and $\lambda \neq 0$, then $\lambda \in \sigma_p(T)$.*

So the spectral theory of compact operators is pretty well-behaved.

Proof. Let $\lambda \in \sigma(T) \setminus \sigma_p(T)$ be nonzero; in particular, T_λ is one-to-one and $R(T_\lambda) \neq X$.

Let's consider the sequence of subspaces $X \supsetneq R(T_\lambda) \supseteq R(T_\lambda^2) \supseteq R(T_\lambda^3) \cdots$. It turns out this sequence must stabilize at some point: if not, then there exist x_n for $n \in \mathbb{N}$ such that $x_n \in R(T_\lambda^n)$, $\|x_n\| = 1$, and $\operatorname{dist}(x_n, R(T_\lambda^{n+1})) \geq 1/2$, by the same argument as in the proof of Proposition 3.3.3. Be aware of this argument: we have used it thrice now.

If $n > m$, let $\widetilde{x} = \lambda x_n + T_\lambda x_n - T_\lambda x_m$, so that

$$Tx_m - Tx_n = T_\lambda x_m - T_\lambda x_n + \lambda(x_m - x_n) = \lambda x_m - \widetilde{x}.$$

In particular, $\widetilde{x} \in R(T_\lambda^{m+1})$, and so too is $(1/\lambda)\widetilde{x}$, so

$$\|\lambda x_m - \widetilde{x}\| = |\lambda| \left\| x_m - \frac{1}{\lambda}\widetilde{x} \right\| \geq \frac{|\lambda|}{2}.$$

So $\{x_n\}$ is a bounded sequence such that $\{Tx_n\}$ has no convergent subsequence (since the distance between Tx_m and Tx_n is approximately $\lambda\|x_n - x\|$), which is a contradiction, since T is compact. Thus, $R(T_\lambda^n)$ has to stabilize; in particular, let n be such that $R(T_\lambda^n) = R(T_\lambda^{n+1})$.

Let $y \in X \setminus R(T_\lambda)$, so that $T_\lambda^n y \in R(T_\lambda^n) = R(T_\lambda^{n+1})$. Thus, there's an $x \in X$ such that $T_\lambda^n y = T_\lambda^{n+1} x$, i.e. $T_\lambda^n(T_\lambda x - y) = 0$. But since T_λ is one-to-one, then so is T_λ^n, and therefore $T_\lambda x = y$, which is a contradiction (we assumed y wasn't in the range). ⊠

Summarizing these results, we have the anticipated spectral theorem.

Theorem 3.4.2 (Spectral theorem for compact operators)**.** *Let T be a compact operator on a Banach space X.*

(1) *The spectrum of T consists of at most countably many eigenvalues.*
(2) *If $\lambda \in \sigma(T)$ is nonzero, then its eigenspace $N(T_\lambda)$ is finite-dimensional.*
(3) *If X is infinite-dimensional, then $0 \in \sigma(T)$.*
(4) *If T has infinitely many eigenvalues, then they converge to 0.*

If you look at the proofs of these results, then you'll notice that a lot of the arguments feel finite-dimensional. The takeaway lesson is that compact operators are "nearly finite-dimensional."

Corollary 3.4.3 (Fredholm alternative[7] for compact operators)**.** *Let X be a Banach space, $\lambda \in \mathbb{C}$ be nonzero, and $T : X \to X$ be compact. If $A = I - (1/\lambda)T$, then exactly one of the following statements is true.*

(1) *For any $y \in X$, there is a unique $x \in X$ such that $Ax = y$.*
(2) *If $y \in X$ is such that $Ax = y$ has a solution, then it has infinitely many solutions.*

Proof. The first one is true iff $\lambda \in \rho(T)$, and otherwise, $\lambda \in \sigma_p(T)$, so since $\lambda \neq 0$, then $N(A) = N(T_\lambda)$ is nonzero. ⊠

Bounded Self-Adjoint Operators. In order to talk about self-adjoint operators, we really need the Riesz representation theorem, so let H be a Hilbert space. The Riesz theorem defines for us an isometric isomorphism $R : H \to H^*$ sending $y \mapsto R_y$, the function $R_y(x) = \langle y, x \rangle$ for all $x \in H$. Thus, if $T : H \to H$, we may think of T^*, which is *a priori* a map $H^* \to H^*$, as a map $H \to H$ as well; if $x, y \in H$, then $(T^* R_y)(x) = R_y(Tx) = \langle Tx, y \rangle$. But the Riesz map is invertible, so $T^* Ry = Rz$ for some $z \in H$, and in particular $R^{-1} T^* Ry = z$. That is,

$$\langle Tx, y \rangle = (T^* R_y)(x) = \langle x, R^{-1} T^* Ry \rangle. \tag{3.3}$$

Definition. The operator $R^{-1} T^* R : H \to H$ is called the *Hilbert adjoint* of T.

In the future, we won't distinguish the Hilbert adjoint (a map $H \to H$) and the regular adjoint (mapping $H^* \to H^*$), since they are identified by way of the Riesz representation theorem. Thus, we may rephrase (3.3) as

$$\langle x, T^* y \rangle = \langle Tx, y \rangle$$

for all $x, y \in H$. In linear algebra, we have already seen this with matrix adjoints over finite-dimensional spaces.

[7] "Fredholm Alternative" would be a really good name for a rock band.

Proposition 3.4.4. *The Hilbert adjoint T^* is a bounded linear operator, and $T^{**} = T$. In particular, it's also true that $\langle T^*x, y\rangle = \langle x, Ty\rangle$ for $x, y \in H$.*

PROOF. These are just a little thinking: taking the Riesz operator introduces a complex conjugation, but then we do it again, so this cancels out, so T^* is linear, rather than conjugate linear. Then, taking the conjugate of $\langle Tx, y\rangle = \langle x, T^*y\rangle$, we get $\langle y, Tx\rangle = \langle T^*y, x\rangle$, and then can flip (which requires taking a conjugate, yes, but we can unconjugate). Thus, for all $x, y \in H$, $\langle x, Ty\rangle = \langle x, T^{**}y\rangle$, which by the Hahn-Banach theorem, implies that $T = T^{**}$. ⊠

Definition. If $T = T^*$, we call T a *self-adjoint* operator.

In terms of inner products, this means $\langle x, Ty\rangle = \langle Tx, y\rangle$ for all $x, y \in H$.

Self-adjoint operators are analogous to symmetric matrices. Remember that symmetric matrices have nicer sets of eigenvalues (e.g. they're always diagonalizable)? Something similar is true of self-adjoint operators.

Theorem 3.4.5. *Let H be a Hilbert space over either $\mathbb{R}$ or $\mathbb{C}$ and $T : H \to H$ be bounded.*

(1) If T is self-adjoint, $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in H$.
(2) If $\mathbb{F} = \mathbb{C}$, then the converse is true: $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in H$ implies T is self-adjoint.

PROOF. For (1), $\langle Tx, x\rangle = \overline{\langle x, Tx\rangle} = \overline{\langle x, T^*x\rangle} = \overline{\langle Tx, x\rangle}$, so $\langle Tx, x\rangle \in \mathbb{R}$.

For (2), we know that for any α,

$$\langle T(x + \alpha y), x + \alpha y\rangle = \langle Tx, x\rangle + |\alpha|^2\langle Ty, y\rangle + \overline{\alpha}\langle Ty, x\rangle \in \mathbb{R},$$

so $\overline{\alpha}\langle Tx, y\rangle + \overline{\overline{\alpha}\langle T^*x, y\rangle} \in \mathbb{R}$. Applying this with $\alpha = 1$ and $\alpha = i$, we see that $\langle Tx, y\rangle = \langle T^*x, y\rangle$ for all $x, y \in H$. ⊠

The TA, Sam Krupa, will give the next two lectures, since the professor will be out of town.

Lecture 27: 10/28/15

The Spectral Theorem for Self-Adjoint Operators.

Today Sam Krupa gave the lecture.

First, recall Theorem 3.4.5: that if T is self-adjoint, then (Tx, x) is real for all $x \in H$, and if H is over $\mathbb{C}$, then the converse is true.

We're going to need the following result at least six times in the next few proofs.

Lemma 3.5.1. *Let X and Y be Banach spaces and $T \in B(X, Y)$ be* bounded below, *i.e. there's a $\gamma > 0$ such that for all $x \in X$, $\|x\|_X \cdot \gamma \le \|Tx\|_Y$. Then, T is injective and $R(T)$ is closed in Y.*

PROOF. Injectivity is obvious: if $Tx = 0$, then $\|x\| \le 0$, so $x = 0$.

To show that $R(T)$ is closed, suppose we have a sequence $\{y_n\}_{n=1}^\infty$ in $R(T)$ that converges to some $y \in Y$. In particular, $\{y_n\}$ is Cauchy, and since $y_n \in R(T)$, then $y_n = Tx_n$ for some $x_n \in X$. We'll show $\{x_n\}$ is Cauchy too.

If you hand me an $\varepsilon > 0$, then there's an N such that if $m, n \ge N$, then $\|y_n - y_m\| < \varepsilon$, i.e. $\|Tx_n - Tx_m\| < \varepsilon$. Since T is bounded below, $\|x_m - x_n\| \le (1/\gamma)\|Tx_m - Tx_n\| < \varepsilon/\gamma$. Since X is Banach, then $\{x_n\}$ converges to some $x \in X$, and T is continuous, then $Tx_n \to Tx$ and $Tx_n = y_n \to y$, so $y = Tx$. Thus, $y \in R(T)$. ⊠

And now, the moment we've all been waiting for.

Theorem 3.5.2 (Spectral theorem for self-adjoint operators, I)**.** *Let H be a Hilbert space and $T \in B(H, H)$ be self-adjoint. Then,*

(1) $\sigma_r(T) = \emptyset$,
(2) $\sigma(T) \subset [r, R] \subset \mathbb{R}$, where $r = \inf_{\|x\|=1}(Tx, x)$ and $R = \sup_{\|x\|=1}(Tx, x)$, and
(3) $\lambda \in \rho(T)$ iff T_λ is bounded below.

PROOF. First, we want to show that $\sigma_p(T) \subset \mathbb{R}$. If $\lambda \in \sigma_p(T)$, so λ is just an eigenvalue, then there's an $x \in H$ such that $Tx = \lambda x$. In particular, because T is self-adjoint,

$$\lambda(x, x) = (\lambda x, x) = (Tx, x) = (x, Tx) = (x, \lambda x) = \overline{\lambda}(x, x).$$

Since $x \neq 0$, then $(x, x) \neq 0$, so $\lambda = \overline{\lambda}$, and thus $\lambda \in \mathbb{R}$.

To prove (3), we'll make a similar argument as on your homework. Suppose $\lambda \in \rho(T)$; then, for all $x \in H$, $\|x\| = \|T_\lambda^{-1} T_\lambda x\| \le \|T_\lambda^{-1}\| \|T_\lambda x\|$, so let $\gamma = 1/\|T_\lambda\|$, making T_λ bounded below.

Conversely, suppose T_λ is bounded below. Then, by Lemma 3.5.1, T_λ is one-to-one and $R(T_\lambda)$ is closed, so to show $\lambda \in \rho(T)$, it suffices to show $R(T_\lambda)$ is dense in H. Since it's closed, we can show that it's all of H: if not, then (using a result from the homework) there's a nonzero $x_0 \in R(T_\lambda)^\perp$. Hence, for all $x \in H$,

$$\begin{aligned} 0 &= (T_\lambda x, x_0) = (Tx - \lambda x, x_0) \\ &= (x, Tx_0) - (x, \overline{\lambda} x_0) \\ &= (x, T_{\overline{\lambda}} x_0), \end{aligned}$$

since T is self-adjoint. This means $\overline{\lambda} \in \sigma_p(T)$, and in particular it's a real number, so $\overline{\lambda} = \lambda$. But we assumed λ was in the resolvent, so this is a contradiction! Thus, T_λ must have dense image, so λ is actually in the resolvent.

Now, let's address (2). Suppose $\lambda = \alpha + i\beta \in \sigma(T)$; first, we'd like to show that $\beta = 0$. Since $(T_\lambda x, x) = (Tx, x) - \lambda(x, x)$ and $\overline{(T_\lambda x, x)} = (Tx, x) - \overline{\lambda}(x, x)$, then their difference is $(T_\lambda x, x) - \overline{(T_\lambda x, x)} = -2i\beta(x, x)$, and in particular

$$|\beta| \|x\|^2 = \frac{1}{2} \left| (Tx, x) - \overline{(T_\lambda x, x)} \right| \le |(Tx, x)| \le \|T_\lambda x\| \|x\|.$$

If $x \ne 0$, we divide by $\|x\|$, showing $|\beta| \le \|T_\lambda x\| / \|x\|$. If β is nonzero, though, then (3) implies $\lambda \in \rho(T)$, but that would be a contradiction, so $\beta = 0$, and therefore $\sigma(T) \subset \mathbb{R}$.

Next, we'll show (1). Suppose $\lambda \in \sigma_r(T)$. We will show it's also an eigenvalue, which contradicts the definition of the residual spectrum.

By definition of the residual spectrum, T_λ is invertible on its range, so we have a $T_\lambda^{-1} : R(T_\lambda) \to H$, but $\overline{R(T_\lambda)} \ne H$. By the same argument as before, there must be a nonzero $y \in \overline{R(T_\lambda)}^\perp$, so for all $x \in H$, $(T_\lambda x, y) = (x, T_\lambda y) = 0$, and therefore $T_\lambda y = 0$, so λ is an eigenvalue, which is a contradiction as we noted.

Finally, we return to (2) and bound the spectrum. This is the crazy part. Pick a $c > 0$ and let $\lambda = R + c > R$; then, let $x \ne 0$ be in H. Then, by the definition of R,

$$(Tx, x) = \|x\|^2 \left(\frac{Tx}{\|x\|}, \frac{x}{\|x\|} \right) \le R\|x\|^2.$$

Next, $-(Tx - T_\lambda x, x) = -(T_\lambda x, x) \le \|T_\lambda x\| \|x\|$, so

$$-(Tx - \lambda x, x) = -(Tx, x) + \lambda \|x\|^2 \ge -R\|x\|^2 + \lambda \|x\|^2 = c\|x\|^2,$$

so $c\|x\|^2 \le \|T_\lambda x\| \|x\|$, and therefore $c\|x\| \le \|T_\lambda x\|$, since $x \ne 0$. Thus, if $\lambda > R$, then $\lambda \in \rho(T)$.

The proof for the lower bound is exactly the same, so we conclude that $\sigma(T) \subset [r, R]$. ⊠

Lecture 28: 10/30/15

The Spectral Theorem for Self-Adjoint Operators, II.

Note: due to the flash flood warnings, I missed this lecture (which Sam also gave). I've added notes corresponding to what he covered, but worked out of the textbook. Throughout this lecture, H is a Hilbert space.

Recall that last time, we proved Theorem 3.5.2, the first part of the spectral theorem for self-adjoint operators. It says that if T is a bounded, self-adjoint operator on H, then

- $\sigma_r(T) = \emptyset$;
- $\sigma(T) \subset \mathbb{R}$, and in fact $\sigma(T) \subset [r, R]$, where $r = \inf_{\|x\|=1}(Tx, x)$ and $R = \sup_{\|x\|=1}(Tx, x)$; and
- $\lambda \in \rho(T)$ iff T_λ is bounded below (see Lemma 3.5.1).

We can actually refine this result slightly, and today we will do so.

Theorem 3.6.1 (Spectral theorem for self-adjoint operators, II)**.** *Let $T : H \to H$ be a bounded, self-adjoint operator, and $r = \inf_{\|x\|=1}(Tx, x)$ and $R = \sup_{\|x\|=1}(Tx, x)$ be as in Theorem 3.5.2. Then, $r, R \in \sigma(T)$ and $\|T\|_{B(X,X)} = \sup_{\|x\|=1} |(Tx, x)| = \max\{|r|, |R|\}$.*

So it's not just that the spectrum is bounded, but that the extremal values of $|(Tx, x)|$ are both actually in the spectrum.

Proof. Let $M = \sup_{\|x\|=1}|(Tx,x)|$. By the Cauchy-Schwarz theorem, for any $x \in H$ with norm 1, $|(Tx,x)| \le \|Tx\|\|x\| = \|Tx\| \le \|T\|$, so passing to the supremum, $M \le \|T\|$.

We can assume without loss of generality that $T \neq 0$ (if that were true, there would be nothing more to prove). Thus, there's a $z \in H$ with $Tz \neq 0$ and $\|z\| = 1$. Thus, if $v = \|Tz\|^{1/2}z$ and $w = \|Tz\|^{-1/2}Tz$, then $\|v\|^2 = \|w\|^2 = \|Tz\|$. Thus, Tv is a multiple of w: $Tv = \|Tz\|w$, so

$$(Tv,w) = \|Tz\|(w,w) = \|Tz\| \cdot \|w\|^2 = \|Tz\| \cdot \|Tz\| = \|Tz\|^2.$$

Since T is self-adjoint, then $(Tw,v) = \overline{(v,Tw)} = \overline{(Tv,w)}$, but this is $\|Tz\|^2$, which is real, so $(Tw,v) = \|Tz\|^2$ too. We'll plug this into

$$\begin{aligned}(T(v+w),v+w)-(T(v-w),v-w) &= (Tv,v)+(Tv,w)+(Tw,v)+(Tw,w)\\ &\quad -(Tv,v)+(Tv,w)+(Tw,v)-(Tw,w)\\ &= 2(Tv,w)+2(Tw,v) = 4\|Tz\|^2. \end{aligned} \tag{3.4}$$

However, by the definition of M,

$$\begin{aligned}|(T(v+w),v+w)-(T(v-w),v-w)| &\le |(T(v+w),v+w)|+|(T(v-w),v-w)|\\ &\le M\|v+w\|^2 + M\|v-w\|^2.\end{aligned}$$

Using the parallelogram law,

$$= 2M\left(\|v\|^2+\|w\|^2\right) = 4M\|Tz\|.$$

Since $a \le |a|$ for all $a \in \mathbb{R}$, then this and (3.4) tell us that $\|Tz\| \le M$. In particular, this is true for all z with $\|z\| = 1$, so we can pass to the supremum: $\|T\| \le M$ too. Thus, $\|T\| = M$.

The rest of the theorem is showing that r and R are in $\sigma(T)$. By definition, $\lambda \in \sigma(T)$ iff $\lambda - \mu \in \sigma(T_\mu)$, so we can translate by $-r$ if necessary to assume that $0 \le r \le R$. In this case, $\|T\| = \sup_{\|x\|=1}(Tx,x) = R$, so there's a sequence $\{x_n\}$ with $\|x_n\| = 1$ for each n and $(Tx_n,x_n) \ge R - 1/n$. Then,

$$\begin{aligned}\|T_R x_n\|^2 &= \|Tx_n - Rx_n\|^2 = \|Tx_n\|^2 - 2R(Tx_n,x_n) + R^2\\ &\le R^2 - 2R\left(R - \frac{1}{n}\right) + R^2 = \frac{2R}{n} \longrightarrow 0.\end{aligned}$$

That is, T_R isn't bounded below, so by Theorem 3.5.2, $R \in \sigma(T)$. A similar argument can be employed to show that $r \in \sigma(T)$ too. ⊠

If T is a bounded, self-adjoint operator, then by Theorem 3.4.5, $(Tx,x) \in \mathbb{R}$ for all x. This can be interpreted as (a scaling of) the angle between x and Tx, so now we can think about direction.

Definition. Let $T \in B(H,H)$. If $(Tx,x) \ge 0$ for all $x \in H$, then T is called a *positive* operator, written $T \ge 0$.

If $T,S \in B(H,H)$ such that $T - S \ge 0$, then one also writes that $T \ge S$. This defines a partial ordering on $B(H,H)$.

Lecture 29: 11/2/15

Positive Operators.

Definition. Let H be a Hilbert space and $T : H \to H$ be a bounded linear operator. If $\langle Tx,x\rangle \ge 0$ for all $x \in H$, then T is called a *positive* operator, written $T \ge 0$. For $R,S \in B(H,H)$, $R \le S$ means that $0 \le S - R$.

A positive operator is sometimes called *positive semidefinite*, and if $\langle Tx,x\rangle > 0$ for all $x \in H \setminus 0$, T is called *positive definite*.

The idea is that the angle between Tx and x is always between 0° and 90° in either direction (since the cosine of the angle comes from the inner product). These tend to be very important in mechanics: one expects forces to be positive, for example.

One might think of these as "nonnegative operators," but English isn't the best language.

Fact. The set of positive operators on a Hilbert space is partially ordered under $\le$.

Proposition 3.7.1. *Let H be a complex Hilbert space and $T : H \to H$ be bounded. Then, T is positive iff $\sigma(T) \ge 0$ and $T = T^*$.*

PROOF. In the forward direction, for every $x \in H$, $(Tx,x) \geq 0$, and in particular is in $\mathbb{R}$, so since H is a complex Hilbert space, then Theorem 3.4.5 tells us that T is self-adjoint. Then, the spectral theorem for self-adjoint operators tells us that $\sigma(T) \geq r$, where $r = \inf_{\|x\|=1}(Tx,x) \geq 0$, so $\sigma(T) \geq 0$.

In the reverse direction, since T is self-adjoint, then $(Tx,x) \in \mathbb{R}$ for all $x \in H$. Then, the spectral theorem for self-adjoint operators tells us that $\sigma(T) \subset [r,R]$, where $r = \inf_{\|x\|=1}(Tx,x)$. Thus, since $\sigma(T) \geq 0$, then for every $x \in H$ with norm 1, $(Tx,x) \geq 0$, and therefore for all nonzero $x \in H$, $(Tx,x) = \|x\|^2(T(x/\|x\|), x/\|x\|) \geq 0$ (and $(T(0),0) = 0$), so T is positive. ⊠

Definition. Suppose $T \geq 0$ and there's another $S \in B(H,H)$ such that $S^2 = T$. Then, we say that S is a *square root* of T. Square roots are not unique, though if $S \geq 0$, it is unique, so one says S is the *positive square root* of T, written $S = T^{1/2}$.

For example, the *Laplace operator* Δ (which we'll define precisely later) is a positive operator, so it has a square root, which is (more or less) the gradient operator: $(-\Delta)^{1/2} \sim \nabla$.

Theorem 3.7.2. *Every positive operator on a complex Hilbert space has a unique positive square root.*

The proof of this theorem is long and tedious, though not difficult, and relies on a generalization of Newton's method applied to compute square roots. As such, we'll skip over it.

Example 3.7.3.

- When $H = L^2(\Omega)$, let $\phi : \Omega \to [0,a]$, where a is finite. Then, $T : H \to H$ sending $f \mapsto \phi f$ is a bounded functional, and is clearly positive, because
$$\langle Tf, f\rangle = \int_\Omega \phi|f|^2\,\mathrm{d}x \geq 0.$$
It probably isn't a great surprise that the square root operator S is $Sf = \sqrt{\phi}f$.
- If $T \in B(H,H)$, then T^*T is positive: $\langle T^*Tx, x\rangle = \langle Tx, Tx\rangle \geq 0$. You've likely seen that for matrices.

Thus far, we have looked at three kinds of nice operators on a Hilbert space: compact operators, self-adjoint operators, and positive operators. If T is a compact, self-adjoint operator, we know a lot about its spectrum:

- $\sigma(T) \subset \mathbb{R}$ and is countable;
- all nonzero spectral values are eigenvalues;
- all eigenspaces are finite-dimensional; and
- if there are infinitely many eigenvalues, then they converge to 0.

We're about to prove a very important structural result for Hilbert spaces, and use it to prove (yet another) spectral theorem, this time for compact, self-adjoint operators.

Theorem 3.7.4 (Hilbert-Schmidt)**.** *Let H be a Hilbert space and $T : H \to H$ be a compact, self-adjoint operator. Then, there exists an orthonormal set $\{u_n\}$ of eigenfunctions of T corresponding to nonzero eigenvalues such that for all $x \in H$, there's a unique collection $\{\alpha_n\}$ such that*
$$x = \sum_n \alpha_n u_n + v,$$
for some $v \in N(T)$.

In other words, we can decompose into eigenspaces, and then 0 is our separate case.

PROOF. The proof will be pretty similar to the case for finite-dimensional vector spaces, leaning on the orthogonal decomposition.

By the spectral theorem for self-adjoint operators, we know there exists an eigenvalue λ_1 of T such that
$$|\lambda_1| = \sup_{\|x\|=1} |\langle Tx, x\rangle|.$$
Let u_1 be an associated eigenvector of norm 1, and let $Q_1 = \{u_1\}^\perp$. Thus, Q_1 is a closed subspace of H, and so is Hilbert in its own right. Moreover, T maps Q_1 onto Q_1: if $x \in Q_1$, then
$$\langle Tx, u_1\rangle = \langle x, Tu_1\rangle = \overline{\lambda}_1\langle x, u_1\rangle = 0,$$
so that $Tx \in Q_1$.

Now, we are right back where we started, so we may recurse. Since T is self-adjoint, we may choose an eigenvalue λ_2 of T that satisfies

$$|\lambda_2| = \sup_{\substack{\|x\|=1 \\ x\in Q_1}} |\langle Tx, x\rangle| \leq |\lambda_1|.$$

Then, let $Q_2 \subseteq Q_1$ be equal to $\{u_2\}^\perp$; within H, this is $\{u_1, u_2\}^\perp$. Again, the same argument shows $T(Q_2) \subseteq Q_2$.

Now, by induction, we have a sequence Q_n of nested, closed linear subspaces such that $Q_n = \{u_1, \dots, u_n\}^\perp$, $\{u_1, \dots, u_n\}$ is orthonormal, and $T(Q_n) \subseteq Q_n$. We also have

$$|\lambda_{n+1}| = \sup_{\substack{\|x\|=1 \\ x\in Q_n}} |\langle Tx, x\rangle| \leq |\lambda_n|.$$

Now, one of two things has to happen: either $|\lambda_n| = 0$ for some n, or the λ_n never reach 0.

Case 1. Suppose $|\lambda_{n+1}| > 0$, but $|\lambda_{n+2}| = 0$. Let $T_1 = T|_{Q_{n+1}}$, which maps $Q_{n+1} \to Q_{n+1}$, and

$$\|T_1\| = \sup_{\substack{\|x\|=1 \\ x\in Q_{n+1}}} |\langle Tx, x\rangle| = |\lambda_{n+2}| = 0.$$

Thus, $T = 0$ on Q_{n+1}, and thus $Q_{n+1} \subseteq N(T)$. We want to show this is all of the null space; we know T doesn't vanish on $\operatorname{span}\{u_1, \dots, u_n\}$ (except, of course, at 0), so if

$$0 = Tx = \sum_j \alpha_j Tu_j = \sum_j \alpha_j \lambda_j u_j,$$

and thus α_j must only be nonzero for $j \geq n+1$, so $x \in Q_{n+1}$.

Since $H = \operatorname{span}\{u_1, \dots, u_n\} \oplus Q_{n+1}$, we can write

$$x = \sum_{j=1}^{n} \alpha_j u_j + v,$$

where $v \in Q_{n+1}$, so we're done.

Case 2. Alternatively, $|\lambda_n| > 0$ for all n, so $\lambda_n \to 0$. Let $H_1 = \overline{\operatorname{span}\{u_1, u_2, \dots\}}$, so $H = H_1 \oplus H_1^\perp$, and if $x \in H$, we can write

$$x = \sum_{j=1}^{\infty} \langle x, u_j\rangle u_j + v,$$

where $v \in H_1^\perp$. It would be easiest if $H_1^\perp = N(T)$, and in fact this happens: choose a nonzero $v \in H_1^\perp$. For all n, $H_1^\perp \subseteq Q_n$, so let's compute the *Rayleigh quotient*:[8] for each n,

$$\frac{\langle Tv, v\rangle}{\|v\|^2} \leq \sup_{x\in Q_n} \frac{\langle Tx, x\rangle}{\|x\|^2} = |\lambda_{n+1}| \longrightarrow 0. \tag{3.5}$$

Thus, $\langle Tv, v\rangle = 0$ for all $v \in H_1^\perp$, so $H_1^\perp \subset N(T)$, because

$$\|T|_{H_1^\perp}\| = \sup_{\substack{\|v\|=1 \\ v\in H_1^\perp}} |\langle Tv, v\rangle| = 0.$$

Now, if $x \in H_1$,

$$Tx = T\left(\sum_{j=1}^{\infty} \beta_j u_j\right) = \sum_{j=1}^{\infty} \beta_j Tu_j = \sum_{j=1}^{\infty} \lambda_j \beta_j u_j,$$

which is in H_1. Thus, $T : H_1 \to H_1$ is one-to-one, so $N(T) \cap H_1 = \{0\}$, and since $H = H_1 \oplus H_1^\perp$, then $N(T) = H_1^\perp$. ⊠

This is an important theorem: it says that there's a nice eigenbasis for a compact, self-adjoint operator, up to the spectral value 0, which may not be an eigenvalue. But the null space is all right.

[8] The Rayleigh quotient of v and T is the leftmost term in (3.5).

Lecture 30: 11/4/15

Compact, Self-Adjoint Operators and the Ascoli-Arzelà Theorem.

Last time, we proved the Hilbert-Schmidt theorem, Theorem 3.7.4, which asserts that for a compact, self-adjoint operator $T : H \to H$ (where H is a Hilbert space), one can decompose any $x \in H$ as

$$x = \sum_{n=1}^{\infty} \alpha_n u_n + v,$$

where $\{u_n\}$ is orthonormal and $v \in N(T)$. Moreover, $\alpha_n = \langle x, u_n \rangle$, which we didn't prove but isn't hard to show.

This helps us prove the following theorem.

Theorem 3.8.1 (Spectral theorem for compact, self-adjoint operators)**.** *Let T be a compact, self-adjoint operator. Then, there exists an orthonormal basis $\{v_\alpha\}_{\alpha\in\mathcal{I}}$ for H such that each v_α is an eigenvector of T, and for any $x \in H$,*

$$Tx = \sum_{\alpha\in\mathcal{I}} \lambda_\alpha (x, v_\alpha) v_\alpha,$$

where λ_α is the eigenvalue associated to v_α.

Proof. We've done a lot of the hard work already. Let $\{u_n\}$ be the orthonormal system that Theorem 3.7.4 buys us; then, we need to complete it. Let $H_1 = \overline{\operatorname{span}\{u_n\}}$ and $\{e_\beta\}_{\beta\in\mathcal{J}}$ be an orthonormal basis for $H_1^\perp$, so that $\{e_\beta\}_{\beta\in\mathcal{J}} \cup \{u_\alpha\}_{\alpha\in\mathcal{I}}$ is an orthonormal basis for H, since $H = H_1 \oplus H_1^\perp$. Moreover, since $Te_\beta = 0$ for all β, then e_β is an eigenvector with eigenvalue 0. ⊠

If two compact, self-adjoint operators commute, that puts a strong condition on what the eigenvectors and eigenvalues are: their eigenspaces have to be related.

Proposition 3.8.2. *Let H be a Hilbert space and $S, T : H \to H$ be compact, self-adjoint operators such that $ST = TS$. Then, there exists an orthonormal basis $\{v_\alpha\}$ of eigenvectors common to both S and T.*

Proof. Let $\lambda \in \sigma_p(S)$ and V_λ be its eigenspace. Then, if $x \in V_\lambda$, $S(Tx) = TSx = T\lambda x = \lambda T x$, so $Tx \in V_\lambda$. This is what we meant by "respecting eigenspaces" just a moment ago.

We've just shown that T is a map $V_\lambda \to V_\lambda$, so V_λ has an orthonormal basis of T-eigenvectors, but these are also S-eigenvectors, so we're done. ⊠

Commuting operators come up a lot in physics; then again, so do non-commuting operators.

The Ascoli-Arzelà Theorem. Perhaps after all of this theory you've been looking for examples. Well, aren't you lucky.

If (M, d) is a compact metric space, then $C(M) = C(M;\mathbb{F})$ denotes the set of continuous functions $M \to \mathbb{F}$. This is a vector space, and under the norm $\|f\| = \max_{x\in M} |f(x)|$, $C(M)$ is a Banach space. We haven't exactly proven this, but it's the same proof as for $C([a, b])$.

Definition. Let $A \subseteq C(M)$; then, A is *equi-continuous* (or *equi-bounded*) if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $f \in A$, $\max_{d(x,y)<\delta} |f(x) - f(y)| < \varepsilon$.

This is a stronger condition that uniform continuity; it can be thought of as "uniformly uniform continuity," if that helps.

Theorem 3.8.3 (Ascoli-Arzelà)**.** *Let M be a compact metric space and $A \subseteq C(M)$ be*

(1) bounded, i.e. there's an $R > 0$ such that $\|f\| < R$ for all $f \in A$; and
(2) equi-continuous.

Then, $\overline{A}$ is compact in $C(M)$.

We'll use this to provide examples of compact operators.

Lemma 3.8.4. *A compact metric space is separable.*

Recall that we defined separability in the context of NLSes, but the definition only ever needed topological information, so it works just fine here.

Proof. For any $n \in \mathbb{N}$, the set $\{B_{1/n}(x) : x \in M\}$ is an open cover for M, so there's a finite subcover:

$$M = \bigcup_{i=1}^{N_n} B_{1/n}(x_i^{(n)}).$$

Then, the set $\{x_i^{(n)} : n \in \mathbb{N}, 1 \le i \le N_n\}$ is dense in M (if $x \in M$ and $\varepsilon > 0$, there's an $N \in \mathbb{N}$ such that $1/N < \varepsilon$, so x is within distance $1/N$ from some $x_i^{(N)}$), and it's a countable union of finite sets, so it's countable. ⊠

Proof of Theorem 3.8.3. Let $\{f_n\}_{n=1}^\infty$ be a sequence in A and $\{x_j\}_{j=1}^\infty$ be dense in M (which we can take by Lemma 3.8.4).

Since A is bounded, it's weak-$*$ compact, and so there's a subsequence $f_{n_k^1}(x_1)$ converging to y_1. Then, there's a subsequence of these $f_{n_k^1}$ such that $f_{n_k^2}(x_2) \to y_2$, and so on, so that for each $\ell \in \mathbb{N}$, we have a subsequence of $f_{n_k^{\ell-1}}$ called $f_{n_k^\ell}$ such that $f_{n_k^\ell}(x_\ell) \to y_\ell$.

We'll want to define $f(x_\ell) = y_\ell$, allowing us to get sequential compactness and thus (since everything is over a metric space) compactness. But we're not done yet.

Let $\varepsilon > 0$, so that there's a $\delta > 0$ with the properties needed for equi-continuity. Since $\{x_j\}$ is dense in M, then there's a subset $\{\widetilde{x}_m\}_{m=1}^N$ such that

$$M \subseteq \bigcup_{m=1}^{N} B_\delta(\widetilde{x}_m).$$

Choose $\widetilde{x}_\ell$ such that $d(x, \widetilde{x}_\ell) < \delta$. Then,

$$\begin{aligned} |f_{n_i}(x) - f_{n_j}(x)| &\le |f_{n_i}(x) - f_{n_i}(\widetilde{x}_\ell)| + |f_{n_i}(\widetilde{x}_\ell) - f_{n_j}(\widetilde{x}_\ell)| + |f_{n_j}(\widetilde{x}_\ell) - f_{n_j}(x)| \\ &\le 2\varepsilon + |f_{n_i}(\widetilde{x}_\ell) - f_{n_j}(\widetilde{x}_\ell)| \\ &\le 2\varepsilon + \max_{1\le m\le N} |f_{n_i}(\widetilde{x}_m) - f_{n_j}(\widetilde{x}_m)|. \end{aligned}$$

That is, this isn't just Cauchy, but it's uniformly so (this bound doesn't depend on x, thanks to compactness), so f_{n_j} converges uniformly (i.e. in norm) to f, where we define f by $f(x_n) = y_n$ and use density of $\{x_j\}$ to extend to all of M. Thus, $\overline{A}$ is sequentially compact, and so compact. ⊠

This has many possible uses; one is to show that integral operators are compact.

Theorem 3.8.5. *Let $\Omega \subset \mathbb{R}^d$ be open and bounded, and let K be a continuous function on $\overline{\Omega} \times \overline{\Omega}$. Let $X = C(\overline{\Omega})$ and $T : X \to X$ by*

$$Tf(x) = \int_\Omega K(x,y) f(y)\,\mathrm{d}y.$$

Then, T is compact.

We studied these kinds of operators in the homework; in any case, because $\overline{\Omega}$ is closed and bounded, then we're integrating a continuous function over a compact set, which means the integral exists. We also showed that if K is L^2, then T maps $L^2(\Omega)$ onto itself.

Corollary 3.8.6. *With Ω as above, if $K \in L^2(\Omega \times \Omega)$, so that $T : L^2(\Omega) \to L^2(\Omega)$. Then, T is compact, and if $K(x,y) = \overline{K(y,x)}$ for almost all $x, y \in \Omega$, then T is self-adjoint.*

This comes directly from the density of $L^2(\Omega)$ in $C(\overline{\Omega})$, and is pretty cool: we don't have to have any smoothness or continuity restrictions for it to hold.

Proof of Theorem 3.8.5. Let $\{f_n\}_{n=1}^\infty$ be a bounded sequence in $C(\overline{\Omega})$. To show compactness of T, we need to find a convergent subsequence of $\{Tf_n\}_{n=1}^\infty$. Since $\overline{\Omega}$ is compact, it suffices to show that $\{Tf_n\}$ is bounded and equi-continuous, by Theorem 3.8.3.

For boundedness, $\|Tf\|_{L^\infty} \le \|f_n\|_{L^\infty(\Omega)} \|K\|_{L^\infty(\Omega\times\Omega)}$, and we took $\{f_n\}$ to be bounded, so $\{Tf_n\}$ is bounded as well.

For equi-continuity,

$$\begin{aligned}|Tf_n(x) - Tf_n(y)| &= \left|\int_\Omega (K(x,z) - K(y,z))f_n(z)\,\mathrm{d}z\right| \\ &\le \|f_n\|_{L^\infty} \sup_{z\in\overline{\Omega}} |K(x,z) - K(y,z)| \int_\Omega \mathrm{d}z.\end{aligned}$$

Since $\{f_n\}$ is bounded, K is uniformly continuous, and Ω is bounded, then this is bounded above independently of x and y. ⊠

Lecture 31: 11/6/15

Sturm-Liouville Theory.

Today, we'll look at an extended application of a whole bunch of ideas, including spectral theory. It's sort of a classical theory, albeit with a functional-analytic perspective, and deals with boundary value problems.

We'll start with an interval $I = [a,b] \subseteq \mathbb{R}$ and three functions $a_j \in C^{2-j}(I)$, for $j = 0,1,2$, where $a_0 > 0$ on I.[9] Then, define $L : C^2(I) \to C^0(I)$ by $Lx(t) = a_0(t)x'' + a_1(t)x' + a_2(t)x$; this is a bounded linear map, since I is closed. One can also write $L = a_0D^2 + a_1D + a_2$, where $D : C^1(I) \to C^0(I)$ is the differentiation operator.

Theorem 3.9.1 (Picard)**.** *Given an $f \in C(I)$ and initial conditions $x_0, x_1 \in \mathbb{R}$, there exists a unique solution $x \in C^2(I)$ to the* initial value problem *$Lx = f$ for $a < t < b$, $x(a) = x_0$, and $x'(a) = x_1$.*

If you've studied differential equations, you're seen this theorem before, and in fact it's possible to weaken the hypothesis to where f is Lipschitz continuous. But we won't prove it.

Corollary 3.9.2. $\dim N(L) = 2$.

This is because if we ignore the initial conditions, two degrees of freedom (choices of $x(a)$ and $x'(a)$) exist in the choice of solution. In particular, L is not invertible.

We want to study the adjoint operator to L, but we want a way to incorporate the initial conditions. This motivates the following definition.

Definition. The *formal adjoint* to L is $L^* = \overline{a}_0D^2 + (2\overline{a}_0' - \overline{a}_1)D + (\overline{a}_0'' + \overline{a}_1' + \overline{a}_1)$.

This seems arbitrary, but the idea is that we like Hilbert spaces, so we're considering the L^2 inner product.

$$\begin{aligned}(Lx, y) &= (a_0x'' + a_1x' + a_2x, y) \\ &= (x'', \overline{a}_0y) + (x', \overline{a}_1y) + (x, \overline{a}_0y).\end{aligned}$$

Now, we can integrate by parts. Let B denote some boundary terms, which we're not too concerned with.

$$\begin{aligned}&= -(x', (\overline{a}_0y)') + \left[x'\overline{a}_0y\right]_a^b - (x, (\overline{a}_1y)') + B + (x, \overline{a}_0y) \\ &= (x, (\overline{a}_0y)'') - (x, (\overline{a}_1y)') + (x, \overline{a}_0y) + B \\ &= (x, \overline{a}_0y'' + 2\overline{a}_0'y' + \overline{a}_0''y - \overline{a}_1y' - \overline{a}_1'y + \overline{a}_0y) + B \\ &= (x, L^*y) + B.\end{aligned}$$

This is why L^* was defined this way: up to some boundary terms, it is the adjoint in L^2. If $L = L^*$, one says that L is *formally self-adjoint*.

Proposition 3.9.3. *If $\mathbb{F} = \mathbb{R}$, then L is formally self-adjoint iff $a_0' = a_1$; in this case, $L = Da_0D + a_2$.*

Proof. The proof is just a calculation: comparing the definitions, $L = a_0D^2 + a_1D + a_2$ and $L^* = a_0D^2 + (2a_0' - a_1)D + (a_0'' - a_1' + a_2)$. The D^1 terms are equal iff $a_0' = a_1$, and in this case $a_0'' = a_1'$, so the D^0 terms agree as well. Then, $(Da_0D)x = (a_0x')' = a_0x'' + a_0'x' = a_0x'' + a_1x'$, so we're good there too. ⊠

[9] There's also the theory where $a_0 \to 0$ at the endpoints; we won't do that, but the theory is very similar and has some important examples.

But L might not be self-adjoint. Too bad; let's make it self-adjoint! The idea is to take a function Q and consider $QLx = Qf$, such that QL is self-adjoint. By the previous proposition, this is equivalent to $(Qa_0)' = Qa_1$, so we just need to solve this differential equation. In a sense, Q is an integrating factor.

$Q'a_0 + a_0'Q = a_1Q$, so

$$\begin{aligned}\frac{Q'}{Q} &= \frac{a_1 - a_0'}{a_0}\\ (\ln|Q|)' &= \frac{a_1}{a_0} - (\ln|a_0|)'\\ \ln|Q| &= \int \frac{a_1}{a_0} - \ln|a_0|,\end{aligned}$$

so if $P(t) = \int_a^t a_1(x)/a_0(x)\,\mathrm{d}x$, then $Qa_0 = \exp(P)$, so our solution is

$$Q(t) = \frac{\exp(P)}{a_0}.$$

So now we can assume L is self-adjoint, and $QL = (Px')' + (a_2/a_0)Px$.

However, we want to solve boundary value problems, rather than initial value problems.

Definition. Let p, q, and w be real-valued functions on $I = [a,b]$ for $a < b$,[10] such that $p \neq 0$ on I and $w > 0$ on I. Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$ be such that $\alpha_1^2 + \alpha_2^2 \neq 0$ and $\beta_1^2 + \beta_2^2 \neq 0$ (i.e. at least one α_i and one β_i are nonzero).

If $f : I \to \mathbb{R}$, the problem of finding an $x(t) \in C^2(I)$ such that

$$\begin{cases} Ax = \dfrac{1}{w}\big((px')' + qx\big) = f \text{ on } (a,b)\\ \alpha_1 x(a) + \alpha_2 x'(a) = 0\\ \beta_1 x(b) + \beta_2 x'(b) = 0\end{cases} \tag{3.6}$$

is called a *(regular) Sturm-Liouville problem* (SL),[11] and if $f = \lambda x$, then the problem is called a *regular Sturm-Liouville eigenvalue problem* (here, we are also looking for $\lambda \in \mathbb{C}$).

The last two equations in (3.6) are the *boundary conditions* that are the whole point of this definition.

Example 3.9.4. One of the most important examples, on $I = [0,1]$, is the eigenvalue problem

$$\begin{cases} Ax = -x'' = \lambda x\\ x(0) = x(1) = 0.\end{cases}$$

These boundary conditions are called the *Dirichlet boundary conditions*. This problem, recast in the form $x'' + \lambda x = 0$, is something you may have solved in undergrad.

When $\lambda > 0$, the solutions are

$$\begin{aligned} x(t) &= A\sin(\sqrt{\lambda}t) + B\cos(\sqrt{\lambda}t)\\ x(0) &= B = 0\\ x(1) &= A\sin(\sqrt{\lambda}) = 0.\end{aligned}$$

Thus, the solutions are when $\sqrt{\lambda} = n\pi$ for $n \in \mathbb{Z}$, so let $\lambda_n = n^2\pi^2$, when $n \in \mathbb{N}$; then, $x_n(t) = \sin(n\pi t)$.

Our goal is to analyze a general Sturm-Liouville problem and determine how to solve it. Since $A : C^2(I) \to C^0(I)$, we would like an inverse $C^0(I) \to C^2(I) \subseteq C^0(I)$, which could be a well-behaved operator on Banach spaces. And if we can make it an integral operator, Theorem 3.8.5 would tell us it's compact, which would be nice.

Definition. Given a Sturm-Liouville problem, a *Green's function* for it is a function $G \in C^0(I \times I)$ satisfying the following conditions.

(1) $G \in C^2(I \times I \setminus D)$, where $D = \{(t,t) : t \in I\}$, i.e. G is C^2 away from the diagonal.
(2) For all $s \in I$, $G(\cdot, s)$ satisfies the boundary conditions.

[10] One can extend this to possibly unbounded intervals, and the theory is similar, though again we're not going to worry about it.
[11] If our coefficient functions are allowed to be 0, then this may be called a *singular Sturm-Liouville problem*.

(3) For all $t, s \in I \times I \setminus D$, $A_t G(t,s) = 0$, so that G is always in the kernel.
(4) The slope jumps in a controlled way at the diagonal:

$$\lim_{s \to t^-} \frac{\partial G}{\partial t}(t,s) - \lim_{s \to t^+} \frac{\partial G}{\partial t}(t,s) = \frac{1}{p(t)}.$$

We'll prove in Theorem 3.10.1 that such a function is necessarily the kernel[12] of an integral operator.

Example 3.9.5. Let's calculate the Green's function for the Sturm-Liouville problem in Example 3.9.4. Specifically, consider

$$G(t,s) = \begin{cases} (1-t)s, & 0 \le s \le t \le 1 \\ (1-s)t, & 0 \le t \le s \le 1. \end{cases}$$

Since this is linear everywhere save maybe the diagonal, G is clearly in $C^2(I \times I \setminus D)$, and since the definitions agree on the diagonal, $G \in C^0(I \times I)$, so G satisfies condition (1).

For (2),

$$G(t,0) = (1-t) \cdot 0 = 0$$
$$G(t,1) = (1-1)t = 0.$$

Great. For (3), $G_{tt}(t,s) = 0$, so $A_t G = 0$ everywhere except perhaps the diagonal, and for part d,

$$\frac{\partial G}{\partial t} = \begin{cases} -s, & 0 \le s \le t \le 1 \\ 1-s, & 0 \le t \le s \le 1. \end{cases}$$

The jump is $-t - (1-t) = -1 = 1/p$. Thus, (4) is satisfied, so this function is a Green's function for this problem.

Next time, we'll show that the integral operator

$$Kf(t) = \int_0^1 G(t,s) f(s)\,\mathrm{d}s \tag{3.7}$$

solves the Sturm-Liouville problem, and since it's compact and formally self-adjoint (see Theorem 3.11.3), this means all sorts of nice things.

Lecture 32: 11/9/15

Solving Sturm-Liouville Problems With Green's Functions.

Recall that last time, we were talking about Sturm-Liouville problems and how to solve them. Specifically, on the interval $I = [a,b]$, we want to solve the problem

$$Ax = \frac{1}{w} Lx = \frac{1}{w}\big((px')' + qx\big) = f, \tag{3.8a}$$

where $p \neq 0$ and $w > 0$, subject to two boundary conditions

$$\alpha_1 x(a) + \alpha_2 x'(a) = 0 \tag{3.8b}$$
$$\beta_1 x(b) + \beta_2 x'(b) = 0, \tag{3.8c}$$

where at least one of α_1 and α_2 is nonzero, and similarly for β_1 and β_2. If $f = \lambda x$ instead, we have an eigenvalue problem, which is slightly different.

We're going to use Green's functions for this problem:[13] these were the functions $G \in C^0$ such that $G \in C^2(I \times I \setminus D)$ (so it's bad only on the diagonal); $G(\cdot, s)$ satisfies the boundary condition; for all $t \neq s$, $A_t G(t,s) = 0$; and finally, $\frac{\partial G}{\partial t}$ jumps $1/p$ at the diagonal. The point is that (3.7) is satisfied, making it much easier to solve the equation. We won't prove this directly; instead, we'll derive it as a consequence of something more general.

Definition. Let $u_1, u_2 \in C^1(I)$; then, their *Wronskian* is

$$W(s) = W(s; u_1, u_2) = u_1(s)u_2'(s) - u_1'(s)u_2(s) = \det\begin{pmatrix} u_1(s) & u_1'(s) \\ u_2(s) & u_2'(s) \end{pmatrix}.$$

[12] Note: this is the analytic notion of kernel, not the algebraic one.

[13] The definition we've given is for a Green's function for these problems; in general, a Green's function is defined to be a function G such that the integral operator (3.7) produces solutions to the problem.

Theorem 3.10.1. *Consider the Sturm-Liouville problem as defined in* (3.8a)*,* (3.8b)*, and* (3.8c)*, such that $p \in C^1(I)$ and $w, q \in C^0(I)$, and suppose* 0 *is* not *an eigenvalue of A. Let u_1 and u_2 be two solutions to $Au = 0$ such that u_1 satisfies* (3.8b) *and u_2 satisfies* (3.8c)*. Define*

$$G(s,t) = \begin{cases} \dfrac{u_2(t)u_1(s)}{p(t)W(t)}, & a \le s \le t \le b \\ \dfrac{u_1(t)u_2(s)}{p(t)W(t)}, & a \le t \le s \le b, \end{cases} \tag{3.9}$$

where $p(t)W(t)$ is a nonzero constant and $W(t)$ is the Wronskian of u_1 and u_2. Then, G is a Green's function for L, and if $\mathscr{G}$ is any Green's function and $f \in C^0(I)$, then

$$u(t) = \int_a^b \mathscr{G}(t,s)f(s)\,\mathrm{d}s$$

is the unique solution to $Lu = f$.

In Example 3.9.5, we had $u_1(t) = t$, $u_2(t) = 1 - t$, $p = -1$, and the Wronskian

$$W = \det\begin{pmatrix} t & 1 \\ 1-t & -1 \end{pmatrix} = -t - (1-t) = -1.$$

We'll prove Theorem 3.10.1 in a couple steps.

Theorem 3.10.2 (Abel)**.** *Let $Lu = (pu')' + qu$, where $p \in C^1(I)$ and $q \in C^0(I)$, $\lambda \in \mathbb{C}$, and $w \in C^0(I)$ be such that $w > 0$; then, if u_1 and u_2 satisfy $Lu = \lambda wu$, then $p(t)W(t;u_1,u_2)$ is constant.*

Proof. Notice that since $W = u_1u_2' - u_1'u_2$, then $W' = u_1u_2'' - u_1''u_2$. Thus,

$$\begin{aligned} 0 &= \lambda w(u_1u_2 - u_2u_1) \\ &= u_1Lu_2 - u_2Lu_1 \\ &= u_1(pu_2'' + p'u_2' + qu_2) - u_2(pu_1'' + p'u_1' + qu_1) \\ &= pW' + p'W' = (pW)'. \end{aligned}$$ ⊠

For a theorem so famous and important, one might have expected the proof to be harder. Not that I'm complaining or anything.

Lemma 3.10.3. *Let $u, v \in C^1$ be such that $W(t_0;u,v) \ne 0$. Then, u and v are linearly independent.*

Proof. Suppose $\alpha, \beta \in \mathbb{R}$ are such that $\alpha u(t) + \beta v(t) = 0$. Differentiating, we get $\alpha u'(t) + \beta v'(t) = 0$. This is a system of equations:

$$\begin{pmatrix} u(t) & v(t) \\ u'(t) & v'(t) \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Plugging in $t = t_0$, the determinant is nonzero, so the matrix is invertible, and therefore the only solution to the above equation is $c_1 = c_2 = 0$. ⊠

Now, we have enough tools to prove Theorem 3.10.1.

Proof of Theorem 3.10.1. First, Picard's theorem (Theorem 3.9.1) shows that such u_1 and u_2 exist.

Since $N(L)$ is two-dimensional (since we're taking two derivatives), then we can write $N(L) = \operatorname{span}\{z_0, z_1\}$, where $z_0(a) = 1$, $z'(a) = 0$, $z_1(a) = 0$, and $z_1'(a) = 1$. Thus, $u_1(t) = -\alpha_2 z_0(t) + \alpha_1 z_1(t)$ (which is how we set our boundary conditions) — in particular, it's nonzero. We can write u_2 in the same way, and thus u_2 isn't zero either.

By Abel's theorem (Theorem 3.10.2), $p(t)W(t)$ is constant. Suppose $p(t)W(t) = 0$; then $W(t;u_1,u_2)$ has to be identically zero, because we specified that $p \ne 0$. Thus, the matrix

$$\begin{pmatrix} u_1(a) & u_1'(a) \\ u_2(a) & u_2'(a) \end{pmatrix}$$

is singular, but since it's not the zero matrix, then its null space is one-dimensional. However, it's also $\operatorname{span}\{\alpha_1, \alpha_2\}$, meaning $\alpha_1 u_2(a) + \alpha_2 u_2'(a) = 0$. Since u_2 already satisfied the boudary conditions for β_1 and β_2, this means u_2 satisfies both boundary conditions. But since 0 isn't an eigenvalue of A and $Au_2 = 0$, then $u_2 = 0$, which is a contradiction.

This means the Wronskian isn't identically zero, so u_1 and u_2 are linearly independent, by Lemma 3.10.3. Thus, the definition of the Green's function in (3.9) is well-defined.

Picard's theorem tells us that G is C^2 off of the diagonal, and on the diagonal $t = s$, so they agree and G is continuous. Then, we need to check the boundary conditions, i.e. that $\alpha_1 G(a,s) + \alpha_2 G_t(a,s) = 0$, but this is just from the definition of G:

$$\alpha_1 G(a,s) + \alpha_2 G_t(a,s) = \alpha_1 \frac{u_1(a)u_2(s)}{pW} + \alpha_2 \frac{u_1'(a)u_2(s)}{pW} = \text{stuff} \cdot \big(\alpha_1 u_1(a) + \alpha_2 u'(a)\big) = 0,$$

so we're set. The proof that G satisfies the other boundary condition proceeds in the same way.

For the third property, that u_1 and u_2 satisfy $Au_i = 0$ leads directly to the calculation that $A_t G(t,s) = 0$. Thus, the most interesting property is the jump condition.

$$\frac{\partial G}{\partial t} = \begin{cases} \dfrac{u_2'(t)u_1(s)}{pW}, & s < t \\ \dfrac{u_1'(t)u_2(s)}{pW} & t < s, \end{cases}$$

so

$$\frac{u_2'(t)u_1(t^-)}{pW} - \frac{u_1'(t)u_2(t^+)}{pW} = \frac{W}{pW} = \frac{1}{p}.$$

Thus, G satisfies the jump condition.

Finally, we need to check that integrating against a Green's function produces the unique solution to $Lu = f$. Uniqueness is simple: 0 isn't an eigenvalue, so L is one-to-one, and the solution is unique.

For existence, let $\mathscr{G}$ be a Green's function and

$$u(t) = \int_a^b \mathscr{G}(t,s) f(s)\,\mathrm{d}s.$$

Since this integral is constant for a given t and $\mathscr{G}$ satisfies the boundary conditions, then u must as well.

Now, we should check that u satisfies (3.8a), but we must be careful, because $\mathscr{G}$ isn't defined on the diagonal. In particular, we must split the integral up into two pieces.

$$\begin{aligned} u'(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \int_a^b \mathscr{G}(t,s) f(s)\,\mathrm{d}s \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_a^t \mathscr{G}(t,s) f(s)\,\mathrm{d}s + \int_t^b \mathscr{G}(t,s) f(s)\,\mathrm{d}s \right) \\ &= \mathscr{G}(t,t) f(t) + \int_a^t \mathscr{G}_t(t,s) f(s)\,\mathrm{d}s - \mathscr{G}(t,t) f(t) + \int_t^b \mathscr{G}_t(t,s) f(s)\,\mathrm{d}s \\ &= \int_a^b \mathscr{G}_t(t,s) f(s)\,\mathrm{d}s. \end{aligned}$$

Now, let's differentiate again. We still have to split along the diagonal, but this time it actually makes a difference.

$$\begin{aligned} (pu')' &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_a^t p(t)\mathscr{G}_t(t,s) f(s)\,\mathrm{d}s + \int_t^b p(t)\mathscr{G}_t(t,s) f(s)\,\mathrm{d}s \right) \\ &= p(t)\mathscr{G}_t(t,t^-) f(t) + \int_a^t (p\mathscr{G}_t)_t f(s)\,\mathrm{d}s - p(t)\mathscr{G}_t(t,t^+) f(t) + \int_t^b (p\mathscr{G}_t)_t f(s)\,\mathrm{d}s. \end{aligned}$$

The jump in $\mathscr{G}_t$ is $1/p$, so it cancels with $p(t)$, and we just get f:

$$= f(t) + \int_a^b (p\mathscr{G}_t)_t f(s)\,\mathrm{d}s.$$

Now, let's put it all together.

$$\begin{aligned}
Lu(t) &= (pu')' + qu \\
&= f(t) + \int_a^b (p\mathscr{G}_t)_t f(s)\,\mathrm{d}s + q\int_a^b \mathscr{G}(s,t)f(s)\,\mathrm{d}s \\
&= f(t) + \int_a^b A_t\mathscr{G}(s,t)f(s)w(t)\,\mathrm{d}s,
\end{aligned}$$

but since the diagonal has measure 0 and $A_t\mathscr{G}$ is zero everywhere else,

$$= f(t).$$

⊠

Next class, we'll rephrase this in terms of operators.

Lecture 33: 11/11/15

Applying Spectral Theorems to Sturm-Liouville Problems.

Recall that we were looking at a Sturm-Liouville problem as defined in (3.8a), (3.8b), and (3.8c); we proved that our solution operator $T : C^0(I) \to C^0(I)$ is the operator

$$Tf(t) = \int_a^b G(t,s)f(s)\,\mathrm{d}s,$$

where G is the Green's function for the problem. Since $C^0(I)$ is dense in $L^2(I)$, then we can view T as an operator on $L^2(I)$.

This T is a bounded linear operator; well, it's clearly linear, and it's bounded because

$$\begin{aligned}
\|Tf\|_{L^2(I)}^2 = \int_a^b |Tf(t)|^2\,\mathrm{d}t &= \int_a^b \left(\int_a^b G(t,s)f(s)\,\mathrm{d}s\right)\mathrm{d}t \\
&\le \int_a^b \int_a^b |G(t,s)|^2\,\mathrm{d}s \int_a^b |f(s)|\,\mathrm{d}s\,\mathrm{d}t \\
&\le \|G\|_{L^2(I\times I)}^2 \|f\|_{L^2(I)}^2.
\end{aligned}$$

By Theorem 3.8.5, a corollary to the Ascoli-Arzelà theorem, T is compact. It's also self-adjoint:

$$\begin{aligned}
(tf,g) &= \int_a^b\int_a^b G(t,s)f(s)\overline{g(t)}\,\mathrm{d}s\,\mathrm{d}t \\
&= \int_a^b f(s)\overline{\int_a^b G(t,s)g(s)\,\mathrm{d}s\,\mathrm{d}t} \\
&= (f, Tg).
\end{aligned}$$

Notice that the original operator A (or L) isn't very nice; its inverse is much nicer.

Now, we want to relate the things we've proven about the Sturm-Liouville problem to things about spectral theory that we already know.

Proposition 3.11.1. *If* 0 *isn't an eigenvalue of the Sturm-Liouville operator L, then it's not an eigenvalue of T.*

Proposition 3.11.2. *If $\lambda \neq 0$, then λ is an eigenvalue of L iff $1/\lambda$ is an eigenvalue of T; moreover, the eigenspaces coincide.*

Proof of Proposition 3.11.1. Suppose $Tf = 0$. Then,

$$\begin{aligned}
0 = (Tf)'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{pW}u_2(t)\int_a^t f(s)u_1(s)\,\mathrm{d}s + \frac{1}{pW}u_1(t)\int_t^b f(s)u_2(s)\,\mathrm{d}s\right) \\
&= \frac{1}{pW}\left(u_2'(t)\int_a^t f(s)u_1(s)\,\mathrm{d}s + u_1'(t)\int_t^b f(s)u_2(s)\,\mathrm{d}s\right).
\end{aligned}$$

In particular, this means

$$0 = Tf = \frac{1}{pW}\bigg(u_2(t)\underbrace{\int_a^t f(s)u_1(s)\,\mathrm{d}s}_{v_1} + u_2(t)\underbrace{\int_t^n f(s)u_2(s)\,\mathrm{d}s}_{v_2}\bigg),$$

so

$$\begin{pmatrix} u_2' & u_1' \\ u_2 & u_1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

but the determinant of the matrix on the left is the Wronskian, which we know is nonzero, so v_1 and v_2 are 0, meaning that for all t,

$$\int_a^t fu_1 = 0 = \int_t^b fu_2,$$

and so $f = 0$. ⊠

Proof of Proposition 3.11.2. Let $f \in C^2(I)$ be an eigenfunction for L corresponding to a nonzero eigenvalue λ, so that f satisfies the boundary conditions and $Lf = \lambda f$. Thus, $f = TLf = \lambda Tf$, so $Tf = (1/\lambda)f$.

Conversely, if $f \in L^2(I)$ is an eigenfunction for T corresponding to the nonzero eigenvalue $1/\lambda$, then $Tf = 1/\lambda f$, so $f \in R(T)$, and therefore f satisfies the boundary conditions. Since G is continuous, then $f \in C^0(I)$. Then, similarly to above, $f = LTf = (1/\lambda)Tf$, so $Lf = \lambda f$. ⊠

These are cool and all that, but we'd really like a version of these propositions for A instead of L. L is nicer and more symmetric (since it is self-adjoint), but A has a $1/w$ term, which is an important part the problem we were originally hoping to solve, but isn't quite as nice.

A is not self-adjoint in the L^2-inner product. This is too bad. So let's change the inner product so that A is self-adjoint. Specifically, we'll put the weight $1/w$ into it. Define

$$\langle f, g\rangle_w = \int_a^b f(t)\overline{g(t)}w(t)\,\mathrm{d}t.$$

This is an inner product, which is easy to check, but the more interesting content is that the norm it defines is equivalent to the L^2-norm. Since w is continuous and I is compact, then it has a minimum m^* and a maximum M^* on I, and so $m^*\|f\|_{L^2} \le \|f\|_w \le M^*\|f\|_{L^2}$.[14]

Now, we define $K : L^2(I) \to L^2(I)$ by

$$Kf(t) = \int_a^b G(t,s)f(s)w(s)\,\mathrm{d}s, \tag{3.10}$$

which is a solution operator for A: if $Lu = f$, then $Au = wf$. Moreover, it's a bounded and even compact operator (since these are preserved under equivalence of norm), and it's self-adjoint: $\langle Kf, g\rangle_w = \langle f, Kg\rangle_w$ (which has the same proof as we saw earlier). In fact, analogues of Propositions 3.11.1 and 3.11.2 exist, with essentially the same proofs.

Proposition 3.11.3. *K as defined in (3.10) is a compact, self-adjoint operator on $(L^2(I), \langle\cdot,\cdot\rangle_w)$. Moreover, $0 \notin \sigma_p(K)$, and $\sigma(K) = \{0\} \cup \{\lambda : 1/\lambda \in \sigma_p(A)\}$ and the eigenspaces of K and A coincide.*

Since K is compact, we know *a priori* that its eigenspaces are finite-dimensional. But we can do better.

Definition. An eigenvalue of an operator T is *simple* if its eigenspace is one-dimensional.

Proposition 3.11.4. *The eigenvalues of a Sturm-Liouville problem are simple.*

Proof. Suppose u and v are eigenvectors for the nonzero eigenvalue λ. The boundary conditions means that

$$\begin{pmatrix} u(a) & u'(a) \\ v(a) & v'(a) \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

[14]An alternate way of thinking about this is that we're defining a different measure on I; rather than the Lesbegue measure $\mathrm{d}t$, we're using the measure $\mathrm{d}\mu = w(t)\,\mathrm{d}t$ and looking at $L^2(I;\mathrm{d}\mu)$.

Since α_1 and α_2 aren't both 0, then this means the determinant of the matrix is 0, but that determinant is $W(a;u,v)$. Since pW is constant, then $p(a)W(a)=0$, so $p(t)W(t)=0$ for all t. Since $p\neq 0$, though, this means $W(t)=0$ for all $t\in I$.

It's also possible to write the Wronskian as the dot product

$$W(t;u,v)=\begin{pmatrix}u\\u'\end{pmatrix}\cdot\begin{pmatrix}v'\\-v\end{pmatrix}=0$$

for all t, so $(u,u')\in\operatorname{span}\{(v,v')\}$ (because $(v,v')\perp(-v',v)$) , and therefore u and v are linearly dependent. ⊠

There's probably a general principle about linear dependence and the Wronskian that you could prove, but we haven't needed it until here.

We can apply the spectral theorem for compact, self-adjoint operators to this problem to learn even more about it.

Theorem 3.11.5. *Suppose* 0 *isn't an eigenvalue for the Sturm-Liouville problem A. Then, A has a countable collection of real eigenvalues* $\{\lambda_n\}_{n=1}^\infty$ *such that* $|\lambda_n|\to\infty$ *and each eigenspace is one-dimensional. If* u_n *is an eigenfunction of* λ_n *with norm* 1*, then* $\{u_n\}_{n=1}^\infty$ *is an orthonormal basis for* $(L^2(I),\langle\cdot,\cdot\rangle_w)$*. Thus, if* $u\in L^2(I)$*, then*

$$u=\sum_{n=1}^\infty\langle u,u_n\rangle_w u_n,$$

and if $Au\in L^2(I)$ *and u satisfies the boundary condition, then*

$$Au=\sum_{n=1}^\infty\lambda_n\langle u,u_n\rangle_w u_n.$$ [15]

This is just a restatement of stuff we know already. But the point is, $L^2(I)$ has an orthonormal basis that is useful for A — but even more, one can generate bases for $L^2(I)$ using Sturm-Liouville problems! This is also useful.

Example 3.11.6. Let's return to the familiar example

$$\begin{cases}-x''=\lambda x\\ x(0)=x(1)=0.\end{cases}$$

Thus, we've seen that $\lambda_n=(n\pi)^2$ for $n\in\mathbb{N}$, and $u_n(t)=\sqrt{2}\sin(n\pi t)$. Thus, if $f\in L^2([0,1])$, then it has a *sine series*

$$f(t)=\sum_{n=1}^\infty\left(2\int_0^1 f(s)\sin(n\pi s)\,\mathrm{d}s\right)\sin(n\pi t). \tag{3.11}$$

In particular, we see that $L^2(I)$ is separable, which is a nice way to prove it.

We can iterate this for higher dimensions. If $f\in L^2(I\times I)$, then

$$\begin{aligned}f(x,y)&=\sum_{n=1}^\infty\left(2\int_0^1 f(x,t)\sin(n\pi t)\,\mathrm{d}t\right)\sin n\pi y\\&=4\sum_{n=1}^\infty\sum_{m=1}^\infty\int_0^1\int_0^1 f(s,t)\sin(n\pi t)\sin(m\pi s)\,\mathrm{d}t\,\mathrm{d}s.\end{aligned}$$

Thus, the set of functions $\sin(n\pi y)\sin(m\pi x)$ for $m,n\in\mathbb{N}$ is an orthonormal basis for $L^2(I\times I)$. So $L^2(I^2)$ is separable, and in the same way so is $L^2(I^d)$. The same kind of proof shows that $L^2(Q)$ for any box $Q\subset\mathbb{R}^d$ is separable, and since any bounded set Ω can be contained in a box, $L^2(\Omega)$ is also separable. Finally, it's possible to extend this off to infinity to prove that if Ω is a measurable subset of $\mathbb{R}^d$, then $L^2(\Omega)$ is separable.

Generalizing Sturm-Lioville theory to higher dimensions is possible, but comes with more involved boundary value problems, and we're not going to worry about that now.

[15] Remember that these equalities are in L^2, so pointwise they're only equal almost everywhere. We're not going to worry about this point, however.

Example 3.11.7. Consider the suspiciously similar example

$$\begin{cases} -x'' = \lambda x \\ x(0) = x'(1) = 0. \end{cases}$$

Now, though, 0 is an eigenvalue; the eigenvalues are $\lambda_n = n^2\pi^2$ for $n = 0, 1, 2, \ldots$, with eigenvectors $u_n(t) = \sqrt{2}\cos(n\pi t)$. These are an orthonormal basis, so you can define *cosine series*, which look pretty similar to (3.11).

You might be wondering where the more famous example, Fourier series, is. I'm glad you asked.

Example 3.11.8. The problem for Fourier series is

$$\begin{cases} -x'' = \lambda x \\ x(0) = x(1) \\ x'(0) = x'(1). \end{cases}$$

These are called *periodic boundary value conditions*. And they're great and all, but they don't form a Sturm-Liouville problem!

However, it turns out you can go through all of the theory again; the solution operator is given by integrating with a Green's function again, and is compact, but the eigenspaces are two-dimensional; the n^{th} eigenspace is spanned by $u_n(t) = \sqrt{2}\cos(2n\pi t)$ and $v_n(t) = \sqrt{2}\sin(2n\pi t)$.

CHAPTER 4

Distributions

Lecture 34: 11/16/15

The Space of Test Functions.

We'll start talking about distributions, sometimes also called *generalized functions*, today.

Consider the perfectly reasonable continuous function

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

Its derivative is the *Heaviside function*

$$H(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0. \end{cases}$$

We don't worry about what this does at 0. We can differentiate this everywhere except 0, and $H'(x) = 0$ wherever it's defined, but this isn't really the right story: at 0, the function jumps, so it sort of makes sense that at $x = 0$, $H'(x) = \infty$. The key here is "sort of."

One way to make this rigorous is with integration by parts; if $\phi \in C^1(-\infty, \infty)$, then

$$\int_a^b u'\phi \, \mathrm{d}x = u\phi\Big|_a^b = \int_a^b u\phi' \, \mathrm{d}x.$$

If $u \in C^0(-\infty, \infty)$ but isn't C^1, then the right-hand side is defined, even if the left-hand side isn't. So we could *define* the left-hand side by the right-hand side, for some appropriate set of *test functions* ϕ: if ϕ is compactly supported, then

$$\int_{-\infty}^{\infty} u'\phi \, \mathrm{d}x = -\int_{-\infty}^{\infty} u\phi' \, \mathrm{d}x.$$

Let's apply this to $u(x) = H(x)$ above. You can directly see that

$$\int_{-\infty}^{\infty} f'(x)\phi(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} H(x)\phi(x) \, \mathrm{d}x = \int_0^{\infty} \phi(x) \, \mathrm{d}x,$$

or use integration by parts:

$$\begin{aligned} \int_{-\infty}^{\infty} f'(x)\phi(x) \, \mathrm{d}x &= -\int_{-\infty}^{\infty} f(x)\phi'(x) \, \mathrm{d}x = -\int_0^{\infty} x\phi'(x) \, \mathrm{d}x \\ &= -x\phi\Big|_0^{\infty} + \int_0^{\infty} x'\phi \, \mathrm{d}x \\ &= \int_0^{\infty} \phi(x) \, \mathrm{d}x. \end{aligned}$$

In this case, we get the same result; everything is nicely well-defined. But if we apply this to $H'(x)$, its derivative at 0 isn't well-defined, so a direct approach doesn't work. Instead, let's try integrating by parts again.

$$\begin{aligned}\int_{-\infty}^{\infty} H'(x)\phi(x)\,\mathrm{d}x &= -\int_{-\infty}^{\infty} H(x)\phi'(x)\,\mathrm{d}x \\ &= -\int_{0}^{\infty} \phi'(x)\,\mathrm{d}x = \phi\Big|_0^\infty \\ &= \phi(0).\end{aligned}$$

The function $H'(\phi) = \delta_0(\phi) = \phi(0)$ is often called a *δ-function*. This is the sense in which $H'(0) = \infty$: if integrating against a test function is like weighting its values, this assigns a maximum possible weight to 0, which would be infinite, I guess.

You can keep going:

$$\int_{-\infty}^{\infty} H''\phi = -\int_{-\infty}^{\infty} H'\phi' = -\phi'(0) = -\delta_0(\phi').$$

But if you integrate against H''', you get $-\delta_0(\phi'')$. Does ϕ have a second derivative?

Integrating against these things, even if they're not perfectly well defined as real-valued functions, is a linear functional from the space of test functions to $\mathbb{F}$.[1] To avoid regularity problems like the one that happened for H''', we take our space of *test functions* to be $\mathscr{D} = C_0^\infty(-\infty,\infty)$, the space of compactly supported smooth functions. Then, functions such as H' are linear operators $\mathscr{D} \to \mathbb{F}$, so they're in the dual space, which is in this context usually denoted $\mathscr{D}'$. We'd like to place a good topology on $\mathscr{D}$.

Our integration by parts formula has a nice reformulation in terms of functionals: the statement $\int u'\phi = -\int u\phi'$ just defines $u'(\phi) = -u(\phi')$. Repeating this process, $u^{(n)}(\phi) = (-1)^n u(\phi^{(n)})$. And now, it's precise what H' is: it's the linear functional $H' : \phi \mapsto \phi(0)$.

We're going to need these functionals to be continuous, and we'd like to take derivatives of arbitrary "functions," even those like H' that aren't really functions (hence the name "generalized function").

Distributions, Rigorously. Let's reframe this more precisely, in terms of functional analysis that we're more familiar with. Let $\Omega \subseteq \mathbb{R}^d$ be a *domain*, i.e. an open set, and $C^0(\Omega)$ denote the space of continuous functions $\Omega \to \mathbb{F}$.

Definition. The *support* of an $f \in C^0(\Omega)$ is $\operatorname{supp}(f) = \overline{\{x \in \Omega : f(x) \neq 0\}}$.

So it's basically, but not quite, the set where f is nonzero.

Definition. A *multi-index* is an $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_d) \in \mathbb{N}^d$.[2] That is, it's an ordered d-tuple of nonnegative integers. We'll write $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d$ and define a differential operator of order $\boldsymbol{\alpha}$ by

$$\partial^{\boldsymbol{\alpha}} = D^{\boldsymbol{\alpha}} = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}.$$

Note that the order doesn't matter in $\partial^{\boldsymbol{\alpha}}$, even though it does for $\boldsymbol{\alpha}$.

This is a clean way to talk about all combinations of partial derivatives at once.

Definition.

- Let $C^n(\Omega) = \{f \in C^0(\Omega) : D^{\boldsymbol{\alpha}} f \in C^0(\Omega) \text{ for all } \boldsymbol{\alpha} \text{ such that } |\boldsymbol{\alpha}| < n\}$.
- Then, let $C^\infty(\Omega)$ be the space of $f \in C^0(\Omega)$ such that $D^{\boldsymbol{\alpha}} f \in C^0(\Omega)$ for all $\boldsymbol{\alpha} \in \mathbb{N}^d$.

Notice that

$$C^\infty(\Omega) = \bigcap_{n=1}^{\infty} C^n(\Omega),$$

which is actually how it's defined sometimes.

We will use $K \Subset \Omega$ to mean that $K \subseteq \Omega$ and K is compact.[3]

[1] It's linear because integration is linear.

[2] In this class, the natural numbers include 0.

[3] On the chalkboard and in the book, the notation is $K \subset\subset \Omega$.

Definition. Finally, our space of *test functions* is $\mathscr{D} = \mathscr{D}(\Omega) = C_0^\infty(\Omega)$, the set of $f \in C^\infty(\Omega)$ such that $\operatorname{supp}(f)$ is a compact set. If $K \Subset \Omega$, define $\mathscr{D}_K = \{f \in C_0^\infty(\Omega) : \operatorname{supp}(f) \subseteq K\}$, the functions supported in K.

Proposition 4.1.1. *The sets $C^n(\Omega)$, $C^\infty(\Omega)$, $\mathscr{D}(\Omega)$, and $\mathscr{D}_K$ are nontrivial vector spaces when $K \Subset \Omega$ has nonempty interior.*

Proof. That these are vector spaces is clear, so to see why they're nontrivial, we will construct a nonzero $\Phi \in \mathscr{D}_K$ given a compact set K.

We start with *Cauchy's infinitely differentiable function* $\Psi : \mathbb{R} \to \mathbb{R}$ defined by

$$\Psi(x) = \begin{cases} e^{-1/x^2}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

We should prove this is infinitely differentiable, but using calculus, we have that

$$\Psi^{(m)}(x) = \begin{cases} R_m(x)e^{-1/x^2}, & x > 0 \\ 0, & x < 0, \end{cases}$$

where $R_m(x)$ is a polynomial over a power of x. This can be proven by induction. Then, using L'Hôpital's rule, one calculates that $e^{-1/x^2}/x^q \to 0$ as $x \to 0$, and so as $x \to 0$, $R_m(x)e^{-1/x^2} \to 0$ too, and so $\Psi^{(m)}(x)$ is continuous. Thus, $\Psi \in C^\infty(\mathbb{R})$.

Let $\phi(x) = \Psi(1-x)\Psi(1+x)$, which is in $C_0^\infty(\mathbb{R})$, because $\operatorname{supp}(\phi) = [-1,1]$. This is kind of strange: at -1 (and 1), it's infinitely smooth, but has a "corner." Before Cauchy found such a function, it wasn't known whether one could exist.

Now, for $x = (x_1, \dots, x_n) \in \mathbb{R}^d$, define $\Phi(x) = \phi(x_1)\phi(x_2)\cdots\phi(x_n)$, which is in $C_0^\infty(\mathbb{R}^d)$ and supported in $[-1,1]^d$. Hence, for any $K \Subset \Omega \subset \mathbb{R}^d$, one can find a box $S \subset K$, and scale and translate Φ to be supported on S, thus producing a nonzero element of $\mathscr{D}_K$. ⊠

Corollary 4.1.2. *There exist infinitely differentiable functions that aren't analytic.*

Proof. Look at $\phi(x)$ in the above proof: its Taylor series at -1 approximates the zero function. ⊠

Our next step is to determine the topology. We'll play the same game again, passing from $C^n(\Omega)$ to $C^\infty(\Omega)$ to $\mathscr{D}$. One might naïvely define the $C^n(\Omega)$ norm to be the one inherited as a subspace of $C^0(\Omega)$, but this makes differentiation an unbounded operator, as we saw in Example 1.12.2. So let's define

$$\|\phi\|_{n,\infty,\Omega} = \sum_{|\boldsymbol{\alpha}| \le n} \|D^{\boldsymbol{\alpha}}\phi\|_{L^\infty(\Omega)}.$$

That is, we sum the norms of the various partial derivatives. This is all right, but it's not always finite; we have to restrict to the subspace $C_B^n(\Omega) = \{\phi \in C^n(\Omega) : \|\phi\|_{n,\infty,\Omega} < \infty\}$. This is clearly a vector space. It's more surprising that it's a Banach space, but we know that the uniform limit of continuous functions is continuous, so if $\{\phi_n\} \subset C_B^n(\Omega)$ converges to a $\phi \in C^0(\Omega)$, then $\phi \in C^n(\Omega)$ (since all of its partials are well-behaved), and $\|D^{\boldsymbol{\alpha}}\cdot\|_{L^\infty(\Omega)}$ is continuous, so $\phi \in C_B^n(\Omega)$.

Notice also that these are nested: if $m \ge n \ge 0$, then $\|\phi\|_{m,\infty,\Omega} \ge \|\phi\|_{n,\infty,\Omega}$. This means it's not so easy to define the norm on $C^\infty(\Omega)$ as a limiting process, because the norm often goes off to ∞ itself. For example, if $\phi \in C_0^\infty(\mathbb{R})$ and $\operatorname{supp}(\phi) \subset [0,1]$, but $\phi(x) > 0$ on $(0,1)$, then the functions

$$\psi_n(x) = \sum_{j=1}^n \frac{1}{j}\phi(x-j)$$

don't have the right convergence properties. We'll figure out how to address this next time.

Lecture 35: 11/18/15

Distributions.

Recall that, in order to talk about distributions, we set up a domain $\Omega \subseteq \mathbb{R}^d$ and a function $\|\cdot\|_{n,\infty,\Omega} : C^n(\Omega) \to [0,\infty]$ defined by

$$\|\phi\|_{n,\infty,\Omega} = \sum_{|\boldsymbol{\alpha}| \le n} \|D^{\boldsymbol{\alpha}}\phi\|_{L^\infty(\Omega)}.$$

Then, we restricted to the subspace $C_B^n(\Omega)$ where his function is finite, producing a norm. We want to move to $C_0^\infty(\Omega)$, the space of compactly supported test functions on Ω, but our first idea for passing to infinity didn't work.

Example 4.2.1. For $\Omega = \mathbb{R}$, consider any $\phi \in C_0^\infty(\mathbb{R})$ with $\operatorname{supp}(\phi) = [0,1]$ and $\phi > 0$ on $(0,1)$. We explicitly produced such a function last lecture. Now, $\phi(x-j)$ is supported in $[j, j+1]$, and so

$$\psi_n(x) = \sum_{j=1}^{n} \frac{1}{j}\phi(x-j)$$

is supported on $[1, n+1]$. Thus, $\operatorname{supp}(\psi_n) \to [1,\infty)$. We can take the limit and define

$$\psi(x) = \sum_{j=1}^{\infty} \frac{1}{j}\phi(x-j),$$

which isn't compactly supported. Nonetheless, for all $m \geq 0$, $\|D^m\psi_n - D^m\psi\|_{L^\infty} \to 0$, so $\|\psi_n - \psi\|_{m,\infty,\mathbb{R}} \to 0$. But ψ isn't compactly supported, which suggests that this isn't the right topology to put on $C_0^\infty(\mathbb{R})$.

Hopefully this provides some motivation for the topology we actually put on $\mathscr{D}(\Omega)$. Well, we're going to define a notion of convergence; we won't actually describe the topology, because it's complicated and we don't need it.

Definition. Given a sequence $\{\phi_n\}_{n=1}^\infty \subset \mathscr{D}$, we say that it *converges* to a $\phi \in \mathscr{D}$, written $\phi_n \xrightarrow{\mathscr{D}} \phi$, if

(1) there exists a $K \Subset \Omega$ such that $\operatorname{supp}(\phi_j) \subseteq K$ for all j, and
(2) for all n, $\|\phi_j - \phi\|_{n,\infty,\Omega} \to 0$.

We say that $\{\phi_n\}$ is *Cauchy* if (1) holds and for all n and $\varepsilon > 0$, there's an N_n such that $\|\phi_j - \phi_k\|_{n,\infty,\Omega} < \varepsilon$ for all $j, k \geq N_n$.

So we've removed the issue with noncompactness by, well, stipulating that things need to be compact.
A good way to get your hands on this topology is to prove the following theorem.

Theorem 4.2.2. *In this topology, $\mathscr{D}$ is complete.*

Fact. However, it's important to note that $\mathscr{D}$ is *not* metrizable! Be careful with notions of convergence.

For example, *a priori* we don't know whether continuity is the same as sequential continuity. We can recover a partial result.

Theorem 4.2.3. *Let $T : \mathscr{D}(\Omega) \to \mathbb{F}$ be linear. Then, T is continuous iff it is sequentially continuous.*

This is *not* true for arbitrary functions. We won't prove this, because it requires delving into the topology, but the point is that we can think of continuity sequentially, and we will.
Now, we can (finally!) define distributions rigorously.

Definition. A *distribution*, or *generalized function*, on Ω is a (sequentially) continuous linear functional $\mathscr{D}(\Omega) \to \mathbb{F}$. The space of distributions is written $\mathscr{D}'(\Omega)$, or sometimes $\mathscr{D}^*(\Omega)$, and if $\Omega = \mathbb{R}^d$, this is sometimes abbreviated to $\mathscr{D}'$.

Theorem 4.2.4. *Suppose $T : \mathscr{D}(\Omega) \to \mathbb{F}$ is linear. Then, T is sequentially continuous iff it's sequentially continuous at $0 \in \mathscr{D}(\Omega)$.*

The proof is the same as for Proposition 1.3.1; suppose $\phi_n \to \phi$, and let $\psi_n = \phi_n - \phi$, so $\psi_n \to 0$. Thus, $T(\psi_n) \to 0$, i.e. $T(\phi_n) \to T(\phi)$.
This next theorem is a bit more interesting. The point is that, as in normed spaces, a linear functional is continuous iff it's bounded, but we need to define bounded a little differently, using compact sets again.

Theorem 4.2.5. *Let $T : \mathscr{D}(\Omega) \to \mathbb{F}$ be linear. Then, $T \in \mathscr{D}'(\Omega)$ iff for all $K \Subset \Omega$, there exist $n_K \geq 0$ and $C_K > 0$ such that for all $\phi \in \mathscr{D}_K$, $|T(\phi)| \leq C_K \|\phi\|_{n_k,\infty,\Omega}$.*

Proof. Suppose otherwise, so that $T \in \mathscr{D}'(\Omega)$ but there exists a $K \Subset \Omega$ such that for all $n \geq 0$ and $m \geq 1$, there's a $\phi_{n,m} \in \mathscr{D}_K$ such that $|T(\phi_{n,m})| > m\|\phi_{n,m}\|_{n,\infty,\Omega}$.
We should first rescale these $\phi_{n,m}$: define

$$\widehat{\phi}_j = \frac{1}{j\|\phi_{j,j}\|_{j,\infty,\Omega}}\phi_{j,j},$$

so that $\widehat{\phi}_j \in \mathscr{D}_K$ and $|T(\widehat{\phi}_j)| > 1$, but $\widehat{\phi}_j \to 0$ in $\mathscr{D}(\Omega)$: after all, $\widehat{\phi}_j \in \mathscr{D}_K$, and when $j > n$,

$$\|\widehat{\phi}_j\|_{n,\infty,\Omega} \le \|\widehat{\phi}_j\|_{j,\infty,\Omega} = 1/j \to 0,$$

so we've satisfied the two conditions for convergence. But this is a contradiction: it means T cannot be continuous, since $|T(\phi_j)| \not\to 0$.

Conversely, suppose $\phi_j \to 0$ in $\mathscr{D}(\Omega)$, so that there's a $K \Subset \Omega$ with $\phi_j \in \mathscr{D}_K$, and $\|\phi_j\|_{n,\infty,\Omega} \to 0$ for all n. By hypothesis, for this K, we have n and C such that $|T(\phi_j)| \le C\|\phi_j\|_{n,\infty,\Omega} \to 0$, so $T(\phi_j) \to 0$, so it's sequentially continuous at 0, so by Theorem 4.2.4, it's continuous everywhere. ⊠

So we have an idea for what a distribution is: it a bounded linear functional, but in a sense of "bounded" that may be less familiar. Let's talk about examples!

Definition. Let $\Omega \subset \mathbb{R}^d$ and define $L^1_{\mathrm{loc}}(\Omega)$ to be the space of functions that are *locally* L^1, the measurable functions $f : \Omega \to \mathbb{F}$ such that for all $K \Subset \Omega$, $\int_K |f(x)|\,\mathrm{d}x$ is finite, where as usual we identify functions that differ only on a set of measure 0.

By comparing the definitions, $L^1(\Omega) \subset L^1_{\mathrm{loc}}(\Omega)$. More interestingly, if Ω is unbounded, polynomials are generally not in $L^1(\Omega)$, but they are all in $L^1_{\mathrm{loc}}(\Omega)$. Lots of functions are locally L^1.

Example 4.2.6. If $f \in L^1_{\mathrm{loc}}(\Omega)$, define a distribution $\Lambda_f \in \mathscr{D}'(\Omega)$ by

$$\Lambda_f(\phi) = \int_\Omega f(x)\phi(x)\,\mathrm{d}x$$

for all $\phi \in \mathscr{D}(\Omega)$. Notice that since ϕ is compactly supported,

$$|\Lambda_f(\phi)| \le \int_{\mathrm{supp}(\phi)} |f(x)|\|\phi\|_{0,\infty,\Omega}\,\mathrm{d}x, \tag{4.1}$$

which is finite. Thus, for any $K \Subset \Omega$, let $n_K = 0$ and $C_K = \int_K |f(x)|\,\mathrm{d}x$, so (4.1) shows that the definition of a distribution is satisfied.

Hence the term "generalized functions:" this rather large space of functions provides us with a large class of distributions. Note that we will often use f to denote the distribution Λ_f.

Lemma 4.2.7 (Lesbegue)**.** *If $f, g \in L^1_{\mathrm{loc}}(\Omega)$, then $\Lambda_f = \Lambda_g$ iff $f = g$ almost everywhere.*

We will prove this next time.

Lecture 36: 11/20/15

Examples of and Operations on Distributions.

Recall that we defined distributions as linear functionals $T : \mathscr{D}(\Omega) \to \mathbb{F}$. Theorem 4.2.5 tells us that $T \in \mathscr{D}'(\Omega)$ iff for all $K \Subset \Omega$, there exist n and C such that for all $\phi \in \mathscr{D}_K$, $|T(\phi)| \le C\|\phi\|_{n,\infty,\Omega}$.

Then, we defined $L^1_{\mathrm{loc}}(\Omega)$ as the space of equivalence classes of locally L^1 functions under equality almost everywhere, where a locally L^1 function is a measurable $f : \Omega \to \mathbb{F}$ such that for all $K \Subset \Omega$, $\int_K |f(x)|\,\mathrm{d}x$ is finite. In particular, there is a map $\Lambda : L^1_{\mathrm{loc}}(\Omega) \to \mathscr{D}'(\Omega)$ sending a function f to

$$\Lambda_f(\phi) = \int_\Omega f(x)\phi(x)\,\mathrm{d}x.$$

The Lesbegue lemma (Lemma 4.2.7) states that Λ is injective.

Proof of Lemma 4.2.7. We want to prove that if $f, g \in L^1_{\mathrm{loc}}(\Omega)$, then $\Lambda_f = \Lambda_g$ iff $f = g$ almost everywhere.

The reverse direction is easy: if $f = g$ almost everywhere, then so do $f\phi$ and $g\phi$, so their integrals are the same.

In the forward direction, we want to show that if $\Lambda_f = \Lambda_g$, then $\int_R (f - g)(x)\,\mathrm{d}x = 0$ for all rectangles $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$; this implies $f = g$ almost everywhere, by taking limits.

We know that if $\Lambda_f = \Lambda_g$, then $\Lambda_{f-g} = 0$, since Λ is linear. We'd like to compute $\Lambda_{f-g}(\chi_R)$ for our rectangle R, where χ_R is the *characteristic function* for our rectangle R:

$$\chi_R(x) = \begin{cases} 1, & x \in R \\ 0, & x \notin R. \end{cases}$$

This isn't even continuous! So it's not a test function. Instead, we'll approximate χ_R by test functions.

Let $\varepsilon > 0$ and define $\phi_\varepsilon(x) = \psi(\varepsilon - x)\psi(x)$, where ψ was Cauchy's nowhere differentiable function. This is in $C_0^\infty(\Omega)$ and is supported on $[0, \varepsilon]$, looking like a smooth bump in that interval. If we think of this as a PDF, its CDF is

$$\Phi_\varepsilon(x) = \frac{\int_{-\infty}^{x} \phi_\varepsilon(\xi)\,\mathrm{d}\xi}{\int_{-\infty}^{\infty} \phi_\varepsilon(\xi)\,\mathrm{d}\xi}.$$

This is a function which is 0 when $x \le 0$, 1 when $x \ge \varepsilon$, and smoothly joins them on $[0, \varepsilon]$. Next, define

$$\Psi_\varepsilon(x) = \prod_{j=1}^{n} \Phi_\varepsilon(x_j - a_j)\Psi_\varepsilon(b_j - x_j),$$

which is in $\mathscr{D}(\Omega)$, and $\operatorname{supp}(\Psi_\varepsilon) \subseteq R$. Pictorially, take a frame of width ε inside the boundary of R; outside of R, Ψ_ε is 0, and inside the frame, it's 1, but it is smooth. In particular, as $\varepsilon \to 0$, $\Psi_\varepsilon(x) \to \chi_R(x)$ pointwise.

By the dominated convergence theorem,

$$0 = \Lambda_{f-g}(\Psi_\varepsilon) = \int_R (f-g)(x)\Psi_\varepsilon(x)\,\mathrm{d}x \longrightarrow \int_R (f-g)(x)\chi_R(x)\,\mathrm{d}x = \int_R (f-g)\,\mathrm{d}x.$$

Thus, $\int f - g\,\mathrm{d}x$ is 0 on every rectangle, so $f = g$ almost everywhere. ⊠

Definition. Let $T \in \mathscr{D}'(\Omega)$.

- If $T = \Lambda_f$ for some $f \in L^1_{\mathrm{loc}}(\Omega)$, then T is called a *regular distribution*.
- If otherwise, T is a *singular distribution*.

Regular distributions are often identified with their associated $L^1_{\mathrm{loc}}(\Omega)$ functions in an abuse of notation, e.g. $f(\phi) = \int f(x)\phi(x)\,\mathrm{d}x$. Even though not all distributions are regular, some people (*cough* physicists *cough*) will write any distribution T as

$$T(\phi) = \int T(x)\phi(x)\,\mathrm{d}x.$$

A more rigorous way to write this is to take advantage of the pairing $\mathscr{D}'(\Omega) \times \mathscr{D}(\Omega) \to \mathbb{R}$ sending $(T, \phi) \mapsto T(\phi)$. This isn't an inner product, but behaves enough like one that we'll write

$$T(\phi) = \langle T, \phi \rangle = \langle T, \phi \rangle_{\mathscr{D}', \mathscr{D}}.$$

Example 4.3.1. If $0 \in \Omega$, define $\delta_0 : \mathscr{D}(\Omega) \to \mathbb{F}$ by $\langle \delta_0, \phi \rangle = \phi(0)$. Clearly this is linear, and $|\langle \delta_0, \phi \rangle| \le \|\phi\|_{0,\infty,\Omega}$, so for $C = 1$ and $n = 0$, this satisfies the condition in Theorem 4.2.5 and therefore is a distribution.

δ_0 goes by many names, including the *Dirac distribution*, *Dirac mass*, *Dirac measure*, and *Dirac delta function*.

Nonetheless, no $f \in L^1_{\mathrm{loc}}(\Omega)$ satisfies $\delta_0 = \Lambda_f$: if it did, f would have to be 0 everywhere except 0, and therefore by continuity, would be 0, but $\delta_0 \ne 0$. This motivates the term "singular distribution," since it's only interesting at a single point.

For any $x \in \Omega$, one can define the distribution δ_x by $\langle \delta_x, \phi \rangle = \phi(x)$. Again, people will write $\delta_x(\xi) = \delta_0(x - \xi) = \delta_0(\xi - x)$, but this is technically wrong. If you think of these as generalized functions, though, you won't get that confused.

Theorem 4.3.2. *$\mathscr{D}(\Omega)$ is not metrizable.*

Proof sketch. Let $K \Subset \Omega$. We can write

$$\mathscr{D}_K = \bigcap_{x \in \Omega \setminus K} \ker(\delta_x),$$

so $\mathscr{D}_K$ is an intersection of closed sets, and therefore is closed.

However, one can show that $\mathscr{D}_K$ has empty interior: for any $f \in \mathscr{D}_K$, one can define f_ε to be equal to f on K, and to have a bump of measure at most ε on a domain outside of K.

The next step is to show that Ω can be approximated by compact sets: there exist compact $K_1 \subset K_2 \subset \cdots \subset \Omega$ such that

$$\bigcup_{n=1}^{\infty} K_n = \Omega \quad\implies\quad \mathscr{D}(\Omega) = \bigcup_{n=1}^{\infty} \mathscr{D}_{K_n}.$$

Now, $\mathscr{D}(\Omega)$ is a countable union of nowhere dense sets, so you can use the Baire category theorem and conclude $\mathscr{D}(\Omega)$ isn't metrizable. ⊠

Example 4.3.3.

(1) Let μ be a complex Borel measure[4] or a positive measure on Ω such that $\mu(K)$ is finite for all $K \Subset \Omega$. Then, one can define a distribution

$$\Lambda_\mu(\phi) = \int_\Omega \phi(x)\,\mathrm{d}\mu.$$

Since $|\Lambda_\mu(\phi)| \le \mu(\operatorname{supp}(\phi))\|\phi\|_{0,\infty,\Omega}$, then this is really a distribution. And unless $\mathrm{d}\mu = f\,\mathrm{d}x$ for an $f \in L^1_{\mathrm{loc}}(\Omega)$, this is a singular distribution, and there are plenty of good examples where this is the case.

(2) $1/x \notin L^1_{\mathrm{loc}}(\mathbb{R})$, but we would still like to integrate against it. Thus, we define its *principal value*, denoted $\mathrm{PV}(1/x)$, to be the distribution

$$\langle \mathrm{PV}(1/x), \phi\rangle = \mathrm{PV}\int_{-\infty}^{\infty} \frac{1}{x}\phi(x)\,\mathrm{d}x = \lim_{\varepsilon\to 0^+}\int_{|x|>\varepsilon} \frac{1}{x}\phi(x)\,\mathrm{d}x.$$

In other words, we're trying to avoid the asymptote as best as we can.

To compute this, we will differentiate ϕ and use integration by parts.

$$\begin{aligned}\int_\varepsilon^\infty \frac{\phi(x)\,\mathrm{d}x}{x} + \int_{-\infty}^{-\varepsilon} \frac{\phi(x)\,\mathrm{d}x}{x} &= \ln|x|\phi(x)\Big|_\varepsilon^{-\varepsilon} - \int_{-\infty}^{\infty} \ln|x|\phi'(x)\,\mathrm{d}x \\ &= \underbrace{2\varepsilon\ln\varepsilon}_{\to 0}\underbrace{\frac{\phi(-\varepsilon)-\phi(\varepsilon)}{2\varepsilon}}_{\to -\phi'(0)} - \lim_{R\to\infty}\int_{-R}^{R} \ln|x|\phi'(x)\,\mathrm{d}x.\end{aligned}$$

In particular, there's some R for which

$$|\langle \mathrm{PV}(1/x), \phi\rangle| \le \int_{-R}^{R} |\ln x|\,\mathrm{d}x\|\phi\|_{1,\infty,\Omega}.$$

So $K = [-R, R]$ depends on f, and for the first time, $n = 1$, but this is a bound, so $\mathrm{PV}(1/x)$ is indeed a distribution.

Notice that the computation for $\mathrm{PV}(1/x)$ uses the one thing we know about test functions: we can always differentiate them. If you get stuck on a problem, it's worth trying to apply this fact.

We want to produce more examples of distributions, but in a general way, by defining operations on $\mathscr{D}'(\Omega)$. First, suppose $T : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ is sequentially continuous and linear; then, look at its adjoint $T^*u = u \circ T$, which is a continuous linear operator $T^* : \mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega)$, which has the property $\langle u, T\phi\rangle = \langle T^*u, \phi\rangle$.

In particular, we get another way to check for distributions: if a mapping $f : \mathscr{D}(\Omega) \to \mathbb{F}$ is in the image of T^* for some $T : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$, then f must be a distribution.

Proposition 4.3.4. *If $u \in \mathscr{D}'(\Omega)$ and $T : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ is a sequentially continuous, linear map, then $T^*u = u \circ T \in \mathscr{D}'(\Omega)$.*

As a very good example, consider multiplication by any smooth function, compactly supported or not: for $f \in C^\infty(\Omega)$, let $T_f : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ be defined by $T_f(\phi) = f\phi$. Then, if $\phi_n \to \phi$ and each ϕ_n is supported in K and $\|\phi_n - \phi\|_{j,\infty,\Omega} \to 0$ for all j, then $T_f(\phi_n) \to Tf(\phi)$.

Thus, for any $u \in \mathscr{D}'(\Omega)$, we can define $fu = T_f^*u = u \circ T_f$. If $u = \Lambda_u$ is a regular distribution, then $fu = \Lambda_{fu}$, because

$$fu(\phi) = u(T_f\phi) = u(f\phi) = \int_\Omega u(x)f(x)\phi(x)\,\mathrm{d}x.$$

[4] A *complex Borel measure* μ on a set Ω is a function $\mathscr{P}(\Omega) \to \mathbb{C}$ that is countably additive. Note that this means it has to be finite on all sets.

The general case of singular distributions is analogous, even if the above equation isn't literally true.

Another very important operation is differentiation: $D^{\boldsymbol{\alpha}} : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ is linear and sequentially continuous — tune in next time to learn more.

Lecture 37: 11/23/15

Differentiation of Distributions.

"I thought I would go over the exam on Wednesday, because some of you will be gone..."

Last time, we talked about Theorem 4.3.4, which states that if $T : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ is linear and sequentially continuous, then for all distributions $u \in \mathscr{D}'(\Omega)$, $u \circ T = T^* u \in \mathscr{D}'(\Omega)$ as well.

Suppose $\phi_n \to \phi$ in $\mathscr{D}(\Omega)$, i.e. there's a $K \Subset \Omega$ such that $\operatorname{supp}(\phi_n) \subset K$ for all n, and $\|\phi_n - \phi\|_{j,\infty,\Omega} \to 0$ for all j as $n \to \infty$.

If $\boldsymbol{\alpha}$ is any multi-index, then $D^{\boldsymbol{\alpha}} : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ is linear. Clearly, $\operatorname{supp} D^{\boldsymbol{\alpha}}\phi \subseteq K$, and

$$\|D^{\boldsymbol{\alpha}}(\phi_n - \phi)\|_{j,\infty,\Omega} \le \|\phi_n - \phi\|_{j+|\boldsymbol{\alpha}|,\infty,\Omega} \longrightarrow 0.$$

Thus, differentiation of distributions $D^{\boldsymbol{\alpha}} u = u \circ D^{\boldsymbol{\alpha}}$ still produces distributions.

If $u = \Lambda_u$ is a regular distribution, then $\langle u \circ D^{\boldsymbol{\alpha}}, \phi\rangle = \langle u, D^{\boldsymbol{\alpha}}\phi\rangle = (-1)^{|\boldsymbol{\alpha}|}(D^{\boldsymbol{\alpha}} u, \phi)$. This motivates the more general definition for all distributions.

Definition. If $u \in \mathscr{D}'(\Omega)$, define its *derivative* $D^{\boldsymbol{\alpha}} u$ to be the distribution acting by $\langle D^{\boldsymbol{\alpha}} u, \phi\rangle = (-1)^{|\boldsymbol{\alpha}|}\langle u, D^{\boldsymbol{\alpha}}\phi\rangle$.

Proposition 4.4.1. *If $u \in C^{|\boldsymbol{\alpha}|}(\Omega)$, then $D^{\boldsymbol{\alpha}}\Lambda_u = \Lambda_{\partial^{\boldsymbol{\alpha}} u}$, i.e. differentiation acts as the ordinary derivative.*

PROOF. Since u is nice, we can integrate by parts. For any $\phi \in \mathscr{D}(\Omega)$,

$$\begin{aligned}
\langle D^{\boldsymbol{\alpha}}\Lambda_u, \phi\rangle &= (-1)^{|\boldsymbol{\alpha}|}\langle \Lambda_u, D^{\boldsymbol{\alpha}}\phi\rangle \\
&= (-1)^{|\boldsymbol{\alpha}|}\int_\Omega u(x) D^{\boldsymbol{\alpha}}\phi(x)\,\mathrm{d}x \\
&= (-1)^{|\boldsymbol{\alpha}|}(-1)^{|\boldsymbol{\alpha}|}\int_\Omega \partial^{\boldsymbol{\alpha}} u(x)\phi(x)\,\mathrm{d}x = \langle \Lambda_{\partial^{\boldsymbol{\alpha}} u}, \phi\rangle. \qquad \boxtimes
\end{aligned}$$

This is good; it means we've probably defined differentiation correctly.

Example 4.4.2. Let's look at some singular examples.

(1) Remember the Heaviside step function? It's the function

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0. \end{cases}$$

Then, for any test function ϕ,

$$\langle H', \phi\rangle = -\langle H, \phi'\rangle = -\int_0^\infty \phi'(x)\,\mathrm{d}x = -\phi\Big|_0^\infty = \phi(0),$$

so $H' = \delta_0$, as distributions, making rigorous the idea we entertained at the start of the chapter.

(2) $\ln|x| \in L^1_{\mathrm{loc}}(\mathbb{R})$, so let's compute its derivative.

$$\begin{aligned}
\langle D\ln|x|, \phi\rangle &= -\langle \ln|x|, \phi'\rangle \\
&= -\int_{-\infty}^\infty \ln|x|\phi'(x)\,\mathrm{d}x \\
&= \lim_{a,b\to 0}\left(\int_a^\infty \ln|x|\phi'(x)\,\mathrm{d}x + \int_{-\infty}^b \ln|x|\phi'(x)\,\mathrm{d}x\right) \\
&= \lim_{a,b\to 0}\left(\int_a^\infty \frac{\phi(x)\,\mathrm{d}x}{x} + \int_{-\infty}^b \frac{\phi(x)\,\mathrm{d}x}{x} - \ln|x|\,\mathrm{d}x\Big|_a^b\right) \\
&= \mathrm{PV}\int \frac{\phi(x)\,\mathrm{d}x}{x},
\end{aligned}$$

so the derivative of $\ln|x|$, in the sense of distributions, is $\mathrm{PV}(1/x)$. This computation might be a little confusing, but ultimately uses an argument from symmetry.

Since mixed partials commute for C^∞ test functions, then they also do so for distributions.

Proposition 4.4.3. *If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are multi-indices and $u \in \mathscr{D}'(\Omega)$, then $D^{\boldsymbol{\alpha}} D^{\boldsymbol{\beta}} u = D^{\boldsymbol{\beta}} D^{\boldsymbol{\alpha}} u = D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} u$.*

Proof. Let's just compute from the definition; let ϕ be a test function.

$$\langle D^{\boldsymbol{\alpha}} D^{\boldsymbol{\beta}} u, \phi \rangle = (-1)^{|\boldsymbol{\alpha}|}(-1)^{|\boldsymbol{\beta}|}\langle u, D^{\boldsymbol{\beta}} D^{\boldsymbol{\alpha}} \phi \rangle,$$

but for test functions, $D^{\boldsymbol{\alpha}} D^{\boldsymbol{\beta}} = D^{\boldsymbol{\beta}} D^{\boldsymbol{\alpha}} = D^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$, so all three derivatives agree on distributions. ⊠

Definition. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are multi-indices, then $\boldsymbol{\alpha}! = \alpha_1!\alpha_2!\cdots\alpha_d!$ and

$$\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = \frac{\boldsymbol{\alpha}!}{(\boldsymbol{\alpha}-\boldsymbol{\beta})!\boldsymbol{\beta}!},$$

and one says that $\boldsymbol{\beta} \le \boldsymbol{\alpha}$ if $\beta_i \le \alpha_i$ for all i.

We haven't used the binomial coefficient much in this class, so recall that if $b > a$, then $\binom{a}{b} = 0$. Also, thinking of it as the number of ways to choose j objects out of n, one can recursively expand this as

$$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}. \tag{4.2}$$

Lemma 4.4.4 (Leibniz rule). *Let $f \in C^\infty(\Omega)$ and $u \in \mathscr{D}'(\Omega)$; then,*

$$D^{\boldsymbol{\alpha}}(fu) = \sum_{\boldsymbol{\beta}\le\boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} D^{\boldsymbol{\alpha}-\boldsymbol{\beta}} f D^{\boldsymbol{\beta}} u.$$

This might be less surprising in one variable, where it boils down to a more familiar-looking product rule.

Proof. We know this is true in one variable, so we're going to induct from that. Without loss of generality, using Theorem 4.4.3 we can assume $\boldsymbol{\alpha} = (n, 0, \dots, 0)$. For $n = 0$ there's nothing to show; for $n = 1$ we calculate for any test function ϕ that

$$\begin{aligned}\langle D_1(fu), \phi\rangle &= -\langle fu, D_1\phi\rangle = -\langle u, fD_1\phi\rangle \\ &= -\langle u, D_1(f\phi) - (D_1 f)\phi\rangle \\ &= \langle D_1 u, f\phi\rangle + \langle u, (D_1 f)\phi\rangle = \langle fD_1 u + (D_1 f)u, \phi\rangle.\end{aligned}$$

For arbitrary n, we don't even need a test function, but the computation looks a little scarier. Suppose it's true for $n-1$; then,

$$\begin{aligned}D_1^n(fu) &= D_1 D_1^{n-1}(fu) \\ &= D_1 \sum_{j=0}^{n-1} \binom{n-1}{j} D_1^{n-1-j} f D_1^j u \\ &= \sum_{j=0}^{n-1} \binom{n-1}{j} \Big(D_1^{n-j} f D + 1^j u + D_1^{n-1-j} f D_1^{j+1} u \Big) \\ &= \sum_{j=0}^{n-1} \binom{n-1}{j} D_1^{n-j} f D_1^j u + \sum_{j=1}^{n} \binom{n-1}{j-1} D_1^{n-j} f D_1^j u \\ &= \sum_{j=0}^{n} \left(\binom{n-1}{j} + \binom{n-1}{j-1} \right) D_1^{n-j} f D_1^j u,\end{aligned}$$

and using (4.2), this simplifies to what we were looking for. ⊠

Hopefully you've noticed the theme: these proofs involve passing to differentiation of test functions, which we understand; doing a bunch of computation; and then returning to the world of distributions.

Example 4.4.5. Let's consider the *distribution* $f(x) = x\ln|x|$. Then, by the Leibniz rule. $D(x\ln|x|) = \ln|x| + x\,\mathrm{PV}(1/x)$. What's $x\,\mathrm{PV}(1/x)$? Well, for any test function ϕ,

$$\begin{aligned}\langle D(x\ln|x|), \phi\rangle &= -\langle x\ln|x|, \phi'\rangle \\ &= -\int_{-\infty}^{\infty} x\ln|x|\phi'(x)\,\mathrm{d}x.\end{aligned}$$

Integrating by parts,

$$= \int_{-\infty}^{\infty}\left(\ln|x| + \frac{x}{x}\right)\phi(x)\,\mathrm{d}x = \langle \ln|x| + 1, \phi\rangle,$$

so $x\,\mathrm{PV}(1/x) = 1$, in the sense of distributions ($x \in C^\infty(\mathbb{R})$, but $\mathrm{PV}(1/x) \in \mathscr{D}'(\mathbb{R})$).

Two more useful operations to have are translation and dilation, which once again will be defined on test functions and passed to distributions. In this setting, we'll need $\Omega = \mathbb{R}^d$; suppose $x \in \mathbb{R}^d$ and $\lambda \in \mathbb{R} \setminus \{0\}$. Then, we have a *translation* operator $\tau_x : \mathscr{D}(\mathbb{R}^d) \to \mathscr{D}(\mathbb{R}^d)$ sending $\phi(y) \mapsto \phi(y - x)$ and a *dilation* operator $T_\lambda : \mathscr{D}(\mathbb{R}^d) \to \mathscr{D}(\mathbb{R}^d)$ sending $T_\lambda\phi(y) \mapsto \phi(\lambda y)$.

Both of these are linear and sequentially continuous, so we can extend them to distributions, defining $\langle \tau_x u, \phi\rangle = \langle u, \tau_{-x}\phi\rangle$, because

$$\int_{\mathbb{R}^d} \tau_x u(y)\phi(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} u(y-x)\phi(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} u(z)\phi(x+z)\,\mathrm{d}z.$$

Analogously, we define $\langle T_\lambda u, \phi\rangle = (1/|\lambda|^d)\langle u, T_{1/\lambda}\phi\rangle$; you can check that this is the right definition in the same way as for translations.

But this is particularly important so that we can define convolutions. Recall that if $f, g : \mathbb{R}^d \to \mathbb{F}$ are integrable functions, then their *convolution* was defined to be

$$f * g(x) = \int_{\mathbb{R}^d} f(y)g(x-y)\,\mathrm{d}y.$$

Substituting in $z = x - y$, this is also $g * f(x)$, so it's symmetric. We can rewrite this definition as

$$f * g(x) = \int_{\mathbb{R}^d} f(y)(\tau_x T_{-1} g)(y)\,\mathrm{d}y,$$

since $\tau_x T_{-1} g(y) = (T_{-1}g)(y - x) = g(x - y)$. This redefinition of convolution allows us to apply it to distributions — sort of.

If $u \in \mathscr{D}'(\mathbb{R}^d)$ and $\phi \in \mathscr{D}(\mathbb{R}^d)$, then we'll let

$$(u * \phi)(x) = \langle u, \tau_x T_{-1}\phi\rangle = \langle T_{-1}\tau_{-x}u, \phi\rangle.$$

Notice that this is a distribution and a test function, *not* two distributions. Sorry.

Example 4.4.6. δ_0 acts as the identity for convolution:

$$\delta_0 * \phi(x) = \langle \delta_0, \tau_x T_{-1}\phi\rangle = (\tau_x T_{-1}\phi)(0) = T_{-1}\phi(-x) = \phi(x),$$

i.e. $\delta_0 * \phi = \phi$.

In general, convolution smooths things out, though δ-functions are an the exception.

Lecture 38: 11/25/15

Midterm 2 Review.

Today we're going to go over the second midterm.

For the first problem, we had Hilbert spaces X and Y, where $X \subseteq Y$ is a linear subspace and the unit ball in X is compact in Y. Recall that the Banach-Alaoglu theorem, Theorem 1.14.2, says that if X is a normed linear space, then the unit ball in X^* is weak-$*$ compact, which was part (a).

For part (b), suppose $x_n \rightharpoonup x$ in X. Then, $\|x_n\|_X$ is bounded by some C, so $\{x_n\}$ is contained in the closed unit ball of radius C in X, which is compact in Y because the unit ball is (you can show this directly by scaling). Thus, there's a subsequence converging strongly: $x_{n_k} \to y$, but this means $x_{n_k} \rightharpoonup y$ in the sense of Y and therefore also in the sense of X (since $X^* \subseteq Y^*$), so since weak limits are unique, then $x = y$. In particular, since the closed ball of radius C in X is compact in Y, hence closed in Y, then $y \in X$.

For part (c), we have $u_n \rightharpoonup u$ in Y and $v_n \to v$ in Y, so we want to determine what $\langle u_n, v_n\rangle$ converges to. The simplest way to do this is to separate them:

$$\begin{aligned}\langle u_n, v_n\rangle_Y - \langle u, v\rangle &= \langle u_n, v_n - v\rangle + \langle u_n - u, v\rangle \\ &\le \|u_n\|\|v_n - v\| + \langle u_n - u, v\rangle \\ &\le C\underbrace{\|v_n - v\|}_{\to 0} + \underbrace{\langle u_n - u, v\rangle}_{\to 0}.\end{aligned}$$

This is an interesting thing: the inner product of two strongly convergent spaces converges, of course, and we've just shown that the inner product of a strongly convergent sequence and a weakly convergent sequence converges. However, the inner product of two weakly convergent sequences does *not* always converge like this; we earlier constructed a sequence $\{e_n\}$ with $\|e_n\| = 1$ and $e_n \rightharpoonup 0$, so $\langle e_n, e_n\rangle = 1$ for all n, which doesn't converge to 0. This is a relatively common mistake.

It's possible to generalize this to a Banach space X, with one sequence in X and one in X^*.

For part (d), we have $T : Y \to X$, $x_n \rightharpoonup x$ in X and $y_n \rightharpoonup y$ in Y. We want to show that there's a subsequence $\langle x_{n_k}, Ty_{n_k}\rangle$ that converges to $\langle x, Ty\rangle$. First, we know that $\langle x_n, Ty_n\rangle_Y = \langle T^*x_n, y_n\rangle_Y$, and since $x_n \rightharpoonup x$, then $T^*x_n \to T^*x$ in Y. Then, by part (b), $T^*x_{n_k} \to T^*x$ in Y, so by part (c), $\langle x_{n_k}, Ty_{n_k}\rangle \to \langle x, Ty\rangle$. The point is that by using the compact embedding, we can work with two weakly convergent subsequences.

In problem 2, we had a Hilbert space bounded self-adjoint operators K and T_n that all commute with each other, and $T_1 \le T_2 \le \cdots \le K$.[5]

For part (a), if $S_n = K - T_n$, then we want to prove that $\langle S_m^2 x, x\rangle \ge \langle S_m S_n x, x\rangle \ge \langle S_n^2 x, x\rangle \ge 0$. First, since K and the T_n all commute and are self-adjoint, then the S_n commute with everything, including each other, and are all positive. The last inequality is thus the simplest: that S_n is self-adjoint means $\langle S_n^2 x, x\rangle = \|S_n x\| \ge 0$.

For the first inequality, we want to understand $S_m^2 - S_m S_n = (K - T_m)(T_n - T_m)$, which is a composition of positive, self-adjoint operators which commute, and by the hint this is a positive operator. The middle inequality is exactly the same: $S_n(S_m - S_n) = S_n(T_n - T_m)$, which once again is a produce of two positive, self-adjoint operators that commute, so it's positive.

We'll use this in part (b), where we want to show that $S_n x$ and $T_n x$ converge for all $x \in H$. You can use one to prove the other, and people started with either, since $T_n = K - S_n$, so if $T_n x$ converges, then $S_n x$ does too. We'll show that $S_n x$ converges: $\langle S_n x\rangle^2 = \langle S_n x, S_n x\rangle = \langle S_n^2 x, x\rangle \ge 0$, so it's a monotone decreasing sequence that's bounded below, and therefore must converge. But we also need convergence directly, so let's show that this is Cauchy:

$$\begin{aligned}\|(S_n - S_m)x\|^2 &= \|S_n x^2\| + \langle S_n S_m x, x\rangle + \|S_m x^2\| \\ &\le \|S_n x\|^2 + \|S_m x\|^2 - 2\langle S_n x, x\rangle \\ &= \|S_m x\| - \|S_n x\| \longrightarrow 0,\end{aligned}$$

since we have convergence of norm, and thus $S_n x$ converges, and $T_n x$ does too.

For part (c), we can define T by $Tx = \lim_{n\to\infty} T_n x$ pointwise. Clearly, T is linear. It's also self-adjoint: $\langle T_n x, y\rangle \to \langle Tx, y\rangle$, but this sequence is also $\langle x, T_n y\rangle \to \langle x, Ty\rangle$, so $T = T^*$. We can also know that $\langle T_n x, x\rangle \le \langle Kx, x\rangle$, since $T_n \le K$, and therefore passing to the limit, $\langle Tx, x\rangle \le \langle Kx, x\rangle$, so $T \le K$. Finally, why is T bounded? We need $\sup_{\|x\|=1}\|T_n x\| \le C$. Here the uniform boundedness principle saves us: $T_n x$ is pointwise bounded, so it has to be uniformly bounded, and therefore T is a bounded linear functional.

For question 3, suppose H is a nontrivial Hilbert space and T is a self-adjoint operator on H. Then, let U be its *Cayley transform*: $U = (T - iI)(T + iI)^{-1}$. This is motivated from a result in complex analysis, but applies to operators too.

Part (a) asks, why is this well-defined? Well, the spectral theorem for self-adjoint operators tells us that $\sigma(T) \subset \mathbb{R}$, so $\pm i \in \rho(T)$, and therefore $(T \pm iI)^{-1}$ is well-defined, so U is too.

For part (b), we want to show that U is *unitary*, i.e. $U^* = U^{-1}$. First, you can calculate $U^{-1} = (T + iI)(T - iI)^{-1}$, and the dual

$$U^* = ((T + iI)^{-1})^*(T - iI)^* = ((T + iI)^*)^{-1}(T - iI)^* = (T - iI)^{-1}(T + iI).$$

[5] Recall that $A \le B$ iff $\langle Ax, x\rangle \le \langle Bx, x\rangle$ for all $x \in H$; using the monotonicity directly is not valid.

We're almost there: $U^*U = (T-iI)^{-1}(T+iI)(T-iI)(T+iI)^{-1}$, but T commutes with itself and with $\pm iI$, so $(T-iI)(T+iI) = T^2+I$, so we can switch them and things cancel, giving us I. Thus, since the inverse is unique, $U^* = U^{-1}$.

In part (c), we want to prove that if V is a unitary operator, then it's an isometry. This isn't hard: $\|Vx\|^2 = \langle Vx, Vx\rangle = \langle V^*Vx, x\rangle = \langle x, x\rangle = \langle x^2\rangle$ for all $x\in H$ (which uses that H is nontrivial). These two parts maybe indicate why unitary operators are pretty important.

For part (d), we want to show that if $\lambda\in\sigma(V)$, then $|\lambda|=1$; since $\|V\|=1$, then the spectrum is bounded by 1. But V^* is also unitary, and the resolvent operator $V-\lambda I$ is $V(I-\lambda V^{-1}) = \lambda V((1/\lambda)I - V^*)$, so the spectral values of V^{-1} are the reciprocals of the spectral values of V.[6] And since V^* is unitary, then $|1/\lambda|\le 1$. There are other ways to solve this more explicitly.

We also need to show that $1\notin\sigma_p(U)$. If it were, then $(U-I)x=0$ for a nonzero x, so $Ux=x$, and therefore $(T-iI)(T+iI)^{-1}x = x$. Letting $y=(T+iI)^{-1}x$, this means $(T+iI)y = (T-iI)y$, so $-y=y$, or $y=0$. Thus, $x=(T-iI)y=0$ too.

Lecture 39: 11/30/15

Convolution of Distributions.

Recall that we defined a translation operator for an $x\in\mathbb{R}^d$: $\tau_x : \mathscr{D}(\mathbb{R}^d)\to\mathscr{D}(\mathbb{R}^d)$ sends $f(t)\mapsto f(t-x)$, and a reflection operator $R=T_{-1}$ sending $g(x)\mapsto g(-x)$. Thus, the convolution of two functions can be expressed as

$$\begin{aligned}(f*g)(x) &= \int_{\mathbb{R}^d} f(y)g(x-y)\,\mathrm{d}y\\ &= (g*f)(x) = (f, \tau_x Rg).\end{aligned}$$

This motivated the notion of the convolution of a distribution $u\in\mathscr{D}'$ and a test function $\phi\in\mathscr{D}$: $(u*\phi)(x) = \langle u, \tau_x R\phi\rangle$. For example, $\delta_0*\phi=\phi$.

Proposition 4.6.1. *Let $u\in\mathscr{D}'$ and $\phi\in\mathscr{D}$.*

(1) Convolution commutes with translation: $\tau_x(u*\phi) = (\tau_x u)*\phi = u*\tau_x\phi$.
(2) $u*\phi\in C^\infty(\mathbb{R}^d)$, and if $\boldsymbol{\alpha}$ is a multi-index, then $D^{\boldsymbol{\alpha}}(u*\phi) = (D^{\boldsymbol{\alpha}}u)*\phi = u*(D^{\boldsymbol{\alpha}}\phi)$.

Some of this may be counterintutive, but a good way to think of it is: what happens to the local average of a function when you transform it in some way?

Proof. For part (1), this follows from a computation: they're both equal to $\langle u, \tau_{y-x}R\phi\rangle$.

For part (2), suppose $h>0$ and $\mathbf{e}\in\mathbb{R}^d$ is a unit vector: $\|\mathbf{e}\|=1$. Then, we can define a *difference quotient operator*

$$T_h = \frac{1}{h}(I-\tau_{h\mathbf{e}}),\quad \text{i.e.}\quad T_hf(x) = \frac{1}{h}(f(x)-f(x-h\mathbf{e})).$$

Then, it's certainly true pointwise that

$$\lim_{h\to 0} T_h\phi = \frac{\partial\phi}{\partial\mathbf{e}}(x),$$

and since ϕ is uniformly continuous, this is also true in the ∞-norm (i.e. uniformly). Here's why: if you pick an $\varepsilon>0$, I can find a $\delta>0$ such that whenever $|x-y|<\delta$, then

$$\left|\frac{\partial\phi}{\partial\mathbf{e}}(x) - \frac{\partial\phi}{\partial\mathbf{e}}(y)\right| < \varepsilon.$$

Thus, if $|h|<\delta$, then

$$\left|\left(T_h\phi - \frac{\partial\phi}{\partial\mathbf{e}}\right)(x)\right| = \left|\frac{1}{h}\int_{-h}^0\left(\frac{\partial\phi}{\partial\mathbf{e}}(x+s\mathbf{e}) - \frac{\partial\phi}{\partial\mathbf{e}}(x)\right)\mathrm{d}s\right| < \varepsilon.$$

Since $D^{\boldsymbol{\alpha}}T_h\phi = T_hD^{\boldsymbol{\alpha}}\phi$, then $\frac{\partial}{\partial\mathbf{e}}D^{\boldsymbol{\alpha}}\phi\to D^{\boldsymbol{\alpha}}\frac{\partial\phi}{\partial\mathbf{e}}$ in the L^∞ norm.

This means that $T_h\phi\to\frac{\partial\phi}{\partial\mathbf{e}}$ in $\mathscr{D}$, since $\operatorname{supp}(T_h\phi)\subseteq\{x\in\mathbb{R}^d : \operatorname{dist}(x,\operatorname{supp}(\phi))\le h\}$, so if we take $|h|\le 1$, then we get a compact set.

[6] We're allowed to have V in front because it's a one-to-one isometry, so it doesn't affect the spectral values.

Then, using part (1), $T_h(u*\phi)(x) = u*(T_h\phi)(x)$, which converges to $u*\frac{\partial\phi}{\partial\mathbf{e}}(x) = \frac{\partial}{\partial\mathbf{e}}(u*\phi)$. Thus, $u*\phi$ is C^1, but we can repeat this over and over, making it C^∞. Then, since $D^{\boldsymbol{\alpha}}$ can go anywhere in $T_h(u*\phi)$, then the second statement holds when we pass to the limit. ⊠

If $\phi,\psi\in\mathscr{D}$, then $\phi*\psi\in\mathscr{D}$, because $\operatorname{supp}(\phi*\psi)\subseteq\operatorname{supp}(\phi)+\operatorname{supp}(\psi)$ (elementwise sum). Note, however, that if we convolve a test function and a distribution, the result might not be compactly supported! Be careful.

Proposition 4.6.2. *Suppose $\phi,\psi\in\mathscr{D}$ and $u\in\mathscr{D}'$. Then, $(u*\phi)*\psi = u*(\phi*\psi)$.*

PROOF. Since $\phi*\psi$ is uniformly continuous, then it can be approximated by a Riemann sum:

$$r_h(x) = \sum_{k\in\mathbb{Z}^d}\phi(x-kh)\psi(kh)h^d,$$

and this converges uniformly (i.e. in L^∞). Since ϕ and ψ are compactly supported, this is actually a finite sum, and therefore $D^{\boldsymbol{\alpha}}$ commutes with the sum. In particular, this means $D^{\boldsymbol{\alpha}}r_h(x)\to(D^{\boldsymbol{\alpha}}\phi)*\psi = D^{\boldsymbol{\alpha}}(\phi*\psi)$ as $h\to 0$ in L^∞.

We also have control over the support, so $r_h\to(\phi*\psi)$ in $\mathscr{D}$. Thus,

$$\begin{aligned}u*(\phi*\psi) &= \lim_{h\to 0^+} u*r_h(x)\\ &= \lim_{h\to 0^+}\sum_{k\in\mathbb{Z}^d}(u*\phi)(x-kh)\psi(kh)h^d\\ &= (u*\phi)*\psi.\end{aligned}$$ ⊠

It's the same theme as before: every time you want to prove something about an operation on a distribution, try it on a test function and sort out what happens.

Since we can take the dual $\mathscr{D}''$ of $\mathscr{D}'$, it makes sense to talk about the weak and weak-$*$ topologies on $\mathscr{D}$.

Fact. $\mathscr{D}$ is reflexive.

We won't prove this, but it means the weak and weak-$*$ topologies are the same, and we will refer to them interchangeably. Note, however, that $\mathscr{D}$ is not reflexive.

In particular, the weak topology on $\mathscr{D}'(\Omega)$ says that $u_j\rightharpoonup u$ if $\langle u_j,\phi\rangle\to\langle u,\phi\rangle$ for all $\phi\in\mathscr{D}(\Omega)$. Sometimes this is denoted $u_j\xrightarrow{\mathscr{D}'(\Omega)}u$, and in today's lecture, it's also just denoted $u_j\to u$; it's the only sense of convergence we'll use in $\mathscr{D}'(\Omega)$ today.

We end up with a nice convergence result here; if $\mathscr{D}(\Omega)$ and $\mathscr{D}'(\Omega)$ were normed spaces, we would have this already, but this isn't true.

Proposition 4.6.3. *If $\{\langle u_n,\phi\rangle\}_{n=1}^\infty$ is Cauchy in $\mathbb{F}$ for all $\phi\in\mathscr{D}(\Omega)$, then the function $u:\mathscr{D}(\Omega)\to\mathbb{F}$ defined by $\langle u,\phi\rangle = \lim_{n\to\infty}\langle u_n,\phi\rangle$ is a distribution.*

The proof relies on a version of the uniform boundedness principle which is more general than the one we proved; many functional analysis textbooks give the most general version, but ours was more concrete. In any case, the proofs are very similar.

Lemma 4.6.4. *If $T:\mathscr{D}\to\mathscr{D}$ is a continuous linear operator and $u_n\to u$ weakly in $\mathscr{D}'(\Omega)$, then $T^*u_n\to T^*u$.*

PROOF. For any $\phi\in\mathscr{D}(\Omega)$, $\langle T^*u_n,\phi\rangle = \langle u_n,T\phi\rangle$, which we know converges to $\langle u,T\phi\rangle = \langle T^*u,\phi\rangle$. ⊠

Corollary 4.6.5. *If $u_n\to u$ in $\mathscr{D}'(\Omega)$, then $D^{\boldsymbol{\alpha}}u_n\to D^{\boldsymbol{\alpha}}u$.*

Proposition 4.6.6. *Let $u\in\mathscr{D}'$ and $\boldsymbol{\alpha}$ be a multi-index with $|\boldsymbol{\alpha}|=1$. Then, $T_h^*u = (1/h)(u-\tau_{h\boldsymbol{\alpha}}u)\to D^{\boldsymbol{\alpha}}u$.*

Now, another result on convergence that every physicist knows.

Proposition 4.6.7. *Let χ_R denote the characteristic function of a set R. Then, if $R_\varepsilon = [-\varepsilon/2,\varepsilon/2]$, then $(1/\varepsilon)\chi_{R_\varepsilon}\to\delta_0$ in $\mathscr{D}'(\mathbb{R})$.*

We'll prove this as a corollary of something more general, Corollary 4.6.10.

Definition. Suppose $\varphi\in\mathscr{D}(\mathbb{R}^d)$ is such that $\varphi\geq 0$ and $\int\varphi(x)\,\mathrm{d}x = 1$. Then, for each $\varepsilon>0$, let $\varphi_\varepsilon(x) = (1/\varepsilon^d)\varphi(x/\varepsilon)$. The collection $\{\varphi_\varepsilon\}_{\varepsilon>0}$ is called an approximation to the identity.

Theorem 4.6.8. *Let φ_ε be an approximation to the identity. Then,*

(1) *if $\psi \in \mathscr{D}$, then $\psi * \varphi_\varepsilon \to \psi$ in $\mathscr{D}$, and*
(2) *if $u \in \mathscr{D}'$, then $u * \varphi_\varepsilon \to u$ in $\mathscr{D}'$.*

Corollary 4.6.9. *$C^\infty(\mathbb{R}^d)$ is a dense subset of $\mathscr{D}'(\mathbb{R}^d)$.*

Corollary 4.6.10. *If φ_ε is an approximation to the identity, then $\varphi_\varepsilon \to \delta_0$ as $\varepsilon \to 0$.*

Proof. $\varphi_\varepsilon = \delta_0 * \varphi_\varepsilon \to \delta_0$ in $\mathscr{D}'$ as above. ⊠

Proof of Theorem 4.6.8. For part (1), we know there's an R such that $\operatorname{supp}(\varphi) \subseteq \overline{B_R(0)}$. Thus, if $0 < \varepsilon < 1$, then $\operatorname{supp}(\psi * \varphi_\varepsilon) \subseteq \operatorname{supp}(\psi) + \operatorname{supp}(\varphi_\varepsilon) \subseteq \operatorname{supp}(\psi) + \overline{B_R(0)}$, which is a compact set K.

Suppose $f \in C_0^\infty(\mathbb{R}^d)$. Then,

$$f * \varphi_\varepsilon(x) = \int_{\mathbb{R}^d} f(x-y)\varphi_\varepsilon(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f(x-y)\frac{1}{\varepsilon^d}\varphi\left(\frac{y}{\varepsilon}\right)\mathrm{d}y.$$

Substitute $z = y/\varepsilon$ to get

$$\begin{aligned} &= \int_{\mathbb{R}^d} f(x - \varepsilon z)\varphi(z)\,\mathrm{d}z \\ &= f(x) + \int_{\mathbb{R}^d} \underbrace{(f(x-\varepsilon z) - f(x))}_{\to 0 \text{ uniformly}}\varphi(z)\,\mathrm{d}z, \end{aligned}$$

so as $\varepsilon \to 0$, this converges to $f(x)$ uniformly in x.

For part (2), suppose $\psi \in \mathscr{D}$, so that $\langle u, \psi\rangle = u * R\psi(0)$, because $u * \psi = \langle u, \tau_x R\psi\rangle$. Then,

$$\begin{aligned} \langle u, \psi\rangle &= u * R\psi(0) \\ &= \lim_{\varepsilon\to 0} u * (R\psi * \varphi_\varepsilon)(0) \\ &= \lim_{\varepsilon\to 0} (u * \varphi_\varepsilon) * R\psi(0) \\ &= \lim_{\varepsilon\to 0} \langle u * \varphi_\varepsilon, \psi\rangle, \end{aligned}$$

and we know this converges to u. ⊠

Lecture 40: 12/2/15

Applications of Distributions to Linear Differential Equations.

"The whole proof is making sense of what I wrote down."

One of the major applications of distributions is in solving differential equations. We're only going to have time to discuss their applications to linear differential equations, however.

Definition. A *differential operator* is an operator $L : C^m(\mathbb{R}^d) \to C^0(\mathbb{R}^d)$ of the form

$$L = \sum_{|\boldsymbol{\alpha}|\le m} a_{\boldsymbol{\alpha}} D^{\boldsymbol{\alpha}},$$

where $m \ge 0$ and $a_{\boldsymbol{\alpha}} \in C^0(\mathbb{R}^d)$.

To talk about distributions, we'll require that $a_{\boldsymbol{\alpha}} \in C^\infty(\mathbb{R}^d)$, so that we can view L as an operator $\mathscr{D}' \to \mathscr{D}'$. This means our problem is, given an $f \in \mathscr{D}'$, find a $u \in \mathscr{D}'$ such that $Lu = f$, or, equivalently, $\langle Lu, \phi\rangle = \langle f, \phi\rangle$ for all $\phi \in \mathscr{D}$.

Definition. Let u be a solution to $Lu = f$ in $\mathscr{D}'$.

- If $u \in C^m(\mathbb{R}^d)$ and $Lu = f$ pointwise, u is a *classical solution*; this is a solution on the level of functions.
- If u is a regular distribution but the equation doesn't hold pointwise, it's said to be a *weak solution*.
- We may also have *singular solutions*, which are singular distributions.

If $d = 1$, we're just dealing with ordinary differential equations, which is a bit more tractable.

Lemma 4.7.1. *If $\phi \in \mathscr{D}(\mathbb{R})$, then $\int \phi(x)\,\mathrm{d}x = 0$ iff there exists a $\psi \in \mathscr{D}(\mathbb{R})$ such that $\phi = \psi'$.*

The proof isn't too hard: the reverse direction is one line, and the forward direction is given by defining

$$\psi(x) = \int_{-\infty}^{x} \phi(s)\,\mathrm{d}s.$$

Definition. If $u \in \mathscr{D}'(\mathbb{R})$, then $v \in \mathscr{D}'(\mathbb{R})$ is a *primitive* or *antiderivative* of u if $v' = u$.[7]

The first equation we should solve is $L = D$, i.e. finding primitives for a distribution. The answer will be almost exactly the same as for functions, but we need to clarify exactly what it means for a distribution to be a constant.

Definition. If $c \in \mathbb{F}$, the *constant distribution* c is defined for a $\phi \in \mathscr{D}$ by $\langle c, \phi\rangle = c\int \phi$; a distribution is said to be *constant* if it's equal to $c \in \mathscr{D}'$ for some $c \in \mathbb{F}$.

The motivation for this definition is that the constant *function* c is in $L^1_{\mathrm{loc}}(\mathbb{R}^d)$, and so the distribution it induces is $\Lambda_c : \phi \mapsto \int c\phi = c\int\phi$, which suggests that we've picked the right definition. The next lemma provides another justification.

Lemma 4.7.2. *If $c \in \mathscr{D}'$ is constant, then $c' = 0$.*

Proof. If ψ is a test function, $\langle c', \psi\rangle = -\langle c, \psi'\rangle = -c\int \psi'(x)\,\mathrm{d}x = 0$, by Lemma 4.7.1. ⊠

Now, we can go back to solving $Dv = u$ for a given $u \in \mathscr{D}'$.

Theorem 4.7.3. *Every distribution has infinitely many primitives, and any two differ by a constant.*

Proof. As always, we will understand differentiation of distributions by reframing it as differentiation of test functions. Accordingly, let $\mathscr{D}_0 = \{\phi \in \mathscr{D} : \phi = \psi' \text{ for some } \psi \in \mathscr{D}\}$: this is just the set of test functions that are derivatives of other test functions. Accordingly, this is a vector space, and $\phi \in \mathscr{D}_0$ iff $\int \phi = 0$, by Lemma 4.7.1.

For any $u \in \mathscr{D}'$, a primitive v for u would satisfy $\langle u, \psi\rangle = \langle v', \psi\rangle = -\langle v, \psi'\rangle$ for all $\psi \in \mathscr{D}$. Thus, we know what v must do to ψ', which suggests defining it on $\mathscr{D}_0$. If $\phi \in \mathscr{D}_0$, then $\psi(x) = \int_{-\infty}^{x} \phi(s)\,\mathrm{d}s$ is an antiderivative for it (by the fundamental theorem of calculus), so define $v : \mathscr{D}_0 \to \mathbb{F}$ by

$$v(\phi) = -\left\langle u, \int_{-\infty}^{x} \phi(s)\,\mathrm{d}s\right\rangle.$$

Since integration is linear, so is v.

Now, we need to extend v from $\mathscr{D}_0$ to $\mathscr{D}$. Fix an $f \in \mathscr{D}$ with total integral 1, and for any $\psi \in \mathscr{D}$, let $\phi = \psi - \langle 1, \psi\rangle f$. In particular, $\phi \in \mathscr{D}_0$, because

$$\int \phi(x)\,\mathrm{d}x = \int \psi(x)\,\mathrm{d}x - \int \psi(x)\,\mathrm{d}x \underbrace{\int f(x)\,\mathrm{d}x}_{=1} = 0,$$

so by Lemma 4.7.1, ϕ has an antiderivative. Now, for any $c \in \mathbb{F}$, define $v_c : \mathscr{D} \to \mathbb{F}$ by $v_c(\psi) = v(\phi) + c\langle 1, \psi\rangle$. Since ϕ depends on ψ linearly, then v_c is a linear function and $v_c|_{\mathscr{D}_0} = v$. We need to show that v_c is continuous, which is equivalent to contunity at the origin, so let $\psi_n \to 0$ in the sense of $\mathscr{D}$, so $\langle 1, \psi_n\rangle \to 0$ as well. Thus, if $\phi_n = \psi_n - \langle 1, \psi_n\rangle\phi_1$, then $\phi_n \to 0$ and

$$\int_{-\infty}^{x} \phi_n(s)\,\mathrm{d}s \longrightarrow 0,$$

both in the sense of $\mathscr{D}$. Thus,[8]

$$\langle v_c, \psi_n\rangle = v(\phi_n) + \langle 1, \psi_n\rangle c = -\left\langle u, \int_{-\infty}^{x} \phi_n(s)\,\mathrm{d}s\right\rangle + \langle 1, \psi_n\rangle c,$$

and both of these terms go to 0, so v_c is continuous and thus is in $\mathscr{D}'$.

It remains to calculate v_c', so let's do that. If $\psi \in \mathscr{D}$ and $\phi = \psi - \langle 1, \psi\rangle f$ as above, then

$$\langle v_c', \psi\rangle = -\langle v_c, \psi\rangle = -v(\phi') - c\langle 1, \psi'\rangle,$$

[7] Just to be clear, $v' = Dv$; there's only one direction of derivative, so there's no ambiguity using v' for it.

[8] Sometimes, people will write $\langle v, \phi_n\rangle$ for $v(\phi_n)$, but since v isn't a distribution, because it's not defined on all of $\mathscr{D}$, then this is technically an abuse of notation.

but $\langle 1, \psi' \rangle = \int \psi' = 0$ because ψ' has an antiderivative, and

$$\begin{aligned} &= -v(\phi') = \left\langle u, \int_{-\infty}^{x} \psi'(s)\,\mathrm{d}s \right\rangle \\ &= \langle u, \psi \rangle. \end{aligned}$$

Thus, we've found infinitely many antiderivatives v_c for u.

Now, suppose $v, w \in \mathscr{D}'$ are any two antiderivatives of u. If $\psi \in \mathscr{D}$, let $\phi = \psi - \langle 1, \psi \rangle f$ as before, and let $c = \langle v - w, f \rangle \in \mathbb{F}$. Since $\phi \in \mathscr{D}_0$, then $\phi = \Phi'$ for some $\Phi \in \mathscr{D}$, so $\langle v - w, \phi \rangle = \langle v - w, \Phi' \rangle = -\langle v' - w', \Phi \rangle$, but since $v' = w' = u$, this is zero. Thus,

$$\begin{aligned} \langle v - w, \psi \rangle &= \underbrace{\langle v - w, \phi \rangle}_{=0} + \langle v - w, \langle 1, \psi \rangle f \rangle \\ &= \langle 1, \psi \rangle \langle v - w, f \rangle = c \langle 1, \psi \rangle = \langle c, \psi \rangle, \end{aligned}$$

so $v - w$ is a constant distribution. ⊠

Corollary 4.7.4. *If $a \in \mathbb{F}$, then the distributional differential equation $u' = au$ has only classical solutions given by $u(x) = Ce^{ax}$.*

PROOF. We could use separation of variables, but we can also use an integrating factor, because why not? We're going to pull the integrating factor e^{-ax} out of a hat. Any two solutions to $(e^{-ax}u)' = 0$ must differ by a constant, so we just get $C = e^{-ax}u$, or $u = Ce^{ax}$. ⊠

Now, we want to understand (and hopefully solve) equations of the form $u' + a(x)u = b(x)$ for functions a and b.

Example 4.7.5. Suppose $xu' = 1$ in $\mathscr{D}'$.

We know $u_0 = \ln|x|$ is a weak solution, by Example 4.4.5. If you think about this in the sense of L^2, you might not care about dividing by x, because the set of singularities has measure zero, but we need to be careful: we get $u_0' = \mathrm{PV}(1/x)$, and $xu_0' = x\,\mathrm{PV}(1/x) = 1$.

Just as in undergrad ODEs, we'll start with the homogeneous part of this nonhomogeneous equation: suppose $xv' = 0$, or even $xw = 0$. Then, δ_0 is a solution: $\langle x\delta_0, \phi \rangle = \langle \delta_0, x\phi \rangle = 0 \cdot \phi(0) = 0$.

Now, suppose $\phi \in \mathscr{D}$ and $\varepsilon > 0$, and let $r \in \mathscr{D}$ be a bump function approximating $\chi_{[-\varepsilon,\varepsilon]}$ (so that it's 1 on $[-\varepsilon, \varepsilon]$). Then,

$$\phi(x) = \phi(0)r(x) + (\phi(x) - \phi(0)r(x)),$$

but

$$\phi(x) - \phi(0)r(x) = \int_0^x \big(\phi'(s) - \phi(0)r'(s)\big)\,\mathrm{d}s.$$

making the substitution $x\eta = s$,

$$= x \underbrace{\int_0^1 \big(\phi'(x\eta) - \phi(0)r'(x\eta)\big)\,\mathrm{d}\eta}_{\psi(x)}.$$

Hence, $\phi(x) = \phi(0)r(x) + x\psi(x)$, and from its definition ψ is compactly supported, and thus $\psi \in \mathscr{D}$.

Thus, $\langle w, x\psi \rangle = \langle xw, \psi \rangle = 0$, so

$$\begin{aligned} \langle w, \phi \rangle &= \langle w, \phi(0)r \rangle + \langle w, x\psi \rangle \\ &= \langle w, \phi(0)r \rangle = \phi(0)\underbrace{\langle w, r \rangle}_{c} \\ &= \langle c\delta_0, \phi \rangle, \end{aligned}$$

so $w = c\delta_0$. This is the homogeneous equation, so if $v' = w = c_2\delta_0$, then $v = c_1 + c_2 H$, where H is the Heaviside step function. Thus, our solution is $u = c_1 + c_2 H + \ln|x|$, and this is a weak solution; if we were careless, we might have missed the term with H.

Now, let's return to $d \geq 1$ and think about partial differential equations. If $\boldsymbol{\alpha}$ is a multi-index in $\mathbb{N}^d$ and $x \in \mathbb{R}^d$, then we define $x^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$. In particular, pick a polynomial $p \in \mathbb{F}[x_1, \ldots, x_d]$ given by

$$p(x) = \sum_{|\boldsymbol{\alpha}| \leq m} c_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}},$$

for $c_{\boldsymbol{\alpha}} \in \mathbb{F}$. This gives us a differential operator $L = p(D)$ (and similarly, differential operators give us polynomials).

The point is, we get a very nice adjoint[9] $\mathscr{L} : \mathscr{D}' \to \mathscr{D}'$ given by

$$\mathscr{L} = \sum_{|\boldsymbol{\alpha}| \leq m} (-1)^{|\boldsymbol{\alpha}|} c_{\boldsymbol{\alpha}} D^{\boldsymbol{\alpha}}. \tag{4.3}$$

This is because $\langle u, \mathscr{L}\phi \rangle = \langle Lu, \phi \rangle$ for all $\phi \in \mathscr{D}$.

Example 4.7.6. Consider the *wave operator*

$$L = \frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2}, \tag{4.4}$$

for a $c > 0$ and $t, x \in \mathbb{R}$.

For every $g \in C^2(\mathbb{R})$, let $f(t,x) = g(x - ct)$, so that $Lf = 0$; if $g \in L^1_{\mathrm{loc}}$, then this gives us a weak solution. If (in some imprecise sense) $g = \delta_0$, then $f(t,x) = \delta_0(x - ct)$ seems like it would be a weak solution, but we need to make this precise.

Let $u \in \mathscr{D}'$ be defined by

$$\langle u, \phi \rangle = \int_{-\infty}^{\infty} \phi(t, ct) \, \mathrm{d}t,$$

which is in some sense a δ-function along the line $\{(t, ct) : t \in \mathbb{R}\}$. u is well-defined and clearly continuous and linear, so it's a distribution. This could be the "$\delta_0(t - ct)$" that we were hoping for, so let's make sure it's a solution.

It turns out L is self-adjoint (just check (4.3)), so

$$\begin{aligned}
\langle Lu, \phi \rangle &= \langle u, L\phi \rangle \\
&= \left\langle u, \left(\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2} \right) \phi \right\rangle.
\end{aligned}$$

Conveniently, this factors:

$$\begin{aligned}
&= \left\langle u, \left(\frac{\partial}{\partial t} + c \frac{\partial}{\partial x} \right) \underbrace{\left(\frac{\partial}{\partial t} - c \frac{\partial}{\partial x} \right) \phi}_{\psi \in \mathscr{D}} \right\rangle \\
&= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial t} + c \frac{\partial}{\partial x} \right) \psi(t, ct) \, \mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial t} \psi(t, ct) \, \mathrm{d}t = 0.
\end{aligned}$$

Definition. If L is a differential operator, then a $u \in \mathscr{D}'$ such that $Lu = \delta_0$ is called a *fundamental solution* for L.

Lecture 41: 12/4/15

Linear Differential Operators with Constant Coefficients.

Recall that if L is a differential operator, a $u \in \mathscr{D}'$ is a fundamental solution for L if $Lu = \delta_0$. In general, these are not unique, but they are fundamental in the following sense.

Theorem 4.8.1. *If E is a fundamental solution for L and $f \in \mathscr{D}$, then $E * f$ is a solution for $Lu = f$.*

Again, this is nonunique, but it's still a useful method for producing solutions.

Proof. Since $LE = \delta_0$, then $(LE) * f = \delta_0 * f = f$, but we have shown that taking derivatives and linear combinations commutes with convolution, and L is just a linear combination of derivatives, so $(LE) * f = L(E * f)$, so $L(E * f) = f$. ⊠

[9] This is the formal adjoint, not the Hilbert adjoint; it lives in the dual space!

Theorem 4.8.2 (Malgrange-Ehrenpreis)**.** *Every linear PDE with constant coefficients has a fundamental solution.*

We won't prove this, but it also illuminates why fundamental solutions are so fundamental: not only do they solve the problem, but they always exist.

In dynamical systems-speak, E is the solution, or *response*, and δ_0 is the *unit impulse*. But since L is a linear operator, then knowing what happens at 0 is sufficient to know what happens everywhere: convolving looks at $E(x-y)f(y)$, shifting to the response to a unit force or impulse at $y=x$; the actual force (or impulse) in the equation is $f(y)$. So if we "sum" (meaning integrate) all of the impulses at a point, we get all the information, which is how integration gives us the solution.

Example 4.8.3. Let's look again at the wave operator (4.4). It factors: $L=(\partial_t - c\partial_x)(\partial_t + c\partial_x)$, giving us the *plus operator* $D_+ = \partial_t + c\partial_x$ and the *minus operator* $D_- = \partial_t - c\partial_x$; hence, $L = D_- D_+$.

It takes an expedition into differential equations to determine a fundamental solution, but we were given

$$E(t,x) = \frac{1}{2c}H(ct-|x|) = \frac{1}{2c}H(ct-x)H(ct+x). \tag{4.5}$$

This is a regular distribution, and is positive on the cone $\{y > x/c\}$, and 0 elsewhere; see Figure 4.1.

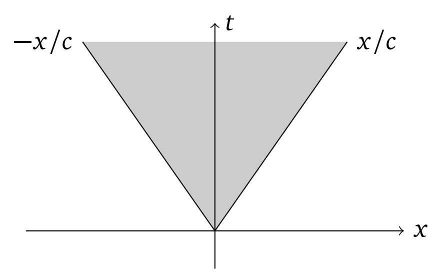


FIGURE 4.1. Depiction of the fundamental solution (4.5) to the wave operator (4.4). The filled-in region is where it takes on the value $1/2c$, and the unshaded region is where it is 0.

Now, why is this a fundamental solution? We want to show that $\langle LE, \phi\rangle = \phi(0,0)$. Since L is self-adjoint,

$$\begin{aligned}
\langle LE, \phi\rangle &= \langle E, L\phi\rangle = \langle E, D_+D_-\phi\rangle \\
&= \iint \frac{1}{2c}H(ct-|x|)D_+D_-\phi(x,t)\,\mathrm{d}t\,\mathrm{d}x \\
&= \frac{1}{2c}\left(\int_0^\infty\int_0^\infty D_+D_-\phi(t,x)\,\mathrm{d}t\,\mathrm{d}x + \int_{-\infty}^0\int_{-x/c}^\infty D_-D_+\phi(t,x)\,\mathrm{d}t\,\mathrm{d}x\right) \\
&= \frac{1}{2c}\left(\int_0^\infty\int_0^\infty D_+D_-\phi\left(t+\frac{x}{c},x\right)\mathrm{d}t\,\mathrm{d}x + \int_0^\infty\int_0^\infty D_-D_+\phi\left(t-\frac{x}{c},x\right)\mathrm{d}t\,\mathrm{d}x\right).
\end{aligned}$$

Next, we have to expand out D_+ and D_-. Specifically,

$$\begin{aligned}
D_+D_-\phi\left(t+\frac{x}{c},x\right) &= c\frac{\partial}{\partial x}(D_-\phi)\left(t+\frac{x}{c},x\right) \\
D_-D_+\phi\left(t-\frac{x}{c},x\right) &= -c\frac{\partial}{\partial x}(D_+\phi)\left(t-\frac{x}{c},x\right),
\end{aligned}$$

so returning to the calculation,

$$\begin{aligned}
\langle LE, \phi\rangle &= -\frac{1}{2}\int_0^\infty (D_-\phi(t,0) + D_+\phi(t,0))\,\mathrm{d}t \\
&= -\frac{1}{2}\int_0^\infty 2\frac{\partial}{\partial t}(t,0)\,\mathrm{d}t \\
&= \phi(0,0),
\end{aligned}$$

so this is indeed a fundamental solution.

More generally, if $Lu = f$, then

$$u = E * f = \frac{1}{2c}\int_{-\infty}^{t}\int_{x-c(t-s)}^{x+c(t+s)} f(s,y)\,\mathrm{d}y\,\mathrm{d}s.$$

This equation implies some interesting physical properties of waves. Most notably, any point x at time 0 can only influence things in an expanding cone as time increases, called its *domain of influence*,[10] which ultimately follows from the cone shape in Figure 4.1. Conversely, any point (x,t) can only be affected by points in an expanding cone as we go further into the past; this is its *domain of dependence*.

Example 4.8.4. Probably the most important example in applied mathematics is the *Laplace operator* or *Laplacian*

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_n^2}.$$

This is also written $\Delta = \nabla\cdot\nabla = \nabla^2$. This shows up in many things, and is usually best thought of as a diffusion operator. It's also true but not obvious that it's coordinate-independent.

Our fundamental solution will be given by

$$E(x) = \begin{cases} \dfrac{1}{2}|x|, & d = 1 \\ \dfrac{1}{2\pi}\ln|x|, & d = 2 \\ \dfrac{1}{d\omega_d}\dfrac{|x|^{2-d}}{2-d}, & d > 2, \end{cases} \tag{4.6}$$

where ω_d denotes the volume of the d-dimensional unit sphere (so cases $d = 1$ and $d = 2$ actually follow the same pattern as the general case).

For $d = 1$, this is easy to check, so let's worry about the general case.

E is actually in L^1_{loc}; since the only singularity is at 0, we need to calculate the integral there, and we'll do so with polar coordinates for $d = 2$. If $\varepsilon > 0$,

$$\int_{B_\varepsilon(0)}\left|\frac{1}{2\pi}\ln|x|\right|\mathrm{d}x = 2\pi\left|\int_0^{\varepsilon}\frac{1}{2\pi}\ln s\right|s\,\mathrm{d}s = \frac{1}{2}\left[s^2\ln s - \frac{s^2}{2}\right]_0^{\varepsilon},$$

which is finite. For general $d > 2$, we can ignore the constant terms, so using d-dimensional spherical coordinates,

$$\int_{B_\varepsilon(0)}\frac{\mathrm{d}x}{|x|^{d-2}} = \int_0^{\varepsilon}\frac{1}{r^{d-2}}r^{d-1}C\,\mathrm{d}r = C\int_0^{\varepsilon} r\,\mathrm{d}r$$

for some constant C; thus, this is finite as well.

Now, we need to show that E is indeed a fundamental solution. Here we need to use the divergence theorem.

Theorem 4.8.5 (Divergence)**.** *If Ω is a submanifold of $\mathbb{R}^d$ and ϕ and ψ are smooth functions on Ω, then*

$$\int_\Omega \nabla\phi\cdot\psi = \int_{\partial\Omega}\phi\psi\cdot\boldsymbol{\nu} - \int_\Omega \phi\nabla\cdot\psi,$$

where $\boldsymbol{\nu}$ denotes the outward-pointing unit normal vector on the boundary $\partial\Omega$.

Now, back to the calculation.

$$\langle \Delta E, \phi\rangle = \langle E, \Delta\phi\rangle = \int E(x)\Delta\phi(x)\,\mathrm{d}x.$$

[10] Yes, this is the positive lightcone in physics!

Now, we use the divergence theorem and the fact that $\Delta\phi = \nabla\cdot(\nabla\phi)$. Let $\varepsilon > 0$ and R be such that $\mathrm{supp}(\phi) \subsetneq \overline{B_R(0)}$, so that

$$= \int_{\varepsilon<|x|<R} E\Delta\phi + \int_{|x|<\varepsilon} E\Delta\phi - \int_{\varepsilon<|x|<R} \nabla E\cdot\nabla\phi + \int_{|x|<\varepsilon} E\nabla\phi\cdot\boldsymbol{\nu}$$
$$= \int_{\varepsilon<|x|<R} \Delta E - \int_{|x|=\varepsilon} \phi\nabla E\cdot\boldsymbol{\nu} + \int_{|x|=\varepsilon} E\nabla\phi\cdot\nu.$$

You can directly calculate that $\Delta E = 0$, so let's attack the remaining two parts. Since $\nabla\phi\cdot\boldsymbol{\nu}$ doesn't blow up as $\varepsilon\to 0$, since ϕ is smooth, then

$$\int_{|x|=\varepsilon} E\nabla\phi\cdot\boldsymbol{\nu} = \int_{S_1(0)} \frac{1}{d\omega_d}\frac{\varepsilon^{2-d}}{2-d}\nabla\phi\cdot\boldsymbol{\nu}\varepsilon^{d-1}\,\mathrm{d}\sigma = O(\varepsilon)\longrightarrow 0$$

as $\varepsilon\to 0$. The last part is

$$-\int_{|x|=\varepsilon} \nabla E\cdot\boldsymbol{\nu}\phi\,\mathrm{d}\sigma = \int_{S_1(0)} \frac{\partial E}{\partial r}\phi\varepsilon^{d-1}\,\mathrm{d}\sigma$$
$$= \int_{S_1(0)} \frac{1}{d\omega_d}|\varepsilon|^{1-d}\phi(\varepsilon,\sigma)\varepsilon^{d-1}\,\mathrm{d}\sigma \longrightarrow \phi(0),$$

since ϕ is smooth.

Notice that since E is supported everywhere, then the domain of dependence of diffusion is everywhere: the speed of propgation is infinite, because things infinitely far away can affect you at arbitrarily small times. This seems a little odd — and so for some other kinds of diffusion, it may be necessary to use a more complicated model (e.g. a nonlinear one). Nonetheless, because E tails off pretty quickly, local influences have a much stronger effect than global ones. Infinite propagation speed also causes all sorts of headaches numerically.

With a bit more work, one can extend this to L^1, subject to a small condition.

Theorem 4.8.6. *Let E be the fundamental solution to the Laplacian given in (4.6). Then, if $f\in L^1(\mathbb{R}^d)$ and $E(x-y)f(y)\in L^1(\mathbb{R}^d)$. Then, $u = E * f$ is a solution to $\Delta u = f$ in $\mathscr{D}'$.*