MATH 205B NOTES

MOOR XU
NOTES FROM A COURSE BY RAFE MAZZEO

Abstract. These notes were taken during Math 205B (Real Analysis) taught by Rafe Mazzeo in Winter 2012 at Stanford University. They were live-TeXed during lectures in `vim` and compiled using `latexmk`. Each lecture gets its own section. The notes are not edited afterward, so there may be typos; please email corrections to `moorxu@stanford.edu`.

1. 1/10

The topic of the course is functional analysis and applications. These include harmonic analysis and partial differential equations. These applications were the reason much of the theory was created.

Today we will review some facts about topological spaces and metric spaces, and state some fundamental theorems.

Definition 1.1. We have *metric spaces* (M, d). Here M is some set, and $d : M \times M \to \mathbb{R}^+$ satisfying

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

The simplest examples are things we know:

Example 1.2. Euclidean space $\mathbb{R}^n$ with $d(x, y) = \sqrt{\sum (x_i - y_i)^2}$

Other examples include the following:

Example 1.3.

- Finite sets with metrics. Here M is a finite set.
- On the interval $a \leq x \leq b$, the real-valued continuous functions are $C([a, b])$. This has a metric $d(f, g) = \sup_{a \leq x \leq b} |f(x) - g(x)|$. We can define $B(f, \varepsilon)$ as the set of all g such that $d(f, g) < \varepsilon$. These are all continuous functions that never stray too far from f, inside a tube around f.
- $C^1([a, b])$ are continuously differentiable functions, so $f \in C^1$ if $f \in C^0$ and $f' \in C^0$. This means that there are no sharp corners, and the derivative doesn't stray too far either. The metric is $d(f, g) = \sup_{x \in [a,b]}(|f(x) - g(x)| + |f'(x) - g'(x)|)$.
- $L^2[a, b]$ are functions with metric $d(f, g) = (\int |f(x) - g(x)|^2 \, dx)^{1/2}$. Here, $d(f, g) = 0$ iff $f = g$ almost everywhere; this notion is one reason why the Lebesgue integral exists. In fact, L^2 functions are actually equivalence classes of functions, where $f \sim g$ if they agree almost everywhere. This eliminates some interesting functions.

- $L^p[a,b]$ for $1 \le p \le \infty$. Here $\|f\|_\infty = \sup|f|$. The triangle inequality holds by the Holder inequality.

All of these are special cases of normed spaces $(V, \|\cdot\|)$. Here, M is the vector space V, and the norm is $\|\cdot\| : V \to \mathbb{R}^+$. By definition, we can write $d(v,w) = \|v-w\|$. A consequence is that $d(v+b, w+b) = d(v,w)$. All of linear functional analysis is the interaction between the vector space and the underlying topology. Slightly more generally, we can talk about topological vector spaces.

We want to apply our finite dimensional intuition, and we can for Hilbert spaces, but not for more exotic cases. But we don't have compactness, even for closures of balls! We'll start with the simplest examples and build up to more generality.

At the bottom, we have Hilbert spaces (i.e. having inner products, like L^2). More generally, there are Banach spaces (having norms such as L^p or C^k), but they are very ubiquitous but harder to deal with. Then there are Frechet spaces (with a countable set of norms like C^∞), and then locally convex topological vector spaces (e.g. distributions).

Let's go back to general metric spaces. The really important property is the notion of completeness.

Definition 1.4. A sequence of elements $\{x_n\}$ in M is a Cauchy sequence if given any $\varepsilon > 0$ there exists N such that $m, n > N$ then $d(x_m, x_n) < \varepsilon$.

Definition 1.5. A space (M,d) is called *complete* if every Cauchy sequence has a limit in M.

Example 1.6. Here are examples of complete spaces.

- $\mathbb{R}$ or $\mathbb{R}^n$ are complete. The rational numbers are not complete, but its completion is all of $\mathbb{R}$. But this makes many elements of $\mathbb{R}$ somewhat intangible – all we know is that they are limits of Cauchy sequences.
- $C^0(I)$ or any $C^k(I)$.
- $L^1(I)$ is complete. (Riesz-Fisher Theorem)

Example 1.7. Here are some counterexamples.

Consider $(C^1, \|\cdot\|_{C^0})$. That's certainly a normed metric space because C^1 is a subset of C^0, but it is not complete; there are differentiable functions that approach a sharp corner. The sup norm doesn't force the derivative to converge.

Theorem 1.8. *$C^0([a,b])$ is complete.*

Proof. Let $\{f_n\}$ be Cauchy with this norm. We want to produce a function which is the limit, and we need to show that it is still in C^0.

If $x \in I$ then $\{f_n(x)\}$ is a sequence in $\mathbb{R}$, and it is still Cauchy, because $|f_n(x) - f_m(x)| \le \|f_n - f_m\|_{C^0} \to 0$. So define $f(x) = \lim_{n\to\infty} f_n(x)$. So now we've shown that this sequence $\{f_n\}$ converges pointwise.

Now, we need to show that f is C^0, i.e. if $\varepsilon > 0$ is fixed, then there exists $\delta > 0$ such that we want $|f(x) - f(y)| < \varepsilon$ if $|x-y| < \delta$. (This is actually uniform continuity, but we're working on a compact interval, so that's ok.) To show this, we have $|f(x) - f(y)| \le |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f(y)|$. Fix $\varepsilon > 0$ and choose N such that $\|f_n - f_N\| < \varepsilon$ if $n \ge N$. Then $\sup_x |f(x) - f_N(x)| = \sup_x \lim_{n\to\infty} |f_n(x) - f_N(x)| \le \sup_x \sup_{n\ge N} |f_n(x) - f_N(x)| = \sup_{n \ge N} \|f_n - f_N\| < \varepsilon$. Now the result follows by choosing N appropriately. ☐

Now, we will head toward the main theorem that we will do today: the Arzela-Ascoli theorem. The question is: What sets of continuous functions are compact?

Definition 1.9. Suppose that $\mathcal{F} \subset C^0(I)$ is some collection of continuous functions. This is called *uniformly equicontinuous* if given any $\varepsilon > 0$ there exists δ such that if $|x-y| < \delta$ and any $f \in \mathcal{F}$ we have $|f(x) - f(y)| < \varepsilon$.

There are a lot of natural examples of equicontinuous functions, and this is the main motivating example.

Example 1.10. Let $\mathcal{F} = \{f \in C^1(I) : \|f\|_{C^1} \leq A\}$. Then $\mathcal{F} \subset C^0$ is uniformly equicontinuous. This is because

$$|f(x) - f(y)| = \left| \int_y^x f'(t)\, dt \right| \leq \int_x^y |f'(t)|\, dt \leq A|y-x|.$$

Here are a couple of preparatory theorems.

Theorem 1.11. *Suppose $\{f_n \in C^0(X,Y)\}$ which is equicontinuous. If $\{f_n\}$ converges pointwise to f, then f is continuous.*

Proof. Left to your vivid imaginations. □

Theorem 1.12. *Let Y be complete, and let $D \subset X$ be a dense set in X. Suppose that f_n is equicontinuous on X and f_n converges pointwise on D. Then, in fact, $f_n \to f$ uniformly. (Uniform needs that X is compact.)*

Proof. First, if $x \in D$, we need to define $f(x)$. Take any sequence $x_k \to x$ with $x_k \in D$. Then the limit $\lim_{n\to\infty} f_n(x_k)$ is well-defined. Define

$$f(x) = \lim_{k\to\infty} \lim_{n\to\infty} f_n(x_k).$$

There's something to check: any other sequence $y_k \to x$ gives the same limit. This comes from equicontinuity. We have $d(x_k, y_l) < \delta$, so therefore $\rho(f_n(x_k) - f_n(y_l)) < \varepsilon$, and this is well-defined.

We still need to show that f is actually continuous, and the convergence is uniform. Go through the details yourself. □

We illustrate this with an example:

Theorem 1.13. *Suppose that $\{f_n\}$ is a uniformly equicontinuous family on $[0,1]$ and $D \subset [0,1]$ is dense with pointwise convergence of f_n on D. Then $f_n \to f$ uniformly.*

Proof. Fix ε. There exists δ such that $|f_n(x) - f_n(y)| < \varepsilon$ if $|x-y| < \delta$. For that δ, choose $y_1, \dots, y_m \in [0,1]$ such that $y_j \in D$ and $B(y_j, \delta)$ cover I. Then choose N so large that $|f_n(y_i) - f(y_i)| < \varepsilon$ if $n \geq N$.

Now, $|f(x) - f(\tilde{x})| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(y_j)| + |f_N(y_j) - f_N(\tilde{x})| + |f_N(\tilde{x}) - f(\tilde{x})|$. □

Theorem 1.14 (Arzela-Ascoli Theorem)**.** *Suppose that $\{f_n\}$ is a family with $f_n : [0,1] \to \mathbb{R}$ uniformly bounded (i.e. there exists A such that $|f_n(x)| \leq A$ for all n, x. Suppose that f_n is equicontinuous. Then there exists f_{n_j} which converges uniformly.*

This is extremely useful and is the template for all compactness theorems in infinite dimensional spaces. The proof uses the diagonalization argument.

Lemma 1.15. *Suppose that $\{f_n(m)\}$ is bounded in m and n uniformly. Then there exists n_j such that $f_{n_j}(m)$ converges for each m.*

Proof. For $m = 1$, $f_n(1)$ is a bounded sequence. Choose $f_{n(1)_j}(1)$ convergent. Take a further subsequence such that $f_{n(2)_j}(2)$ converges.

So we get numbers $n(k)_j$. Then define $n_j = n(j)_j$. This is the diagonalization. □

Proof of Theorem 1.14. Let $D = \mathbb{Q} \cap I$ be dense. Call $D = \{x_k\}$. For every rational, we have $f_n(x_k)$ is uniformly bounded in n, k, so we can apply the diagonalization lemma above to choose $n_j \to \infty$ so $f_{n_j}(x_k) \to f(x_k)$ for all x_k. Now, using Theorem 1.12 gives the desired result. □

Remark. This is in fact a very general theorem about maps between general metric spaces.

This is a nice proof. Let's give a different proof that is easier to visualize.

First, here's a characterization of compactness.

Theorem 1.16. *(K, d) is compact if and only if for each $\delta > 0$ there is a finite δ-net $y_1, \dots, y_N \in K$ such that $B_\delta(y_j)$ cover K.*

This allows a picture proof of Arzela-Ascoli. Draw a grid, ranging from $-A$ to A in the vertical direction and along some interval in the horizontal direction. Equicontinuous means that the function stays within each box and doesn't move too much. Take every function that connects vertices and never jumps more than one step and is linear in between. Every function is deviating very much because it always sticks to one of these piecewise linear functions.

2. 1/12

Today we will discuss inner product spaces. We start with a vector space V over $\mathbb{R}$ or $\mathbb{C}$. Call the field $\mathbb{F}$.

Definition 2.1. A map $(\cdot, \cdot) : V \times V \to \mathbb{F}$ is called an inner product if it is

(1) positive definite, i.e $(v, v) \geq 0$ for all $v \in V$ and $(v, v) > 0$ unless $v = 0$.
(2) linear in the second entry: $(v, w_1 + w_2) = (v, w_1) + (v, w_2)$ and $(v, cw) = c(v, w)$
(3) Hermitian symmetry: $(v, w) = \overline{(w, v)}$.

(If $\mathbb{F} = \mathbb{R}$ then we can ignore the complex conjugation.)

Note that (2) and (3) show that the inner produce is conjugate linear in the first variable, i.e. $(v_1 + v_2, w) = (v_1, w) + (v_2, w)$ and $(cv, w) = \bar{c}(v, w)$. If $\mathbb{F} = \mathbb{R}$, the inner product is linear in the first variable.

Example 2.2. In $\mathbb{R}^n$, we can take the standard inner product: $(x, y) = \sum_{j=1}^n x_j y_j$. Or we can take a different inner product: $(x, y) = \sum_{j=1}^n \lambda_j x_j y_j$.

In $\mathbb{C}^n$, we have $(x, y) = \sum_{j=1}^n \overline{x}_j y_j$.

Let ℓ^2 be sequences (real or complex valued) $\{a_n\}_{n=1}^\infty$ with $\sum_{n=1}^\infty |a_n|^2 < \infty$. Then we have $(\{a_n\}, \{b_n\}) = \sum_{n=1}^\infty \overline{a}_n b_n$.

In the case $L^2([0, 1])$, we have $(f, g)_{L^2} = \int_0^1 \overline{f(x)} g(x)\, dx$.

In the case $C([0, 1]; \mathbb{F})$, we have $(f, g)_{L^2} = \int_0^1 \overline{f(x)} g(x)\, dx$.

We'll soon see that inner products give norms, and $L^2([0, 1])$ is complete as a normed space, while $C([0, 1])$ is not (and in fact has L^2 as completion).

The metric space completion of an inner product space is always an inner product space.

Definition 2.3. $\|x\| = \sqrt{(x,x)}$ for any $x \in V$. This yields a map $\|\cdot\| : V \to [0,\infty)$.

We certainly know that this is positive definite by the positive definiteness of the inner product. Also, it is absolutely homogeneous, i.e. $\|cx\| = \sqrt{(cx,cx)} = |c|\sqrt{(x,x)} = |c|\,\|x\|$. To show that this is a norm, we need to check the triangle inequality. We'll do this instructively in a slightly roundabout way.

Definition 2.4. We say that x and y are *orthogonal* if $(x,y) = 0$.

Proposition 2.5 (Pythagoras)**.** *If* $(x,y) = 0$ *then* $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2$.

Proof. Indeed, $\|x \pm y\|^2 = (x \pm y, x \pm y) = (x,x) \pm (x,y) \pm (y,x) + (y,y) = \|x\|^2 + \|y\|^2$. □

Proposition 2.6 (Parallelogram Law)**.** $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$.

Proposition 2.7. *If* $y \in V$ *and* $y \neq 0$*, and we have* $x \in V$*, then we can write* $x = cy + w$ *such that* $(w,y) = 0$.

Proof. Notice that $(y,x) = (y,cy) + (y,w) = c\|y\|^2$, so set $c = \frac{(y,x)}{\|y\|^2}$. So to actually show that c and w exist, just let c be given as above, and let $w = x - cy$. We just need to compute $(y,w) = (y, x-cy) = (y,x) - c\|y\|^2 = 0$. □

We call $cy = \frac{(y,x)}{\|y\|^2} y$ the orthogonal projection of x to $\operatorname{span}(y)$.

Theorem 2.8 (Cauchy-Schwarz)**.** *If* $x, y \in V$ *then* $|(x,y)| \leq \|x\|\,\|y\|$*, with equality if and only if* x *and* y *are collinear, i.e. one is a multiple of the other.*

Proof. The linearity and conjugate linearity (sesquilinearity) of the inner product implies that if $y = 0$ then $(x,y) = 0$ for all $x \in V$, and so in particular $(y,y) = 0$. So if $y = 0$ then Cauchy-Schwarz holds and both sides are zero.

Suppose that $y \neq 0$. Then write $x = cy + w$ and $(y,w) = 0$, so Pythagoras gives that $\|x\|^2 = \|cy\|^2 + \|w^2\| \geq c^2\|y\|^2$, so $\|x\|^2\|y\|^2 \geq |(y,x)|^2$. And the inequality is strict unless $\|w\|^2 = 0$, i.e. x and y are collinear. □

So far we've just used the inner product structure, and haven't done any real analysis.
Now we can prove the triangle inequality.

Theorem 2.9. $\|\cdot\|$ *is a norm.*

Proof. We only need to show the triangle inequality, i.e. $\|x+y\| \leq \|x\| + \|y\|$. Equivalently, we have $\|x+y\|^2 \leq (\|x\| + \|y\|)^2 = \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2$. But the left hand side is $\|x+y\|^2 = \|x\|^2 + (x,y) + (y,x) + \|y\|^2 \leq \|x\|^2 + 2|(x,y)| + \|y\|^2$, so Cauchy-Schwarz completes the proof. □

Therefore, any inner product space is a normed space in a canonical manner.

Definition 2.10. A *Hilbert space* is a complete inner product space.

Example 2.11. The spaces $\mathbb{R}^n$, $\mathbb{C}^n$, $L^2([0,1])$, ℓ^2 are complete. $C([0,1])$ is not complete. Any inner product space can be completed to an inner product space. Whenever you can an inner product space that's not complete, complete it to get a Hilbert space!

Definition 2.12. A subset S of an inner product space is orthonormal if for all $x, y \in S$, $x \neq y$ implies $(x, y) = 0$, and $\|x\| = 1$.

Proposition 2.13. *If $\{x_1, \ldots, x_n\}$ is an orthonormal set in V and $y \in V$ then we can write $y = \sum_{j=1}^n c_j x_j + w$ such that $(w, x_j) = 0$ for all j.*

Proof. This is the same argument as before. Indeed, if so, $(x_k, y) = c_k \|x_k\|^2 = c_k$, and now we use this to check that this works. Let $c_k = (x_k, y)$ and $w = y - \sum c_k x_k$. Then $(x_k, w) = (x_k, y) - c_k = 0$. $\square$

Corollary 2.14 (Bessel's inequality)**.** *If $\{x_1, \ldots, x_n\}$ is orthonormal then for all $y \in V$ we have*

$$\|y\|^2 = \sum_{j=1}^n |c_j|^2 + \|w\|^2 \geq \sum_{j=1}^n |c_j|^2.$$

There are some general constructions that we can do.

Definition 2.15. If V and W are two inner product spaces, then we can take $V \oplus W$ is an inner product space with $\langle (v_1, w_1), (v_2, w_2) \rangle_{V \oplus W} = \langle v_1, v_2 \rangle_V + \langle w_1, w_2 \rangle_W$. If V and W are complete, so is $V \oplus W$.

In fact, we aren't limited to two. We could even do this uncountably many times.

Definition 2.16. If V is an inner product space and $A \neq \emptyset$ is a set, let $\ell^2(A; V)$ be maps $A \to V$ such that $\sum_{a \in A} \|f(a)\|_V^2 < \infty$. This is an inner product space $\langle f, g \rangle = \sum_{a \in A} \langle f(a), g(a) \rangle_V$.

If A is a set and $c : A \to [0, \infty)$ then one can define

$$\sum_{a \in A} c(A) = \sup \left\{ \sum_{\substack{a \in B \\ B \subset A \\ B \text{ finite}}} c(a) \right\}.$$

(This could be $+\infty$, if it is not bounded above.) This agrees with the usual notion when A is finite, or $A = \mathbb{N}$.

Also, if this sum is finite, then $\{a \in A : c(a) \neq 0\}$ is countable because each set of the sets $\{a : c(a) > \frac{1}{n}\}$ for $n \in \mathbb{N}$ is finite. So we can take uncountable sums of Hilbert spaces.

The Hilbert spaces are the best kind of infinite dimensional vector spaces.

Definition 2.17. If M is a subspace of an inner product space V, let

$$\begin{aligned} M^\perp &= \{v \in V : (v, w) = 0 \text{ for all } w \in M\} \\ &= \bigcap_{w \in M} \ker(v \mapsto (v, w)). \end{aligned}$$

Also, the map $v \mapsto (v, w)$ are continuous because they are bounded conjugate linear maps: $|(v, w)| \leq \|v\| \|w\|$, so $M^\perp$ is closed. Also, $M \subset M^{\perp\perp}$. (This is true in any inner product space.)

Theorem 2.18. *Suppose H is a Hilbert space, and M is a closed subspace. Then for $x \in H$, there exists unique $y \in M$ and $z \in M^\perp$ such that $x = y + z$.*

Thus, $H = M \oplus M^\perp$ with the sum being orthogonal.

The key lemma in proving this is as follows:

Lemma 2.19. *Suppose that H is a Hilbert space, and M is a convex closed subset (i.e. $x, y \in M$, $t \in [0,1]$ implies $tx + (1-t)y \in M$). Then for $x \in H$ there is a unique point $y \in M$ which is closest to x, i.e. for all $z \in M$, $\|z - x\| \geq \|y - x\|$, with strict inequality if $z \neq y$.*

Proof. Let $d = \inf_{z \in M} \|z - x\| \geq 0$. Then there exists some $y_n \in M$ such that $\lim_{n\to\infty} \|y_n - x\| = d$. The main claim is that y_n is Cauchy, and thus it converges to some $y \in M$ since M is complete. Once this is done, the distance function being continuous, we have $d = \lim_{n\to\infty} \|y_n - x\| = \|y - x\|$. So $z \in M$ implies that $\|z - x\| \geq \|y - x\|$. Further, if y and y' are both distance minimizers, then $y, y', y, y', \ldots$ would have to be Cauchy, so $y = y'$, giving uniqueness.

We will work out the main claim (Cauchyness of y_n) using the parallelogram law. Applying the parallelogram law to $y_n - x$ and $y_m - x$ yields

$$2(\|y_n - x\|^2 + y_m - x^2) = \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2 = \|y_n - y_n\|^2 + 4\left\|\frac{y_n + y_m}{2} - x\right\|^2,$$

so

$$\|y_n - y_m\|^2 = 2(\|y_n - x\|^2 + \|y_m - x\|^2) - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 \geq 2(\|y_n - x\|^2 + \|y_m - x\|^2) - 4d^2.$$

As $n, m \to \infty$, we see that $\|y_n - y_m\|^2 \to 0$.

This proves this closest point lemma. ∎

Note that a subspace M is always convex.

Proof of Theorem 2.18. Suppose M is a closed subspace and $x \in H$. Let y be the closest point in M to x by Lemma 2.19. Let $z = x - y$, so $x = y + z$. We need to show that $z \in M^\perp$, i.e. for all $w \in M$, $(w, z) = 0$. But for $t \in \mathbb{R}$, we have $\|y + tw - x\|^2 \geq \|y - x\|^2$ since y is a distance minimizer. Then expanding the left hand side yields $2t \operatorname{Re}(w, x - y) + t^2 \|w\|^2 \geq 0$. This can only happen if $\operatorname{Re}(w, x - y) = 0$. If the field is $\mathbb{C}$, repeat with it in place of t to get $\operatorname{Im}(w, x - y) = 0$. In any case, $(w, z) = 0$ for all $w \in M$, so $z \in M^\perp$, and we are done.

For uniqueness, if $y + z = y' + z'$ with $y, y' \in M$ and $z, z' \in \perp M$, then $y - y' = z' - z = v \in M \cap M^\perp$, so $(y - y', y - y') = 0$ so $y - y' = 0$ and hence $y = y'$. ∎

This theorem is what makes Hilbert spaces really good. A consequence will be the Riesz representation theorem, and completeness is important.

3. 1/17

Recall the definition of Hilbert spaces. For infinitely dimensional Hilbert spaces, finite dimensional intuition holds extremely well. The only issue is that infinite dimensional Hilbert spaces are not locally compact.

Example 3.1. Suppose we take $B_1(0) = \{x : \|x\| \leq 1\}$. From the sequential definition of compactness, this is not compact.

Example 3.2. As a concrete example, consider the space ℓ^2. This is the basic example of a Hilbert space.

Proposition 3.3. *Suppose that $x^{(k)} \in \ell^2$ are Cauchy. There exists $x \in \ell^2$ such that $x^{(k)} \to x$. This shows that ℓ^2 is complete.*

Proof. We have $|x_j^{(k)} - x_j^{(l)}| \leq \left\| x^{(k)} - x^{(l)} \right\|$. Hence each $\{x_j^{(k)}\}_k$ is Cauchy, so $x_j^{(k)} \to x_j$. We need to check that $\sum |x_j|^2 < \infty$. ☐

We are interested in showing that ℓ^2 is not locally compact. Consider $e_1 = (1, 0, 0, \dots)$, $e_2 = (0, 1, 0, \dots)$, ... These are all in the unit ball, but $\|e_j - e_k\| = \sqrt{\langle e_j - e_k, e_j - e_k \rangle} = \sqrt{2}$, so the closed unit ball in ℓ^2 is not compact.

Recall the notions of Cauchy-Schwarz inequality, and orthonormality. There is also the notion of separability: H has a *countable* dense subset.

Example 3.4. Why is ℓ^2 separable? Take the subspace

$$V = \{x : x_j = 0 \text{ for } j \text{ large, all } x_j \in \mathbb{Q}\}.$$

We need that $\{x : x_j = 0 \text{ for } j \text{ large}\}$ is dense in ℓ^2.

We also discussed Bessel's inequality: Suppose $\{e_j\}$ is an infinite orthonormal set. For $v \in H$, set $\langle v, e_j \rangle = a_j$. Then $\sum_{j=1}^N |a_j|^2 \leq \|v\|^2$.

Now, $\{e_j\}$ is called an orthonormal basis if

- it is orthonormal
- $\{\sum_{\text{finite}} a_j e_j\}$ is dense in H.

To make that precise, consider $W = \overline{\{\sum_{\text{finite}} a_j e_j\}} \subset H$, i.e. $w \in W$ means that there exists a_j such that

$$\lim_{N \to \infty} \left\| w - \sum_{j=1}^N a_j e_j \right\| = 0.$$

Why is W a subspace? There are two things to do:

- $\{\sum_{\text{finite}} a_j e_j\}$ is a subspace
- V is a subspace implies that $\overline{V}$ is a subspace.

Suppose that we have an infinite orthonormal set $\{e_j\}$. The question is: Is $W = H$? This is true in some cases but not always. If $W = H$, then we've found a basis. Suppose that $W \subsetneq H$. Then the claim is that there exists $w \neq 0$ where $w \perp W$.

Proof. To prove this, find any $v \notin W$, and write $v = w + \tilde{w}$. Arrange for $\|v - \tilde{w}\|$ minimizes among all $\tilde{w} \in W$, which then means $v - \tilde{w} = w \perp W$. ☐

Now, take $\{\tilde{e}_j\} = \{e_j\} \cup \{w\}$. This proves that if $\{e_j\}$ is any orthonormal set, with closure of its span $= W \subsetneq H$, then there exists a countable set $\{\tilde{e}_j\}$ with closure of its span $\tilde{W} \supsetneq W$.

Proposition 3.5. *Suppose that H is separable. Then H is isomorphic to ℓ^2.*

Proof. Suppose that $\{e_j\}$ is a basis. If we have any $v \in H$, with $v = \sum a_j e_j$, and $\|v\| = \sum |a_j|^2$. Then we have a map $H \to \ell^2$ via $v \to (a_j)$. ☐

So why do we care about other Hilbert spaces? Because they are other representations of ℓ^2 with their own interesting features and motivations.

Consider the space $L^2(S^1) = \{f(\theta)$ is periodic of period 2π, with $\int_0^{2\pi} |f(\theta)|^2 \, d\theta\}$, where $\langle f, g\rangle = \int_0^{2\pi} f(\theta)\overline{g(\theta)} \, d\theta$. This is a separable Hilbert space by taking step functions with rational heights and endpoints.

This has an interesting basis: e_n with $n \in \mathbb{Z}$ where

$$e_n = \frac{e^{in\theta}}{\sqrt{2\pi}}.$$

Then

$$\langle e_n, e_m\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{i(n-m)\theta} \, d\theta = \begin{cases} 0 & n \neq m \\ 1 & n = m. \end{cases}$$

Why does this have dense span?

Theorem 3.6 (Stone-Weierstrass Theorem)**.** *If $f \in C^0(S^1)$ then there exists $\sum a_n e^{in\theta}$ such that*

$$\lim_{N\to\infty} \sup_{\theta\in S^1} \left| f(\theta) - \sum_{-N}^{N} a_n e^{in\theta} \right|.$$

That's density in C^0 but not in L^2. But the point is that C^0 is dense in L^2, which can be done slickly via mollification.

Consider $f \in L^2$ and pick $\chi \in C_0^\infty((-\pi, \pi))$ with $\chi \geq 0$ and $\int x = 1$. This is a bump function. Then define $\chi_\varepsilon(\theta) = \varepsilon^{-1}\chi(\frac{\theta}{\varepsilon})$. Now, we have $\int \chi_\varepsilon(\theta) \, d\theta = 1$. Now define $f_\varepsilon(\theta) = (f \star \chi_\varepsilon)(\theta) = \int_0^{2\pi} f(\tilde{\theta})\chi_\varepsilon(\theta - \tilde{\theta}) \, d\tilde{\theta}$. Then $|f_\varepsilon(\theta)| \leq \|f\|_{L^2} \|\chi_\varepsilon\|_{L^2}$, which is bounded for each ε (but not uniformly). We need to check that $f_\varepsilon(\theta) \in C^\infty$ for all $\varepsilon > 0$, and that $\|f_\varepsilon - f\|_{L^2} \to 0$.

The first fact is true because $\partial_\theta^j f_\varepsilon = f \star (\partial_\theta^j \chi_\varepsilon)$. For the second fact, we have

$$\int_{S^1} \left| \int_{S^1} f(\tilde{\theta})\chi_\varepsilon(\theta - \tilde{\theta}) - f(\theta) \right|^2 d\theta \leq \int_{S^1} \left| \int_{S^1} (f_\varepsilon(\theta) - f(\theta))\chi_\varepsilon(\theta - \tilde{\theta}) \, d\tilde{\theta} \right| d\theta$$

$$= \int_{S^1} \left| \int_{S^1} \chi_\varepsilon(\hat{\theta})(f(\theta - \hat{\theta}) - f(\theta)) \, d\hat{\theta} \right|^2 d\theta \to 0$$

as a simple estimate. We can apply a class of inequalities known as Young's inequalities.

Proposition 3.7. $\|u \star v\|_{L^p} \leq \|u\|_{L^1} \|v\|_{L^p}$.

This implies that $f_\varepsilon \in L^2$ uniformly in ε, which is already much better than what we had before.

Now, we've shown that C^0 is dense in L^2. Now, we have that $\sum_{\text{finite}} a_n e^{in\theta}$ is dense in C^0 and hence L^2.

So now we have that for $f \in L^2$, we can find $g \in C^0$ such that $\|f - g\|_{L^2} < \varepsilon$ and we can find $\sum_{-N}^N a_n e^{in\theta}$ such that $\left\|g - \sum a_n e^{in\theta}\right\|_{C^0} < \varepsilon$, and hence

$$\|g - h\|_{L^2} = \left(\int |g - h|^2 \, d\theta\right)^2 \leq \|g - h\|_{C^0} \left(\int 1\right)^{1/2} = C\|g - h\|_{C^0} \leq C\varepsilon.$$

This closely relates to Parseval's Theorem.

Theorem 3.8 (Parseval's Theorem)**.** *If $f \in L^2(S^1)$ has coefficients a_n then $\|f\|^2 = \sum_{-\infty}^{\infty} |a_n|^2$.*

Now we can discuss the Riesz Representation Theorem. Given a Hilbert space H, we have the dual space H^* given by continuous and linear $\ell : H \to \mathbb{C}$.

Theorem 3.9 (Riesz Representation Theorem)**.** *Given any $\ell \in H^*$ there exists $w \in H$ such that $\ell(v) = \langle v, w \rangle$.*

Fix w. Then the map $v \to \langle v, w \rangle$ satisfies $| \langle v, w \rangle | \leq \|v\| \|w\| = C \|v\|$. The claim is that these are the only ones.

Proof. Fix any $\ell \in H^*$. Define $W = \ell^{-1}(0) = \{w \in H : \ell(w) = 0\}$. This is a closed subspace. Then the projection theorem that we proved last time says that if $W \subsetneq H$ then there exists $v_0 \perp W$. We can even assume that $\|v_0\| = 1$.

We then see that $F(v) = \ell(v_0) \langle v, v_0 \rangle \in H^*$. We claim that $F = \ell$. Note that F and ℓ both vanish on W and they agree on v_0. So they agree on $W \oplus \operatorname{span} v_0$.

The final thing to check is that $W \oplus \mathbb{C}v_0 = H$. Suppose that $v_1, v_2 \perp W$ are independent. Then we have find $\alpha\ell(v_1) + \beta\ell(v_2) = 0$, but then $\ell(\alpha v_1 + \beta v_2) = 0$, which is a contradiction. □

This theorem in fact gives us something even slightly stronger: a norm on H^*. We have $\|\ell\|_{H^*} = \sup_{\|v\| \leq 1} |\ell(v)|$.

Corollary 3.10. *Given any ℓ, we associated to it a vector w. Then $\|\ell\|_{H^*} = \|w\|_H$.*

Proof. We have

$$\|\ell\| = \sup_{\|v\| \leq 1} |\ell(v)| = \sup_{\|v\| \leq 1} | \langle v, w \rangle | \leq \|w\|_H .$$

The other inequality is

$$\|\ell\|_H = \sup |\ell(v)| \geq \left| \ell\left(\frac{w}{\|w\|}\right) \right| = \frac{| \langle w, w \rangle |}{\|w\|} = \|w\| .$$ □

What is the broader context of the Riesz Representation Theorem? Already, last week we discussed the general idea of the normed linear space or Banach space $(V, \|\cdot\|)$. In this context, we can talk about the dual space in exactly the same way. The question is: Can we identify the dual space V^*?

In the case where V is a Hilbert space, we have $V^* = V$. As a more interesting example, we have $(L^1([0,1]))^* = L^\infty([0,1])$ and more generally, we have $(L^p([0,1]))^* = L^q([0,1])$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Here's the whole point of studying measure theory: $C^0([0,1])^* = \mathcal{M}([0,1])$ is the set of signed Borel measures. This is the Riesz-Fisher theorem. The point is that $\ell(u) = \int u \, d\mu$.

Now, we have $C^k([0,1])^*$ is the collection of things that are k derivatives of measures, which makes sense using distributions.

Next time, we'll talk more about dual spaces and give nice examples of the Riesz Representation Theorem.

4. 1/19

Today, we will discuss a few small extensions of the Riesz Representation Theorem 3.9, and do lots of examples.

Recall that the Riesz Representation Theorem says that if H is a Hilbert space and H^* is the dual space, then $H \cong H^*$, i.e. give any $y \in H$, we can define $\ell_y \in H^*$ such that

$\ell_y(x) = \langle y, x\rangle$ and $\|ly\| = \|y\|$ because $|\langle x, y\rangle| \le \|x\| \|y\|$. Conversely, given any $\ell \in H^*$, there exists unique $y \in H$ such that $\ell = \ell_y$.

Here's an alternate formulation:

Theorem 4.1. *Suppose $B : H \times H \to \mathbb{C}$ is sesquilinear, satisfying $|B(x, y)| \le C \|x\| \|y\|$. Then there exists a linear operator $A : H \to H$ bounded ($A \in \mathcal{B}(H)$) i.e. $\|Ax\| \le C \|x\|$ such that $B(x, y) = \langle x, Ay\rangle$ for all $x, y \in H$.*

Proof. Fix x and note that $y \mapsto \overline{B(x, y)}$ is in H^*, i.e. $\overline{B(x, y)} = \overline{\langle x, v_y\rangle}$. This defines $Ay = v_y$. We just need to check that $y \to v_y$ is linear and bounded.

- $B(x, y_1 + y_2) = \langle x, v_{y_1+y_2}\rangle = B(x, y_1) + B(x, y_2) = \langle x, v_{y_1}\rangle + \langle x, v_{y_2}\rangle$.
- $\|Ay\| = \|v_y\| \le C \|y\|$, or $|\langle x, v_y\rangle| = |B(x, y)| \le C \|x\| \|y\|$.

□

Here is a slight generalization with a bit more content:

Theorem 4.2 (Lax-Milgram Lemma)**.** *Take $B : H \times H \to \mathbb{C}$ sesquilinear such that $|B(x, y)| \le C \|x\| \|y\|$ and $B(x, x) \ge c_1 \|x\|^2$ (with $c_1 > 0$). This property is often called* coercivity*. Then, given any $\ell \in H^*$, there exists a unique $v \in H$ so that $\ell(u) = B(u, v)$ for all $u \in H$.*

Remark. If $B(x, x) \ge -c_2 \|x\|^2$ then we can study $\hat{B}(x, y) = B(x, y) + (c_2 + 1) \langle x, y\rangle$.

This is just Riesz representation with respect to this other inner product.

Proof. First, $B(x, y) = \langle x, Ay\rangle$. We know that $\|Ay\| \le C \|y\|$, and $c_1 \|y\|^2 = B(y, y) = \langle y, Ay\rangle \le \|y\| \|Ay\|$, so $\|Ay\| \ge c_1 \|y\|$. Then $c_1 \|y\| \le \|Ay\| \le C \|y\|$.

This tells us that as a bounded linear transformation, $A : H \to H$ is injective. We want it to be invertible. Also, A is surjective. For if not, we can define $V = \text{im}(A) \subsetneq H$ is a *proper* closed subspace. Now, assume that V is closed. Choose $z \in H$ with $z \ne 0$ and $z \perp V$. Then $\langle z, Ay\rangle = 0$ for all y, so $B(z, y) = 0$ for all y. Choose $y = z$ to see that $z = 0$, which is a contradiction.

We claimed that V is closed, and we will prove it now. Suppose $w_j \in V$ and $w_j \to w$ in H. So we can write $w_j = Ay_j$. Then $c_1 \|y_j - y_k\| \le \|A(y_j - y_k)\| = \|w_j - w_k\| \to 0$, so we have a Cauchy sequence in a complete space, so we win.

We've now proved that $A : H \to H$ is an isomorphism. Here, $c_1 \|y\| \le \|Ay\|$ if and only if $c_1 \|A^{-1}w\| \le \|w\|$, so the inverse is bounded.

Now, $\ell \in H^*$. Ordinary Riesz gives us $\ell(x) = \langle x, w\rangle = B(x, A^{-1}w)$. □

Our basic example is that $(\ell^2)^* = \ell^2$. We want to define a family of separable Hilbert spaces $\ell^2_s = \{x : \|x\|_s^2 := \sum_{j=1}^\infty |x_j|^2 j^{2s} < \infty\}$. This is what is called a natural pairing: $\ell^2_s \times \ell^2_{-s} \to \mathbb{C}$ via $x, y \to \sum_{j=1}^\infty x_j \overline{y_j}$. Why is this well-defined? We have

$$\left|\sum x_j \overline{y_j}\right| \le \left(\sum |x_j|^2 j^{2s}\right)^{1/2} \left(\sum |y_j|^2 j^{-2s}\right)^{1/2}.$$

We can think of this as $H_1 \times H_2 \to \mathbb{C}$ is a *perfect pairing.* If we have $\ell : H_1 \to \mathbb{C}$ is bounded then the claim is that there exists $y \in \ell^2_{-s}$ with $\ell(x) = \langle x, y\rangle_{\ell^2}$, which is not on ℓ^2_s.

So now, we take $\ell(x) = \langle x, \hat{y}\rangle_{\ell^2_s} = \sum x_j \overline{\hat{y}_j} j^{2s}$. Define $y_j = \hat{y}_j j^{2s}$, The claim is that $\hat{y} \in \ell^2_s$. This implies that $y \in \ell^2_{-s}$.

Now, consider $L^2(S^1)$ with orthonormal basis $e^{inx}/\sqrt{2\pi}$. Note that any $f \in L^2(S^1)$ corresponds to an infinite sequence $\{a_n\}_{-\infty}^{\infty}$, where we can write

$$a_n = \left\langle f, e^{in\theta}/\sqrt{2\pi} \right\rangle_{L^2} = \frac{1}{2\pi} \int_{S^1} f(\theta) e^{-in\theta}\, d\theta.$$

Here, $f = \sum a_n e_n$ implies that $f = \sum_{-\infty}^{\infty} \frac{a_n e^{in\theta}}{\sqrt{2\pi}}$.

Then $S_N(f) = \sum_{-N}^{N} \frac{a_n e^{in\theta}}{\sqrt{2\pi}} \to f$ in L^2. The problem is that this convergence in L^2 is not so great. Can we do better? There are many questions here.

Definition 4.3. Define $H^s(S^1) = L^2_s(S^1)$ given by $\{f : (a_n) \in \ell^2_s,$ i.e. $\sum_{-\infty}^{\infty} |a_n|^2 (1+n^2)^s < \infty\}$. If $s > 0$, we have $H^s \subset L^2 \subset H^{-s}$.

This may look artificial, but it comes from something very natural. Take the case $s = 1$. We have

$$H^1(S^1) = \left\{ f : \sum |a_n|^2 (1+n^2) < \infty \right\}.$$

Then

$$f(\theta) = \frac{1}{\sqrt{2\pi}} \sum a_n e^{in\theta}$$

and

$$f'(\theta) = \frac{1}{\sqrt{2\pi}} \sum i n a_n e^{in\theta}.$$

Suppose that both $f, f' \in L^2$. Then $\sum |a_n|^2 (1+n^2) < \infty$. So H^s consists of f such that $f, f' \in L^2$. These are called *Sobolev spaces*, and they are very fundamental examples of Hilbert spaces.

Now, what is H^{-1}? We have a pairing $H^1 \times H^{-1} \to \mathbb{C}$. We have f and g corresponding to $\{a_n\}$ and $\{b_n\}$. It makes sense to consider $\sum a_n \overline{b_n}$,

Then we have

$$\int_0^{2\pi} f(\theta) \overline{g(\theta)}\, d\theta.$$

Here, f is better than L^2, so we integrate it by something worse than L^2. Also g really should be thought of as a distribution: it doesn't converge as a usual function.

If $f \in L^2$, does $S_N(f) \to f$ almost everywhere? If f is continuous, does $S_N(f) \to f$ in C^0? What if $f \in C^k$, or $f \in H^s$?

One of the famous examples is that the answer to the first question is no! This convergence is extraordinarily badly behaved.

We will use Cesaro summation to define

$$\sigma_N(f) = \frac{S_0(f) + \cdots + S_N(f)}{N+1}.$$

These are called the Cesaro partial sums and the Fejer partial sums. It turns out that this is

$$\sigma_N(f) = \frac{1}{\sqrt{2\pi}} \sum_{-N}^{N} \left(1 - \frac{|n|}{N+1}\right) a_n e^{in\theta}.$$

($2\pi = 1$ is the basic assumption of harmonic analysis.)

Now the point is that we're summing with a smoother filter, making the properties for σ_N much better behaved than for s_N.

So let's ask the same questions for σ_N instead of s_N. For the first question, the answer is now yes!

There are two very good references:

- Katznelson: Harmonic analysis
- Pinsky: Fourier analysis and wavelets

Proposition 4.4. *There exist functions $D_N(\theta)$ and $K_N(\theta)$ such that $S_N(f)(\theta) = (D_N \star f)(\theta)$ and $\sigma_N(f)(\theta) = (K_N \star f)(\theta)$.*

Here we think of the circle as a group $S^1 \cong \mathbb{R}/2\pi\mathbb{Z}$ in order to think about convolution.

Proof. Write

$$S_N(f) = \frac{1}{2\pi}\sum_{-N}^{N} e^{in\theta}\int_{S^1} f(\tilde{\theta})e^{-in\tilde{\theta}}d\tilde{\theta} = \int f(\tilde{\theta})\sum_{-N}^{N}\frac{1}{2\pi}e^{in(\theta-\tilde{\theta})}d\tilde{\theta}.$$

For σ_N, we do the same thing.

$$\sigma_n(f)(\theta) = \int f(\tilde{\theta})\frac{1}{2\pi}\sum_{-N}^{N}\left(1 - \frac{|n|}{N+1}\right)e^{in(\theta-\tilde{\theta})}d\tilde{\theta}.$$

□

In fact, we can write

$$D_N(\theta) = e^{-iN\theta} + \cdots + e^{iN\theta} = \frac{1}{2\pi}\frac{\sin(N+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}.$$

This is an even function that oscillates, and it has lots of nice properties. For example, $\int_{-\pi}^{\pi} D_N(\theta) = 1$. This is the Dirichlet kernel.

With a bit more calculation, we have

$$K_N(\theta) = \frac{1}{N+1}\left(\frac{\sin(N+\frac{1}{2})\theta}{\sin\frac{\theta}{2}}\right)^2.$$

This has $\int K_N(\theta) = 1$ as well, and it is positive. It is called the Fejer kernel.

Our question is now: $D_N \star f \to f$ in L^2? Does $K_N \star f \to f$ in L^2? Both answers are yes.

We can now finish a calculation that we started last time.

$$\begin{aligned}
\|f \star K_N(\theta) - f(\theta)\|_{L^2}^2 &= \int \left|\int K_N(\hat{\theta})(f(\theta-\hat{\theta}) - f(\theta))d\hat{\theta}\right|^2 d\theta \\
&\leq \int \left(\int K_N(\hat{\theta})d\hat{\theta}\right)\left(\int K_N(\hat{\theta})(f(\theta-\hat{\theta}) - f(\theta))\,d\hat{\theta}\right)d\theta \\
&\leq \int\int K_N(\hat{\theta})(f(\theta-\hat{\theta}) - f(\theta))\,d\theta\,d\hat{\theta}.
\end{aligned}$$

Now we split the integral up into two pieces, and the two pieces are small in different ways. One is because the Fejer kernel is uniformly small. The other piece needs that L^2 norm is translation invariant, which is because the continuous integrals are dense, and we can do that by Lusin's theorem. So that's a very fundamental technique.

5. 1/24

Suppose that H is a Hilbert space with orthonormal basis $\{e_n\}$. As an example, we can consider the theory on $L^2(S^1)$. Last time, we discussed Plancherel's theorem, giving a norm-preserving map $L^2 \to \ell^2$. We also saw two approximations: $S_N(u) = D_n \star u(\theta)$ and $\sigma_N(u) = K_n \star u(\theta)$. We clearly have $S_N(u), \sigma_N(u) \to u$ in L^2. Here's the question we'll discuss today: Suppose $u \in C^0(S^1) \hookrightarrow L^2(S^1)$. Does $S_N(u) \to u$ and $\sigma_N(u) \to u$ in C^0? The answer is no and yes respectively.

Proposition 5.1. *If $u \in C^0$ then $\sigma_N(u) \to u$ in C^0.*

Proof. We need to verify that

$$|K_N \star u(\theta) - u(\theta)| \to 0$$

as $N \to \infty$. Recall that

$$K_N = \frac{1}{N+1}\left(\frac{\sin((N+\frac{1}{2})\theta)}{\sin\frac{\theta}{2}}\right)^2.$$

Then we need to estimate

$$\begin{aligned}\left|\int K_N(\theta-\theta')(u(\theta')-u(\theta))\,d\theta'\right| &\le \int_{S^1} K_N(\theta-\theta')|u(\theta')-u(\theta)|\,d\theta' \\ &= \left(\int_{|\theta-\theta'|<\eta} + \int_{|\theta-\theta'|>\eta}\right) K_N(\theta-\theta')|u(\theta')-u(\theta)|\,d\theta',\end{aligned}$$

and each piece is small. □

A general principle is that regularity of u corresponds to delay of the Fourier coefficients a_n. For example, $u \in C^\infty$ corresponds to $|a_n| \le C_N(1+|u|)^{-N}$ for any N.

Proof. We have

$$\begin{aligned} a_n &= \int e^{-in\theta}u(\theta)\,d\theta \\ u(\theta) &= \sum a_n e^{in\theta}. \end{aligned}$$

$\Leftarrow$: We can differentiate under the summation sign to get that all series $\sum a_n n^k e^{in\theta}$ are absolutely convergent, so $u \in C^\infty$.

$\Rightarrow$: We have $\int e^{-in\theta}u^{(k)}(\theta)\,d\theta = (in)^k a_n$, so that $\{n^k a_n\} \in \ell^2$ for any k. □

We now mention some other interesting orthonormal bases.

(1) Take $L = -\frac{d^2}{d\theta^2} + q(\theta)$ on S^1. This is an ordinary differential operator on S^1. If $u \in C^2(S^1)$ then $Lu \in C^0(S^1)$. We want to consider $L : L^2(S^1) \to L^2(S^1)$. This is linear, but *not* continuous. This is defined on the dense subspace $H^2 \hookrightarrow L^2$.

We are interested in the spectrum of this operator. There exists $u_k(\theta) \in C^\infty(S^1)$ and $\lambda_k \to +\infty$ where $\langle u_k, u_l\rangle = \delta_{kl}$, and $Lu_k = \lambda_k u_k$. These are eigenfunctions and eigenvalues. Moreover, $\{u_n\}$ are dense in L^2, so $u = \sum a_n u_n(\theta)$. This is a generalization of Fourier series.

In the special case $q = 0$, we have $L_0 e^{in\theta} = n^2 e^{in\theta}$. So all of Fourier series is a special case of this. We can also do this on any compact manifold, and this is only the tip of the iceberg. There is a general theorem that says that we always have orthonormal sequences of eigenfunctions.

(2) Wavelet bases. We want an orthonormal basis for $L^2([0,1])$. For $u \in L^2$, we can write $u = \sum a_n e^{2\pi i n\theta}$.

Many signals are messy in some parts of the interval and pretty smooth for most of the time. How do we detect the complicated but localized geometric properties? With Fourier series, we cannot, and the Fourier series converges only through lots of mysterious cancellations. To study this type of signal, we need different bases.

We start with a function that is $1/2$ on $[0,1/2]$ and $-1/2$ on $[1/2,1]$. Then $\psi_{n,k}(x) = \psi(2^n x - k)$. This sort of basis allows us to isolate the interesting behavior in a signal and identify where a signal is flat and boring.

Finally, we have one more topic about Hilbert spaces. This is an application of the Riesz representation theorem 3.9 or its slight generalization, the Lax-Milgram lemma 4.2: Given a bilinear function $B(x,y)$, we require $|B(x,y)| \leq \|x\| \|y\|$ and $c\|x\|^2 \leq B(x,x)$ (coercivity). Then, given any $\ell \in H^*$, there exists v such that $\ell(u) = B(u,v)$ for all $u \in H$.

Take a domain $\Omega \subset \mathbb{R}^n$. This is a bounded smooth domain, e.g. $|x| \leq 1$. Consider a positive definite matrix of functions $(a_{ij}(x)) > 0$ with $a_{ij} \in C^\infty$. Then

$$B(u,v) = \int_\Omega a_{ij}(x)\frac{\partial u}{\partial x_i}\overline{\frac{\partial v}{\partial x_j}} + \sum b_k(x)\frac{\partial u}{\partial x_k}\overline{v} + c(x)u\overline{v}$$

for $u, v \in H^1$ and $u, v, \nabla u, \nabla v \in L^2$.

We verify that this satisfies the conditions of the Lax-Milgram lemma. That is,

$$|B(u,v)| \leq C\left(\int |u|^2 + |\nabla u|^2\right)^{1/2}\left(\int |v|^2 + |\nabla v|^2\right)^{1/2}.$$

Also,

$$c\int |u|^2 + |\nabla u|^2 \leq \int_\Omega a_{ij}(x)\frac{\partial u}{\partial x_i}\overline{\frac{\partial u}{\partial x_j}} + \sum b_k(x)\frac{\partial u}{\partial x_k}\overline{u} + c(x)|u|^2$$

Assume that $c(x) \geq A \gg 0$. Then

$$\int a_{ij}(x)\frac{\partial u}{\partial x_i}\overline{\frac{\partial u}{\partial x_j}} \geq \inf_{x\in\Omega} \lambda_{\min}(x)\|\nabla u\|^2.$$

Also

$$\int b_k(x)\frac{\partial u}{\partial x_k}\overline{u} \geq -\eta\|\nabla u\|^2 - \frac{C}{\eta}\|u\|^2.$$

To see that, note that we can write

$$|\langle f, g\rangle| \leq \|f\|\|g\| \leq \frac{\eta}{2}\|f\|^2 + \frac{1}{2\eta}\|g\|^2.$$

Then

$$\int c(x)|u|^2 \geq A\|u\|^2.$$

Now, Lax-Milgram says that given any $\ell \in (H^1)^*$, there exists $v \in H^1$ such that $\ell(u) = B(u,v)$.

For example, $\ell(u) = \int u\overline{f}$ for $f \in L^2$, and $|\int u\overline{f}| \leq \|u\|_{H^1}\|f\|_{L^2}$.

This tells me that

$$\int u\overline{f} = \int a_{ij}(x)\frac{\partial u}{\partial x_i}\frac{\partial \overline{v}}{\partial x_j} + \sum b_k(x)\frac{\partial u}{\partial x_k}\overline{v} + c(x)u\overline{v}$$
$$= \int u\left(-\sum \overline{\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial v}{\partial x_j}\right)} - \overline{\frac{\partial}{\partial x_k}(b_k(x)v)} + \overline{c(x)v}\right)$$

for all u, so we end up finding that

$$f = \sum -\frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial v}{\partial x_j}\right) - \sum \frac{\partial}{\partial x_k}(b_k(x)v) + c(x)v.$$

This is an abstract theorem and the motivating example for Lax and Milgram.

Now, we will move on to the theory of *Banach spaces*. There are complete normed vector spaces $(X, \|\cdot\|)$.

Example 5.2. Here are some examples:

(1) $\ell^p = \{x : \|x\|_p = (\sum |x_j|^p)^{1/p} < \infty\}$
(2) $L^p(\Omega, d\mu)$ for $1 \leq p \leq \infty$
(3) $C^0(\mathbb{R}^n)$
(4) $C^k(\mathbb{R}^n)$

However,

$$\{x : \|x\|_{2,N} = \sum |x_j|^2 j^{2N} < \infty \text{ for some } N\}.$$

This is a slightly more general type of space, and it is not a Banach space, as it is regulated by a countable set of norms instead of just one norm.

Let X and Y be Banach spaces. Then define the space $\mathcal{B}(X, Y) = \{A : X \to Y, \|Ax\| \leq C\|x\| \text{ for all } x\}$. This has the *operator norm*

$$\|A\| = \sup_{x\neq 0}\frac{\|Ax\|}{\|x\|}.$$

Theorem 5.3. *$(\mathcal{B}(X, Y), \|\cdot\|)$ is a Banach space.*

When $Y = \mathbb{C}$, $\mathcal{B}(X, \mathbb{C}) = X^*$ is called the *dual space*.

Proof. We only need to prove completeness. That is, let $A_n \in \mathcal{B}(x, y)$ be Cauchy with respect to the operator norm: $\|A_n - A_m\| \to 0$ for $n, m \to \infty$.

Effectively, if X is the unit ball, then AX is a generalized squashed ellipsoid. A is continuous if this image is bounded, and the norm is the diameter of the image.

First, for $x \in X$, then $A_n x$ is a sequence in Y. This is Cauchy because $\|A_n x - A_m x\| \leq \|A_n - A_m\|\|x\| \to 0$, so therefore we can define $Ax = \lim_{n\to\infty} A_n x$. It is straightforward to check that A is linear.

We need to check that A is bounded, and then that $\|A - A_n\| \to 0$.

First, using the triangle inequality, we have $\|A_n\| \leq \|A_n - A_m\| + \|A_m\|$, so $|\,\|A_n\| - \|A_m\|\,| \leq \|A_n - A_m\|$, i.e. $\{\|A_n\|\}$ is a Cauchy sequence too. Then $\|Ax\| = \lim\|A_n x\| \leq \lim\|A_n\|\|x\| \leq C\|x\|$.

Finally,

$$\frac{\|(A - A_n)x\|}{\|x\|} = \lim_{n\to\infty}\frac{\|(A_m - A_n)x\|}{\|x\|} \leq \lim_{n\to\infty}\|A_m - A_n\| \to 0$$

as $n \to \infty$. $\square$

Now, what are dual spaces? We have $(L^p)^* = L^q$ if $1 \le p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. This is true even for general measure spaces. This is related to the fact that

$$\left| \int f\overline{g} \right| \le \|f\|_p \|g\|_q$$

for $f \in L^p$ and $g \in L^q$. However, $(L^\infty)^*$ is complicated and we don't think about it.

Now, for $1 < p < \infty$, we can map $L^p \to (L^p)^* = L^q \to (L^p)^{**} = (L^q)^* = L^p$. This is a special property, and it means that L^p is *reflexive*.

It is true in general, however, that $X \hookrightarrow X^{**}$ injects into its double dual.

Given $x \in X$ and $\ell \in X^*$, we can take $x(\ell) = \ell(x)$. Then $\ell \to x(\ell)$ is a continuous linear functional on X^*, and $|\ell(x)| \le \|\ell\|_{X^*} \|x\|_X$. Then $x \mapsto \mu_x \in X^{**}$. This shows us that $\|\mu_x\|_{X^{**}} \le \|x\|_X$.

6. 1/26

Last time, we talked about the general notion of a Banach space, and we defined a dual space. An interesting question is: Can we identify the dual of any given space?

Proposition 6.1. *Suppose that (M, μ) is a measure space, and $\mu(M) = 1$. Suppose $1 < p < 2$. Then $(L^p(M, d\mu))^* = L^q(M, d\mu)$ where $\frac{1}{p} + \frac{1}{q} = 1$.*

Note that for $p = 2$ this is given by the Riesz representation theorem 3.9.

Proof. Use the Holder inequality to see that

$$\|f\|_{L^p}^p = \int |f|^p \le \left(\int 1\right)\left(\int |f|^2\right)^{2/p} \le \left(\int |f|^2\right)^{2/p},$$

so hence $\|f\|_p \le \|f\|_2$. This means that we have inclusions $L^2 \hookrightarrow L^p$.

Take $\ell \in (L^p)^*$ is a bounded linear functional $\ell : L^p \to \mathbb{C}$. Then $|\ell(f)| \le C\|f\|_p \le C\|f\|_2$. Then $L^2 \hookrightarrow L^p$ is dense, so ℓ_{L^2} is continuous with respect to the L^2 norm, so we can think of $\ell \in (L^2)^*$. Therefore, $\ell(f) = \int f\overline{g}$, $g \in L^2$. The last trick is to show that $g \in L^q$, which we do not get for free.

Define the truncations $|g_k|(x) = \min\{|g(x)|, k\}$. Then $|g_n| \le L^r$ for all r. We will show that $\||g_k|\|_q \le C$ uniformly in k.

For ease of notation, assume g is real-valued (i.e. take real and imaginary parts). Write $f_k = |g_k|^{q-1} \operatorname{sgn} g$. Then

$$\ell(f_k) = \int |g_k|^{q-1}(\operatorname{sgn} g)g = \int |g_k|^{q-1}|g| \ge \int |g_k|^q.$$

Also,

$$\|f_k\|_p^p = \int |f_k|^p = \int |g_k|^{p(q-1)} = \int |g_k|^q.$$

(Note that $\frac{1}{p} + \frac{1}{q} = 1$, so that $q = p(q-1)$.) Note, we have

$$\int |g_k|^q \le C\left(\int |g_k|^q\right)^{1/p},$$

which shows that

$$\left(\int |g_k|^q\right)^{1/q} \leq C.$$

Therefore, $g \in L^q$ because it is the limit of truncations. This concludes the proof. $\square$

We have a mapping $X \to X^*$, and we can do it again: $X \to X^* \to X^{**}$. We claim that there exists a natural mapping $F : X \to X^{**}$ that is an isometric injection.

Definition 6.2. X is reflexive if this F is onto, i.e. $X = X^{**}$.

For $x \in X$, we can write the map F as $F(x)(\ell) = \ell(x)$. Then

$$|F(x)(\ell)| \leq \|\ell\|_{X^*} \|x\|_X .$$

Therefore, $\|F(x)\|_{X^{**}} \leq \|x\|_X$. Then

$$\frac{\|F(x)(\ell)\|}{\|\ell\|} \leq \|x\| .$$

Given any $x \in X$, we can choose an $\ell \in X^*$ so that $\|\ell\|_{X^*} = 1$ and $|\ell(x)| = \|x\|$.

Note that

$$\|F(x)\|_{X^{**}} = \sup \frac{\left\|F(x)(\hat{\ell})\right\|}{\left\|\hat{\ell}\right\|} \geq \|x\|$$

which implies that this is norm-preserving and hence F is an isometry.

Here is a remarkable geometric characterization of reflexivity:

Theorem 6.3. *X is reflexive if and only if it is uniformly convex.*

Take the unit ball, and take points y, z in the ball. Then the line $\alpha y + \beta z$ for $\alpha + \beta = 1$ and $0 < \alpha, \beta$, so then $\|\alpha y + \beta z\| \leq \alpha \|y\| + \beta \|z\| \leq \alpha + \beta = 1$. Let $\|y\| = 1 = \|z\|$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left\|\frac{y+z}{2}\right\| > 1 - \delta$$

implies that $\|y - z\| < \varepsilon$. This is the meaning of uniform convexity.

How can that fail? Take $\|x\|_\infty = \sup_{j=1,2} |x_j|$. Then the unit ball with respect to this norm is a square. Alternatively, $\|x\|_1$ gives a diamond. These are clearly far from uniformly convex.

Fact 6.4. L^p is reflexive if $1 < p < \infty$.

This means that $(L^p)^* = L^q$.

We would like to construct a lot of linear functionals. This doesn't depend on completeness, and it holds in great generality. We will prove three different versions of this.

Theorem 6.5 (Hahn-Banach)**.** *Suppose that X is a real Banach space. Choose a function $p : X \to \mathbb{R}^+$ (gauge) to be subadditive and positively homogeneous. (This means that $p(\alpha x + \beta y) \leq \alpha p(x) + \beta p(y)$ and $p(ax) = ap(x)$ for $a > 0$.)*

Suppose we have $\ell : Y \to \mathbb{R}$ where Y is a closed subspace of X with $|\ell(y)| \leq p(y)$ for all $y \in Y$. Then there exists $\tilde{\ell} : X \to \mathbb{R}$ bounded such that $|\tilde{\ell}(x)| \leq p(x)$ for all $x \in X$.

This is an extension theorem. Given a linear function on a small space (such as a line), we can extend without increasing the norm.

Proof. The proof is squishy.

Take $Y \subset X$ and pick any $z \in X \setminus Y$. Define $\tilde{\ell}(y + az) = \ell(y) + a\tilde{\ell}(z) \le p(y + az)$. Using positive homogeneity, we can scale this to factor out a factor of a. It is enough to check this for $a = \pm 1$.

Then $\ell(y) + \tilde{\ell}(z) \le p(y + z)$ and $\ell(y) - \tilde{\ell}(z) \le p(y - z)$. This means that

$$\ell(y) - p(y - z) \le \tilde{\ell}(z) \le p(y' + z) - \ell(y').$$

for any y and y'. In order for this to work, we need

$$p(y - z + y' + z) \ge \ell(y - z + y' + z) = \ell(y + y') = \ell(y) + \ell(y') \le p(y - z) + p(y' + z),$$

so those are true for any y, y'. We've shown that if $Y \subsetneq X$ then ℓ can be extended nontrivially.

So look at

$$\mathcal{E} = \{(\tilde{Y}, \tilde{\ell}) : \tilde{Y} \supsetneq Y,\ \tilde{\ell} : \tilde{Y} \to \mathbb{R} \text{ bounded, st } |\tilde{\ell}(y)| \le p(y) \text{ for all } y \in \tilde{Y},\ \tilde{\ell}|_Y = \ell\}.$$

This is a poset, so Zorn's Lemma implies that there exists a maximal set $(\overline{Y}, \bar{\ell})$. If $\overline{Y} \subsetneq X$ then there exists an extension, so $\overline{Y} = X$. □

Here's the geometric version.

Definition 6.6. Take some $S \subset X$. x_0 is called an interior point of S if $x_0 + ty \in S$ for $|t| < \varepsilon$ (depending on y).

Choose a convex set K and suppose $0 \in K$. Then define $p_K(x) = \inf\{a > 0 : \frac{x}{a} \in K\}$. This is positively homogeneous, and $p_K(x) < \infty$ always. Subadditivity uses convexity: Pick $x, y \in K$ and any numbers $a, b > 0$ such that $\frac{x}{a} \in K$ and $\frac{y}{b} \in K$. Now,

$$\frac{x + y}{a + b} = \frac{a}{a + b}\frac{x}{a} + \frac{b}{a + b}\frac{y}{b} \in K,$$

so $p_K(x + y) \le a + b$, and therefore $p_K(x + y) \le p_K(x) + p_K(y)$.

Now, note that $K = \{x : p_K(x) \le 1\}$. So convex sets and these gauge functions are exactly the same thing.

Theorem 6.7. *Suppose we take K so that it is nonempty and convex, and all points of K are interior.*

If $y \notin K$, we can choose a bounded linear functional $\ell \in X^*$ such that $\ell(y) = c$ and $\ell(x) < c$ for all $x \in K$.

This is some sort of separation theorem. We have a level set $\{\ell = c\}$ so that the convex set is on one side of it.

Proof. We have $0 \in K$ and define p_K as before. Then $p_K(x) < 1$ for all $x \in K$. Then we have a point $y \notin K$, and we define ℓ on $\{ay\}$ for $\ell(y) = 1$ and $\ell(ay) = a$. Therefore $\ell(y) \le p_K(y)$. Then by homogeneity, we have $\ell(ay) \le p_K(ay)$ for all $a \in \mathbb{R}$.

Extend to $\tilde{\ell} : X \to \mathbb{R}$. Then $\tilde{\ell}(x) \le p_K(x)$ for all $x \in K$. If $x \in K$ and $p_K(x) < 1$ then $\tilde{\ell}(x) < 1$. □

This is also a complex version of this. We don't have this picture; separation results do not make sense. This is an exercise.

There are many applications of the Hahn-Banach theorem.

Proposition 6.8. *Suppose that $Y \subset X$ is closed and $\ell \in Y^*$. Then there exists a continuous linear functional $L \in X^*$ so that $L|_Y = \ell$ so $\|L\|_{X^*} = \|\ell\|_{Y^*}$.*

Proof. Let $p(x) = \|\ell\|_{Y^*} \|x\|_X$. Then $|\ell(y)| \leq p(y)$. Choose extensions $L : X \to \mathbb{R}$ so that $|L(x)| \leq \|\ell\|_{Y^*} \|x\|_X$, so $\|L\|_{X^*} \leq \|\ell\|_{Y^*}$. □

Proposition 6.9. *Fix $y \in X$. Then there exists $\Lambda \neq 0$ on X^* such that $\Lambda(y) = \|\Lambda\|_{X^*} \|y\|_X$.*

Proof. Let $Y = \{ay\}$. Then let $\ell(ay) = a\|y\|$. Choose Λ an extension with $\|\Lambda\| = \|\ell\|$. Then $\|\ell\| = 1$ implies that $\|\Lambda\| = 1$. Now we can extend Λ to the whole space. □

Proposition 6.10. *If $W \in X$ is any subspace then*

$$\overline{W} = \{x \in X : \ell(x) = 0 \text{ for all } \ell \in X^*,\ \ell(W) = 0\}.$$

Proof. If $x \notin \overline{W}$, choose any $\ell \in X^*$ so that $\ell(x) \neq 0$ and $\ell|_{\overline{W}} = 0$. □

Here is a beautiful result. Suppose that $\{\zeta_j\}_{j=1}^\infty$ is a discrete set in $\mathbb{C}$. Consider $e^{ix\zeta_j} \in C^0([-\pi, \pi])$. When is their span dense? As an example, if ζ_j were the integers, it would be dense.

Theorem 6.11. *Define $N(t) = \#\{j : |\zeta_j| < t\}$. If $\limsup \frac{N(t)}{2t} > 1$ then $\operatorname{span}\{e^{ix\zeta_j}\}$ is dense.*

Proof. If the span is not dense, then there exists $\ell \in (C^0)^\infty$ such that $\ell \neq 0$ and $\ell(e^{ix\zeta_j}) = 0$, i.e. there exists measure μ such that

$$\int_{-\pi}^{\pi} e^{ix\zeta_j}\, d\mu = 0$$

for all j. Then

$$F(\zeta) = \int_{-\pi}^{\pi} e^{ix\zeta}\, d\mu.$$

Now, F is holomorphic in $\mathbb{C}$ and $F(\zeta_j) = 0$ for all j.

Let $N_1(t)$ be the number of zeros of F in $|\zeta| < t$. Then $|F(\zeta)| \leq Ce^{\pi|\operatorname{Im}\zeta|}$. A theorem in complex analysis says that

$$\frac{N_1(t)}{2t} \to \frac{\operatorname{diam}\operatorname{supp}\mu}{2\pi}.$$

But on the other hand, $\frac{N(t)}{2t} \leq \frac{N_1(t)}{2t}$, which is a contradiction. □

7. 1/31

Last time, we were talking about two notions: When X is a Banach space, we wanted to identify dual spaces X^*, and we were interested in reflexivity, when $X^{**} = X$.

The central problem in analysis is solving equations, and we want to find substitutes for compactness. Working with duals will allow us to do this.

Reflexivity is closely tied to the geometry of the closed unit ball. Last time, given the interior of a compact set, we could define a gauge function, which was effectively a norm. So we are talking about some flavor of convex set. The geometry of a convex set gives us information about the gauge function. We stated last time that reflexivity is equivalent to uniform convexity.

Another remarkable geometric characterization is: Reflexivity is equivalent to for all $\ell \in X^*$ there exists $x \in X$ with $\|x\| = 1$ such that $\ell(x) = \|\ell\|$.

Recall that given $x \in X$, there exists $\ell \in X^*$ such that $\ell(x) = \|x\|$ with $\|\ell\| = 1$. This is the dual statement, which is related to the result about level sets that we derived from the Hahn-Banach theorem. Reflexivity means: If the unit ball is rotund enough, it touches its support planes rather than just getting very close. This is due to James.

Proposition 7.1. *If X^* is separable then X is separable.*

Example 7.2. The converse is not true. L^1 is separable, but its dual is L^∞, which is not separable.

Proof. Choose $\{\lambda_n\}$ dense in X^*. For each n, choose $x_n \in X$ with $\|x_n\| = 1$ such that $\lambda_n(x_n) \geq \frac{1}{2}\|\lambda_n\|$.

Then define $\mathcal{D} = \{\sum_{\text{finite}} a_n x_n, a_n \in \mathbb{Q}\}$. We claim that $\mathcal{D} \subset X$ is dense. If not, then $\overline{\mathcal{D}} \subsetneq X$, and we can choose a $\lambda \in X^*$ such that $\lambda|_{\overline{D}} = 0$ and $\lambda \neq 0$.

We can now choose $\lambda_{n_k} \to \lambda$ in X^*. Then

$$\|\lambda - \lambda_{n_k}\|_{X^*} \geq \|(\lambda - \lambda_{n_k})(x_{n_k})\| = |\lambda_{n_k}(x_{n_k})| \geq \frac{1}{2}\|\lambda_{n_k}\|.$$

Hence $\lambda_{n_k} \to 0$ in X^*, so $\lambda = 0$, and we are done. The key here was the Hahn-Banach theorem 6.5. □

Now, we will move on to the three big theorems of Banach space theory.

Here is an illustrative theorem:

Theorem 7.3. *Then $T : X \to Y$ is bounded (i.e. $\|Tx\| \leq C\|x\|$ for some C independent of x) if and only if $T^{-1}(\{\|y\| \leq 1\})$ has nontrivial interior.*

Proof. Note that $\Rightarrow$ is obvious. We do $\Leftarrow$.

Suppose that $\|T(a)\| \leq 1$ and $\{\|x - a\| < \varepsilon\} \subset T^{-1}\{\|y\| \leq 1\}$. Then if $\|x\| < \varepsilon$ then it doesn't get distorted too much: $\|Tx\| = \|T(-a) + T(x + a)\| \leq \|T(a)\| + 1 \leq 2$. The point is that it doesn't get infinitely distorted when we apply T.

Take any $z \in X$ then $\frac{\varepsilon}{2}\frac{z}{\|z\|} \in B_\varepsilon(0)$ implies that $\left\|T(\frac{\varepsilon}{2}\frac{z}{\|z\|})\right\| \leq 2$, so therefore linearity gives $\|T(z)\| \leq \frac{4}{\varepsilon}\|z\|$. □

Here is the first of the big theorems:

Theorem 7.4 (Baire Category Theorem)**.** *Let (M, d) be any complete metric space with an infinite number of points. Then $M \neq \bigcup_{j=1}^\infty A_j$ where each A_j is nowhere dense, i.e. where $\overline{A_j}$ has no interior.*

Remark. The key is that we only allow ourselves a countable number of sets A_j. We will see lots of applications of this, and we will primarily apply this when M is a Banach space.

Proof. Suppose $M = \bigcup_{j=1}^\infty A_j$ where each A_j is nowhere dense. We will construct a Cauchy sequence with limit (using completeness) outside every one of the A_j.

Pick $x_1 \in M \setminus \overline{A_1}$, and choose a ball B_1 so that $x_1 \in B_1 \subset M \setminus \overline{A_1}$ with radius $r_1 < 1$.

Pick $x_2 \in B_1 \setminus \overline{A_2}$, and choose a ball B_2 so that $x_2 \in B_2 \subset B_1 \setminus \overline{A_2}$ with radius $r_2 < \frac{1}{2}$.

We continue in this way: Pick $x_n \in B_n \subset \overline{B_n} \subset B_{n+1}$ and $B_n \subset B_{n-1} \setminus \overline{A_n}$ with radius $r_n < 2^{1-n}$. So $\{x_n\}$ is defined, and $d(x_n, x_m) \leq 2^{1-n}$ if $m > n$. So this is a Cauchy sequence. Finally, let $x = \lim x_n$. Note that $x_n \in B_N$ if $n \geq N$, so therefore $x \in \overline{B_N} \subset B_{N-1}$, disjoint from A_N. This is true for all A_N, so we have a contradiction. □

We will now discuss various corollaries of the Baire Category Theorem.

- Uniform boundedness principle (Banach-Steinhaus)
- Open mapping theorem
- Closed graph theorem.

Let $\mathcal{T}$ be some collection of linear operators $T : X \to Y$. Recall that there are three ways of measuring the size of an operator:

- $\|T\|$ operator norm
- $\|T(x)\|$ for any x (strong boundedness)
- $|\ell(T(x))|$ weak boundedness for any x, ℓ fixed.

The uniform boundedness principle connects the first two ideas.

Theorem 7.5 (Uniform boundedness principle)**.** *If, for each $x \in X$, $\{\|T(x)\|\}_{T\in\mathcal{T}}$ is bounded (i.e. $\|T(x)\| \leq C(x)$), then $\{\|T\|\}_{T\in\mathcal{T}}$ is bounded, i.e. $\|T(x)\| \leq C\,\|x\|$ for C independent of $x \in X$ and $T \in \mathcal{T}$.*

Remark. We are looking at a uniform modulus of continuity: The unit ball doesn't get squished arbitrarily much.

Proof. Define $A_n = \{x : \|T(x)\| \leq n \text{ for all } T \in \mathcal{T}\}$. Our hypothesis implies that $\bigcup_{n=1}^{\infty} A_n = X$. The Baire Category Theorem 7.4 implies that some A_n has nonempty interior.

Then the argument of the proof of theorem 7.3 implies that $\{\|T\| : T \in \mathcal{T}\}$ is bounded, by picking $x \in B_\varepsilon(0)$ and showing that $\|T(x)\| \leq 2$. □

Here is an application of this.

Proposition 7.6. *There exists $f \in C^0(S^1)$ such that $S_N(f)(0)$ are not bounded.*

Proof. Recall $S_N f(\theta) = D_n \star f(\theta)$. We can write

$$S_N f(0) = \int f(\theta) D_N(\theta)\, d\theta = \ell_N(f).$$

These are continuous linear functionals. If $|\ell_N(f)| \leq C$ (depending on f), then the $\|\ell_n\| \leq C$ by the uniform boundedness principle 7.5. Here, strongly bounded means that the operators are bounded.

The claim now is that $\int_{S^1} |D_N(\theta)|\, d\theta = L_N \to \infty$. These L_N are called *Lebesgue numbers.* We sketch a proof of this. $D_N(\theta) = 0$ when $(N + \frac{1}{2})\theta = k\pi$, so when $\theta_k = \frac{k\pi}{N+\frac{1}{2}}$. Take $\alpha_k \in (\theta_k, \theta_{k+1})$ with $\alpha_k = \frac{(k+\frac{1}{2})\pi}{N+\frac{1}{2}}$, and then we approximate each bump by a triangle. This is $|\theta_{k+1} - \theta_k| \leq \frac{\pi}{N}$ and

$$\frac{\sin((N + \frac{1}{2})\alpha_k)}{\sin(\frac{1}{2}\alpha_k)} \lesssim \frac{\sin((k + \frac{1}{2})\pi)}{\frac{k\pi}{2N}} \approx \frac{N}{k}.$$

Therefore, the area of each triangle is $\approx \frac{1}{k}$, so the sum of areas of triangles is approximately $\sum_{k=1}^{N} \frac{1}{k} \approx \log N$, which is a contradiction to uniform boundedness. □

Remark. This depends on D_N not being positive. The Fejer kernel does not suffer from this issue.

Now, we discuss two more basic theorems about linear transformations.

Theorem 7.7 (Open Mapping Theorem)**.** *Suppose we have $T : X \to Y$ is bounded and surjective. Then T(open set) is an open set.*

Corollary 7.8. *If T is bijective, then it is an isomorphism, i.e. T^{-1} is bounded.*

Example 7.9. Suppose $X \supset Y_1, Y_2$. Suppose $Y_1 \cap Y_2 = \{0\}$ and every $x \in X$ can be written as $x = y_1 + y_2$, i.e. $X = Y_1 + Y_2$. Now, think of $Y_1 \oplus Y_2$ with norm $\|(y_1, y_2)\| = \|y_1\| + \|y_2\|$. Then $T : (y_1, y_2) \to y_1 + y_2$ is surjective and bounded since $\|T(y_1, y_2)\| = \|y_1 + y_2\| \leq \|y_1\| + \|y_2\|$. Also T^{-1} is bounded and $\|y_1\| + \|y_2\| \leq C\|x\|$. This is the Open Mapping theorem.

Proof of Open Mapping Theorem 7.7. We will use linearity and translation invariance. It is enough to check that $T(B_r(0)) \supset B_{r'}(0)$. In fact, it's enough to check for a single radius r (by scaling).

Now, take $Y = \overline{\bigcup_{n=1}^{\infty} T(B_n(0))}$ (by surjectivity). Then the Baire Category Theorem 7.4 says that some $\overline{T(B_n(0))}$ has nonempty interior. Assume $B_\varepsilon(0) \subset \overline{T(B_n(0))}$. Then the claim is that $\overline{T(B_1)} \subset T(B_2)$, which would be what we wanted.

Take $y \in \overline{T(B_1)}$ and show that $y = Tx$ for $x \in B_2$. Pick $x_1 \in T(B_1)$ such that $\|y - T(x_1)\| < \frac{\varepsilon}{2}$. Then $y - T(x_1) \in \overline{T(B_{1/2})}$.

Choose $x_2 \in T(B_{1/2})$ such that $\|y - T(x_1) - T(x_2)\| < \frac{\varepsilon}{4}$ and $y - T(x_1) - T(x_2) \in T(B_{1/4})$.

We continue doing this to get points $x_1, x_2, \dots$ so that $x = \sum_{j=1}^{\infty} x_j$ converges since $\|x_j\| \leq 2^{-j}$. Then $T(x) = \sum T(x_j) = y$. In addition, clearly $\|x\| < 2$, so $y \in T(B_2)$, which is what we wanted. ☐

8. 2/2

Last time, we discussed theorems relating to uniformity. We want to finish the last of these theorems. This is yet another criterion for continuity or boundedness.

Theorem 8.1 (Closed Graph Theorem)**.** *We have two Banach spaces X and Y, and $A : X \to Y$ is a continuous mapping. If* graph(A) *is a closed subspace of $X \times Y$ then A is bounded, and conversely.*

Here, graph$(A) = \{(x, Ax) : x \in X\}$ and $\|(x, y)\| = \|x\|_X + \|y\|_Y$.

Remark. In real life, many maps (e.g. in quantum mechanics) are not bounded, and are defined on a dense subspace of X. So many of our theorems aren't true in that context.

Here, graph(A) closed means that for $x_j \in X$ and $y_j = Ax_j$, suppose that $(x_j, y_j) \to (\overline{x}, \overline{y}) \in X \times Y$ then $\overline{y} = A\overline{x}$. Note that this is obvious if we know that A is continuous. The other direction contains the content of the theorem.

Remark. It is possible to define $A : X \to Y$ define on all of X such that the graph is not closed. Such an A would need to be badly discontinuous and we would need the axiom of choice.

Remark. It is possible that graph(A) is closed but A is not bounded. (A is only densely defined.)

Example 8.2. Let $X = Y = L^2(S^1)$. Then take $A = \frac{\partial}{\partial\theta}$. Strictly, $A : L^2(S^1) \to L^2(S^1)$ doesn't make dense. This means that the domain of A is $\text{dom}(A) = \{u \in L^2(S^1) : Au \in L^2(S^1)\} \subsetneq L^2$. Actually, we have $\text{dom}(A) = H^1(S^1) = \{u = \sum a_n e^{in\theta} : \sum n^2|a_n|^2 < \infty\}$. This is a dense subspace (e.g. consider sequences with only finitely many nonzero entries).

Fact: $\text{graph}(A) = \{(u, \frac{\partial u}{\partial\theta})\}$ is closed, i.e. $(u_j, f_j) \in \text{graph}(A)$ and if $u_j \to u$ and $f_j \to f$ then $Au = f$ (and $u \in H^1$). Here, we are saying that $\int |u_j - u_k|^2 \to 0$ and $\int |u_j' - u_k'|^2 \to 0$. This means that the u_j converge in a sense that is a bit better than in L^2.

Here, $A = \frac{\partial}{\partial\theta}$ is called a *closed* operator. These are the nice generalizations of bounded operators. Most natural examples of operators are closed.

Proof of closed graph theorem 8.1. We have $A : X \to Y$ and $\text{dom}(A) = X$ and graph(A) is closed. Then graph(A) is a Banach space.

Now, there are two natural projections. Let $\pi_1 : \text{graph}(A) \to X$. This is surjective, and by the open mapping theorem 7.7, π_1 is invertible. Hence $\pi_1^{-1} : X \to \text{graph}(A)$ is continuous. Now, we can factor $A : X \to Y$ through graph(A) , i.e. $A = \pi_2 \circ \pi_1^{-1}$ is the composition of bounded operators, so A is bounded too. □

We've now finished the three big theorems about uniformity and boundedness.

We want to review some basic theorems in point set topology.

Definition 8.3. We're interested in $(X, \mathcal{F})$ where X is a set and $\mathcal{F}$ is a collection of subsets of X. This is the definition of a general topological space, where $\mathcal{F}$ are the open sets.

Here, $\emptyset, X \in \mathcal{F}$. If $\{U_\alpha\}_{\alpha\in A}$ and $U_\alpha \in \mathcal{F}$ then $\bigcup_{\alpha\in A} U_\alpha \in \mathcal{F}$. Also $U_{\alpha_1} \cap \cdots \cap U_{\alpha_N} \in \mathcal{F}$.

The main examples for us are as follows:

Example 8.4. Metric spaces (X, d). Here, $U \in \mathcal{F}$ if for any $p \in U$ there is some $r > 0$ such that $B_\rho(r) \subset U$. The open sets are arbitrary unions of open balls.

As a special case, Banach spaces $(X, \|\cdot\|)$.

Example 8.5. Banach spaces X but with the weak topology. Here $\mathcal{F}$ is generated by $\ell^{-1}(V)$ where $\ell \in X^*$ and $V \in \mathbb{C}$ is open.

Suppose that $\ell : X \to \mathbb{R}$. Then $\ell^{-1}(a, b)$ is some open set contained between two parallel hyperplanes $\ell^{-1}(a)$ and $\ell^{-1}(b)$. We can take arbitrary unions and finite intersections.

In ℓ^2, we might have $\ell_j(x) = x_j$. Then we might have $U_1 \times \cdots \times U_N \times \ell_N^2$ where ℓ_N^2 has first n entries 0.

Example 8.6. Frechet spaces. This is a vector space X with topology generated by a countable number of norms (or seminorms): $(X, \{\|\cdot\|_k\}_{j=0}^\infty)$. Taking any one norm gives a Banach space, but open sets only need to be open under any one norm.

As an example, consider $C^\infty(S^1)$ with $\|u\|_k = \sup_{j\le k} |\partial_\theta^j u|$.

This is a generalization of Banach spaces, and it's clearly something that we should care about. We can define a metric

$$d(x, y) = \sum_{j=0}^\infty 2^{-j} \frac{\|x - y\|_j}{1 + \|x - y\|_j}.$$

A nasty example is $C^\infty(S^1)^* = \mathcal{D}^*(S^1)$ is the space of distributions on S^1, and it is *not* Frechet.

The topology of Frechet spaces is in some sense the intersection of countably many topologies, so this is very strong. What would the topology on the dual space be? Consider $\ell : C^\infty(S^1) \to \mathbb{C}$ is continuous. If $u_j \to u$ in C^∞ then $\ell(u_j) \to \ell(u)$. Then the topology on the dual looks like the union of all of these topologies, and it's not Frechet.

The key idea is compactness. Here's the primitive definition:

Definition 8.7. $K \subset X$ is called compact if every open cover $\{U_\alpha\}$ of K (i.e. $K \subset \bigcup_\alpha U_\alpha$) has a finite subcover $K \subset \bigcup_{j=1}^N U_j$.

An alternate formulation is this. Suppose we have a sequence of open sets $\{W_\alpha\}$. Every collection of open sets in $\mathcal{F}$ which does not finitely subcover K does not cover K.

There are some subclasses of topological spaces.

Definition 8.8. $\mathcal{T}_2$ are spaces where for all $x \neq y \in X$, there are open sets $U_x \ni x$ and $U_y \ni y$ so that $U_x \cap U_y = \emptyset$. This is called *Hausdorff*, and it can be thought of as separating points.

A more restrictive condition is $\mathcal{T}_3$: We can separate a point and a closed set (complement is open). This is also called *regular*.

Finally, we have $\mathcal{T}_4$, which can separate disjoint closed sets. This is called *normal*.

Definition 8.9. Recall that $f : X \to Y$ is called continuous if $f^{-1}(\text{open})$ is open.

Note that weakening the topology in X makes it much harder for $f : X \to Y$ to be continuous. On the other hand, weakening the topology on Y makes it easier for f to be continuous.

Proposition 8.10. *Take $f : X \to Y$ is a continuous bijection between compact Hausdorff spaces X and Y. Then f is a homeomorphism, i.e. f^{-1} is continuous.*

This is some sort of generalization of the open mapping theorem 7.7. If we used the weak topology, do we still separate points? Yes, by Hahn-Banach 6.5.

Proof. Suppose that $Z \subset X$ is closed, then Z is compact. To see this, take an open cover of Z, and add $X \setminus Z$, so there's a finite subcover, and $X \setminus Z$ can't cover Z, so we have a finite subcover of Z.

Continuity of f says that f^{-1} maps open to open, or equivalently, f^{-1} maps closed to closed. Then continuity of f^{-1} says that f takes closed sets to closed sets.

Now, $f(\text{compact})$ is compact and hence closed. This a general fact about continuous functions. □

There are three main theorems:

(1) Tychonoff
(2) Urysohn
(3) general form of Weierstrass approximation.

Theorem 8.11 (Tychonoff)**.** *Consider some compact topological spaces $\{A_\alpha\}_{\alpha \in I}$. Define $\mathcal{A} = \times_{\alpha \in I} A_\alpha$ with the weak topology: We want natural projection map $\ell_\alpha : \mathcal{A} \to A_\alpha$ to be continuous, so $\ell_\alpha(U_\alpha)$ generates the weak topology.*

If all A_α are compact, then $\mathcal{A}$ itself is compact in the weak topology.

This theorem is equivalent to the axiom of choice.

Proof. If X and Y are compact, then we want to show that $X \times Y$ is compact.

Suppose we have $W = \{W_\gamma\}$ is some collection of open sets. Suppose that this does not finitely subcover $X \times Y$.

Step 1: There exists $x_0 \in X$ such that no product $U \times Y$ with $x_0 \in U$ is finitely covered by W. If not, then every $x \in X$ has a neighborhood U_x such that $U_x \times Y$ is finitely covered by W, so choose $x_1, \ldots, x_N$ such that $X = U_{x_1} \cup \cdots \cup U_{x_N}$ by compactness of X, so therefore $X \times Y$ is finitely covered, which is a contradiction.

Step 2: There exists $y_0 \in Y$ such that no rectangle $U \times V$ for $x_0 \in U$ and $y_0 \in V$ is finitely covered. If not, then for all $y \in Y$, can find $U_y \times V_j$ with $x_0 \in U_y$ which are finitely covered. Choose $y_1, \ldots, y_N \in Y$ so that $V_{y_1} \cup \cdots \cup V_{y_N} = Y$. Then $U = U_{y_1} \cap \cdots \cap U_{y_N} \ni x_0$, then $U \times Y$ is finitely covered by $U_{y_j} \times V_{y_j}$.

We have a reformulation. Given $z_0 = (x_0, y_0)$, no open rectangle $U \times V$ containing z_0 is finitely covered.

Step 3: Now, we do the product of three spaces $X \times Y \times Z$. Suppose that there exists $x_0 \in X$ such that no $U \times Y \times Z$ is finitely covered by $\{W_\gamma\}$. If Y is compact, there exists $y_0 \in Y$ such that no $U \times V \times Z$ is finitely covered. This is the same argument as before. We've explicitly found a $z_0 \notin W_\gamma$.

Suppose we have a countable collection $X = \times_{j=1}^\infty X_j$ where each X_j is compact. Then $\{W_\gamma\}$ is some collection of open sets with no finite subcover. We find a point $x \in X$ for $x = (x_j)_{j=1}^\infty$ and $x \notin W_\gamma$ for any γ.

First, find $x_1 \in X_1$ such that no tube $U_1 \times \times_{j=2}^\infty X_j$ is finitely subcovered.

Now, find $x_2 \in X_2$ such that no $U_1 \times U_2 \times \times_{j=3}^\infty X_j$ is finitely subcovered.

Now, by induction, having chosen $x_1, \ldots, x_{n-1}$ so $(X_1 \times \cdots \times X_{n-1}) \times X_n \times \times_{j=n+1}^\infty X_j$. Define $x = (x_j)$.

Any open set around x contains $U_1 \times \cdots \times U_N \times \times_{N+1}^\infty X_j$. Hence, this cannot be finitely subcovered, so $x \in \bigcup W_\gamma$.

The last step is arbitrary covers. If we have $\times_{\alpha \in I} X_\alpha$, choose a well-ordering of the index set I. We now do exactly the same thing. If $\beta \in I$, then $\times_{\alpha < \beta} X_\alpha$ is compact, and we have $\times_{\alpha<\beta} X_\alpha \times X_\beta \times \times_{\gamma > \beta} X_\gamma$. This is transfinite induction, and we use the axiom of choice. ☐

Next, we will discuss Urysohn, which says that in a normal space, there are lots of continuous functions. This is an analog of the Hahn-Banach theorem.

9. 2/7

Suppose we have spaces $P = \times_{\alpha \in J} X_\alpha$ and each X_α is compact. Then P is compact in the weak topology. We have projection maps $\pi_\alpha : P \to X_\alpha$. The weak topology is the weakest topology so that these projections are continuous. Then for any $U_\alpha \subset X_\alpha$ open, we need $\pi_\alpha^{-1}(U_\alpha)$ are open.

In general, if X is any linear space, with a locally convex topology. Here, the locally convex topology means that we have a collection of open convex sets around 0 (by translation invariance) that generate the topology. Then we can define the weak topology on X, with base as all $\ell^{-1}(U)$ for $U \subset \mathbb{C}$ is open.

Recall that we had an isometric embedding $X \hookrightarrow X^{**}$. For each x, we have $\varphi_x \in X^{**}$ such that $\varphi_x(\ell) = \ell(x)$. On X^*, there are two weak topologies, namely, the one from X^{**} with the above weak topology, or the one from X (called the weak* topology, where we restrict to functionals coming from X itself).

It turns out that as we weaken the topology, it's more likely that something is compact. The danger is that we weaken too far and get something stupid. The weak* topology allows us to prove theorems, but it is still interesting.

Note that when X is reflexive ($X^{**} = X$), then the weak and weak-* topologies are the same.

We now prove the first application of the Tychonoff theorem.

Theorem 9.1 (Banach-Alaoglu)**.** *If X is Banach, then the unit ball $B \subset X^*$ is weak* compact.*

As an application, let X be a Hilbert space. Then the unit ball is weakly compact. For example, take Ω as a domain in $\mathbb{R}^n$ (or a compact manifold). The space we are interested in is $X = H_0^1(\Omega)$ is the closure of $C_0^\infty(\Omega)$ with respect to $\|u\|_{H^1}^2 = \int |u|^2 + |\nabla u|^2$. Then $u_j \in H_0^1$, so $\|u\|_j \leq C$, and then there exists u_{j_k} such that $u_{j_k} \to u$.

Here, weak convergence means that there exists $v \in H_0^1$ such that $\int u_j v + \nabla u_j \cdot \nabla v \to \int uv + \nabla u \cdot \nabla v$.

If we have a $u \in H_0^1$ satisfying $\int uv + \nabla u \cdot \nabla v = 0$ for all v, then $\int(-\Delta u + u)v = 0$ for all v, so then we should have $-\Delta u + u = 0$, so this gives us a way to find weak solutions of such equations. This is an application of our generality.

Proof of Theorem 9.1. Define a map $X^* \to \mathbb{R}^X = \times_{x \in X} \mathbb{R}$ via $\ell \mapsto (\ell(x))_{x \in X}$. If $\ell \in B$, then we have $|\ell(x)| \leq \|x\|$, so in fact $\Phi : B \to P = \times_{x \in X}[-\|x\|, +\|x\|]$, we have $\ell \mapsto (\ell(x))_{x \in X} \in P$. Now, P is a product of compact spaces, so it is compact by the Tychonoff theorem 8.11. Then compactness of P implies that B is compact.

We need to check that Φ is injective. Here, $\Phi(\ell) = 0$ means that $\ell(x) = 0$ for all $x \in X$, so $\ell = 0$. Linearity is obvious.

We want to say that Φ is continuous. The open sets in P are generated by projections $\pi_x : P \to [-\|x\|, \|x\|]$, i.e. $U \subset [-\|x\|, \|x\|]$ with $\Phi^{-1}(U) = \{\ell : \ell(x) \in U\} = \pi_x^{-1}(U)$. There are exactly the weak* open sets.

We need to check that $\Phi(B)$ is a closed set in P. If $p \in P$, $p \in \overline{\Phi(B)}$, we want to say that $p = \Phi(\ell)$ for some $\ell \in B$.

If $p = \Phi(\ell)$ then $p_{x+y} = p_x + p_y$. Similarly, $p_{ax} = ap_x$. But these are relationships between a finite number of components of P, so they are preserved under the (weak) topology on P. These conditions mean that the image $\Phi(B)$ is therefore closed. Now we're done by Tychonoff's theorem and compactness. This is using the full power of the transfinite induction and logic that we've done. $\square$

Now, we want to return to general topology, and come back to weak and weak* topology.

Lemma 9.2 (Urysohn's Lemma)**.** *If $(X, \mathcal{T})$ is normal (T_4), then for all $A, B \subset X$ closed and disjoint, there exists a continuous function $f : X \to \mathbb{R}$ such that $f|_A = 0$ and $f|_B = 1$.*

Proof. Enumerate $\mathbb{Q} \cap [0, 1] = \{p_i\}$. For all n, define U_n and V_n closed and disjoint such that $A \subset U_n$ and $B \subset V_n$. We require that if $m < n$ and $p_m < p_n$ then $\overline{U_m} \subset U_m$ and $\overline{V_m} \subset V_m$, while if $p_m > p_n$ then $\overline{U_n} \subset U_m$ and $\overline{V_n} \subset V_m$.

Define f on X by $f(x) < p_n$ if $x \in U_n$. This is continuous, and the details are left as an exercise. $\square$

Theorem 9.3 (Stone-Weierstrass)**.** *Let X be a compact Hausdorff space. Then the real valued continuous functions $C(X)$ is a Banach space. Suppose $B \subset C(X)$ is a subspace with the following properties:*

(1) subalgebra
(2) closed
(3) separates points
(4) $1 \in B$.

Then $B = C(X)$.

Remark. This is a density theorem, because if we have subspace satisfying these but isn't closed, then we can take its closure and say that the closure is all of $C(X)$.

We also have a complex version, with B as before, but with another property: the space is closed under complex conjugation $\overline{B} = B$.

Definition 9.4. $S \subset C_{\mathbb{R}}(X)$ is called a lattice if for all $f, g \in S$, we can define $f \wedge g = \min\{f, g\}$ and $f \vee g = \max\{f, g\}$. (This is why we want real-valued.)

Lemma 9.5. *If B is as in theorem 9.3, then it is a lattice.*

Proof. Note that $f \vee g = \frac{1}{2}|f - g| + \frac{1}{2}(f + g)$ and $f \wedge g = -((-f) \vee (-g))$. It's enough to show that $f \in B$ implies that $|f| \in B$. This isn't free from algebra-ness because it isn't a linear statement.

We can assume that $\|f\|_\infty \leq 1$. Choose polynomials $P_n(x) \in C([-1, 1])$ such that $P_n(x) \to |x|$ on $[-1, 1]$. This is the version of Weierstrass approximation that we've proved before. Since B is an algebra, then $P_n(f) \in B$, and $P_n(f) \to |f| \in B$ also. ☐

Theorem 9.6 (Kakutani-Krein)**.** *Suppose X is compact and Hausdorff, and suppose $\mathcal{L} \subset C_{\mathbb{R}}(X)$ is a lattice which is closed, separates points, with $1 \in \mathcal{L}$. Then $\mathcal{L} = C_{\mathbb{R}}(X)$.*

Proof. Fix any $h \in C_{\mathbb{R}}(X)$, and find $f \in \mathcal{L}$ such that $\|f - h\| < \varepsilon$.

For each $x \in X$, find $f_x \in \mathcal{L}$ such that $f_x(x) = h(x)$ and $h < f_x + \varepsilon$ on all of X.

Choose a neighborhood U_x so that $f_x - \varepsilon < h$ on U_x. Define $X = \bigcup U_x = U_{x_1} \cup \cdots \cup U_{x_N}$ by compactness, and choose $f = f_{x_1} \wedge \cdots \wedge f_{x_N} \in (h - \varepsilon, h + \varepsilon)$.

Then $f(y) + \varepsilon = \min_i\{f_{x_i}(y) + \varepsilon\} > h(y)$. If $y \in U_{x_i}$, then $h(y) > f_{x_i}(y) + \varepsilon > f(y) + \varepsilon$, which means that $f \in (h - \varepsilon, h + \varepsilon)$, so $\|f - h\| < \varepsilon$.

How do we find these various f_x's? Now, for $x, y \in X$, choose $f_{x,y} \in \mathcal{L}$ with $f_{x,y}(x) = h(x)$ and $f_{x,y}(y) = h(y)$. Now, for any $\varepsilon > 0$, let $V_y = \{z : f_{x,y}(z) + \varepsilon \geq h(z)\}$. These V_y's cover, so $V_{y_1} \cup \cdots V_{y_k} = X$. Then $f_x = f_{x,y_1} \vee \cdots \vee f_{x,y_k}$. Why should this work? Firstly, $f_x(x) = h(x)$. For any z, we have $f_x(z) + \varepsilon = \max\{f_{x,y_i}(z) + \varepsilon\} \geq h(z)$.

The idea is that first we go above h and then below it by taking maximums and minimums, using the lattice property, to squeeze within an ε-neighborhood of h. ☐

Now, we will return to weak convergence.

Let X be a Banach space. What do weakly convergent sequences look like? We do this by giving a lot of examples. The answer depends on the Banach space – in some cases, this is restrictive, not so much on others.

Example 9.7. In ℓ^2, let $x \xrightarrow{\pi_n} x_n$.

Then $x^{(k)} \to 0$ means that $x_n^{(k)} \to 0$ for each n. So we are only controlling one component at a time.

Now, take $x^{(k)} = (0, 0, \ldots, 0, 1, 0, 0, \cdots, 0)$ with a 1 in the kth position and 0s elsewhere. This converges to zero. Geometrically, take the unit ball, and take the intersections of coordinate axes with the unit ball. These converge to zero.

Example 9.8. In $L^2(S^1)$, take $e^{in\theta}$. We claim that these converge to 0 weakly. Then for any $f \in L^2(S^1)$, we have $\int e^{in\theta} f(\theta)\, d\theta \to 0$ as $n \to \infty$. The point is that $e^{in\theta}$ oscillates so much that it sees f as essentially constant, and we get lots of cancellation.

Example 9.9. In ℓ^1, weak convergence is no weaker than strong convergence!

10. 2/9

We want to give some intuition about the weak topology. We discussed two examples: ℓ^2 and $L^2(S^1)$. In the second case, the point was that high oscillations imply that the mean converges to 0.

Example 10.1. Consider $C([0,1])$ and consider functions $u_n(t)$ that are triangles of height 1 and width $\frac{2}{n}$, and 0 elsewhere. This converges weakly to 0, but it definitely does not converge to 0 in the sup norm.

To prove this, choose any $\ell \in C([0,1])^*$. Suppose that $\ell(u_n) \to 0$. Assume that $\ell(u_n) \geq \delta > 0$. Pass to a subsequence n_k such that $n_{k+1} \geq 2n_k$. Define $v_N(t) = \sum_{j=1}^N u_{n_j}$. We can check that $v_N(t) \leq 4$ (Exercise). Then $N\delta \leq \ell(v_N) \leq 4\,\|\ell\|$, which is a contradiction.

Theorem 10.2. *Suppose $x_n \in X$ is a sequence in a Banach space satisfying $\|x_n\| \leq C$ and there exists a dense set $D \subset X$ such that $\ell(x_n) \to \ell(x)$ for all $\ell \in D$, then $x_n \to x$.*

Proof. Take any $\hat{\ell} \in X^*$, and chose $\ell \in D$ so $\left\|\ell - \hat{\ell}\right\| < \varepsilon$. Then we have $\hat{\ell}(x_n) = \ell(x_n) + (\hat{\ell} - \ell)(x_n)$, so therefore $|\hat{\ell}(x_n) - \hat{\ell}(x)| < 2\varepsilon$ for n large. □

We want to head toward the converse. Consider a collection $\{f_\alpha\}_{\alpha \in A}$ of subadditive, real-valued, positively homogeneous functions on X, i.e. $f_\alpha(x+y) \leq f_\alpha(x) + f_\alpha(y)$ and $f_\alpha(\lambda x) = |\lambda| f_\alpha(x)$. For example, $|\ell|$ satisfies this.

Proposition 10.3. *Assume that $\sup_{\alpha \in A} f_\alpha(x) \leq C(x)$ for all $x \in X$. Then $|f_\alpha(x)| \leq C\,\|x\|$ for all $x \in X$.*

Proof. There exists a ball $B_z(r) \subset X$ so that if $y \in B_z(r)$ then $|f_\alpha(y)| \leq C$.

If $y = z + x$ for $|x| < r$, then $|f_\alpha(x)| = |f_\alpha(y)| + |f_\alpha(z)| \leq C'$. Then for any x, $x = \frac{2\|x\|}{r}(\frac{x}{\|x\|}\frac{r}{2})$, so therefore $|f_\alpha(x)| \leq C''\,\|x\|$. □

Theorem 10.4. *If $\{\ell_\alpha\} \subset X^*$ such that $|\ell_\alpha(x)| \leq C(x)$ for each $x \in X$, then $\|\ell_\alpha\| \leq C$.*

Here is a weak* version of the same thing:

Theorem 10.5. *If $\{x_\alpha\}_{\alpha \in A} \subset X \subset X^{**}$ such that for all $\ell \in X^*$. Then $|\ell(x_n)| \leq C(\ell)$ implies that $\|x_\alpha\| \leq C$.*

Corollary 10.6. *If $x_n \to x$ weakly then $\ell(x_n) \to \ell(x)$ for all $\ell \in X^*$, so therefore $|\ell(x_n)| \leq C(\ell)$. This tells us that $\|x_n\| \leq C$.*

This should look like Fatou:

Theorem 10.7. *If $x_n \in X$ and $x_n \to x$ weakly, then $\|x\| \leq \liminf_{n\to\infty} \|x_n\|$.*

(For example, points on the sphere can converge weakly to 0.)

Proof. Choose $\ell \in X^*$ such that $\|\ell\| = 1$ and $\ell(x) = \|x\|$ by the Hahn-Banach theorem. Then $\|x\| = \ell(x) = \lim \ell(x_n)$. Since $|\ell(x_n)| \le \|\ell\| \|x_n\| = \|x_n\|$, we are done after taking $\liminf$. □

Theorem 10.8. *Suppose that X is reflexive. Then take the ball $B = \{x : \|x\| \le 1\}$. Then B is weakly sequentially compact, i.e. for any $x_n \in B$ there exists $x_{n_j} \to \overline{x}$ weakly.*

Proof. Take a sequence $\{y_n\} \subset B$. Take $Y = \overline{\text{span}\{y_n\}}$. By construction, this is a separable closed subspace. Then X is reflexive implies that Y is reflexive. Then Y^{**} is separable, and hence Y^* is separable.

Choose a sequence $\{\ell_k\} \in Y^*$ that is dense. Choose a subsequence of y_n, so y_{n_j} such that $\ell_k(y_{n_j})$ converges for each k.

All $\|y_{n_j}\| \le 1$. Hence for all $\ell \in Y^*$, $\ell(y_{n_j})$ converges to a limit.

We've now passed to a subsequence such that $\ell(y_{n_j})$ converges for all $\ell \in Y^*$. Define $\varphi \in Y^{**}$ so that $\varphi(\ell) = \lim_{j\to\infty} \ell(y_{n_j})$. Then $\varphi(\ell) = \ell(y)$ for some $y \in Y$.

We've shown that $\ell(y_{n_j}) \to \ell(y)$ for all $\ell \in Y^*$. But we want to do this for X. If we have $\hat{\ell} \in X^*$, define $\ell \in Y^*$ by $\ell = \hat{\ell}|_Y$. □

Theorem 10.9. *If ℓ_n is weak* convergent, then $\|\ell_n\| \le C$.*

Theorem 10.10. *If X is separable, then the unit ball $B \subset X^*$ is weak* sequentially compact.*

Proof. Suppose that $\ell_n \le X^*$ and $\|\ell_n\| \le 1$. Choose a dense sequence $x_n \in X$ and a subsequence ℓ_{n_j} such that $\ell_{n_j}(x_k)$ has a limit as $j \to \infty$. This defines a number $\lim \ell_{n_j}(x)$ for each $x \in X$, so that $\ell \in X^*$ and $\ell(x) = \lim \ell_{n_j}(x)$. □

Now we move on to discussing a broader class of spaces. The reason is that there are interesting and useful applications.

Definition 10.11. A *locally convex topological vector space* means a vector space X with a topology $\mathcal{T}$. If $U \in \mathcal{T}$ then we demand that $U + x \in \mathcal{T}$ and $xU \in \mathcal{T}$, i.e. the topology should be translation and dilation invariant.

We require that there exists $\{U_\alpha\}_{\alpha\in A}$ which generates $\mathcal{T}$, satisfying $0 \in U_\alpha$ and U_α convex. We need these to be absorbing: the union of all dilates covers the entire space.

Example 10.12. For example, $L^p(I)$ for $0 < p < 1$ is *not* a locally convex topological vector space. The open neighborhoods of 0 are not convex!

For each U_α, associate a function ρ_α (seminorms), so that $\rho_\alpha(x) = \inf\{\lambda : x \in \lambda U_\alpha\}$. This makes since because U_α is absorbing.

The open sets U_α satisfy three properties:

- convexity
- balancing: For a point $x \in U$, we have $\lambda x \in U$ for all $|\lambda| = 1$.
- absorbing: $\bigcup_{\lambda \ge 0} \lambda U = X$.

Remark. The Hahn-Banach theorem extends to this situation. To make sense of this, we need to define X^* to be the continuous linear functionals. Go back and look at this: the proof didn't use completeness or norms!

Now, we specialize:

Definition 10.13. X is *Frechet* if the topology $\mathcal{T}$ is generated by a countable family of U_j, or seminorms ρ_j.

Suppose we have $T : X \to Y$ where X and Y are both Frechet. Suppose X and Y have seminorms ρ_j and d_i. Then for each i, there exist $j_1, \dots, j_N$ such that $d_i(T(x)) = C(\rho_{j_1}(x) + \cdots + \rho_{j_N}(x))$. Then any $\ell : X \to \mathbb{C}$ satisfies $|\ell(x)| \leq C \sum_{j=1}^{N} \rho_{\alpha_j}(x)$.

Example 10.14.
(1) $C^\infty(S^1)$
(2) Schwartz space: $\mathcal{S}$.

Example 10.15. This is not Frechet but is a locally convex topological vector space: $C_0^\infty(\mathbb{R}^n)$.

Example 10.16. In the case $C^\infty(S^1)$, we use norms $\|u\|_k = \sup_{j\leq k} |\partial_\theta^j u|$.

Exercise 10.17. Take any sequence $a_n \in \mathbb{R}$. Then there exists $u \in C_0^\infty(\mathbb{R})$ such that $u^{(n)}(0) = a_n$. The derivatives can be as bad as we like. We can control some finite number of derivatives, but we cannot control all derivatives. This is called *Borel's lemma*. The idea of the proof is to take the Taylor series and try to make it converge: $\sum \frac{a_n}{n!} x^n \chi(x/\varepsilon_n)$ for some cutoff function χ. We need this sum to converge, along with all of its derivatives.

This has a dual space $(C^\infty(S^1))^* = \mathcal{D}'(S^1)$, which is the space of generalized functions or distributions.

Suppose we have some $\ell(u) \in \mathbb{C}$ for all $u \in C^\infty(S^1)$. Then continuity means that $|\ell(u)| \leq C \|u\|_k$, so it only sees some finite number of derivatives.

Example 10.18. Suppose $f \in L^1(S^1)$. Then $\langle f, u \rangle = \int fu$ for all $u \in C^\infty(S^1)$. Then $|\langle f, u \rangle| \leq \|f\|_{L^1} \|u\|$, so any function is a distribution.

There is also $\delta \in \mathcal{D}'$, with $\delta(u) = u(0)$ and $|\langle \delta, u \rangle| \leq \|u\|$. Here, $\delta^{(k)}(u) = (-1)^k u^{(k)}(0)$.

Another important example is the principal value function: $p.v. \frac{1}{x}$:

$$\lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} \frac{u(x)}{x}\, dx.$$

This is a distribution:

$$\int_{|x|\geq\varepsilon} \frac{u(x)}{x}\, dx = \int_{x=\varepsilon}^{\pi} \frac{u(x) - u(-x)}{x}\, dx.$$

Note that $|u(x) - u(-x)| \leq \|u'\|_0 2 \|x\|$, so therefore $|\langle p.v. \frac{1}{x}, u \rangle| \leq C \|u\|_1$.

There are all sorts of exotic distributions, and we will talk about this later.

Example 10.19. $\mathcal{S}(\mathbb{R}^n) = \{|x^\alpha \partial_x^\beta u| \leq C_{\alpha\beta}\}$, i.e.

$$|u| \leq \frac{C_N}{1 + |x|^N}$$

for any N, and

$$|\partial_x^\beta u| \leq \frac{C_{N,\beta}}{(1 + |x|)^N}$$

for any N and β. This is the space of rapidly decreasing functions, e.g. Gaussians $e^{-|x|^2}$ or $e^{-(1+|x|^2)^{\alpha/2}}$.

We have a dual space $\mathcal{S}'$ of *tempered distributions*, which means that they are slowly growing. If $f \in \mathcal{S}'$ then

$$|\langle f, u\rangle| \leq C \sum_{|\alpha|\leq k, |\beta|\leq l} \sup |x^\alpha \partial_x^\beta u|.$$

Now, consider $C_0^\infty(\mathbb{R}^n)$.

Definition 10.20. $u_j \in C_0^\infty(\mathbb{R}^n)$ is convergent to u if $u_j \to u$ in any C^k and $\operatorname{supp} u_j \subseteq K$.

The condition on the support is to avoid a bump that runs away to infinity. Its dual space is the space of all distributions.

11. 2/14

We've discussed some main structural theorems and some applications. We'll say a bit more about distributions, and then we'll jump into the study of linear operators, which will occupy us for the remainder of the course.

We had two main examples of Frechet spaces:

- $C^\infty(\Omega)$ where Ω is any compact manifold or bounded domain in $\mathbb{R}^n$ with smooth boundary.
- Schwartz space $\mathcal{S}$ of rapidly decreasing functions on $\mathbb{R}^n$.

  Recall: for nonnegative integers k and l,

$$\|u\|_{k,l} = \sup_{x\in\mathbb{R}^n} \sup_{\substack{|\alpha|\leq k \\ |\beta|\leq l}} |x^k \partial_x^\beta u|$$

The Schwartz space is nice because we can do anything we like to them without things going wrong.

To contrast, consider $C_0^\infty(\mathbb{R}^n)$, which is not a Frechet space. This is called the space of *test functions*.

Definition 11.1.

$$C_0^\infty(\mathbb{R}^n) = \bigcup_K C_0^\infty(K).$$

If $\varphi_j \to \tilde{\varphi}$ in $C_0^\infty(\mathbb{R}^n)$ then

(1) $\varphi_j \to \tilde{\varphi}$ in every C^k
(2) $\operatorname{supp} \varphi_j$ all lie in a *fixed* compact K.

This prevents $\varphi_j(x) = \varphi(x-j)$, which is a bump that runs away.

Definition 11.2. The dual space $C_0^\infty(\mathbb{R}^n)'$ is the space of all *distributions* on $\mathbb{R}^n$, and $\mathcal{S}'$ is the space of all *tempered distributions*.

What does it mean to be an element of $\mathcal{S}'$?

Definition 11.3. $u \in \mathcal{S}'$ means that there exists k and l such that $|\langle u, \varphi\rangle| \leq C \|\varphi\|_{k,l}$ for all $\varphi \in \mathcal{S}$.

Definition 11.4. $u \in \mathcal{D}'$ means that for each K, there exists k such that $|\langle u, \varphi\rangle| \leq C_K \|\varphi\|_k$ for all $\varphi \in C_0^\infty(K)$.

Think of a distribution as a black box that spits out a number given certain rules.

Here are the main examples:

Example 11.5. If $u \in L^1_{loc}$, i.e. $\int_K |u| < \infty$ for any compact K, then we can define $\langle u, \varphi \rangle = \int_{\mathbb{R}^n} u\varphi$.

Then if $\operatorname{supp} \varphi \subset K$, we have $|\langle u, \varphi \rangle| \leq (\int_K |u|) \|\varphi\|_0$, so $u \in \mathcal{D}'$. Is such a u in $\mathcal{S}'$? No. Take $u = e^{|x|}$. Then for $\varphi \in \mathcal{S}$, $\int u\varphi$ is not defined in general.

Suppose $u \in L^1_{loc}$ and $|u| \leq C(1+|x|)^N$ for some fixed N. Then

$$\left| \int u\varphi \right| \leq C \int (1+|x|)^{Nq} \leq C' \|\varphi\|_{q,0}$$

and $|\varphi| \leq C_q(1+|x|)^{-q}$ for any q. Here $C_q = \|\varphi\|_{q,0}$.

Slightly more generally, define

$$\langle u, \varphi \rangle = \sum_{|a| \leq l} \int a_\alpha(x) \partial_x^\alpha \varphi$$

(1) $a_\alpha \in L^1_{loc}$ implies $u \in \mathcal{D}'$ and $|\langle u, \varphi \rangle| \leq C \|\varphi\|_l$.
(2) $|a_\alpha| \leq C(1+|x|)^N$ for fixed N implies that $u \in \mathcal{S}'$.

Example 11.6. $\delta \in \mathcal{D}'$ and $\mathcal{S}'$. We take $\langle \delta, \varphi \rangle = \delta(0)$ and $\left\langle \delta^{(\alpha)}, \varphi \right\rangle = (-1)^{|\alpha|}\varphi^{(\alpha)}(0)$.

We can also take $u = \sum c_j \delta_{p_j}$.

Case 1: p_j is discrete in $\mathbb{R}^n$. In this case, $\langle u, \varphi \rangle = \sum c_j \varphi(p_j)$. Then $u \in \mathcal{D}'$ because the φ are test functions, so compactly supported, so we only consider finite sums. We don't have $u \in \mathcal{S}'$ because the c_j might grow exponentially. The same is true if we instead consider $u = \sum c_j \delta_{p_j}^{\alpha(j)}$.

Case 2: $u = \sum c_j \delta_{p_j}$ and $p_j \in B$ (not discrete), and $\sum |c_j| < \infty$. Can we make sense of $\langle u, \varphi \rangle = \sum c_j \varphi(p_j)$? Yes: $|\langle u, \varphi \rangle| \leq \sup |\varphi| \sum |c_j|$. Then $u \in \mathcal{D}'$ and $u \in \mathcal{S}'$.

Example 11.7. Suppose we have a sequence of elements $u_N \in C^0(\overline{B})^*$, e.g. $u_N = \sum_{j=1}^N c_j \delta_{p_j}$. When does u_N lie in the unit ball in this dual space $C^0(\overline{B})^*$?

$$\sup_{\varphi \in C^0, \|\varphi\| \leq 1} \left| \sum_{j=1}^N c_j \varphi(p_j) \right| \leq 1$$

for each N. This is true if $\sum_{j=1}^\infty |c_j| \leq 1$. Is this necessary? Weak* compactness from the Banach-Alaoglu theorem 9.1 implies that $u_{N_l} \to u$ in the weak* sense, i.e. $\langle u_{N_l}, \varphi \rangle \to \langle u, \varphi \rangle$ and $u = \sum c_j \delta_{p_j} \in C^0(\overline{B})^*$.

Theorem 11.8. *Let μ_N be a sequence of probability measures, and fix some measure space $(X, d\nu)$, so that $\int_X d\mu_N = 1$ or $\sup_{\|\varphi\| \leq 1} \int \varphi \, d\mu_N = 1$. This implies that $\mu_N \xrightarrow{w*} \mu$ is a probability measure.*

There are various structure theorems. We want to solve some equation, and we look for a solution in the most general sense, like a distribution. But distributions are like black boxes. We want to say that it is nice.

Suppose that $u \in \mathcal{D}'$ and $u|_U \in C^\infty(U)$. This means that for all $\varphi \in C_0^\infty(U)$, $\langle u, \varphi \rangle = \int f\varphi$ for $f \in C^\infty(U)$.

Example 11.9. $\operatorname{supp} u = \{0\}$ means that $\langle u, \varphi \rangle = 0$ for all φ with $0 \notin \operatorname{supp} \varphi$. Then by the problem set, $u = \sum_{|a| \le N} c_\alpha \delta^{(\alpha)}$.

Here's a more substantial structure theorem.

Theorem 11.10. *Suppose that u is a distribution $u \in \mathcal{D}'(\mathbb{R}^n)$. For each $R > 0$,*

$$u|_{\overline{B_R(0)}} = \sum_{|\alpha| \le N} c_\alpha \partial_x^\alpha u_\alpha$$

where $u_\alpha \in C^0(\overline{B_R})$.

Definition 11.11. If $u \in \mathcal{D}'$, $\partial_{x_j} u \in \mathcal{D}'$ defined by $\left\langle \partial_{x_j} u, \varphi \right\rangle = -\left\langle u, \partial_{x_j} \varphi \right\rangle$.

So this theorem tells us that if $\operatorname{supp} \varphi \in B_R$ then

$$\langle u, \varphi \rangle = \int_{B_R} \sum_{|\alpha| \le N} (-1)^{|\alpha|} c_\alpha(x) \partial_x^\alpha \varphi \, dx.$$

Example 11.12. How does the δ function arise this way? Define the *Heaviside function*

$$H(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0. \end{cases}$$

Then

$$\langle H, \varphi \rangle = \int_{-\infty}^{\infty} H \cdot \varphi = \int_0^\infty \varphi,$$

so

$$\langle H', \varphi \rangle = -\langle H, \varphi' \rangle = -\int_0^\infty \varphi'(x) \, dx = \varphi(0),$$

i.e. $H' = \delta$. Now, taking

$$x_+ = \int_0^x H(t) \, dt = \begin{cases} x & x \ge 0 \\ 0 & x < 0. \end{cases}$$

We have $x_+' = H$ because

$$\left\langle x_+', \varphi \right\rangle = -\langle x_+, \varphi' \rangle = -\int_0^\infty x\varphi'(x) \, dx = \int_0^\infty \varphi(x) \, dx = \langle H, \varphi \rangle.$$

Here we just followed the definitions.

What does $u_j \to u$ in $\mathcal{D}'$ mean? This means that for all $\varphi \in C_0^\infty(\mathbb{R}^n)$, we have $\langle u_j, \varphi \rangle \to \langle u, \varphi \rangle$.

Example 11.13. Suppose that $u_j(x) = \chi(jx) j^N$ where χ is a standard bump function, i.e. $\chi \in C_0^\infty(\mathbb{R}^n)$, $\chi \ge 0$, $\operatorname{supp} \chi \subset B$, $\int \chi = 1$. Note that $u_j \in B_{1/j}$ and $\int u_j(x) \, dx = \int \chi(jx) j^N \, dx = 1$.

Proposition 11.14. $u_j \to \delta_0$, *i.e.* $\langle u_j, \varphi \rangle \to \varphi(0)$.

Proof.

$$\int j^N \chi(jx) \varphi(x) \, dx - \varphi(0) = \int_{|x| \le 1/j} j^N \chi(jx)(\varphi(x) - \varphi(0)) \, dx.$$

Given any $\varepsilon > 0$, choose j such that $|\varphi(x) - \varphi(0)| < \varepsilon$ for $|x| < 1/j$. Then the above quantity has size $\le \varepsilon \int \chi_j(x) \, dx = \varepsilon$. □

This means that the Fejer kernel $K_N \to \delta$. This is not true for the Dirichlet kernel $D_n \not\to \delta$ because it is not positive.

Theorem 11.15. *$C^\infty(\mathbb{R}^n) \subset \mathcal{D}'(\mathbb{R}^n)$ is a dense subspace.*

Proof. We use convolution. Choose χ and χ_j exactly as before. For $u \in \mathcal{D}'$, we claim that $\chi_j \star u \in \mathcal{D}'$, $\langle \chi_j \star u, \varphi\rangle \to \langle u, \varphi\rangle$, and $\chi_j \star u \in C^\infty$.

Define $\chi_j \star u$ by duality.

$$\langle \chi_j \star u, \varphi\rangle = \int \chi_j(y) u(x-y)\varphi(x)\, dy\, dx = \int u(z)\chi_j(y)\varphi(z+y)\, dz\, dy.$$

Then φ becomes (where $\tilde{\chi}_j(y) = \chi_j(-y)$)

$$\tilde{\chi}_j \star \varphi = \int \tilde{\chi}_j(y)\varphi(x-y)\, dy = \int \chi_j(y)\varphi(x+y)\, dy.$$

So $\langle \chi_j \star u, \varphi\rangle = \langle u, \tilde{\chi}_j \star \varphi\rangle$.

We still need $\tilde{\chi}_j \star \varphi \in C_0^\infty(\mathbb{R}^n)$. We know that $\tilde{\chi}_j \star \varphi \to \varphi$ in $C_0^\infty(\mathbb{R}^n)$.

Note that $\operatorname{supp} \tilde{\chi}_j \star \varphi \subset \operatorname{supp} \varphi + B_{1/j}(0)$. Also, for all α, $\partial_x^\alpha \tilde{\chi}_j \star \varphi \to \partial_x^\alpha \varphi$, which means that $\tilde{\chi}_j \star \partial_x^\alpha \varphi \to \partial_x^\alpha \varphi$. This shows that $\chi_j \star u \in \mathcal{D}'$.

Next, we show $\chi_j \star u \in C^\infty$.

$$\langle \chi_j \star u, \varphi\rangle = \int u(y)\chi_j(x-y)\varphi(x)\, dx\, dy.$$

We have a family of test functions $\chi_{j,x} = \chi_j(x-y)$. Then $\langle u, \chi_{j,x}\rangle$ depends continuously on $\chi_{j,x}$. So now we claim that $\langle u(\cdot), \chi_j(x-\cdot)\rangle$ is smooth in x. Suppose we take

$$\frac{\langle u, \chi_j(x+h-\cdot)\rangle - \langle u, \chi_j(x-\cdot)\rangle}{h} = \left\langle u, \frac{\chi_j(x+h-\cdot) - \chi_j(x-\cdot)}{h}\right\rangle.$$

Taking limits, and noting that continuity of u implies that we can take the limit inside, implies that this goes to $\langle u, \partial_x \chi_j\rangle$. Hence $x \to \langle u, \chi_j(x-\cdot)\rangle$ is C^∞.

Finally,

$$\langle \chi_j \star u, \varphi\rangle = \langle u, \tilde{\chi}_j \star \varphi\rangle \to \langle u, \varphi\rangle. \qquad \square$$

12. 2/16

Today's topic is the Fourier transform.

(1) Definitions and properties
(2) Two consequences
(3) Fourier inversion
(4) Plancherel's theorem; extension to L^2
(5) Tempered distributions; examples

The Fourier transform is an essential tool in PDEs. We need it to define Sobolev spaces and analyze linear PDEs.

Definition 12.1. If $f(x) \in C_0^\infty(\mathbb{R}^n)$, then let

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\, dx.$$

What kind of function can we plug into it? The domain extends beyond C_0^∞ to $\mathcal{S}$ and in fact to all of L^1. Note that

$$\left|\hat{f}(\xi)\right| \leq \int_{\mathbb{R}^n} |f(x)|\, dx = \|f\|_1 .$$

So the $\left\|\hat{f}\right\|_\infty \leq \|f\|_1$. For now, let the domain be L^1. It will be really nice to try to extend this to L^2. It can't be L^2 because if $n = 1$, the function $f(x) = x^{-2/3}$ on $[-1, 1]$ will be bad; f is integrable but f^2 is not integrable.

Proposition 12.2. *If $f \in L^1$ then $\hat{f} \in C^0$ is continuous.*

Proof. Pick any sequence $\xi_i \to \xi_0$. We will show that $\hat{f}(\xi_i) \to \hat{f}(\xi_0)$. Then the question is:

$$\int e^{ix\cdot\xi_i} f(x)\, dx \overset{?}{\to} \int e^{in\cdot\xi_0} f(x)\, dx.$$

The integrands converge. The absolute values are bounded uniformly by $\int |e^{-ix\cdot\xi_i} f(x)|\, dx = \|f\|_1$, so we are done by dominated convergence. $\square$

We now state some properties of the Fourier transfoirm.

Properties 12.3.

(1) Fix $y \in \mathbb{R}^n$. Let $f_y(x) = f(x - y)$. Then

$$\widehat{f_y}(\xi) = \int e^{iy\cdot\xi} e^{-ix\cdot\xi} f(x)\, dx$$

after changing variables, and hence $\widehat{f_y}(\xi) = e^{iy\cdot\xi}\widehat{f}(\xi)$.

(2) Suppose $f \in L^1$; then $xf \in L^1$. What is $\widehat{xf}$? We have

$$\widehat{xf}(\xi) = \int x e^{-ix\cdot\xi} f(x)\, dx = \int i\frac{\partial}{\partial\xi} e^{-ix\cdot\xi} f(x)\, dx = i\frac{\partial}{\partial\xi}\widehat{f}(\xi).$$

(3) Suppose that $f \to 0$ at ∞, $f \in L^1$, and $\frac{\partial f}{\partial x} \in L^1$, then

$$\widehat{\frac{\partial f}{\partial x}}(\xi) = \int e^{-ix\cdot\xi}\frac{df}{dx}\, dx = i\xi\widehat{f}(\xi)$$

by integration by parts.

(4) Compute

$$\begin{aligned}(\widehat{f \star g})(\xi) &= \int e^{-ix\cdot\xi}\left(\int f(y) g(x-y)\, dy\right) dx \\ &= \iint g(x-y) e^{-i(x-y)\cdot\xi} f(y) e^{-iy\cdot\xi}\, dx\, dy = \widehat{f}(\xi)\widehat{g}(\xi).\end{aligned}$$

(5) For $f(x)$ and $g(x)$, we have

$$\int f(\xi)\widehat{g}(\xi)\, d\xi = \int f(\xi)\int g(x) e^{-ix\cdot\xi}\, dx\, d\xi = \iint f(\xi) g(x) e^{-ix\cdot\xi}\, dx\, d\xi.$$

This is symmetric in x and ξ, so therefore $\int f\widehat{g} = \int \widehat{f} g$.

Here are some consequences of these basic properties:

Proposition 12.4. *$\mathcal{F}(\mathcal{S}) \subset \mathcal{S}$ and $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is continuous.*

Proof. Let $f \in \mathcal{S}$. For any α and β, $x^\alpha \partial^\beta f \in \mathcal{S} \subset L^1$, so $\left\| \widehat{x^\alpha \partial^b f} \right\|_\infty$ is bounded. So therefore $\widehat{x^\alpha \partial^\beta f} = (i \frac{\partial}{\partial \xi})^\alpha (i\xi)^b \widehat{f}$ is bounded. This precisely means that $\widehat{f} \in \mathcal{S}$.

Remark. This is why Schwartz space is space. The Fourier transform exchanges smoothness and decay at infinity. Schwartz space has both, so it behaves well under the Fourier transform.

As for continuity, take

$$\begin{aligned} |\mathcal{F}(f)|_{\alpha,\beta} = \left\| \partial^\alpha (\xi^\beta \widehat{f}) \right\|_\infty \leq \left\| x^\alpha \partial^\beta f \right\|_1 = \int |x^\alpha (\partial^b f)| \, dx \\ \leq \int_{|x|<1} |x^\alpha (\partial^b f)| \, dx + \int_{|x|>1} \frac{1}{|x|^{n+1}} |x^{\alpha+n+1} (\partial^\beta f)| \, dx \end{aligned}$$

The second integral is bounded by a constant and some seminorms, and the first integral is bounded by some constant times $\left\| \partial^\beta f \right\|_\infty$. So we're done. ☐

Lemma 12.5 (Riemann-Lebesgue Lemma)**.** *If $f \in L^1$ then $\lim_{|\xi| \to \infty} \widehat{f}(\xi) = 0$.*

Proof. First, suppose that $f \in C_0^1$. Then $\frac{\partial f}{\partial x} \in L^1 \cap C^0$, so $|\widehat{\frac{\partial f}{\partial x}}|$, so $|\xi \widehat{f}(\xi)|$ is bounded. But also $f \in L^1$, so $|\widehat{f}(\xi)|$ is bounded. So $\lim_{|\xi| \to \infty} |\widehat{f}(\xi)| = 0$.

To extend to the whole space, take $f_n \in C_0^1 \cap L^1$ converging to $f \in L^1$. So $\widehat{f_n} \to \widehat{f}$ in L^∞, and $\lim_{|\xi| \to \infty} |\widehat{f}_n(\xi)| = 0$ for each n. This is enough. ☐

There's a key computation that we will need: What is $\widehat{f}(\xi)$ when $f(x) = e^{-a|x|^2}$? Note that $|x|^2 = \sum x_j^2$, this reduces to the 1-D case. Then

$$\widehat{f}(\xi) = \int e^{-ix\xi} e^{-ax^2} \, dx = \sqrt{\frac{\pi}{a}} e^{-|\xi|^2/4a}$$

by computing the square in the exponent.

Theorem 12.6 (Fourier inversion)**.** *Given $v(\xi) \in L^1(\mathbb{R}^n_\xi)$, define*

$$v^\vee(x) = (2\pi)^{-n} \int e^{ix \cdot \xi} v(\xi) \, d\xi.$$

If $f, \widehat{f} \in L^1$ then $(\widehat{f})^\vee = f$ almost everywhere.

Proof. We attempt

$$(\widehat{f})^\vee(x) = \iint f(y) e^{-2\pi i y \cdot \xi} e^{2\pi i x \cdot \xi} \, dy \, d\xi.$$

But we can't use Fubini! So instead we use an approximate identity.

For $t > 0$, let $\varphi_t(\xi) = e^{ix \cdot \xi - \frac{t^2}{2}|\xi|^2}$. By computation, we have (using the Gaussian formula)

$$\mathcal{F}\left(e^{-\frac{t^2}{2}|\xi|^2}\right)(y) = (2\pi)^{n/2} t^{-n} e^{-|y|^2/2t^2}.$$

So $\widehat{\varphi}_t(y) = (2\pi)^{n/2} t^{-n} e^{-|x-y|^2/2t^2}$.

Consider

$$\lim_{t\to 0}\int \widehat{f}\varphi_t = \lim_{t\to 0}\int e^{ix\cdot\xi-\frac{t^2}{2}|\xi|^2}\widehat{f}(\xi)\,d\xi = \int e^{ix\cdot\xi}\widehat{f}(\xi) = (2\pi)^n(\widehat{f})^\vee(x).$$

On the other hand,

$$\lim_{t\to 0}\int \widehat{f}\varphi_t = \lim_{t\to 0}\int f\widehat{\varphi}_t = \lim_{t\to 0}\int f(y)(2\pi)^{n/2}t^{-n}e^{-|x-y|^2/2t^2} = \lim_{t\to 0} f\star g_t$$

where $g_t(x) = t^{-n}g(x/t)$ and $g(x) = (2\pi)^{n/2}e^{-|x|^2/2}$.

Here's a fact: For any $g \in L^1$, we can define $g_t(x) = t^{-n}g(x/t)$. As $t \to 0$, g_t turns into a delta function. Then $\lim_{t\to 0}(f\star g_t) = f\cdot\|g\|_{L^1}$.

So this means that $\lim_{t\to 0}\int f\widehat{\varphi}_t = f\cdot\|g\|_{L^1} = (2\pi)^n f$. Therefore, $(\widehat{f})^\vee = f$. □

We now want to extend the domain of $\mathcal{F}$ to L^2.

Theorem 12.7 (Plancherel's Theorem). *If $f \in L^1\cap L^2$ then $\widehat{f} \in L^2$, and $\mathcal{F}|_{L^1\cap L^2}$ extends uniquely to an isomorphism of L^2.*

(In fact, it is actually a constant times isometry.)

Proof. Let $\mathcal{H} = \{f \in L^1 : \widehat{f} \in L^1\}$. Of course $\widehat{f} \in L^1$ implies $f \in L^\infty$ by Fourier inversion, so $\mathcal{H} \subset L^2$. Also since $\mathcal{S} \subset \mathcal{H} \subset L^2$ we know that $\mathcal{H} \subset L^2$ is dense.

Let $f, g \in \mathcal{H}$ and let $h = \overline{\widehat{g}}$. We have

$$\widehat{h}(\xi) = \int e^{-ix\cdot\xi}\overline{\widehat{g}}(x)\,dx = (2\pi)^n\overline{g(\xi)}.$$

So

$$\langle f, g\rangle_{L^2} = \int f\overline{g} = (2\pi)^n\int f\widehat{h} = (2\pi)^n\int \widehat{f}h = (2\pi)^n\int \widehat{f}\overline{\widehat{g}} = (2\pi)^n\left\langle \widehat{f},\widehat{g}\right\rangle_{L^2}$$

In particular, $\|f\|_{L^2} = (2\pi)^{n/2}\left\|\widehat{f}\right\|_{L^2}$. So $\mathcal{F}$ extends by continuity from $\mathcal{H}$ to an isomorphism of L^2. □

There's one technical question that we need to check. Does the extension agree with $\mathcal{F}$ on $L^1\cap L^2$?

Proof. Let $f \in L^1\cap L^2$ and let g_t be an approximate identity.

We claim that $f\star g_t \in \mathcal{H}$. This is because $g \in \mathcal{S}$, so $(\widehat{f\star g_t})(\xi) = \widehat{f}(\xi)\widehat{g}_t(\xi)$ has infinite-order decay at $|\xi| = \infty$ (since it is bounded times Schwartz).

Now $f\star g_t \to f$ in L^1 and L^2 by a fact about approximate identities. In addition, $\widehat{f\star g_t} \to \widehat{f}$ pointwise (directly, as $\widehat{g}_t \to 1$) and in L^2 (by the isomorphism). □

How do we find $\widehat{f}$ when $f \in L^2$? We can just plug in the integral, but we can approximate by taking $f_n \in \mathcal{S}$ converging to f (in L^2). This allows us to make sense of some divergent integrals.

Proposition 12.8. *We know that $f : \mathcal{S} \to \mathcal{S}$ is an isomorphism. We claim that it actually extends to a map from $\mathcal{S}' \to \mathcal{S}'$. This map is continuous in $\sigma(\mathcal{S}',\mathcal{S})$, which means that $\ell_i \in \mathcal{S}' \to \ell \in \mathcal{S}'$ if for all $g \in \mathcal{S}$, $\ell_i(g) \to \ell(g)$ in $\mathbb{C}$; this is the weak* topology.*

Proof. For $\ell \in \mathcal{S}'$, define (for any $g \in \mathcal{S}$) $[\mathcal{F}(\ell)](g) = \ell(\widehat{g})$. Why? If $\varphi \in \mathcal{S}$, let ℓ be integration against φ, and use $\int \widehat{\varphi} g = \int \varphi \widehat{g}$.

Why is $\mathcal{F} : \mathcal{S}' \to \mathcal{S}'$ continuous? Suppose $\ell_i \to \ell$ in $\mathcal{S}'$. Then for each $g \in \mathcal{S}$, $[\mathcal{F}(\ell_i)]g \to \ell_i(\widehat{g}) \to \ell(\widehat{g}) = [\mathcal{F}(\ell)]g$ in $\mathbb{C}$.

This is more or less by definition, and there's a general method to extend maps $\mathcal{S} \to \mathcal{S}$ to maps $\mathcal{S}' \to \mathcal{S}'$. ☐

Example 12.9. $[\mathcal{F}(\delta_0)](g) = \delta_0(\widehat{g}) = \widehat{g}(0) = \int g(x)\, dx$, so $\mathcal{F}(\delta_0)$ is integration against 1.

Example 12.10.

$$[\mathcal{F}(x)](g) = \int \xi \widehat{g}(\xi)\, d\xi = \int i \widehat{\frac{\partial g}{\partial x}}(\xi)\, d\xi = \left(i \widehat{\frac{\partial g}{\partial x}} \right)^{\vee} (0) \cdot (2\pi)^n = (2\pi)^n i \frac{\partial g}{\partial x}(0).$$

So $\mathcal{F}(p(x)) = (2\pi)^n (p(i\frac{\partial}{\partial x})g)(0)$.

13. 2/21

We will discuss bounded linear operators. We've been talking about various types of spaces, and the point of all of this is to understand continuous linear operators between these spaces.

Definition 13.1. Consider two spaces X, Y (Banach or Hilbert). Then

$$\mathcal{B}(X, Y) = \{A : X \to Y \text{ linear and continuous}\}.$$

Recall that continuous means that for all $x \in X$, $\|Ax\| \leq C \|x\|$. When $Y = X$, we write $\mathcal{B}(X)$.

There are several different themes:

1. Examples and explicit operators:
   We look at some "natural" operator A. Is it bounded?
2. General structure theory:
   Given $A \in \mathcal{B}(X)$, understand its "internal structure", e.g. "Jordan form".
3. Interesting subclasses of operators.

Here are some key examples to keep in mind.

Example 13.2.

1. Infinite matrices. Consider operators $A : \ell^2 \to \ell^2$, $A = (a_{ij})$, where $(Ax)_i = \sum a_{ij} x_j$. We ask for general criteria on (a_{ij}) such that A is bounded.
2. Integral operators. $u(x) \mapsto \int K(x, y) u(y)\, dy$. Here, K is a function on $M \times M$ for some measure space M. Is K bounded as a map $L^p \to L^q$? Is is true that

$$\left(\int \left| \int K(x, y) u(y)\, dy \right|^q dx \right)^{1/q} \leq C \left(\int |u(x)|^p\, dx \right)^{1/p}.$$

3. Convolution kernels. Consider some $k(z)$ and $K(x, y) = k(x - y)$. Then $\|k \star u\|_q \leq C \|u\|_p$. This relates to Young's inequality.

Everything is an integral kernel.

Theorem 13.3 (Schwartz kernel theorem)**.** *Let $A : C_0^\infty(\mathbb{R}^n) \to \mathcal{D}'(\mathbb{R}^n)$ be a continuous linear mapping. Then there exists a distribution $K_A \in \mathcal{D}'(\mathbb{R}^n \times \mathbb{R}^n)$ such that $Au(x) = \int K_A(x,y)u(y)\,dy$.*

This really means $\langle Au, \varphi\rangle = \langle K_A, \varphi \otimes u\rangle$ where $\varphi \otimes u = \varphi(x)u(y)$, for all $\varphi \in C_0^\infty(\mathbb{R}^n)$.

Sketch of proof. Define K_A as an element of $C_0^\infty(\mathbb{R}^n \times \mathbb{R}^n)^*$. We need to "define" $\langle K_A, \psi\rangle$ for $\psi \in C_0^\infty(\mathbb{R}^{2n})$.

But we define K_A as an operator acting on $C_0^\infty(\mathbb{R}^n) \otimes C_0^\infty(\mathbb{R}^n)$.

We need to check the following:

- Continuity of A implies that K_A is continuous with respect to the topology of $C_0^\infty(\mathbb{R}^{2n})$ restricted to $C_0^\infty(\mathbb{R}^n) \otimes C_0^\infty(\mathbb{R}^n)$. This is a matter of writing down the definitions.
- $C_0^\infty(\mathbb{R}^n) \otimes C_0^\infty(\mathbb{R}^n)$ is dense in $C_0^\infty(\mathbb{R}^{2n})$. □

We have the completed tensor product $C_0^\infty(\mathbb{R}^n)\widehat{\otimes} C_0^\infty(\mathbb{R}^n) = C_0^\infty(\mathbb{R}^{2n})$ is the limits of elements in the tensor product, by the result above. This means that if $\psi(x,y) \in C_0^\infty(\mathbb{R}^{2n})$ then we can find $\sum_{j=1}^N u_j(x)v_j(y) = \psi_N(x,y)$ so that $\psi_N \to \psi$.

Consider ψ compactly supported in $\mathbb{R}^{2n}$, so $\operatorname{supp}\psi \subset [-L,L]^{2n}$. Extend this periodically to $\psi \in C^\infty((\mathbb{R}/2L\mathbb{Z})^{2n})$. This is then a function on the torus, and we can take its Fourier series. This gives

$$\psi(x,y) = \sum e^{ix\cdot p + iy\cdot q}\psi_{pq}$$

for some $\psi_{pq} \in \mathbb{C}$ where $p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_n)$ and $p_i, q_j \in \frac{\pi}{L}\mathbb{Z}$.

Example 13.4. Here are more examples of integral operators. Consider

$$u(y) \mapsto \int \frac{u(y)}{|x-y|^{n-2}}\,dy$$

is the Newtonian potential, or the solution operator for Δ. Note that $\Delta(Nu) = u$.

Another example is the Fourier transform.

$$u(x) \mapsto \int e^{-ix\xi}u(x)\,dx = \widehat{u}(\xi) = \mathcal{F}u.$$

We also have Fourier multipliers $m(\xi)$. Define

$$A_m u = \mathcal{F}^{-1}(m\mathcal{F})u = (2\pi)^{-n}\int e^{ix\xi}m(\xi)e^{-iy\xi}u(y)\,dy\,d\xi.$$

Is $A_m : L^2 \to L^2$ bounded? Equivalently, when is $\widehat{u} \mapsto m\widehat{u}$ bounded on L^2. We claim that this is true if and only if $m \in L^\infty$.

Consider pseudodifferential operators

$$u \mapsto \int e^{i(x-y)\cdot\xi}a(x,y,\xi)u(y)\,dy\,d\xi = \int\left(\int a(x,y,\xi)e^{i(x-y)\cdot\xi}\,d\xi\right)u(y)\,dy.$$

Finally, consider $L^2(D) \supset \mathcal{H} = \{\text{holomorphic } L^2 \text{ functions on } D\}$.

Proposition 13.5. *$\mathcal{H}$ is a closed subspace. This means that $u_j \in \mathcal{H}$, i.e. $\frac{\partial u_j}{\partial \bar{z}} = 0$ for all j, i.e. $\int |u_j|^2 < \infty$ and $u_j \to u$ in L^2, then $\frac{\partial u}{\partial \bar{z}} = 0$*

This uses the Cauchy integral formula.

Define $\pi : L^2 \to \mathcal{H}$ as the orthogonal projector. For any $u \in L^2(D)$, $\pi u(z) \in \mathcal{H}$, and $(I - \pi)u \perp \mathcal{H}$. Then $\pi u(z) = \int_D B(z, w)u(w)\, dw$. In fact, $B \in C^\infty(D \times D)$ called the *Bergman kernel*.

Now, we discuss structure theory and special classes of operators. Given a Banach space X, take $\mathcal{B}(X)$. Inside, we have $\mathcal{B}(X)^0 = \{A \in \mathcal{B}(X) \text{ invertible}\}$.

Theorem 13.6. *$\mathcal{B}(X)^0 \subset \mathcal{B}(X)$ is an open set.*

Proof. If $A \in \mathcal{B}(X)^0$, we wish to show that there exists $\varepsilon > 0$ so that $\|B\| < \varepsilon$ implies that $A + B$ is invertible.

Here is a fact: $\|B_1 \circ B_2\| \le \|B_1\| \|B_2\|$. This means that $\mathcal{B}(X)$ is actually a Banach algebra.

We guess that the inverse is A^{-1}. Then we have $(A+B)A^{-1} = I + BA^{-1}$, and $\|BA^{-1}\| \le \|B\| \|A^{-1}\|$. Choose $\varepsilon = \frac{1}{2}\|A^{-1}\|^{-1}$. Then $\|BA^{-1}\| \le \frac{1}{2}$.

Now, we've reduced the problem to showing that $I+K$ with $\|K\| \le \frac{1}{2}$ is invertible. We've shifted this problem to being centered around the identity.

To do this, we have the Neumann series $(I+K)^{-1} = \sum_{j=0}^\infty (-1)^j K^j = I - K + K^2 - \cdots$. Why does this make sense? Let $K_N = \sum_{j=0}^N (-1)^j K^j$. Then $\|K_N\| = \sum_{j=0}^N \|K\|^j$ is uniformly bounded in N since $\|K^j\| \le \|K\|^j$. In fact, K_N is Cauchy, so $K_N \to (I+K)^{-1}$.

Now, we can write $(A+B)A^{-1}(I+K)^{-1} = I$. Strictly speaking, this is a right inverse. Now, take $A^{-1}(A+B) = I + A^{-1}B = I + K'$. Then write $(I+K')^{-1}A^{-1}(A+B) = I$, and we get a left inverse (with the same ε).

Given an operator C, we can find left and right inverses $CD_1 = I$ and $D_2C = I$. Then $D_2CD_1 = D_2 = D_1$. □

This gives us a large class of invertible operators. Any operator in the open ball $B_1(I)$ is invertible. Products of invertible operators are also invertible, and we can fill out a lot of space this way.

Suppose we take any $A \in \mathcal{B}(X)$. Then for $\lambda \in \mathbb{C}$ we can perturb it: $A - \lambda I$. We would like to perturb in a way that makes this invertible. We claim that $A - \lambda I$ is invertible if λ is large. As before, $A - \lambda I = (-\lambda)(I - \lambda^{-1}A)$. Now $\|\lambda^{-1}A\| = \frac{1}{\|\lambda\|}\|A\| < 1$, and $|\lambda| > \|A\|$ implies that $(A - \lambda I)^{-1}$ exists. $(A - \lambda I)^{-1}$ is called the resolvent family.

Definition 13.7. The *spectrum of* A, written $\operatorname{spec}(A)$, is defined as

$$\{\lambda \in \mathbb{C} : A - \lambda I \text{ is not invertible}\}.$$

Note that $\operatorname{spec}(A) \subseteq B_{\|A\|}(0) \subset \mathbb{C}$. This is a closed set, since its complement is open.

In fact, there exists A such that $\operatorname{spec}(A) = \{0\}$ or $\operatorname{spec}(A) = \overline{B}$. Does there exist A with empty spectrum?

Remark. $(\lambda I - A)^{-1}$ is a holomorphic function on $\mathbb{C} \setminus \operatorname{spec}(A)$.

Let $\mathcal{K} = \{A \in \mathcal{B}(X) \text{ compact operators}\}$. This means that A maps any bounded set to a precompact set. Alternatively, if $\{x_j\} \subset X$ has $\|x_j\| < C$ then Ax_j has a convergent subsequence.

Note that finite rank operators are compact.

Theorem 13.8. $\mathcal{K} \subset \mathcal{B}(X)$ *is a closed space.*

Proof. Consider $A_j \in \mathcal{K}$ where $A_j \to A$. We want to show $A \in \mathcal{K}$.

Suppose x_k are bounded. Choose a subsequence x_{k_l} such that $A_j(x_{k_l})$ converges as $\ell \to \infty$ for each j. Now, $Ax_{k_l} = A_j x_{k_l} + (A - A_j)x_{k_l}$. The first term converges, and the second term has small norm $\|(A - A_j)x_{k_l}\| \le \|A - A_j\| C$. We can pick another subsequence and diagonalize and get that the second term goes away. □

Proposition 13.9. *If X is a separable Hilbert space, then $\mathcal{K}$ is the closure of finite rank operators.*

Proof. Choose $\{\varphi_n\}$ is a countable orthonormal basis. Define

$$\lambda_n = \sup_{\substack{\psi \perp \varphi_j, j=1,\dots,n \\ \|\psi\|=1}} \|\mathcal{K}\psi\| .$$

We impose more restrictions, so we have a decreasing sequence $\lambda_1 \ge \lambda_2 \ge \cdots \to \overline{\lambda} \ge 0$.

If $\overline{\lambda} > 0$, then there exist $\psi_n \perp \varphi_j$ for $j \ge n$. Then $\|K\psi_n\| \ge \frac{1}{2}\overline{\lambda}$. But $\psi_n \to 0$. This means that $K\psi_n \to 0$, which is a contradiction, so therefore $\overline{\lambda} = 0$.

We also claim that

$$\sum_{j=1}^{N} \langle \varphi_j, \cdot \rangle K\varphi_j \to K.$$

□

14. 2/23

Last time, we talked about compact operators $\mathcal{K}$, and we showed that $\mathcal{K}$ is closed in the operator norm.

Example 14.1. Consider a diagonal operator A on ℓ^2 and $Ae_n = \lambda_n e_n$. We claim that $A \in \mathcal{K}$ if and only if $\lambda_n \to 0$. If $A \in \mathcal{K}$ then $\{e_n\}$ is bounded, so $Ae_n = \lambda_n e_n$ must have a convergent subsequence, so $\lambda_n \to 0$. Conversely, if $\lambda_n \to 0$, then define A_N be a truncation so that $A_N e_n = \lambda_n e_n$ for $n \le N$ and $A_N e_n = 0$ otherwise. Then A_N is finite-rank, so hence compact. Then $\|A_n - A\| \to 0$. Then $\|A_N - A\| = \sup_{n \ge N+1} |\lambda_n|$ is the largest eigenvalue.

Proposition 14.2. *If X is a separable Hilbert space, then $\mathcal{K}$ is the closure of finite rank operators.*

Proof. We finish the proof from last time. So we have $K = \sum_{j=1}^{\infty} \langle \varphi_j, \cdot \rangle K\varphi_j$. Then $K\varphi_l = K\varphi_l$, so this does the right thing on the basis elements. Then $\|K\varphi_j\| \le \lambda_j$, and hence $K_N = \sum_{j=1}^{N} \langle \varphi_j, \cdot \rangle K\varphi_j$ satisfies $K_n \to K$. □

There are also trace class operators: $\sum |\lambda_n| < \infty$. Also Hilbert-Schmidt operators: $\sum |\lambda_n|^2 < \infty$. These are subclasses of compact operators.

Example 14.3. Suppose $X = C^0(I)$. Then define the integral operator $Gu(x) = \int G(x,y)u(y)\,dy$. Suppose $G \in C^0(I \times I)$. We have a sequence of bounded continuous functions $|u_j(x)| \le C$ for all j. Then $\{Gu_j(x)\}$ satisfies $|Gu_j(x)| \le \int |G(x,y)||u_j(y)|\,dy \le C \int |G(x,y)|\,dy \le C'$, and

$$|Gu_j(x) - Gu_j(\tilde{x})| \le \int |G(x,y) - G(\tilde{x},y)||u(y)|\,dy \le \varepsilon$$

has uniform equicontinuity. Therefore, Arzela-Ascoli applies, and hence G is compact.

Proposition 14.4. *If K is compact then K maps weakly convergent sequences to strongly convergent ones.*

Proof. Suppose we have $x_n \rightharpoonup x$ weakly. Then $\|x_n\| \leq C$. Then $Kx_{n_j} = y_j \to y$. But $\ell(x_{n_j}) \to \ell(x)$ for every ℓ. But now take some $\tilde{\ell} \in X^*$, and

$$\tilde{\ell}(y_j - y_i) = \tilde{\ell}(Kx_{n_j} - Kx_{n_i}) = (K^*\tilde{\ell})(x_{n_j} - x_{n_i}) \to 0.$$

where $\tilde{\ell}(Kx) = K^*(\tilde{\ell})(x)$. Therefore, y_j is weakly Cauchy and $y_j \rightharpoonup \tilde{y} = y$. □

Since $\mathcal{K} \subset \mathcal{B}(X)$, we can take the quotient $\mathcal{B}(X)/\mathcal{K}$. Note that $\mathcal{B}(X)$ is an algebra, and $\mathcal{K}$ is an ideal. This means that if $A \in \mathcal{B}(X)$ and $K \in \mathcal{K}$ then $KA, AK \in \mathcal{K}$. Why? Think about what happens to bounded sequences and subsequences. This means that $\mathcal{B}(X)/\mathcal{K}$ is also an algebra, with coset operations $(A + \mathcal{K})(B + \mathcal{K}) = AB + \mathcal{K}$. This is called the *Calkin algebra*.

What does it mean for $\overline{A} \in \mathcal{B}(X)/\mathcal{K}$ to be invertible? This is the basic theorem of Fredholm theory:

Theorem 14.5. *The following are equivalent:*

(1) Given $A \in \mathcal{B}(X)$, $\overline{A}$ is invertible
(2) There exists $B \in \mathcal{B}(X)$ and $K_1, K_2 \in \mathcal{K}$ such that $AB = I + K_1$ and $BA = I + K_2$.
(3) A is Fredholm*:*
  (a) $\ker A$ *is finite dimensional*
  (b) $\operatorname{ran} A$ *is closed*
  (c) $X/\operatorname{ran} A$ *is finite dimensional.*

Example 14.6. What are Fredholm operators? Given $\ker A$, this has a complementary subspace W that is mapped into $\operatorname{ran} A$, with complementary subspace W'. Ignoring finite dimensional subspaces, these are invertible, and $A : W \to \operatorname{ran} A$ is an isomorphism.

Choose a basis $\overline{v}_1, \dots, \overline{v}_k$ for $X/\operatorname{ran} A$, which corresponds to linearly independent $v_1, \dots, v_k$ missing $\operatorname{ran} A$, and their span covers W'.

Proof. (1) $\implies$ (2): Let $\overline{B} = \overline{A}^{-1}$. Choose a representative $B \in \mathcal{B}(x)$ with $[B] = \overline{B}$. Then $[AB] = \overline{I}$, so $AB = I + K_1$, and we get (2).

(2) $\implies$ (1): This is trivial.

(3) $\implies$ (2): $A|_W : W \to \operatorname{ran} A$ is an isomorphism. This means that we can find an inverse for this. Choose $B' : \operatorname{ran} A \to W$. Extend B' to $B : X \to X$ such that B is B' on $\operatorname{ran} A$ and 0 on W'. Do this by projection. Then AB is I on $\operatorname{ran} A$ and 0 on W', so $AB = I - \Pi_{W'}$. Similarly, $BA = I - \Pi_{\ker A}$. These projections are finite rank operators, and hence compact. We've therefore found inverses up to compact error.

(2) $\implies$ (3): We know that $AB = I - K_1$ and $BA = I - K_2$. We prove (a). If $x \in \ker A$, then $BA(x) = x - K_2 x$, so $x = K_2 x$, and hence $I|_{\ker A} = K_2|_{\ker A}$, so $\ker A$ is finite dimensional.

Now we do (b): If $x_m \in X$ then $Ax_n = y_n \in \operatorname{ran} A$. Suppose $y_n \to y$. Then our goal is to show $y = Ax$. Let $\tilde{x}_n$ be the projection of x_n onto W. So we want to show that $\tilde{x}_n$ converges; we still have $A\tilde{x}_n = y_n$. So then $BA(\tilde{x}_n) = \tilde{x}_n - K_2\tilde{x}_n = By_n$, so that $\tilde{x}_n = By_n + K_2\tilde{x}_n$. If $\|\tilde{x}_n\| \leq C$ then $K_2\tilde{x}_n$ converges (up to subsequence), so therefore $\tilde{x}_n$ converges. If not, suppose $\|\tilde{x}_n\| = c_n \to \infty$. Then $A(\tilde{x}_n/c_n) = y_n/c_n \to 0$, so that $(\tilde{x}_n/c_n) = B(y_n/c_n) + K_2(\tilde{x}_n/c_n)$, and therefore $\tilde{x}_n/c_n \to \widehat{x}$ with $\|\widehat{x}\| = 1$. Then $A(\tilde{x}_n/c_n) = y_n/c_n$ and

$A\widehat{x} = 0$. But then $\widehat{x} \in \ker A$, which is a contradiction, since it is in the complement of the kernel.

Finally, we prove (c). We take a side route. Suppose that X is a Hilbert space. Then we have perpendicular complements. We also have the adjoint A^* defined via $\langle Ax, y\rangle = \langle x, A^*y\rangle$. Given y, define A^*y as an element of X^*, and we have $(A^*y)(x) = \langle Ax, y\rangle$, so that $|(A^*y)(x)| \le \|Ax\| \, \|y\| \le \|A\| \, \|x\| \, \|y\|$. This defines the adjoint operator. Also, recall from linear algebra that $\ker A^* = (\operatorname{ran} A)^\perp$ and $\operatorname{ran} A^* = (\ker A)^\perp$. For example, $A^*y = 0$ if and only if $\langle A^*y, x\rangle = \langle y, Ax\rangle = 0$ for all x.

Now, A satisfies (2) if and only if A^* does. That is, $AB = I - K_1$ and $BA = I - K_2$. Then $B^*A^* = I - K_1^*$ and $A^*B^* = I - K_2^*$. So what we actually need is this:

Proposition 14.7. *K is compact if and only if K^* is compact.*

In fact, this is true in Banach spaces, and we will do this in generality.

Proof. Suppose that $K \in \mathcal{K}$. Then $K^* : X^* \to X^*$ so that $(K^*\ell)(x) = \ell(Kx)$. Then let $B^* \subset X^*$ be the unit ball. Then we show that $K^*(B^*)$ is precompact. We consider $\overline{K(B)}$ is a compact set. Suppose we have $\ell_n \in B^*$. Then $u \in K(B)$, and we have $\{\ell_n|_{\overline{K(B)}}\}$ is uniformly bounded and uniformly equicontinuous. Indeed, for $u = Kx$, we have $\ell_n(u) = \ell_n(Kx)$, so $|\ell_n(u)| \le \|\ell_n\| \, \|x\| \, \|K\| \le C$. In addition, $|\ell_n(u) - \ell_n(v)| \le \|u - v\|$, so hence ℓ_n has a uniformly convergent subsequence as an element of $C^0(\overline{K(B)})$. This now says that $|\langle \ell_n - \ell_m, Kx\rangle| = |\langle K^*\ell_n - K^*\ell_m, x\rangle| < \varepsilon$, which implies that $\|K^*\ell_n - K^*\ell_m\| < \varepsilon$. This is what we wanted to show. □

In the Hilbert space setting, this finishes this Fredholm theorem. It remains to show that if (2) holds, i.e. $BA = I - K_2$ and $AB = I - K_1$ for compact K_1, K_2, then $\operatorname{ran} A$ has finite codimension. The problem is that we cannot use duality.

Suppose we have $Ax = y$. We guess $\overline{x} = By$, leading to $A\overline{x} = y - K_2 y$, so this isn't good enough. So what if $\overline{x} = Bz$? Then $A\overline{x} = z - K_2 z = y$, so we want to solve $(I - K_2)z = y$. We've reduced this to a nicer problem, with a simpler Fredholm operator. □

15. 3/1

Recall that we have $\mathcal{B}(X, Y)$ is a space of bounded linear operators $X \to Y$, and a space of invertible operators $\mathcal{I}$. We proved that $\mathcal{I}$ is an open set. There is also $\mathcal{K}$ is the space of compact operators. This is an ideal. Finally, we have the space $\mathcal{F}$ of Fredholm operators, which we will continue to discuss.

Definition 15.1. $A \in \mathcal{F}$ if

(1) $\ker A$ is finite dimensional
(2) $\operatorname{ran} A$ is closed
(3) $\operatorname{ran} A$ has finite codimensional.

Equivalently, $A \in \mathcal{F}$ if and only if there exist operator B and compact operators K_1 and K_2 such that $AB = I - K_1$ and $BA = I - K_2$.

We showed that if K is compact then K^* is compact. Note that $A^*B^* = I - K_1^*$ and $B^*A^* = I - K_2^*$, so therefore $\ker A^*$ is finite dimensional, and $\ker A^* = (\overline{\operatorname{ran} A})^\circ$.

There is a mapping $\mathcal{F} \to \mathbb{Z}$ called the *index*.

Definition 15.2. $\operatorname{Ind}(A) = \dim \ker A - \dim(Y/\operatorname{ran} A)$.

First, we will check that this is a very stable object that is constant on big open sets. This makes it easily computable and a useful object.

Example 15.3. Note that $\dim \ker A$ does not have these nice stability properties. For example, if $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$. Then $A - \lambda I$ has kernel with dimension that changes in a noncontinuous way.

Proposition 15.4. *If* $A = I + K \in \mathcal{F}(X)$ *then* $\operatorname{Ind}(A) = 0$.

Proof. First, assume that $\ker A = \{0\}$. Then look at $X_1 = A(X) \subseteq X$. If $X_1 = A(X) \subsetneq X$, then take a descending series $X_j = A(X_{j-1})$ to get $\cdots \subset X_3 \subset X_2 \subset X_1 \subset X$. We claim that this series has to stabilize. Here's a Banach space trick (in Hilbert spaces, this is easier): Choose $x_k \in X_k$ with $\|x_k\| = 1$ and $\operatorname{dist}(x_k, X_{k-1}) > \frac{1}{2}$. Note that $A^j = (I+K)^j = I + K + 2K^2 + \cdots + K^j$, which is still identity plus compact. Now, if $m < n$ then $Kx_m - Kx_n = (A-I)x_m - (A-I)x_n = -x_m + X_{m+1}$, which means that $\|Kx_m - Kx_n\| > \frac{1}{2}$, which is a contradiction since there must be a Cauchy sequence. Hence the sequence stabilizes. Hence, $X_1 = X$ so $\operatorname{Ind} A = 0$.

Now, suppose that $N_j = \ker A^j$. Notice that $N_1 \subset N_2 \subset \cdots \subset N_j \subset N_{j+1}$. Then if all inclusions are proper, choose $x_j \in N_j$ and $\operatorname{dist}(x_j, N_{j-1}) > \frac{1}{2}$ and $\|x_j\| = 1$. Then just as before, the sequence must stabilize. So we cannot have any infinite Jordan blocks. Now, suppose $N_j = N_{j+1} = \cdots$. Then $A : N_j \to N_j$, so it passes to a mapping of the quotient space $\tilde{A} : X/N_j \to X/N_j$. Similarly, $\tilde{K}$ is the operator induced by K, so $\tilde{A} = I + \tilde{K}$. Hence, for all points $y \in X$, there exists $z \in N_j$ such that $Ax = y + z$ has a solution, so $\operatorname{ran} A + N_j = X$.

Suppose we consider $\operatorname{ran} A \cap N_j = \{w = Aw', A^j w = 0\} = W$. Then $A^j w = A^{j+1} w'$, so hence $A^j w' = 0$. Then $w' \in \ker A_j$ so $W = \operatorname{ran}(A|_{N_j})$. So we've reduced our problem to a finite dimensional thing. So we can now think about $A|_{N_j} = N_j \to N_j$. Then $\dim(\ker A|_{N_j}) = \dim(\operatorname{coker} A|_{N_j})$ by the rank-nullity theorem. Since $X = \operatorname{ran} A + N_j$, so $\dim \operatorname{coker} A = \dim(N_j) - \dim W = \dim(A|_{N_j}) = \dim \ker A$. This is what we wanted. □

Proposition 15.5. *If* $A \in \mathcal{F}(X)$ *and* $K \in \mathcal{K}$ *then* $\operatorname{Ind}(A+K) = \operatorname{Ind}(A)$.

This depends on the following lemma.

Lemma 15.6. *If* $A_1, A_2 \in \mathcal{F}$ *then* $A_2 A_1 \in \mathcal{F}$ *and* $\operatorname{Ind}(A_2 A_1) = \operatorname{Ind}(A_2) + \operatorname{Ind}(A_1)$.

Proof. We can write $B_1 A_1 = I - K_1^1$, $A_1 B_1 = I - K_2^1$, $B_2 A_2 = I - K_1^2$, $A_2 B_2 = I - K_2^2$. Then $(B_1 B_2)(A_2 A_1) = B_1(I - K_1^2)A_1 = I - K_1^1 - B_1 K_1^2 A_1$ and similarly for the other composition. So $A_2 A_1 \in \mathcal{F}$.

Sublemma 15.7. *Take an exact sequence* $0 \to V_0 \xrightarrow{A_0} V_1 \xrightarrow{V_1} V_2 \xrightarrow{A_2} \cdots \xrightarrow{A_{N-1}} V_N \to 0$. *Then* $\sum (-1)^j \dim V_j = 0$.

Proof. This is because $V_j = \operatorname{ran} A_j \oplus W_j$. Just write down exactness. □

Now, we can write down the sequence

$$0 \to \ker A_1 \hookrightarrow \ker A_2 A_1 \xrightarrow{A_1} \operatorname{im} A_1 \cap \ker A_2 \to X/\operatorname{ran} A_1 \to \xrightarrow{A_2} X/\operatorname{ran}(A_2 A_1) \to X/\operatorname{ran} A_2 \to 0.$$

Exercise 15.8. Fix this.

Then we can observe that it is exact and finish the proof. Note that this requires no topology at all. It is a purely algebraic lemma. □

Now we can go back to the proposition.

Proof of Proposition 15.5. Suppose that $AB = I + K_1$. Then $\operatorname{Ind} A + \operatorname{Ind} B = 0$. Also, $(A+K)B = I + K_1 + KB$, so therefore $\operatorname{Ind}(A+K) = -\operatorname{Ind} B$, so we're done. ☐

This shows that $\mathcal{F} + \mathcal{K} = \mathcal{F}$.

Proposition 15.9. *$\mathcal{F} \subset \mathcal{B}(X)$ is open.*

Proof. This means that if $A \in \mathcal{F}$ then for all C with $\|C\| < \varepsilon$ we have $A + C \in \mathcal{F}$. If $AB = I + K$ then $(A+C)B = I + K + CB$, so therefore $I + CB + K$, so $(I+CB)^{-1} = I - CB + (CB)^2 - \cdots$ if $\|CB\| < 1$. Then $(A+C)B(I+CB)^{-1} = I + K(I+CB)^{-1}$. ☐

Proposition 15.10. *The index is also stable under small norm perturbations.*
This means that $\operatorname{Ind}(A+C) = \operatorname{Ind}(A)$.

We won't prove this now.

Now, we have a homomorphism $\operatorname{Ind} : \mathcal{F} \to \mathbb{Z}$. They form huge open components, and the index tells you which component you are in.

Fact 15.11. The index "counts" the components of $\mathcal{F}$. This means that $\mathcal{F} = \sqcup_{j\in\mathbb{Z}} \mathcal{F}_j$ where each $\mathcal{F}_j$ is connected and open, and $\operatorname{Ind}|_{\mathcal{F}_j} = j$.

We still haven't produced any operators of nonzero index, so to show this we would need to do this.

Example 15.12. Consider the shift operator $A : \ell^2 \to \ell^2$. Then $(Ax)_j = x_{j+1}$ where $(x_1, x_2, \dots) \to (x_2, x_3, \dots)$. Then $\operatorname{Ind} A = 1$.

Consider $L^2(S^1) = \operatorname{span}\{e^{in\theta}, n \in \mathbb{Z}\}$. Then $u(\theta) = \sum_{-\infty}^{\infty} a_n e^{in\theta}$, which corresponds to $(a_n) \in \ell^2$.

Consider $\mathcal{H} = \operatorname{span}\{e^{in\theta}, n \geq 0\}$. We claim that $\mathcal{H}$ is the set of $\overline{u}(\theta)$ such that there exists $u(r,\theta)$ holomorphic in the unit disk, with $\sup_{r<1} \int |u(r,\theta)|^2 < \infty$ and $u(1,\theta) = \overline{u}(\theta)$. This is called a *Hardy space*. Then $\mathcal{H} \subset L^2$ is a closed subspace.

Here is an interesting Fredholm operator. Choose $f : S^1 \to \mathbb{C}$ with $f \in C^\infty$. Then define $Mf : \mathcal{H} \to \mathcal{H}$ via $M_f u = \Pi f u$ where $\Pi : L^2 \to \mathcal{H}$ is a projection.

Theorem 15.13 (Toeplitz Index Theorem)**.** *M_f is Fredholm if and only if $f \neq 0$ everywhere on S^1. If this is the case,* $\operatorname{Ind}(M_f) = -$*(winding number of f).*

This is very easy to prove given what we've done.

Sketch of proof. If $f \neq 0$, then we want to construct a pseudoinverse for M_f; we claim that this is $M_{f^{-1}}$. Then $M_{f^{-1}} M_f = \Pi f^{-1} \Pi f u$ where $u \in \mathcal{H}$. If $f \in \mathcal{H}$ then this wouldn't be so bad, since $fu \in \mathcal{H}$. For general f, let $f_N = \sum_{-N}^{\infty} b_n e^{in\theta} = \sum_{-N}^{-1} + \sum_0^\infty = f_N' + f_N''$. Write $u = \sum_0^\infty a_n e^{in\theta}$. Then $f^{-1}\Pi f u = u$ if $u = \sum_N^\infty a_n e^{in\theta}$, so this is invertible on a space of finite codimension. Then $M_{f_N^{-1}} M_{f_N} = I - K_N$, where $M_f = M_{f_N} + C_N$ where $\|C_N\| \ll 1$.

If $f(\theta_0) = 0$, then there exists a sequence $u_j \in \mathcal{H}$ such that $M_f u_j \to h$ but $u_j \not\to h$. ☐

Now, if f does not vanish, it can be homotoped to some Z^n for $n \in \mathbb{Z}$, and $\operatorname{Ind} M_{Z^n} = -n$. This is because M_{Z^n} sends $u = \sum_0^\infty a_n e^{in\theta} \to \sum a_n e^{i(n+N)\theta}$ for $N \geq 0$, so the index is $-N$. This is important because it shows that the indices can be reasonably computed.

Recall the notion of the spectrum of $A \in \mathcal{B}(X)$. If $\lambda \in \mathbb{C}$ then $\lambda \notin \operatorname{spec} A$ if and only if $\lambda I - A$ is invertible. Note that if $|\lambda - \lambda'| < \varepsilon$ then $\|(\lambda I - A) - (\lambda' I - A)\| < \varepsilon$, so in fact $\mathbb{C} \setminus \operatorname{spec}(A)$ is an open set, and hence $\operatorname{spec}(A)$ is closed.

Proposition 16.1. *If $|\lambda| > \|A\|$ then $\lambda \notin \operatorname{spec} A$.*

Proof. $\lambda I - A = \lambda(I - \lambda^{-1}A)$. Then $(\lambda I - A)^{-1} = \lambda^{-1}\sum_{j=0}^{\infty} \lambda^{-j}A^j$. □

Proposition 16.2. $\operatorname{spec}(A) \neq \emptyset$.

Proof. $R(\lambda) = (\lambda I - A)^{-1}$ is holomorphic away from $\operatorname{spec} A$. Note that $R(\lambda)$ holomorphic in Ω means that (equivalently)

(1) If $\lambda_0 \in \Omega = \mathbb{C} \setminus \operatorname{spec} A$ then $R(\lambda) = \sum_{j=0}^{\infty} R_j(\lambda - \lambda_0)^j$ for $|\lambda - \lambda_0| < \varepsilon$.
(2) For all $f \in X$, $R(\lambda)f = \sum f_j(\lambda - \lambda_0)^j$ for $f_j \in X$.
(3) For all $f \in X$, $\ell \in X^*$, then $\ell(R(\lambda)f) = \sum a_j(\lambda - \lambda_0)^j$.

These three definitions are equivalent. One of the advantages of working with holomorphic functions is that we can use all of the machinery of complex analysis. So we can use the Cauchy integral formula:

$$\int_{|\lambda|=R} (\lambda I - A)^{-1}\, d\lambda = I + O(1/R) \to I$$

because only the leading term of the Neumann series is important. If R were holomorphic in $|\lambda| \leq R$ then the integral would be zero, which isn't true. □

In fact, any closed bounded nonempty set can be the spectrum of some operator. We will concentrate on nicer operators.

Proposition 16.3. *If $K \in \mathcal{K}(X)$ is compact, then*

(1) $\operatorname{spec}(K) = \{\lambda_j\} \cup \{0\}$ *for λ_j discrete in $\mathbb{C} \setminus \{0\}$.*
(2) Each $\lambda_j \in \operatorname{spec}(K)$ is an eigenvalue of finite multiplicity, i.e. $\ker(\lambda_j I - K)$ *is finite dimensional, and* $\ker(\lambda_j I - K)^\ell$ *is finite dimensional and stabilizes for large ℓ.*

If we look at $(\lambda I - K)^{-1} = \sum_{n \geq -N} C_n(\lambda - \lambda_j)^n$, the Laurent series only has finitely many terms with negative n, and $C_{-N}, \ldots, C_{-1}$ are finite rank operators.

Proof. $\lambda I - K = \lambda(I - \lambda^{-1}K)$. Suppose that $\lambda \neq 0$. This is noninvertible if and only if $(I - \lambda^{-1}K)$ is noninvertible, if and only if $\ker(I - \lambda^{-1}K) \neq 0$.

We now have that $\lambda_j \in \operatorname{spec} K$ if and only if there exists $w_j \in X$ such that $Kw_j = \lambda_j w_j$. Now, let $Y_n = \operatorname{span}\{w_1, \ldots, w_n\}$. Then we have a sequence $Y_1 \subset Y_2 \subset \cdots \subset Y_n \subset \cdots$. Choose an element $y_n \subset Y_n$ with $\|y_n\| = 1$ and $\operatorname{dist}(y_n, Y_{n-1}) > \frac{1}{2}$. We can write $y_n = \sum_{j=1}^{n} a_j w_j$. Then $(K - \lambda_n I)y_n = \sum_{j=1}^{n} a_j(\lambda_j - \lambda_n)w_j \in Y_{n-1}$. When $n > m$, we have $Ky_n - Ky_m = (K - \lambda_n I)y_n - \lambda_n y_n - Ky_m \in \lambda_n(y_n + Y_{n-1})$. Therefore $\|Ky_n - Ky_m\| \geq \frac{1}{2}|\lambda_n|$. So necessarily $|\lambda_n| \to 0$ or else $\{Ky_n\}$ wouldn't have a Cauchy subsequence. □

Why is $0 \in \operatorname{spec} K$? Because $\operatorname{spec} K$ is an infinite closed sequence accumulating at 0.

We will discuss an important special case.

Theorem 16.4. *Let H be a separable Hilbert space, and let K be a self-adjoint compact operator $K^* = K$. Under these conditions, there exist $\{x_j\}$ and $\{\lambda_j\}$ so that $\{\lambda_j \in \mathbb{R}\}$ and $\lambda_j \to 0$ so that $Ax_j = \lambda_j x_j$ and $\{x_j\}$ is an orthonormal basis for H.*

Proof. Define $Q(x) = \frac{\langle Kx,x\rangle}{\|x\|^2}$ for $x \neq 0$. (Alternatively, work with $\tilde{Q}(x) = \frac{\|Kx\|^2}{\|x\|^2}$, and the argument works almost identically.)

First, Q is bounded for $x \neq 0$. Actually, $\frac{|\langle Kx,x\rangle|}{\|x\|^2} \leq \|K\|$.

Now, we try to maximize A. Choose $y_l \in H$ so $\|y_l\| = 1$ and $Q(y_l) \to \sup Q$ or $\to \inf Q$. Note that $-\|K\| \leq Q(x) \leq \|K\|$. The claim is that there exists a convergent subsequence of the y_l. Since y_l are bounded and K is compact, we have a convergent $Ky_{l_j} \to z$. Then $Q(y_{l_j}) = \langle Ky_{l_j}, y_{l_j}\rangle$. Passing to a subsequence if necessary, we can assume $y_{l_j} \rightharpoonup \tilde{y}$ weakly. Drop the index j. We want the y_l to converge strongly. We can write $\langle Ky_l, y_l\rangle = \langle Ky_l - z, y_l\rangle + \langle z, y_l\rangle \to \langle z, \tilde{y}\rangle = \langle K\tilde{y}, \tilde{y}\rangle$. Therefore, $Q(y_l) \to \sup Q = Q(\tilde{y})$.

Then $\tilde{y}$ must be an eigenvalue, and $Q(\tilde{y}+tw) \leq Q(\tilde{y})$ for all t. We claim that $\frac{d}{dt}|_{t=0} Q(\tilde{y} + tw) = 0$. Then

$$\left.\frac{d}{dt}\right|_{t=0} \frac{\langle K(\tilde{y}+tw), \tilde{y}+tw\rangle}{\|\tilde{y}+tw\|^2} = 2\langle K\tilde{y}, w\rangle - (\sup Q) 2 \langle \tilde{y}, w\rangle = 0.$$

Therefore, $\langle K\tilde{y} - \lambda_1 \tilde{y}, w\rangle = 0$ for all w where $\lambda_1 = \sup Q$. Set $x_1 = \tilde{y}$, so then $Kx_1 = \lambda_1 x_1$ and $\lambda_1 = \sup Q \in \mathbb{R}$. So we've found a single eigenvector.

Now, let $H_1 = \{x \in H : x \perp x_1\}$. Then $Kx_1 = \lambda_1 x_1$ implies that $\mathbb{C}x_1$ is invariant by K, so $(\mathbb{C}x_1)^\perp$ is also invariant by K. Then $\langle x, x_1\rangle = 0$ and $\langle Kx, x_1\rangle = \langle x, Kx_1\rangle = \lambda_1 \langle x, x_1\rangle = 0$. So now restrict to $Q|_{H_1}$ and apply exactly the same argument. Find a maximizer x_2 for $Q|_{H_1}$. We want $Q(x_2) \geq Q(w)$ for all $w \in H_1$. Then $Q(x_2 + tw) \leq Q(x_2)$ if $w \in H_1$. Running the argument gives us $\langle Kx_2 - \lambda_2 x_2, w\rangle = 0$ for all $w \in H_1$. Therefore, $Kx_2 = \lambda_2 x_2 + cx_1$, and hence actually $Kx_2 = \lambda_2 x_2$. Notice that by definition, $\lambda_2 \leq \lambda_1$.

We continue inductively. Define $x_1, x_2, \ldots, x_N, \ldots$, each satisfying $Kx_j = \lambda_j x_j$. These are orthonormal, $\lambda_j \in \mathbb{R}$, and $|\lambda_1| \geq |\lambda_2| \geq \cdots$. This is a monotone decreasing sequence, so it converges to $\overline{\lambda}$. Is this zero? By the previous theorem, yes; more primitively, otherwise we would contradict compactness.

Finally, we should check completeness. Let $Y = \overline{\operatorname{span}\{x_j\}}$. Let $Y^\perp \subset H$ be its orthogonal complement. We want $K|_{Y^\perp} = 0$. Suppose not. Then $Q|_{Y^\perp}$ has nonzero supremumm μ. This is impossible because $\mu > |\lambda_j|$ for j large. That's the end of the proof. □

This is a spectral theorem in the strongest possible sense.

Corollary 16.5. *Consider $A = I + K$ where K is compact and self-adjoint, so A is Fredholm and $A^* = A$. Then all of the spectral accumulation is at* 1.

If $Kx_j = \lambda_j x_j$, then $Ax_j = (1+\lambda_j)x_j$. If we want to solve $Ax = y$, we write $y = \sum b_j x_j$, then we hope that $x = \sum a_j x_j$. But we want $Ax = \sum (1+\lambda_j) a_j x_j$, so $a_j = \frac{b_j}{1+\lambda_j}$.

There is nothing too exotic happening yet, but we'll see more general results next time.

17. 3/8

Recall from last time: A is a symmetric operator, compact on H.

Proposition 17.1. *If $A^* = A$ then* $\operatorname{spec}(A) \subset \mathbb{R}$.

Proof. If $Ax = \lambda x$ then $\lambda \|x\|^2 = \langle Ax, x\rangle = \langle x, A^* x\rangle = \overline{\lambda} \|x\|^2$. But the spectrum is more than just the eigenvalues.

We really need $\lambda \notin \mathbb{R}$ implies that $\lambda I - A$ is invertible. Then if $\lambda = a + ib$, define $B(x, y) = \langle (A - \lambda)x, y \rangle$. We have $|B(x,x)|^2 = |\langle (A - aI)x, x\rangle - ib \langle x, x\rangle|^2 = |\langle (A - aI)x, x\rangle|^2 + b^2 \|x\|^2 \geq b^2 \|x\|^2$. This means that this B is still coercive (recall the Lax-Milgram 4.2), i.e. $|B(x, y)| \leq C \|x\| \|y\|$ and $|B(x, y)|^2 \geq c \|x\|^2$. This means that there is some version of the Riesz representation theorem.

Hence given z, there exists $x \in H$ such that $y \mapsto B(x, y) = \langle z, y\rangle$, which is the statement that $\langle (A - \lambda I)x, y\rangle = \langle z, y\rangle$, which is true if and only if $(A - \lambda I)x = z$. □

Proposition 17.2. *If $A^* = A$ and A is compact, then* $\operatorname{spec}(A)$ *is discrete in* $\mathbb{R} \setminus \{0\}$.

Proof. The proof was fairly constructive. Given $y \mapsto \langle Ay, y\rangle$, we realize the supremum $\sup_{\|y\|=1} \langle Ay, y\rangle$ by choosing y_n with $\|y_n\| = 1$ to get that $\langle Ay_n, y_n\rangle \to \sup \langle Ay, y\rangle$. Using some of our standard theorems about weak convergence, we saw that $\langle Ay_n, y_n\rangle \to \langle A\tilde{y}, \tilde{y}\rangle$. To do this, we have $y_n \rightharpoonup \tilde{y}$ but not strong convergence. If $\|\tilde{y}\| < 1$ then

$$\left\langle A\frac{\tilde{y}}{\|\tilde{y}\|}, \frac{\tilde{y}}{\|\tilde{y}\|} \right\rangle = \frac{\langle A\tilde{y}, \tilde{y}\rangle}{\|\tilde{y}\|^2} \sup_{\|y\|=1} \langle Ay, y\rangle,$$

which is bad. Hence, $y_n \to \tilde{y}$ and $\|y\| = 1 = \lim \|y_n\|$. This implies (see the homework) that $y_n \to \tilde{y}$. □

Definition 17.3. Suppose $L = -\Delta + V$ on $T^n = S^1 \times S^1 \times \cdots \times S^1$, where $\Delta = \sum_{j=1}^n \frac{\partial^2}{\partial \theta_j^2}$. Then

$$\langle -\Delta u, u\rangle = \sum_{j=1}^n \left\| \frac{\partial u}{\partial \theta_j} \right\|^2 = \|\nabla u\|^2$$

for $u \in C^\infty(T^n)$ and V real-valued in C^∞.

We want to think of L as a compact operator.

There are two points of view.

Definition 17.4. Firstly, consider $H^s \subset L^2$ where

$$H^s = \left\{ u = \sum_{k=(k_1,\ldots,k_n)} a_k e^{ik\theta} : \sum_{k\in\mathbb{Z}^n} (1 + |k|^2)^s |a_k|^2 < \infty \right\},$$

and $k\theta = k_1\theta_1 + \cdots + k_n\theta_n$.

An easy fact is this:

Proposition 17.5. *If $u \in H^s$ then $\partial_{\theta_j} u \in H^{s-1}$.*

Proof. If $u = \sum a_k e^{ik\theta}$ then $\partial_{\theta_j} u = \sum a_k (ik_j) e^{ik\theta}$, so then

$$\sum (1 + |k|^2)^{s-1} |(ik_j a_k)|^2 \leq C \sum (1 + |k|^2)^s |a_k|^2.$$ □

In particular, look at $s = 2$, i.e. the space of functions where $\sum (1 + |k|^2)^2 |a_k|^2 < \infty$. Applying the fact twice, we see that $\Delta : H^2(T^n) \to L^2(T^n)$.

Also, we see that $L = -\Delta + V : H^2(T^n) \to L^2(T^n)$ is a bounded mapping. One way to do this is to follow the definition: $(Vu)_k = \sum_{\ell\in\mathbb{Z}^n} V_\ell a_{k\ell}$. Then we need to compute

$$\|Vu\|^2 = \sum_{k\in\mathbb{Z}^n} |(Vu)_k|^2 = \sum_k \left| \sum_\ell V_\ell a_{k\ell} \right|^2 < C \sum |a_k|^2 = C \|u\|_{L^2}^2.$$

This is a standard fact about convolutions but it is a bit of a mess.

Instead, to make this easier, think of H^2 in a different way:

Proposition 17.6. $H^2 = \{u \in L^2 : u, \partial_{\theta_j} u, \partial^2_{\theta_j \theta_\ell} \in L^2\}$.

Proposition 17.7. *This L is Fredholm.*

Proof. We start with a reference operator: $L_0 = -\Delta + 1 : H^2 \to L^2$ is invertible. This is because

$$L_0 u = (-\Delta + 1)u = \sum (1 + |k|^2) a_k e^{ik\theta} = \sum b_k e^{ik\theta}$$

is satisfied for $a_k = \frac{b_k}{1+|k|^2}$. Then $u = \sum a_k e^{ik\theta} \in H^2$ because we have $\sum |a_k|^2 (1 + |k|^2)^2 = \sum |b_k|^2 < \infty$.

Now $(-\Delta+V)(-\Delta+1)^{-1} = (-\Delta+1+V-1)(-\Delta+1)^{-1} = I+(V-1)(-\Delta+1)^{-1} = I+K$. We need to show that $(V-1)(-\Delta+1)^{-1} : L^2 \to L^2$ is compact. It suffices to show that $(-\Delta+1)^{-1}$ is compact.

We just checked that $(-\Delta + 1)^{-1} : L^2 \to H^2$ is bounded. How about $(-\Delta + 1)^{-1} : L^2 \to H^2 \hookrightarrow L^2$? The claim is that $H^2 \hookrightarrow L^2$ is a compact inclusion. This means that if we have $\{u_j : \|u_j\|_{H^2} \le c\}$ then there exists a convergent subsequence in L^2. So we want $\sum |u(j)_k|^2 (1 + |k|^2)^2 \le C$.

Given any $\varepsilon > 0$, there exists N such that

$$\left\| u_j - \sum_{|k_l| \le N} u(j)_k e^{ik\theta} \right\|^2_{L^2} < \varepsilon.$$

We know that the $|u(j)_k| \le C'$ uniformly, so therefore the tail of the series is $< \varepsilon$, i.e. $\sum_{\text{some } k_l > N} |u(j)_k|^2 \le \varepsilon$.

The upshot is what we wanted: $(-\Delta+V)(-\Delta+1)^{-1} = I+K$. We can do the same thing on the other side: $(-\Delta + 1)^{-1}(-\Delta + V) : H^2 \to H^2$. Write this as $(-\Delta + 1)^{-1}(-\Delta + V) = I+(-\Delta+1)^{-1}(V-1)$. Now, we have $(-\Delta+1)^{-1}(V-1)$ represents mappings $H^2 \to H^4 \hookrightarrow H^2$, where the last map is a compact inclusion. This requires a slight generalization of before, and then we are done. ∎

Then we've shown that $-\Delta + V : H^2 \to L^2$ is Fredholm. But we can't talk about spectrum because it is a map between different spaces. But this is great for discussing solvability: $(-\Delta + V)u = f$. What is the index? The index is invariant under deformation, so we deform this so that V is carried to 1, where the operator is invertible. So therefore $L = -\Delta + V$ is Fredholm of index 0. So as long as there is trivial null space, it is surjective, so we are reduced to studying $(-\Delta + V)u = 0$.

If $V = 0$, then $-\Delta(1) = 0$, and in fact that's the only solution, since $-\Delta u = 0$ means $\langle -\Delta u, u\rangle = \int |\Delta u|^2 = 0$, so u is constant. So when can we solve $-\Delta u = f$? The dimensional of the null space is 1, so the range has codimensional 1. Then $(\operatorname{ran} L)^\perp = \ker \Delta =$ constant. Therefore, $-\Delta u = f$ is solvable if and only if $\int f = 0$.

Equivalently, if $u = \sum a_k e^{ik\theta}$ then $-\Delta u = \sum |k|^2 a_k e^{ik\theta} = \sum b_k e^{ik\theta} = f$ requires $a_k = \frac{b_k}{|k|^2}$, which means that $b_0 = 0$, so then $\int f = 0$. This is actually very concrete.

Suppose that $V > 0$. Then what happens to $(-\Delta + V)u = 0$? Expanding by Fourier series would be a mess, but we can write $\langle(-\Delta + V)u, u\rangle = \int |\Delta u|^2 + V|u|^2 = 0$ implies that

$u = 0$, so $\ker(-\Delta + V) = 0$, so hence $\operatorname{ran}(-\Delta + V) = L^2$, and hence $(-\Delta + V)u = f$ is always solvable!

Here's a second point of view. Think of $-\Delta + V : L^2 \to L^2$ as an *unbounded operator*. Then $C_0^\infty(T^n) \hookrightarrow L^2$. For all $u \in C_0^\infty$, we have $(-\Delta + V)u \in L^2$. We can now consider the graph $\{(u, Lu)\}$ for $u \in C_0^\infty$. This is definitely not the whole thing, and it isn't even closed. So consider

$$\overline{\{(u, Lu) : u \in C_0^\infty(T^n)\}} = \{(u, f) : u_j \in C_0^\infty, u_j \xrightarrow{L^2} u, Lu_j \xrightarrow{L^2} f\}.$$

Is this still a graph? Are there some points $(u, f_1), (u, f_2)$ in the closure of the graph, or alternatively, can there be $(0, f_1 - f_2)$ in this closure? Can we have $u_j \xrightarrow{L^2} 0$ and $Lu_j \xrightarrow{L^2} f \neq 0$? It turns out that this cannot happen, and this is still a graph.

Then

$$\{(u, f) : u_j \in C_0^\infty, u_j \xrightarrow{L^2} u, Lu_j \xrightarrow{L^2} f\} = \operatorname{Gr}(A)$$

for some A defined on $\mathcal{D}_A \subset L^2$. Then $C_0^\infty \subset \mathcal{D}_A \subset L^2$ is dense, and we call $(A, \mathcal{D}_A)$ a *closed extension* of L on C_0^∞, and we denote this $(L, \mathcal{D})$.

Now, we can consider $\mathbb{C} \setminus \operatorname{spec}(L, \mathcal{D}) = \{\lambda : (L - \lambda I)^{-1}$ exists as a bounded operator$\}$. Choose λ_0 so large that $-\Delta + V + \lambda_0$ where $V + \lambda_0 \geq 1$. What we showed earlier is that $(-\Delta + V + \lambda_0)^{-1} : L^2 \to L^2$ is bounded. Despite the operator being unbounded, the inverse is bounded. So then $-\lambda_0 \notin \operatorname{spec}(L)$.

It turns out that $L^* = L$ is still true, which we will do next time. Then $\operatorname{spec}(L, \mathcal{D}) \subset \mathbb{R}$, and there is nothing that is too negative. Instead consider $\operatorname{spec}((-\Delta + V + \lambda_0)^{-1}) = \{\mu_j\}$ is some positive sequence converging to zero. Then we have $(-\Delta + V + \lambda_0)^{-1}u_j = \mu_j u_j$. Then $(-\Delta + V)u_j = (-\lambda_0 + \frac{1}{\mu_j})u_j$. Then $\lambda_j = -\lambda_0 + \frac{1}{\mu_j} \to +\infty$, which gives us the spectral resolution of this unbounded operator via considering the spectral resolution of its inverse. We also know that the u_j form an orthonormal basis for L^2.

18. 3/13

Recall that we were talking about the spectral theorem. We did the case of compact self-adjoint operators, and we were doing the case of unbounded operators in the special case of $A = -\Delta + V$ for V real-valued and C^∞. We can talk about this on $T^n = S^1 \times \cdots \times S^1$ or on some smoothly bounded $\Omega \subset \mathbb{R}^n$. Our goal is to show that we have eigendata $\{\varphi_j, \lambda_j\}_{j=0}^\infty$ where $A\varphi_j = \lambda_j \varphi_j$ and $\lambda_j \to \infty$ without accumulation points.

Recall that we have $(A, \mathcal{D})$ where $\mathcal{D} \subset L^2$ is a dense subspace and $\operatorname{Gr}(A, \mathcal{D}) = \{(u, Au), u \in \mathcal{D}\}$ is a closed subspace of $L^2 \times L^2$. This means that A is a closed operator. Think of $\mathcal{D} \supset C_0^\infty(\Omega)$ or $C_0^\infty(T^n)$.

Remark. We cannot just work with the unbounded operator A; we need to work with A and its domain $\mathcal{D}$ together.

Definition 18.1. We wish to define $(A, \mathcal{D})^*$.

We have $v \in \mathcal{D}^*$ if $u \mapsto \langle Au, v \rangle$ for $u \in \mathcal{D}$ extends to an element of $(L^2)^*$. If $|\langle Au, v \rangle| \leq C \|u\|_{L^2(\Omega)}$ for all $u \in \mathcal{D}$ then we can extend to a continuous linear functional on L^2. Then $\langle Au, v \rangle = \langle u, A^* v \rangle$ by the Riesz representation theorem.

Example 18.2. $A = -\frac{\partial^2}{\partial x^2}$ on $C_0^\infty((0,1))$. The graph is not closed, so consider $\overline{\operatorname{Gr}(A)} = \overline{\{(u, -u'') : u \in C_0^\infty\}}$ in $L^2(I) \times L^(I)$.

We claim that $(u, f) \in \overline{\mathrm{Gr}(A)}$ if and only if $u \in H_0^2(I)$, which is the closure of C_0^∞ with respect to the H^2 norm $\|u\|_{L^2} + \|u'\|_{L^2} + \|u''\|_{L^2}$.

We have $(u, f) \in \overline{\mathrm{Gr}(A)}$ if there exist $u_j \in C_0^\infty$ such that $u_j \xrightarrow{L^2} u$ and $u_j'' \xrightarrow{L^2} f$. Then $u \in H_0^2$. Why? The u_j' satisfy $\int u_j' u_j' = -\int u_j u_j''$, so $\int |u_j'|^2 \le C \|u_j\| \left\|u_j''\right\|$.

Exercise 18.3. u_j' is also Cauchy in L^2.

If $u' \in L^2$, then

$$|u(x) - u(\tilde{x})| \le \left| \int_x^{\tilde{x}} u'(t)\, dt \right| \le \left(\int |u'(t)|^2 \right)^{1/2} \left(\int_x^{\tilde{x}} 1 \right)^{1/2} \le \|u'\|_{L^2} \sqrt{|x - \tilde{x}|}.$$

Now, this shows that $H_0^2 = \{u : u, u', u'' \in L^2, u(0) = u(1) = u'(0) = u'(1) = 0\}$ is the closure that we want.

What we've shown is that

$$\overline{\mathrm{Gr}\left(-\frac{\partial^2}{\partial x^2}, C_0^\infty \right)} = \left(-\frac{\partial^2}{\partial x^2}, H_0^2 \right).$$

We want the adjoint to satisfy $\langle Au, v \rangle = \langle u, A^* v \rangle$. Completely formally, we have

$$\int_0^1 -u'' v\, dx = \int_0^1 u(-v'')\, dx + (u(1)v'(1) - u'(1)v(1)) - (u(0)v'(0) - u'(0)v(0)).$$

We need these boundary terms to vanish, because $v \in \mathcal{D}^*$ if and only if $|\langle Au, v \rangle| \le C \|u\|_{L^2}$. We need $v'' \in L^2$. Also, when $u \in H_0^2$, all boundary terms are zero. This proves that

$$\left(-\frac{\partial^2}{\partial x^2}, H_0^2 \right)^* = \left(-\frac{\partial^2}{\partial x^2}, H^2 \right).$$

This means that $(-\frac{\partial^2}{\partial x^2}, H^2)$ is a bigger closed extension of $(-\frac{\partial^2}{\partial x^2}, C_0^\infty)$. Here, we write $(-\frac{\partial^2}{\partial x^2}, H_0^2) \subset (-\frac{\partial^2}{\partial x^2}, H^2)$.

Definition 18.4. $(A, \mathcal{D})$ is called self-adjoint if $(A, \mathcal{D})^* = (A, \mathcal{D})$.

Proposition 18.5. $(-\frac{\partial^2}{\partial x^2}, H^2 \cap H_0^1)$ *is self-adjoint.*

Here, $H^2 \cap H_0^1 = \{u \in H^2, u(0) = u(1) = 0\} \supset C_0^\infty$.

Proof. Take $u \in H^2 \cap H_0^1$. Then

$$\langle -u'', v \rangle = \langle u, -v'' \rangle + (u(1)v'(1) - u'(1)v(1)) - (u(0)v'(0) - u'(0)v(0)).$$

$v \in \mathcal{D}^*$ means that $|\langle -u'', v \rangle| \le C \|u\|_{L^2}$ can only happen if there are no boundary terms. So v is in the adjoint domain if and only if $v \in H^2$ and $v(0) = v(1) = 0$, i.e. $v \in H^2 \cap H_0^1$. □

So now we have $(-\frac{\partial^2}{\partial x^2}, C_0^\infty) \subset (-\frac{\partial^2}{\partial x^2}, H_0^2) \subset (-\frac{\partial^2}{\partial x^2}, H^2 \cap H_0^1) \subset (-\frac{\partial^2}{\partial x^2}, H^2)$. Somehow we wanted to impose half of the boundary conditions. Picking different boundary conditions give different extensions.

How do we connect this to spectral theory?

Proposition 18.6. *Consider* $A = -\frac{\partial^2}{\partial x^2} + V(x)$ *on* $L^2((0,1))$*, and consider* $(A, H^2 \cap H_0^1)$ *as a self-adjoint operator. Then there is an orthonormal set of eigenfunctions* $\{\varphi_j\}$ *where* $(-\frac{\partial^2}{\partial x^2} + V)\varphi_j = \lambda_j \varphi_j$ *with* $\varphi_j(0) = \varphi_j(1) = 0$ *and* $\lambda_j \to \infty$.

Proof. Shift the operator so that $-\frac{\partial^2}{\partial x^2} + V(X) + C > 0$. Then $\langle (A+C)u, u\rangle \geq \|u\|^2$ for all $u \in \mathcal{D}$.

Define $(-\frac{\partial^2}{\partial x^2} + V(x) + C)^{-1}$. For $(A+C)u = f \in L^2$, we can find $u \in \mathcal{D}$. The Lax-Milgram lemma 4.2 was a good way to think about this. This tells us that there exists $u \in H_0^1$ such that $\int vf = B(u,v)$ for all $v \in H_0^1$. This means that

$$\int \nabla u \nabla v + (V+C)uv = \int fv,$$

so for all $u \in H_0^1$ we have $u \in H^2 \cap H_0^1$.

So given f we can find $u = (-\frac{\partial^2}{\partial x^2} + V + C)^{-1} f \in H^2 \cap H_0^1$. Since $K = (-\frac{\partial^2}{\partial x^2} + V + C)^{-1}$ is compact on L^2. We have $K : L^2 \to H^2 \cap H_0^1 \hookrightarrow L^2$ where the second map is a compact inclusion by Arzela-Ascoli. That is, for u_j with $\int |u_j|^2 + |u_j'|^2 + |u_j''|^2 \leq C$ we have

$$|u_j(x) - u_j(\tilde{x})| \leq C\sqrt{|x - \tilde{x}|}$$

which gives uniform equicontinuity. Also, $|u_j(x) - 0| \leq C\sqrt{x} \leq C$ for all j, giving uniform boundedness.

Therefore, K has orthonormal eigenfunctions and discrete spectrum, i.e. $K\varphi_j = \mu_j \varphi_j$ with $\mu_j > 0$ since we have a positive operator, and $(-\frac{\partial^2}{\partial x^2} + V + C)\varphi_j = 1/\mu_j \varphi_j$, so therefore $A\varphi_j = (-C + \frac{1}{\mu_j})\varphi_j$. Setting $\lambda_j = -C + \frac{1}{\mu_j}$ we see that we have $\lambda_j \to \infty$. □

Remark. Similarly, we can do this for $-\Delta + V$ on T^n or on $\Omega \subset \mathbb{R}^n$ to get a complete basis of eigenfunctions.

Now, we go back to bounded self-adjoint operators, where there is no unpleasantness with adjoint domains. Take $A \in \mathcal{B}(\mathcal{H})$ with $A^* = A$. In fact, any closed subset can be a spectrum.

First, we discuss *functional calculus*. Given an operator A, we can compute A^n and hence take polynomials $p(A) = \sum_{k=0}^{N} c_k A^k \in \mathcal{B}(\mathcal{H})$. In fact, if we have $f \in C^0(\text{spec}(A))$, we can define $f(A)$. If $A\varphi = \lambda\varphi$ then $p(A)\varphi = (\sum c_k A^k)\varphi = p(\lambda)\varphi$. Here, the eigenvectors are the same but the eigenvalues have changed by p. This is what we want for continuous functions.

Theorem 18.7. *There is a unique map* $\varphi : C^0(\text{spec}(A)) \to \mathcal{B}(\mathcal{H})$ *such that*

(1) $\varphi(fg) = \varphi(f)\varphi(g)$, $\varphi(\lambda f) = \lambda\varphi(f)$, $\varphi(1) = I$, $\varphi(\overline{f}) = \varphi(f)^*$.
(2) $\|\varphi(f)\|_{\mathcal{B}(\mathcal{H})} \leq C\|f\|_{L^\infty}$.
(3) $\varphi(x) = A$.
(4) $A\psi = \lambda\psi$ *implies that* $\varphi(f)(\psi) = f(\lambda)\psi$.
(5) $\text{spec}(\varphi(f)) = f(\text{spec}(A))$.
(6) $f \geq 0$ *implies that* $\varphi(f) \geq 0$ *(i.e.* $\langle \varphi(f)u, u\rangle \geq 0$ *for all* $u \in \mathcal{H}$*).*
(7) $\|\varphi(f)\|_{\mathcal{B}(\mathcal{H})} \leq \|f\|_{L^\infty}$.

We will actually prove (v) and (vii) and extend by continuity.

Lemma 18.8. $\text{spec}(P(A)) = P(\text{spec}(A))$.

Proof. If $\lambda_0 \in \text{spec}(A)$, then $p(x) - p(\lambda_0)$ has root $x = \lambda_0$, so $p(x) - p(\lambda_0) = (x - \lambda_0)q(x)$. Then $p(A) - p(\lambda_0)I = (A - \lambda_0 I)q(A)$, so therefore $p(A) - p(\lambda_0)I$ is not invertible, so $p(\lambda_0) \in \text{spec}(p(A))$.

Conversely, if $\mu \in \text{spec}(p(A))$. Then factor the polynomial $p(x) - \mu = a(x - \lambda_1)\cdots(x - \lambda_n)$. Now, $p(A) - \mu I = a(A - \lambda_1 I)\cdots(A - \lambda_n I)$, which implies that some $A - \lambda_j I$ is not invertible, so therefore $p(\lambda_j) = \mu$ and hence $\lambda_j \in \text{spec}(A)$. □

Lemma 18.9. $\|p(A)\| = \sup_{\lambda\in\operatorname{spec}(A)} |p(\lambda)| = \|p\|_\infty$.

From this, we can claim that the properties above are all preserved under operator limits, so that they are true for continuous f.

Next, given f, we can get $\langle f(A)x, y\rangle = \int f(\lambda) d\mu^A_{x,y}$ by the Riesz-Markov theorem. So we'll be able to talk about the spectral measure.

19. 3/15

Recall that we are working with $A \in \mathcal{B}(H)$ where H is a separable Hilbert space, and $A^* = A$.

The first goal was to develop the continuous functional calculus. Suppose we have any $f \in C_0^\infty(\operatorname{spec} A)$ and $f(A) \in \mathcal{B}(H)$. Using Weierstrass approximation, choose a sequence of polynomials so that $p_n(\lambda) \to f(\lambda)$ in $C^0(\operatorname{spec} A)$.

We proved two lemmas last time:

Lemma 19.1. *If $P(\lambda)$ is a polynomial then* $\operatorname{spec} P(A) = P(\operatorname{spec} A)$.

Lemma 19.2. $\|P(A)\|_{\mathcal{B}(H)} = \sup_{\lambda\in\operatorname{spec} A} |P(\lambda)| = \|P\|_\infty$.

Proof. First recall that $\|B^*B\| = \|B\|^2$. This is because $\|B^*B\| \le \|B^*\|\,\|B\| = \|B\|^2$, and on the other hand, $\|B\|^2 = \sup_{\|x\|=1} \|Bx\|^2 = \sup_{\|x\|=1} |\langle B^*Bx, x\rangle| \le \|B^*B\|$.

$\|P(A)\|^2 = \|P(A)^*P(A)\| = \left\|(\overline{P}P)(A)\right\| = \sup_{\lambda\in\operatorname{spec}\overline{P}P(A)} |\lambda| = \sup_{\lambda\in\operatorname{spec} A} |\overline{P}P(\lambda)| = \sup_{\lambda\in\operatorname{spec} A} |P(\lambda)|^2$. ☐

This allows us to carry out our scheme.

Theorem 19.3. *$\varphi : C^0(\operatorname{spec} A) \to \mathcal{B}(H)$ via $f \mapsto f(A)$. Here, $fg \mapsto f(A)g(A)$ and $\overline{f} \mapsto f(A)^*$. If $A\psi = \lambda\psi$ then $f(A)\psi = f(\lambda)\psi$.*

One corollary of this whole construction is the following:

Corollary 19.4. *$(A - \lambda I)^{-1} = R_A(\lambda)$ is defined when $\lambda \notin \operatorname{spec}(A)$. Then $\|R_A(\lambda)\| = 1/\operatorname{dist}(\lambda, \operatorname{spec}(A))$.*

This means that given $(A - \lambda I)^{-1}$ for $\lambda \approx \lambda_0$, then this blows up at most like $1/(\lambda - \lambda_0)$. We see that $\lambda \to \langle R_A(\lambda)x, y\rangle$ is holomorphic, so we can expand it in a Laurent series to see that it has at worst a simple pole, with $R_A(\lambda) = \frac{A_{-1}}{\lambda-\lambda_0} + \tilde{A}(\lambda)$.

What is A_{-1}? Note that $(A - \lambda I)R_A(\lambda) = I$. Expanding this gives

$$(A - \lambda I)R_A(\lambda) = ((A - \lambda_0 I) - (\lambda - \lambda_0)I)\left(\frac{A_{-1}}{\lambda - \lambda_0} + \tilde{A}(\lambda)\right) = \frac{(A - \lambda_0 I)A_{-1}}{\lambda - \lambda_0} + \widehat{A}(\lambda) = I$$

where $\widehat{A}$ is regular. Then $(A - \lambda_0 I)A_{-1} = 0$. So if λ_0 is isolated in $\operatorname{spec}(A)$ (and $A^* = A$), then $A_{-1} = -\Pi_{\lambda_0}$ is a projector onto an eigenspace $\{x : Ax = \lambda_0 x\}$. Now, checking that $A_{-1}A_{-1} = A_{-1}$ and $A_{-1}^* = A_{-1}$ yields what we want.

What goes wrong if we lose self-adjointness? Consider $A = \begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}$. Then

$$R_A(\lambda) = \begin{pmatrix} -1/\lambda & 1/\lambda^2 \\ 0 & -1/\lambda \end{pmatrix}.$$

This is a feature of having Jordan blocks that are not diagonal.

In general, the spectrum might not have isolated points and it could just be a mess. What do we do then?

Pick any $x \in H$. Then $C^0(\operatorname{spec} A) \ni f \mapsto \langle f(A)x, x\rangle$ is a continuous linear functional. This means that this has to be representable by a measure $\langle f(A)x, x\rangle = \int_{\operatorname{spec} A} f(\lambda) d\mu_x(\lambda)$. In general, we don't have a good way to identify the measure, however. Let's think about a couple of special cases.

Suppose $A^* = A > 0$ is compact. This means that we have a basis $A\psi_n = \lambda_n \psi_n$ where $\{\psi_n\}$ form an orthonormal basis for $\mathcal{H}$ and $\lambda_1 \geq \lambda_2 \geq \cdots \to 0$. For $x = \psi_n$, we have $\langle f(A)x, x\rangle = f(\lambda_n) \|x\|^2 = \int f(\lambda) d\mu_x = \left\langle \delta(\lambda - \lambda_n) \|x\|^2, f \right\rangle$, so here the measure is just some δ-measure at that point.

For a general x, we have $x = \sum x_n \varphi_n$. Then

$$\langle f(A)x, x\rangle = \sum f(\lambda_n) |x_n|^2 = \left\langle \sum_n \delta(\lambda - \lambda_n) \left|\langle x, \varphi_n\rangle\right|^2, f \right\rangle,$$

and this sum of delta-measures is $d\mu_x(\lambda)$.

There are different ways to think about $d\mu_x$. Fix x, and define $H_x = \overline{\operatorname{span}\{A^n x\}} \subset H$.

Definition 19.5. x is called cyclic in H' if the $A^n x$ span a dense subspace in H'.

Lemma 19.6. *Clearly,* $A : H_x \to H_x$*.*

There exists a unitary mapping $U : H_x \to L^2(\operatorname{spec} A, d\mu_x)$ *such that* UAU^{-1} *is multiplication by* λ*:*

$$\begin{array}{ccc} H_x & \xrightarrow{A} & H_x \\ {\scriptstyle U}\big\downarrow & & \big\downarrow{\scriptstyle U} \\ L^2 & \xrightarrow[M_\lambda]{} & L^2 \end{array}$$

Proof. Take any $f(A)x \in H_x$. Then U sends $f(A)x$ to $f(\lambda)$.

Then $\|f(A)x\|^2 = \langle f(A)^* f(A)x, x\rangle = \int_{\operatorname{spec} A} |f(\lambda)|^2 \, d\mu_x$.

This defines a mapping $H_x \to C^0(\operatorname{spec} A)$ that intertwines $\|\cdot\|_{H_x}$ and $\|\cdot\|_{L^2(\operatorname{spec} A, d\mu_x)}$. Finally, use that $C^0 \hookrightarrow L^2$ is dense. $\square$

This means that $A|_{H_x}$ is unitarily equivalent to multiplication by κ on $L^2(\operatorname{spec} A, d\mu_x)$. This is a very simple and concrete operator. This is now like the spectral theorem. We have $A(U^{-1}f) = U^{-1}(\lambda f)$.

We can write

$$H = \bigoplus_{n=1}^{N} H_{x_j}$$

where $N \leq \infty$. So we can write the entire separable Hilbert space as a direct sum of these special subspaces, where $A|_{H_{x_j}}$ looks like multiplication by λ.

In general, if A corresponds to M_λ then $f(A)$ corresponds to $M_{f(\lambda)}$. When does $M_{f(\lambda)}$ exist as a bounded operator on $L^2(\operatorname{spec} A, d\mu_x)$? This is still ok if f is a Borel function in L^∞. Hence, for such functions, $f(A)$ makes sense.

What are interesting characteristic functions? Here's the upshot: Let $\Omega \subset \operatorname{spec} A$ be any Borel subset. We've defined $\chi_\Omega(A)$, and we have some properties: $\chi_\Omega(A) \circ \chi_\Omega(A) = \chi_\Omega(A)$

and $\chi_\Omega(A)^* = \chi_\Omega(A)$. This means that $\chi_\Omega(A) = P_\Omega$ is an orthogonal projector on H. If we have any bounded operator so that $P^2 = P$ and $P^* = P$, then we have $P(I - P) = 0$, so $\operatorname{ran}(P) = \ker(I - P)$, so this P is just projecting onto a closed subspace.

Then P_Ω is the orthogonal projector onto the "sum of the eigenspaces with eigenvalues in Ω". This is exactly correct if we have discrete spectrum. Then $P_\Omega = \Pi_{\varphi_n} + \Pi_{\varphi_{n+1}}$ and $P_\Omega x = \langle x, \varphi_n\rangle \varphi_n + \langle x, \varphi_{n+1}\rangle \varphi_{n+1}$.

In general, suppose that we have the spectrum that decomposes into Ω_1 and Ω_2 so that $\operatorname{spec}(A) = \Omega_1 \cup \Omega_2$ and $\Omega_1 \cap \Omega_2 = 0$. and we have P_{Ω_1} and P_{W_2}. Then $P_{\Omega_1} P_{\Omega_2} = 0$ and $P_{\Omega_1} + P_{\Omega_2} = I$. So therefore $\operatorname{ran} P_{\Omega_1}$ and $\operatorname{ran} P_{\Omega_2}$ are orthogonal subspaces that fill out H.

So we have this family of projectors P_Ω satisfying some characteristic properties:

(1) $P_\emptyset = 0$
(2) $P_{(-a,a)} = I$
(3) If $\Omega = \bigcup_{n=1}^\infty \Omega_n$ where $\Omega_n \cap \Omega_m = \emptyset$ then $P_\Omega = \lim \sum_{n=1}^N P_{\Omega_n}$ where the limit is a strong limit, i.e. $\sum_{n=1}^N P_{\Omega_n} x \to P_\Omega x \in H$, and not an operator norm limit; operator norm corresponds to L^∞ and this is an L^2 statement.
(4) $P_{\Omega_1 \cap \Omega_2} = P_{\Omega_1} P_{\Omega_2}$.

This looks like properties of an ordinary measure.

Given any family $\{P_\Omega\}$, think of $\Omega \to \langle P_\Omega x, x\rangle$. For any f Borel, we can look at $\int f(\lambda) \langle dP_\lambda x, x\rangle = \langle f(A)x, x\rangle$.

Finally, we can think of $f(A) = \int f(\lambda) dP_\lambda$. Let's think about what this means in cases that we understand.

If $f = 1$ then $\int 1 \, dP_\lambda = I$. Here $P_\lambda = P_{(-\infty,\lambda]}$. In general, it doesn't make sense to project onto a certain λ because the spectrum can be messy and there might not be an eigenspace there. But it makes sense to project onto everything $\leq \lambda$, and take the sum of those eigenspaces.

Example 19.7. If A is compact, $A \geq 0$ and $A^* = A$. If $\lambda > 0$, we have an infinite rank projector. This is a nice example.

Example 19.8. This is the motivation for all of this. We'll have to go back to unbounded operators. Consider $-\Delta$ on $L^2(\mathbb{R}^n)$. We can still talk about $\operatorname{spec}(-\Delta) = [0, \infty)$. We have some unitary transformation

$$\begin{array}{ccc} L^2(\mathbb{R}^n) & \xrightarrow{-\Delta} & L^2(\mathbb{R}^n) \\ \downarrow \mathcal{F} & & \downarrow \mathcal{F} \\ L^2(\mathbb{R}^n) & \longrightarrow & L^2(\mathbb{R}^n) \end{array}$$

with $\mathcal{F} \circ (-\Delta) \circ \mathcal{F} = M_{|\xi|^2}$ and $(-\Delta f)^\wedge = |\xi|^2 \widehat{f}(\xi)$. Then $\operatorname{spec}(-\Delta) = \operatorname{spec}(-M_{|\xi|^2})$. Then $(M_{|\xi|^2} - \lambda I)^{-1} = M_{(|\xi|^2 - \lambda)^{-1}}$ makes sense if and only if $\lambda \in \mathbb{C} \setminus \overline{\mathbb{R}^+}$.

Now, we have

$$f(x) = (2\pi)^{-n} \int e^{ix\cdot\xi} \widehat{f}(\xi) \, d\xi.$$

The mysterious thing is that $e^{ix\cdot\xi} \notin L^2$, but $-\Delta_x(e^{ix\cdot\xi}) = |\xi|^2 e^{ix\cdot\xi}$. So we have a perfectly nice set of eigenvectors, but they don't lie in our space.

Now, consider $\{\xi : |\xi|^2 \leq \lambda\}$. The spectral projector here is then

$$\int_{|\xi|\leq\sqrt{\lambda}} e^{ix\cdot\xi}\, d\xi \in L^2.$$

This is the projector onto $\{f : \widehat{f}(\xi) = 0, |\xi| > \sqrt{\lambda}\}$.

So in other words, we can take any function and project it onto $f \to P_{(-\infty,\lambda]}f \xrightarrow{L^2} f$ for $\lambda \to +\infty$. This is the Fourier inversion formula. We have

$$\int e^{ix\cdot\xi}\chi_{(-\infty,\sqrt{\lambda}]}(\xi)\widehat{f}(\xi)\, d\xi \to f$$

as $\lambda \to \infty$.

We have this family of measures $d\mu_x$. This decomposes into

(1) an atomic part (supported at points)
(2) absolutely continuous part (with respect to Lebesgue measure)
(3) singular continuous part (invisible with respect to Lebesgue measure).

Then we can write $H = H_{pp} \oplus H_{ac} \oplus H_{sc}$. In the first piece, this is most familiar, where $A|_{H_{pp}}$ has only eigenvalues and eigenvectors in the usual sense. The absolutely continuous part morally looks like what we have for the Laplacian. The final piece is just weird.

This was all developed for quantum mechanics, for studying $L = -\Delta + V$ acting on $L^2(\mathbb{R})$ where V is a real-valued potential function. In general, we get some complicated picture with the spectrum, and mathematical physics studies this.

E-mail address: `moorxu@stanford.edu`