Math 624 Lecture Notes

Taught by Robin Young, Notes by Patrick Lei

University of Massachusetts, Amherst

Spring 2019

Abstract

Continuation of Math 623. Introduction to functional analysis; elementary theory of Hilbert and Banach spaces; functional analytic properties of Lp-spaces, applications to Fourier series and integrals; interplay between topology, and measure, Stone-Weierstrass theorem, Riesz representation theorem. Further topics depending on instructor.

Contents

1   Organizational

A webpage (`http://people.math.umass.edu/~young/analysis/index.html`) has been put up with one topic per week, which is impractical. We will aim to get through ten topics. Formal office hours will not be scheduled because this is the only class Robin is teaching. The understanding is that we will visit his office. At some point during lecture, the electrical engineering student walked into the room anad had the following conversation with Robin:

Robin Are you Bo?
Bo Yes
Robin Everyone, this is Bo. Bo, this is everyone.
Bo I am in the electrical engineering department. This will help me with my research.

2   Review of Last Semester

2.1   Lecture 1 (Jan 22) We will review measure theory from an abstract perspective and treat the Lebesgue measure as an example.[1] Let X be a locally compact topological space.

Definition 1 (Sigma-Algebra)**.** A σ-algebra Σ to be a collection of subsets closed under complement and countable union.

---

[1]It happens to the unique Haar measure on $\mathbb{R}^n$

We denote by $\mathcal{B}$ the Borel σ-algebra, which is the smallest such containing all open sets. μ is a (non-negative) measure if μ is defined on Σ, the empty set has measure 0, and μ is countably additive.

Remark 2 (Notation). Indexing with a Roman letter means countable and indexing with a Greek letter means uncountable.

Definition 3 (Sigma-finite). A measure μ is σ-finite if $\mu(X) = \sum_{i=1}^{\infty} \mu(E_i)$ where each E_i has finite measure.

Definition 4 (Completeness). A measure space (X, Σ, μ) is complete if all subsets of measure zero sets are measurable: $\mu(B) = 0$ and $A \subset B$ implies $A \in \Sigma$.

Definition 5 (Measurable function). A function $f : X \to R$ is measurable if $f^{-1}(\alpha, \infty)$ is measurable for all real α.

Given the notion of a measurable function, we can integrate:

Definition 6 (Simple Function). A simple function is a finite linear combination of indicator functions

$$\varphi = \sum_{i=1}^{k} a_i \chi_{E_i}.$$

We define the integral of a simple function $\varphi = \sum_{i=1}^{k} a_i \chi_{E_i}$ to be

$$\int \varphi \; \mathrm{d}\mu = \sum_{i=1}^{k} a_i \mu(E_i).$$

Given a nonnegative measurable function f, we define the integral

$$\int f \; \mathrm{d}\mu = \sup \left\{ \int \varphi \; \mathrm{d}\mu \right\},$$

where φ ranges over all nonnegative simple functions less than or equal to f almost everywhere.

We also set, for a measurable set E,

$$\int_E f \; \mathrm{d}\mu = \int f \chi_E \; \mathrm{d}\mu.$$

Example 7. Given the trivial sigma-algebra, only constant functions are simple.

Lemma 8. *Any measurable f can be approximated from below by $\phi_k \to f$ with $\int \varphi_k \; \mathrm{d}\mu \to \infty$.*

We observe that Fatou's Lemma, monotone convergence, and dominated convergence hold.

Lemma 9 (Convergence theorems). *If $\mu_n \to \mu$ setwise on Σ and if $f_n \to f$ almost everywhere (with respect to μ) on X, then*

$$\int f_n \; \mathrm{d}\mu \to \int f \; \mathrm{d}\mu.$$

The ability to exchange limits and integrals is the motivation behind measure theory, and more broadly, analysis is about exchanging limits.

Definition 10 (Mutually Singular). Two measures μ, ν are mutually singular, written $\mu \perp \nu$ if there exist disjoint A, B such that $X = A \bigcup B$ such that $\mu(B) = \nu(A) = 0$.

Example 11 (Cantor). Consider the Cantor set C and Cantor function f_C. Recall that the Cantor function is a continuous surjection from $[0, 1]$ to itself and has derivative zero almost everywhere. In particular, f_C violates the fundamental theorem of calculus. Define the Cantor measure on the Borel sets in $[0, 1]$ the usual way by $\mu_C((a, b]) = f_C(b) - f_C(a)$. We see that $\mu_C(C) = 1$, but $\lambda(C) = C$ where λ is the Lebesgue measure. Therefore, $\mu_C \perp \lambda$ on $[0, 1]$.

Definition 12 (Absolute Continuity). Given two measure μ, ν, ν is absolutely continuous ($\nu \ll \mu$) with respect to μ if $\mu(A) = 0$ implies that $\nu(A) = 0$.

Example 13. Given any nonnegative integrable f with respect to μ, we can define a measure $\mu_f(E) = \int_E f \, d\mu$. Then $\mu_f \ll \mu$.

Theorem 14 (Radon-Nikodym). *If $\nu \ll \mu$ on a σ-finite X (with respect to μ), then there exists a non-negative measurable f such that $\nu = \mu_f$. We write $f = \frac{d\nu}{d\mu}$.*

Example 15. Let f be monotone increasing. Recall that left and right limits exist everywhere. Then observe that f has at most countably many discontinuities x_j and

$$\sum_{j=1}^{\infty} f(x_j+) - f(x_j-) < \infty$$

for x_j restricted to a compact set. We can write $f = f_j + f_c$ where f_j encodes the jumps and f_c is continuous. Then f_c decomposes into $f_c = f_a + f_C$, where f_a is absolutely continuous with respect to the Lebesgue measure, and f_C is the "Cantor" part. To extract f_a we will consider the points at which f is differentiable. In the end, we will have $f = f_a + f_C + f_j$, where f_a is absolutely continuous, f_C is "Cantor," and f_j denotes the jumps. The three induced measures will be mutually singular.

2.2　Lecture 2 (Jan 24)

2.2.1　Inanities Robin came in late and we had a conversation about Hagoromo Fulltouch chalk and chalk holders. Apparently those are like mechanical pencils. Robin would like to continue with the review and do a calculation.

2.2.2　Review Continued Recall that we have a monotone increasing function F, which we will assume is right-continuous. Recall that we can decompose $F = F_j + F_a + F_C$ as we did yesterday. There are several questions:

1. How do we find this decomposition?
2. What do we do with it?

To find F_j, the jump part, we can locate jumps as the discontinuities. Let J be the set of jumps. Then

$$F_j = \sum_{x_i \in J} \left(F(x_i+) - F(x_i-) \right) \cdot H(x - x_i),$$

where H is the Heaviside function. Then we write the measure

$$dF_j = \sum_i \left(F(x_i+) - F(x_i-) \right) \cdot \delta_{x_i},$$

where δ is the Dirac mass.

Now to find F_a, we differentiate F. Recall that we can compute the Dini derivatives everywhere. We also proved in the fall that the set of points of nondifferentiability has measure zero with respect to Lebesgue. Then, we can integrate F' with respect to λ to get

$$F_a = \int F' \, \mathrm{d}x.$$

We know that $F(b) - F(a) \geq \int_a^b F'(x) dx$, so F_a is the absolutely continuous part of F. Finally, the "Cantor" part F_C is the remainder $F_C = F - F_j - F_a$.

The bottom line is that it is hard to locate the "Cantor" part of a function.

Example 16. Suppose $F(x) = \begin{cases} x & x < 0 \\ 1 & 0 \leq x < 2 \\ x^2 & x \geq 2 \end{cases}$ with jumpts at $0, 2$. Then we see that $F_j(x) = H_0(x) + 3H_2(x)$ We can differentiate to find $F'(x) = \begin{cases} 1 & x < 0 \\ 0 & 0 \leq x < 2 \\ 2x & x \geq 2 \end{cases}$. Then, integrating, we find

that F has no "Cantor" part. Now we see that the Lebesgue-Stieltjes measure is $\mathrm{d}\mu = F' \, \mathrm{d}\lambda + \delta_0(x) + 3\delta_2(x)$ and integrating some function $g(x)$ is something like

$$\int_{[-4,7]} g(x) \, \mathrm{d}\mu = \int_{-4}^0 g(x) \, \mathrm{d}x + g(0) + \int_0^2 0 + 3g(2) + \int_2^7 g(x) \cdot 2x \, \mathrm{d}x.$$

Conversely, given a measure μ, we can define the corresponding function F satisfying $\mu(a, b] = F(b) - F(a)$. Define F by $F(y) = F(0) + \mu(0, y]$ for positive y and similarly for negative y

Recall again the Radon-Nikodym Theorem:

Theorem 17 (Radon-Nikodym)**.** *Let (X, Σ, μ) be a σ-finite measure space and ν another σ-finite signed measure on Σ. Then there exist unique measures ν_s, ν_a, where $\nu_s \perp \mu, \nu_a \ll \mu$, such that $\nu = \nu_s + \nu_a$. Moreover, there exists f_a integrable such that $\nu_a(E) = \int_E f_a \, \mathrm{d}\mu$ for all measurable E. We write $f_a = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ the Radon-Nikodym derivative.*

Now recall signed measures:

Definition 18 (Signed Measure)**.** A signed measure is a countably additive function on Σ that vanishes on the empty set and does not take on both positive and negative infinity.

Lemma 19. *There exist $A, B \subset X$ such that $\nu(E) \geq 0$ for all $E \subset A$ and $\nu(E) \leq 0$ for all $E \subset B$.*

We can set $\nu_+(E) = \nu(E \cap A)$ and $\nu_-(E) = -\nu(E \cap B)$. Then we have $\nu(E) = \nu_+(E) - \nu_-(E)$ and $|\nu(E)| = \nu_+(E) - \nu_-(E)$. In addition, the decomposition $\nu = \nu_+ - \nu_-$ is unique. Similarly, we can define complex measures. Then $|\nu|$ is the "total variation" of ν.

3   PRODUCT MEASURES AND FUBINI

3.1   Lecture 2 continued Our first issue: are products of measurable sets measurable? What is the correct product σ-algebra? We will simplify the structure.

Definition 20 (Semi-algebra)**.** A semi-algebra $\mathcal{C}$ satisfies: if $A, B \in \mathcal{C}$, then $A \cap B \in \mathcal{C}$ and A^C can be written as a finite disjoint union $A^c = \sqcup_{i=1}^{n} B_i$, where $B_i \in \mathcal{C}$.

This is very convenient for Cartesian products because $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$ and $(A \times B)^C = (A \times B^C) \sqcup (A^C \times B) \sqcup (A^C \times B^C)$.

Generate an algebra $\mathcal{A}$ from our semi-algebra $\mathcal{C}$ by $\varnothing$ plus finite unions of elements of A. Now suppose we have an additive set function μ defined on $\mathcal{C}$, this extends to a measure on A provided that:

1. If $C \in \mathcal{C}$ is a finite union of disjoint union of elements of $\mathcal{C}$ $C = \sqcup_{i=1}^{k} B_i$, then $\mu(C) = \sum_{i=1}^{k} \mu(B_i)$.
2. μ is countably subadditive on $\mathcal{C}$.

Given (X, Σ_x, μ) and (Y, Σ_y, ν), we develop the product measure $\mu \times \nu$ on $X \times Y$. First let $\mathcal{C} = \Sigma_x \times \Sigma_y$. Obviously, $\mathcal{C}$ is a semialgebra, so it extends to an algebra $\mathcal{A}$. Now define $\mu \times \nu$ on $\mathcal{C}$ by $\mu \times \nu(A \times B) = \mu(A)\nu(B)$.

Lemma 21. *$\mu \times \nu$ satisfies the above two conditions.*

3.2   Lecture 3 (1/29)

3.2.1   An Apology Robin would like to offer an apology. He realizes that this is the first time he's teaching the class, so we are all guinea pige here. He has come to realize that Stein's book is not very good. It's boring, encyclopedic, and not very well arranged. It has too much measure theory, and he apologizes to us for spending too much time in the fall on measure and integration, which is really boring. After we summarize Fubini in a general setting, we will take measure theory for granted and move on to something more interesting.

Robin says that the cardinal rule of teaching is to never end on time.

3.2.2   Product Measures/Fubini Continued Recall the definition of a semi-algebra from last time.

We restate the lemma from last time:

Lemma 22. *Suppose we have an additive set function μ defined on $\mathcal{C}$, this extends uniquely to a measure on A provided that:*

1. *If $C \in \mathcal{C}$ is a finite union of disjoint union of elements of $\mathcal{C}$ $C = \sqcup_{i=1}^{k} B_i$, then $\mu(C) = \sum_{i=1}^{k} \mu(B_i)$.*
2. *μ is countably subadditive on $\mathcal{C}$.*

Proof. If $C = \sqcup C_i = \sqcup D_i$, write each C_i as $C_i = \sqcup(C_i \cap D_j)$, so $\mu(C_i) = \sum \mu(C_i \cap D_j)$. For the second part, we show that $\mu(\sqcup C_i) = \sum \mu(C_i)$. We see that $\sqcup_{i=1}^{n} C_i \subset C = \sqcup C_i$ for all n. Then we see that

$$\sum_{i=1}^{n} \mu(C_i) = \mu(\sqcup_{i=1}^{n} C_i) \leq \mu(C),$$

so letting $n \to \infty$, we have our desired result. $\square$

Now we return to our product construction from last time.

Lemma 23. *$\lambda = \mu \times \nu$ satisfies the above two conditions.*

Proof. First we show that if $\{A_i \times B_i\}$ is a countable disjoint collection with $A \times B = \bigcup A_i \times B_i$, then $\lambda(A \times B) = \sum \lambda(A_i \times B_i)$. To see this, fix $x \in A$. For all $y \in B$, we see that (x, y) is in only one $A_i \times B_i$, so $B = \bigsqcup_{\{i \mid x \in A_i\}} B_i$, so as a function on X, $\nu(B)\chi_A(x) = \sum \nu(B_i)\chi_{A_i}(x)$ by countable additivity. Integrating with respect to μ and using monotone convergence, we have

$$\sum_i \int \nu(B_i)\chi_{A_i}(x) \, \mathrm{d}\mu = \int \nu(B)\chi_A(x) \, \mathrm{d}\mu,$$

and therefore $\sum_i \nu(B_i)\mu(A_i) = \nu(B)\mu(A)$. $\qquad \square$

However, to achieve our goal, we need to extend this algebra $\mathcal{R}'$ to a σ-algebra so we can take limits. In order to do this, we need to perform Caratheodory's construction and use outer measures:

$$\lambda_*(E) = \inf \left\{ \sum_{i=1}^n \lambda(R_i) \; \middle| \; E \subset \bigcup_{i=1}^n R_i \right\}.$$

Then we declare E to be measurable if $\lambda_*(A) = \lambda_*(A \cap E) + \lambda_*(A \cap E_C)$ for all A. Then conclude that there is a complete σ-algebra. $\mathcal{S} \supset \mathcal{R}'$ on which the product measure is defined.

Now we integrate with respect to $\mu \times \nu$. In order to prove Fubini, as before we denote the slice $E_x = \{y \mid (x, y) \in E\}$.

Lemma 24. *E_x is measurable if $E \in \mathcal{R}_{\sigma\delta}$, where $\mathcal{R} = \mathcal{R}'$, $\mathcal{R}_\sigma = \{\bigcup_{i=1}^\infty R_i \mid R_i \in \mathcal{R}\}$ and $\mathcal{R}_{\sigma\delta} = \{\bigcap_{j=1}^\infty R_j \mid R_j \in \mathcal{R}_\sigma\}$. In addition, $\mathcal{R}_{\sigma\delta}$ is a σ-algebra containing $\mathcal{R}'$, and $\mu \times \nu$ can be defined on the completion of $\mathcal{R}_{\sigma\delta}$.*

Lemma 25. *If $E \in \mathcal{R}_{\sigma\delta}$ and $\mu \times \nu(E)$ is finite, then $g : X \to \mathbb{R}$ defined by $g(x) = \nu(E_x)$ is measurable and $\int g \, \mathrm{d}\mu = \mu \times \nu(E)$.*

Lemma 26. *If $\mu \times \nu(E) = 0$, then for almost every x, $\nu(E_x) = 0$.*

Theorem 27. *Let E be a measurable subset of $X \times Y$ with $\mu \times \nu(E)$ finite. Then for almost every x, E_x is measurable and $g(x) = \nu(E_x)$ is a measurable function and $\int g(x) \, \mathrm{d}\mu = \mu \times \nu(E)$.*

Theorem 28 (Fubini).   *1. $f_x : Y \to \mathbb{R}$ defined by $f_x(y) = f(x, y)$ is integrable on Y for almost every X. The same statement holds if we switch the roles of X and Y.*
    2. $\int_Y f_x(y) \, \mathrm{d}\nu(y)$ is an integrable function on X. Again, the same statement holds for Y.
    3. $\int_X \left(\int_Y f_x \, \mathrm{d}\nu\right) \mathrm{d}\mu = \int_Y \left(\int_X f_y \, \mathrm{d}\mu\right) \mathrm{d}\nu = \int_{X \times Y} f \, \mathrm{d}(\mu \times \nu)$.

Theorem 29 (Tonelli). *The same conditions in Fubini's theorem hold for non-negative measurable functions on $X \times Y$.*

Example 30 (Counterexample if non-negativity is not required). Consider $f = \frac{x^2 - y^2}{x^2 + y^2}$ (see the homework).

We are going to assume now that integration over a general measure has been mastered. After this, we will talk about convergence of functions. Because spaces of functions are infinite-dimensional, we don't have nice convergence results that exist in the finite-dimensional case. We will see that the most important norms are L^1, L^2, L^∞, which correspond to integration, Euclidean, and uniform, respectively. We will treat all of our Banach spaces abstractly. "Abstraction makes things easier

generally. However, when applying this stuff, we have to remove ourselves from this abstract into real stuff."

"We stopped early and banked three minutes so I don't have to stop early again."

4   NOTIONS OF CONVERGENCE

4.1   Lecture 4 (Jan 31)   In undergrad analysis and last semester, we always think about $\mathbb{R}^n$. Now, we're going to always think about infinite-dimensional spaces.

Recall that in $\mathbb{R}^d$, we can have sequences $x_n \to x$ This is really the only form of convergence in a finite dimensional space. Functions can converge in several ways:

Pointwise This is easy but not very helpful. Functions are only defined up to sets of measure zero. This is essentially convergence in $\mathbb{R}$ at each $x \in X$.

Uniform This is much better than pointwise convergence. For example, a uniform limit of uniformly continuous f is uniformly continuous. This is the same thing as convergence in the L^∞ norm $\|f\|_\infty = \sup_{x \in X}|f(x)|$.

Convergence in Measure, L^1, etc. These were considered last semester.

We will consider abstract convergence in metric spaces. Let (X, d) be a metric space. Then X inherits a topology given by the metric.

Lemma 31. *Let (X, d) be a metric space and $F \subset X$. Then F is closed if and only if every convergent sequence in F has a limit in F.*

This gives us a characterization of closure using sequences.

Example 32 (Continuity)**.** Continuity on a metric space can be defined the same way as continuity in calculus. Alternative equivalent definitions include sequential continuity (f is continuous if it commutes with limits of sequences)and topological continuity. The first one is the most tangible and connected to $\mathbb{R}$, the second is the easiest to use in practice (but highly dependent on the notion of completion), and the third is the most abstract.

Definition 33 (Density)**.** A is dense in X if $\overline{A} = X$.

Remark 34. If A is dense, then for all x there exists a sequence in A converging to x. Typically, if we can establish a property on A, we can extend to X by passing to the limit.

Example 35. C^∞ is dense in L^1. We saw this last semester when we discussed convolution. Recall that if K_δ is a nice kernel and $u \in K^1$, then $u * K_\delta \in C^\infty$ and the convolutions converge to u. Thus, if we can define an operator on smooth functions, we can pass to the limit to define the operator on a much bigger class.

Example 36 (Heat equation)**.** The heat equation is $\Delta_t = \Delta u$. This "looks like" $\frac{\mathrm{d}u}{\mathrm{d}t} = Au$ where $A = \Delta$ and $u = e^{tA}u_0$. So we'd like to write $u(t) = e^{t\Delta}u_0$ where Δ is the Laplacian. The operator $e^{t\Delta}$ doesn't make sense for general functions $u_0(x)$, but we can define it on C^∞ using the power series. Then we can extend this to L^1 by taking limits.

4.1.1 Completeness This is the same as the topological completion of $\mathbb{Q}$ to give $\mathbb{R}$. Recall that there are two constructions: Dedekind cuts, which are easier, and Cauchy sequences, which are better.[2] Here, we will use Cauchy sequences which can be done in any metric space.

Definition 37. $\tilde{X}$ is a completion of X if:

1. There exists an isometry $i : X \to \tilde{X}$;
2. $i(X)$ is dense in $\tilde{X}$;
3. $\tilde{X}$ is complete.

Theorem 38. *Any metric space X has a unique completion $\tilde{X}$ up to isometry.*[34]

Example 39 (Equivalence of metrics)**.** Let X be a positive definite symmetric matrix. Then $d_A(x, y) = x^T A y$ is a metric with the "unit ball" the ellipsoid with axes given by λ_i. Then we see that d_A is equivalent to the Euclidean metric, but not isometric.

Remark 40. All isometries are injective. An invertible isometry is an isomorphism in the category of metric spaces.

Proof of Theorem 38. Suppose we have two completions $\tilde{X}_1, \tilde{X}_2$ of X. Then we find an isomorphism between $\tilde{X}_1, \tilde{X}_2$.

Because X is dense in $\tilde{X}_1$, we use the definition of i_2 on X to derive $\hat{i}$. Given any $x \in \tilde{X}_1$, there exists a sequence $\{ x_k \}$ in X where $x_k \to z$. Then we define $\hat{i}z = \lim i_2 x_k$.

We see that clearly this is well-defined because two sequences that converge to the same point must approach each other. To check that $\hat{i}$ is an isometry, $d(z_1, z_2) = \lim d(x_k \overline{x}_k) = \lim d_2(i_2 x_k, i_2 \overline{x}_k) = d_2(\hat{i}z_1, \hat{i}z_2)$. Also, note that $\hat{i}i_1 x = i_2 x$ for all $x \in X$. Finally, note that if $z_2 \in \tilde{X}_2$, choose $x_k \in X$ such that $i_2 x_k \to z_2$. Then $z = \lim i_1 x_k$ satisfies $\hat{i}z = z_2$.

Existence is seen using Cauchy sequences. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

5   COMPACTNESS

5.1   Lecture 5 (Feb 5) The notion of compactness distinguishes the two cases of finite and infinite dimensions. There are two notions of compactness: sequential and topological. A prototypical example of a compact set in $\mathbb{R}$ is the closed interval $[0, 1]$.

Definition 41 (Sequentially Compact)**.** Let X be a metric space. $K \subset X$ is sequentially compact if every sequence $\{x_k\} \subset K$ has a convergent subsequence with limit in K.

Theorem 42 (Heine-Borel)**.** *$K \subset \mathbb{R}^d$ is sequentially compact if and only if it is closed and bounded.*

This follows easily from the Bolzano-Weierstrass theorem, which states that every bounded sequence in $\mathbb{R}^d$ has a convergent subsequence. Robin explained the idea of $B - W$, but I will assume that you already know it. In the context of a metric space, we introduce the notion of "total boundedness."

---

[2]There was a question about Dedekind cuts, but please refer to Wikipedia instead.

[3]We are not defining isometries to be bijective

[4]Isometry and equivalence of metrics are not the same. Two metrics are equivalent if they generate the same topology.

Definition 43 (ε-net)**.** Given a set B, an ε-net N_ε is a subset of points of B such that

$$B \subset \bigcup_{x \in N_\varepsilon} B_\varepsilon(x).$$

Definition 44 (Totally Bounded)**.** A set A is totally bounded if and onlg if A has a finite ε-net for all $\varepsilon > 0$.

Lemma 45. *If X is totally bounded, then it is separable.*

Proof. For all n, we extract a finite $1/n$-net A_n. Now let $A = \bigcup_{n=1}^\infty A_n$, which is countable. We show that A is dense in X.

Given $x \in X$, for any ε, there exists $n > \frac{1}{\varepsilon}$, but then there exists $y \in A_n$ such that $d(x,y) < \frac{1}{n} < \varepsilon$. $\square$

Theorem 46. *$K \subset X$ is sequentially compact if and only if it is complete and totally bounded.*[5] *Cauchy subsequence.*

Proof. First assume that K is complete and totally bounded. Then given any sequence $\{x_n\} \subset K$ we extract a subsequence converging in K. By completeness, we extract a Cauchy subsequence.

Choose x_{n_1} and suppose inductively we have the subsequence $x_{n_1} \ldots, x_{n_k}, y_1^k, \ldots, y_m^k, \ldots$ chosen at the k-th step. We want the subsequence $\{x_{n_k}\}$ to be Cauchy. Each subsequence is totally bounded. At the k-th inductive step, set $\varepsilon = \frac{1}{2^{k+1}}$. Then the set $Y_k = \{y_1^k, \ldots, y_m^k, \ldots\}$ is totally bounded, so Y_k has a finite ε-net. Then there exists a finite ε-net $\{x_1, \ldots, z_m\}$ such that $Y_k \subset \bigcup_{i=1}^m B_\varepsilon(z_j)$. Thus there exists one $B_\varepsilon(z_j)$ containing infinitely many $\{y_k\}$. Choose $x_{n_{k+1}} = z_j$ and $y_1^{k+1}, \ldots, y_m^{k+1}, \ldots$ which are in $B_\varepsilon(z_j)$. Clearly $x_{n_{k+1}}$ is Cauchy by construction.

For the converse, let $\{x_k\}$ be Cauchy in K. Then there exists a convergent subsequence $x_{n_k} \to x \in K$. The limit of any subsequence is the limit of the full sequence because the sequence is Cauchy, so K is complete.

Now assume K is not totally bounded. We find a sequence with no convergent subsequence. Then there exists $\varepsilon > 0$ such that K has no fintie ε-net, i.e.for any finite ε-net, there exists $x \in K$ not within ε of the net. For each n, having chosen $\{x_1, \ldots, x_n\}$, choose x_{n+1} such that $x_{n+1} \in K$ but $x_n + 1 \notin \bigcup_{j=1}^n B_\varepsilon(x_j)$. By construction, for all $n > m$, $d(x_n, x_m) > \varepsilon$, so this sequence has no Cauchy subsequence. $\square$

Definition 47 ((Topological) Compactness)**.** K is compact if every open cover has a finite subcover.

In a metric space, compactness and sequential compactness are equivalent. However, there are examples of compact sets not sequentially compact and sequentially compact sets not compact.

Theorem 48. *Let X be a metric space. $K \subset X$ is compact if and only if it is sequentially compact.*

Proof. Let K be sequentially compact. Given G_α an open cover of K, we wish to extract a finite subcover.

---

[5]This is the metric space analogue of Heine-Borel.

First we use total boundedness to extract a countable subcover. We know K is totally bounded, so it is separable. Let A be a countable dense subset of K. Let $\mathcal{B} = \{ B_r(a) \mid r \in \mathbb{Q}, a \in A \}$, which is countable. Then let $\mathcal{C} = \{ B_i \in \mathcal{B} \mid B_i \subset G_\alpha$ for some $\alpha \}$. We use $\mathcal{C}$ to extract a countable subcover $\{G_n\}$ of K. Let $x \in K$. Then there exists α such that $x \in G_\alpha$. Then there exists ε such that $B_\varepsilon(x) \subset G_\alpha$. Now there exists $y \in A$ such that $d(x,y) < \frac{ep}{3}$. Also, $x \in B_{\frac{2\varepsilon}{2}}(y) \subset B_\varepsilon(x) \subset G_\alpha$. Now choose $q \in \mathbb{Q}$ such that $\frac{ep}{3} < q < \frac{2\varepsilon}{3}$. Then $B_q(y) \in \mathcal{C}$. Thus $\mathcal{C}$ is a countable open cover of X. For any $B \in \mathcal{C}$, there exists α_B such that $B \subset G_{\alpha_B}$. These G_{α_B} form a countable subcover of $\{G_\alpha\}$. We relabel them as G_n.

Suppose there exists no finite subcover of G_n. Then for all N, there exists $x \in K \smallsetminus \bigcup_{n=1}^{N} G_n$. Choose $x_1 \in K$. Then there exists N_1 such that $x_1 \in G_{N_1}$, but there exists $x_2 \in K$ such that $x_2 \notin \bigcup_{n=1}^{N_2} G_n$. There must exists N_2 such that $x_2 \in G_{N_2}$. Inductively, choose a sequence $\{x_k\}$ and a strictly increasing sequence $\{N_{k+1}\}$ of integers. For all n, there exists M_n such that whenever $m > M_n$, $x_m \notin G_n$. Therefore if $x \in G_n$, there is no subsequence $x_{k_n} \to x$. But the G_n cover K, $\{x_k\}$ has no subsequence converging to any $x \in K$, so K is not sequentially compact.

In the other direction, suppose K is not sequentially compact. Then there exists a sequence $\{x_k\}$ containing no convergent subsequence. Without loss of generality, suppose the x_k are distinct. Then let $x \in K$. If $B_\varepsilon(x)$ cpmtaoms an $x_k \neq k$ for all ε, then that subsequence converges to x. Thus there exists ε_x such that $B_{\varepsilon_x}(x)$ has no element of $\{x_k\}$ or $x = x_k$ for some k. But then $B_{\varepsilon_x}(x)$ is an open cover. Thus there exists a finite subcover, which implies that there are only finitely many distinct elements of the sequence, which is a contradiction. $\qquad\square$

5.2 Lecture 6 (Feb 7) Robin actually arrived to lecture early today. This has never happened before in the history of this analysis class with the exception of the first lecture last semester.

Today we will carry on with compact sets and prove some theorems. One remark is that if you lose a derivative, you gain compactness.

Theorem 49. *Let $f : K \to Y$ is continuous with K compact and Y a metric space. Then $f(K)$ is also compact.*

Proof. Let $\{G_\alpha\}$ be an open cover of $f(K)$. We pull back to an open cover of K, extract a finite subcover, and then send this open cover back to the image. $\qquad\square$

Theorem 50. *If $f : K \to Y$ from a compact set K is continuous, then it is uniformly continuous.*

Proof. There exists ε such that for any n, we can find x_n, y_n such that $d(x_n, y_n) < \frac{1}{n}$, but $d(f(x_n), f(y_n)) > \varepsilon$. Now extract convergent subsequences $\{x_{n_k}\}$ and $\{y_{n_k}\}$. Then $x_{n_k} \to x$, y_{n_k}, but $d(f(x), f(y)) \geq \varepsilon$, which contradicts continuity because $x = y$. $\qquad\square$

Theorem 51. *Let $f : K \to \mathbb{R}$ be continuous on a compact K. Then f attains a maximum and minimum.*

Proof. Let $\beta = \sup_{z \in K} f(z)$. Then for all n, there exists x_n such that $\beta - \frac{1}{n} < f(x_n) \leq \beta$. Extract a convergent subsequence $x_{n_k} \to x \in K$. By continuity $f(x) = \lim_{k \to \infty} f(x_{n_k}) = \beta$. $\qquad\square$

Corollary 52. *If $f : K \to \mathbb{R}$ is upper semi-continuous (resp. lower semi-continuous), then it is bounded above (resp. below) and attains its maximum (resp. minumum).*

Remark 53. The theorem says nothing about the uniqueness of optimizing sequences. An example is $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = |x| \left(1 - \cos \left(\frac{1}{x} \right) \right)$. Then there are concentric circles of minima.

5.2.1 Compactness in Infinite Dimensions We introduce our metric space as $C(X)$ the set of continuous real-valued functions on compact X. Use the norm $\|f\| = \sup_{x \in X} |f(x)|$. Let $C_b(X)$ denote the set of bounded functions $X \to \mathbb{R}$. Then $C_b(x)$ is a Banach space (a complete normed vector space). In fact $C_b(K)$ is a Banach algebra: $\|fg\| \le \|f\| \|g\|$.

Theorem 54. *If K is compact, then $C(K)$ is complete and therefore Banach.*

Idea of Proof. Given a Cauchy sequence f_n in $C(K)$, define the limit pointwise. Then each f_n is uniformly continuous and f_n is uniformly Cauchy (so converges uniformly to f), so $f \in C(K)$. $\square$

Remark 55. In general, the sup norm is not a norm on $C(X)$ but is a norm on $C_b(X)$.

We need compactness to obtain completeness. Define $C_c(X)$ be the set of continuous functions with compact support.

Theorem 56. *Let $C_0(X)$ be the closure of $C_c(X)$ in $C_b(X)$. This is a Banach space.*

Sketch of Proof. Given a Cauchy sequence in $C_0(X)$, we need to show $f = \lim f_n$ is in $C_0(X)$. We use a diagonal argument. Each $f_n(X)$ is in $C_0(X)$, so each $f_n = \lim_{m \to \infty} f_{n,m}$ where each $f_{n,m}$ has compact support. Then we see that each $f = \lim_{n \to \infty} f_{n,n} \in C_0(X)$. $\square$

We sau that $C_0(X)$ are the functions that "vanish at infinity." Consider the prototype $C(K)$ is $C([a,b])$.

Theorem 57 (Weierstrass approximation theorem)**.** *Any function $f \in C([a,b])$ can be uniformly approximated by a polynomial. More precisely, the set of polynomials is dense in $C([a,b])$.*

Proof. Without loss of generality, take $[a,b] = [0,1]$. Recall the binomial theorem $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$. In particular $1 = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}$ for all x, n. Thus $f(x) = \sum_{k=0}^{n} f(x) \binom{n}{k} x^k (1-x)^{n-k}$. We then find the maximum of $x^k (1-x)^{n-k}$, which is $x = \frac{k}{n}$. Therefore, we define $B_n f = \sum_{k=0}^{n} f \left(\frac{n}{k} \right) \binom{n}{k} x^k (1-x)^{n-k}$.

We show that $\|B_n f - f\| \to 0$ as $n \to \infty$. Note $(B_n f - f)(x) = \sum_{k=0}^{n} \left(f \left(\frac{n}{k} \right) - f(x) \right) \binom{n}{k} x^k (1-x)^{n-k}$. Choose $\varepsilon > 0$. Because f is uniformly continuous, there exists δ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Set $I(X) = \left\{ k \mid \left| x - \frac{k}{n} \right| < \delta \right\}$ and $J(X) = \left\{ k \mid \left| x - \frac{k}{n} \right| \ge \delta \right\}$.

Then

$$|B_n f(x) - f(x)| \le \sum_{I(x)} + \sum_{J(x)} \left| f \left(\frac{n}{k} \right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k}.$$

On $I(x)$, $\left|f(\frac{k}{n}) - f(x)\right| < \varepsilon$, so

$$\|B_n f - f\| \leq \sup_x \sum_{I(x)} \varepsilon \binom{n}{k} x^k (1-x)^{n-k} + \sup_x \sum_{J(x)} 2\|f\| \binom{n}{k} x^k (1-x)^{n-k}$$

$$\leq \varepsilon + 2\|f\| \sup_x \sum_{J(x)} \binom{n}{k} x^k (1-x)^{n-k}.$$

Because δ has been fixed, on $J(x)$, $\left(x - \frac{k}{n}\right)^2 \geq \delta^2$, or $x^2 - 2\frac{k}{n}x + \left(\frac{k}{n}\right)^2 \geq \delta^2$. Therefore

$$\sum_{J(x)} \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{1}{\delta^2} \left(\sum_{J(x)} \binom{n}{k} x^k (1-x)^{n-k} \right) \left(x^2 - 2\frac{k}{n}x + \left(\frac{k}{n}\right)^2 \right),$$

which is at most

$$\frac{1}{\delta^2} \left(x^2 \sum_{J(x)} \binom{n}{k} x^k (1-x)^{n-k} - 2x \sum_{J(x)} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k} + \sum_{J(x)} \left(\frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k} \right).$$

Therefore,

$$\sup_x \sum_{J(x)} \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{1}{\delta^2} \sup \left[x^2 B_n(1) - 2x B_n(x) + B_n(x^2) \right].$$

Now we compute $B_n(x), B_n(x^2)$. Note that $(x+y)^n = \sum \binom{n}{k} x^k y^{n-k}$. Differentiate with respect to x and get $n(x+y)^{n-1} = \sum \binom{n}{k} k x^{k-1} y^{n-k}$. Therefore $n = \sum \binom{n}{k} k x^{k-1}(1-x)^{n-k}$. Therefore

$$x = \sum \binom{n}{k} x^k (1-x)^{n-k} = B_n(x).$$

Again, we now get $n(n-1)(x+y)^{n-2} = \sum \binom{n}{k} k(k-1) x^{k-2} y^{n-k}$. We attain that $n(n-1) = \sum \binom{n}{k} k(k-1) x^{k-2}(1-x)^{n-k}$. Multiplying through by $\left(\frac{n}{n}\right)^2$, we obtain that $x^2 \left(1 - \frac{1}{n}\right) = \sum \binom{n}{k} \frac{k^2}{n^2} x^k (1-x)^{n-k} - \frac{1}{n} \sum \binom{n}{k} \frac{k}{n} x^k (1-x)^{n-k} = B_n(x^2) - \frac{1}{n} B_n(x)$. Therefore,

$$x^2 - 2x B_n(x) + B_n(x^2) = x^2 - 3x^2 + x^2 \left(1 - \frac{1}{n}\right) + \frac{x}{n} = \frac{1}{n}(x - x^2).$$

We conclude that $\|B_n f - f\| \leq \varepsilon + \frac{2\|f\|}{\delta^2} \cdot \frac{1}{n} \sup x - x^2$. $\qquad\qquad\square$

Lecture 7 (Feb 12) When Robin walked into lecture, Jared asked him if he knew what a circumgalactic medium was, and Robin said it sounds like the aether. Robin has also posted homework, which is due a week from today. We will finish the proof of the Weierstrass approximation theorem, which is in the notes for last time.

Recall that in $\mathbb{R}^d$, K is compact if and only if it is closed and bounded.

Definition 58 (Precompact). M is precompact if its closure $\overline{M}$ is compact.

Note that this means $M \subset \mathbb{R}^d$ is precompact if and only if it is bounded.

However, observe that $C(K)$ is infinite-dimensional, so closed and bounded is not enough. Examples will come later. The concept that provides compactness in $C(K)$ is equicontinuity. This is an analogue of uniformity across the functions.

Definition 59 (Equicontinuity)**.** Let $\mathcal{F}$ be a set of functions $X \to Y$. Then $\mathcal{F}$ is equicontinuous if given any $\varepsilon > 0$ and $x_0 \in X$, there exists δ such that if $d(x, x_0) < \delta$, then $d(f(x), f(x_0)) < \varepsilon$ for all $f \in \mathcal{F}$.

Definition 60 (Uniform Equicontinuity)**.** $\mathcal{F}$ is uniformly equicontinuous if δ is independent of x_0 (Think the definition of uniform continuity or uniform convergence).

Theorem 61. *If $\mathcal{F}$ is equicontinuous on K where K is compact, and $\mathcal{F} \subset C(K)$, then $\mathcal{F}$ is uniformly equicontinuous.*[6]

Proof. Suppose not. Then there exists ε such that for all n, there exist x_n, z_n, f_n such that $d(x_n, z_n) < \frac{1}{n}$, but $d(f_n(x_n), f_n(z_n)) \geq \varepsilon$. The goal is to show that $\mathcal{F}$ is not equicontinuous.

We know that K is compact, so from $\{x_n\}$ we can extract $x_{n_k} \to x$. Then the z_{n_k} also converge to x.[7] However, because $d(f_n(x_n), f_n(z_n) \geq \varepsilon)$ for all n, we must have either $d(f_n(x_n), f_n(x)) \geq \varepsilon/n$ or $d(f_n(z_n), f_n(x)) \geq \varepsilon/n$. This contracts equicontinuity. $\square$

Theorem 62 (Arzela-Ascoli (1880s?))**.** *Let K be compact. A subset $\mathcal{F} \subset C(K)$ is compact if and only if it is closed, bounded, and equicontinuous.*

Proof. There are three steps:

1. If $\mathcal{F}$ is unbounded, then it is not precompact.
2. If $\mathcal{F}$ is precompact, then it is equicontinuous.
3. If $\mathcal{F}$ is bounded and equicontinuous, then it is precompact.

Now we prove each part:

1. This one is easy. Build a sequence f_n such that $\|f_{n+1}\| \geq \|f_n\| + 1$. This sequence is not Cauchy, so there is no convergent subsequence.
2. Assume $\mathcal{F}$ is precompact. Then write $\overline{\mathcal{F}} \subset \bigcup_{f \in \mathcal{F}} B_{\varepsilon/3}(f)$. We can extract a finite subcover $\overline{\mathcal{F}} \subset \bigcup_{i=1}^n B_{\varepsilon/3}(f_i)$. Because each f_i is uniformly continuous, then there exists δ_i such that if $d(x, y) < \delta_i$, then $d(f_i(x), f_i(y)) < \varepsilon/3$ for all $x, y \in K$. Choose $\delta = \min \delta_i > 0$. Now if $d(x, y) < \delta$, then for any $f \in \mathcal{F}$, choose i such that $\|f - f_i\| < \varepsilon/3$. Therefore, $d(f(x), f(y)) \leq d(f(x), f_i(x)) + d(f_i(x), f_i(y)) + d(f_i(y), f(y)) < \varepsilon$.
3. Let $\mathcal{F}$ be bounded and equicontinuous. We will show that it is sequentially pre-compact. Given a bounded sequence, extract a Cauchy subsequence. Because $C(K)$ is complete, this implies sequential compactness. Let $\{f_n\} \subset \mathcal{F}$ be a bounded sequence. Because K is compact, it has a countable dense subset $\{x_1, \ldots, x_n, \ldots\}$.
   Extract a subsequence $f_{1,n}$ of f_n such that $f_{1,n}(x)$ converges. Then extract a subsequence $f_{2,n}$ of $f_{1,n}$ such that $f_{2,n}(x_2)$ converges. Then we extract a subsequence f_{k+1}, n of $f_{k,n}$ such that $f_{k+1}(x_{k+1})$ converges. Continue this process to infinity. Now extract the diagonal sequence

---

[6] Robin put his notes down here in a show of mathematical dominance.

[7] Robin checked his notes here, forfeiting his mathematical dominance.

of functions $g_k = f_{k,k}$. We know that at each x_i, $g_k(x_i)$ is Cauchy. We will show that g_k is uniformly Cauchy.

We will use equicontinuity of $\mathcal{F}$. Because K is compact, $\mathcal{F}$ is uniformly equicontinuous. Thus there exists δ such that if $d(x,y) < \delta$, then $d(g_k(x) - g_k(y)) < \varepsilon/3$ for all k. Because K is compact with $\{x_i\}$ dense, $K \subset \bigcup_{i=1}^{\infty} B_\delta(x_i)$. We extract a finite subcover $K \subset \bigcup_{j=1}^{N} B_\delta(z_j)$. For each z_j, the sequence $g_k(z_j)$ is Cauchy. Therefore, there exists N_j such that if $k, \ell > N_j$, then $d(g_k(z_j), g_\ell(z_j)) < \varepsilon/3$. Choose $N = \max N_j < \infty$. Then given any $x \in K$, for $k, \ell > N$, choose z_j such that $d(x, z_j) < \delta$. Then $d(g_k(x), g_\ell(x)) \leq d(g_k(x), g_k(z_j)) + d(g_k(z_j), g_\ell(z_j)) + d(g_\ell(z_j), g_\ell(x)) < \varepsilon$. $\qquad\square$

5.3 Lecture 8 (Feb 14 ♡) Last time we saw that equicontinuity and boundedness implies precompactness, so what is equicontinuity?

Example 63 (Not equicontinuous). Let h_n be the skew hat function, which is piecesise linear and continuous with $h_n(\frac{1}{n}) = 1, h_n(\frac{1}{n\pm 1}) = 0$. Notice that the $\{h_n\}$ are not equicontinuous because $\|h_n - h_m\| = 1$ for all $n \neq m$. Therefore this sequence has no convergent subsequence, so it is not precompact. However, this sequence is clearly bounded.

Example 64. If $\mathcal{F}$ is a uniformly Lipshitz family, then it is equicontinuous. $\mathcal{F}$ is uniformly Lipshitz if there exists K such that for all $f \in \mathcal{F}$, $|f(x) - f(y)| \leq K|x - y|$.

We can weaken this statement, but not by much.

Example 65. Suppose our f are represented by power series of the form $f = \sum_{j=0}^{\infty} a_j x_j$. Then $f(x) - f(y) = \sum_j a_j(x^j - y^j)$. Recalling that $(x^j - y^j) = (x - y)(x^{j-1} + \cdots + y^{j-1})$, we see that

$$f(x) - f(y) = (x - y) \sum_{j=0}^{\infty} a_j \sum_{k=0}^{j-1} y^k x^{j-1-k}.$$

Then note that $\sum_{k=0}^{j-1}$ can only be bounded by jL^{j-1} if $x, y \in [-L, L]$. To use the factorization, the only possible bound is

$$|f(x) - f(y)| \leq |x - y| \sum_{j=1}^{\infty} |a_j| jL^{j-1}.$$

This morally looks like $\sup_{[-L,L]} |f'(x)|$, so the moral is that we need some sort of derivative to claim precompactness.

Remark 66. We will see later that functions with k derivatives are precompactly contained in functions with $k-1$ derivatives. Thus we can find a convergent subsequence of a bounded sequence, but the limit can only be expected to be C^{k-1}.

5.3.1 Differential Equations

Example 67 (Existence theorem for ODEs). Consider the ODE $\dot{u} = f(t, u)$ with initial condition $u(t_0) = u^0$. We will take $u \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$. We will explore the existence of solutions ot this equation. Assume that $f(t, u)$ is continuous.

Recall Euler's method, which is to set $u(t + h) = u(t) + \int_t^{t+h} f(u, s)\, ds$ and approximate the integral term, taking n steps of constant size. Given h, we will build u_h^0 and then $u_h^{k+1} = u_h^k + hf(t_k, u_h^k)$. Then we take $u_h(t)$ to be the piecewise linear interpolation of these points.

Now we extract a convergent subsequence from the u_h. We call $u_\varepsilon(t)$ an ε-approximate solution to the initial value problem if

1. $u_\varepsilon(t_0) = u_0$;
2. $u_\varepsilon(t)$ is continuous, and it is differentiable at all but finitely many points of I;
3. If $\dot{u}_\varepsilon$ exists at t, then $|\dot{u}_\varepsilon(t) - f(t, u_\varepsilon(t))| \le \varepsilon$.

Then for $t \in (t_k, t_{k+1})$, we have $|\dot{u}_h(t) - f(t, u_h)| = |f(t_k, u_h^k) - f(t, u_h(t))|$. We will use continuity of f to make this small:

Theorem 68. *There exists an interval I and a continuously differentiable function $u : I \to \mathbb{R}^d$ such that the IVP holds in I.*

Note that the above is not a global existence theorem. Let T_1 be our initial guess for the time of existence. Choose $L > 0, T \le T_1$ such that for each u_ε with $|t - t_0| \le T_1$ satisfies $|u_\varepsilon(t) - u_0| \le L$. If $|f| \le M$, then set $T = \min\{T_1, L/M\}$. Call $R = \{(t, u) \mid |t - t_0| \le T, |u - u_0| \le L\}$. Then R is compact, so f is uniformly continuous in R. Thus u is uniforly Lipshitz continuous. Therefore u_ε is equicontinuous and bounded. By Arzela-Ascoli, we extract a convergent subsequence. Let $u(t) = \lim_{\varepsilon_k \to 0} u_{\varepsilon_k}(t)$. We know that $u(t) \in C[t_0 - T, t_0 + T]$.

We need to show that the limit is continuously differentiable and satisfies the IVP. We know that $u_\varepsilon(t)$ satisfies

$$u_\varepsilon(t) = u_\varepsilon(t_0) + \int_{t_0}^{t} \dot{u}_\varepsilon(s) \, \mathrm{d}s$$

$$= u_0 + \int_{t_0}^{t} f(s, u_\varepsilon(s)) \, \mathrm{d}s + \int_{t_0}^{t} (\dot{u}_\varepsilon(s - f(s, u_\varepsilon(s)))) \, \mathrm{d}s$$

Now we see that $\left| \int_{t_0}^{t} \dot{u}_\varepsilon(s) - f(s, u_\varepsilon(s)) \, \mathrm{d}s \right| \le |t - t_0| \cdot K \cdot \varepsilon$, which approaches 0 by construction. Therefore $u(t) = u_0 + \int_{t_0}^{t} f(s, u(s)) \, \mathrm{d}s$. Because f is continuous, then $\int f$ is differentiable and thus $u \in C^1$.

We have the following theorem for these solutions.

Theorem 69. *Let $R = \{(t, u) \mid |t - t_0| \le T, |u - u_0| \le L\}$ and suppose $|f(t, u)| \le M$ on R. If $\delta = \min\{T, L/M\}$. Then $u(t)$ solution is defined ans satisfies $|u(t) - u_0| \le L$ for all $t \le \delta$ and if f is Lipshitz in u and uniform in t, then the solution is unique.*

Example 70. Consider the initial value problem $\dot{u} = u^{1-\alpha}, u(0) = 0$, where $\alpha > 0$. Then $u(t) = 0$ is a solution. Also we see that $\frac{\mathrm{d}u}{\mathrm{d}t} = u^{1-\alpha}$, so separating variables and integrating yields $u = (\alpha t)^{1/\alpha}$, which satisfies the initial condition. Therefore uniqueness fails.

The first part of Theorem 69 is the statement of time of existence. While f remains continuous, the only way for solutions to fail to exist is for them to "escape to ∞," or to leave any compact set.

5.3.2 Digression Consider a 4-body problem. We consider gravitational attraction. We want to find out where solutions do not exist, where a solution is a continuous trajectory for each planet. The question is: how do solutions fail to be globally defined?

1. Collision: f has a singularity and loses continuity. For the 3-body, problem, this is all;

2. For the 4-body problem, there are initial conditions for which solutions escape to ∞ in finite time!

Example 71. Consider a 5-body problem with two pairs of two overlapping elliptical orbits. Then the fifth body is an oscillator between the two pairs of elliptical orbits.

5.4   Lecture 9 (Feb 21) Last time we discussed Peano's proof of the existence of solutions to first order differential equations using the Euler approximation. The idea is that we construct an equicontinuous sequence, so we extract a convergent subsequence, which is the solution.

Robin then proceeded to tell a story about a time when he visited Hershey Park in Hershey, PA and went on a rollercoaster.[8] This somehow relates to the five-body problem where solutions escape to infinity in finite time. There is a Cantor set of initial conditions under which planets escape to infinity in finite time. Jared asked him what that would mean physically, and Robin started talking about shockwaves, which he claims to do for a living. He argues that it is more physically relevant to study integral equations than differential equations.

5.4.1   Differential Equations Continued

Theorem 72. *Let R be the rectangle $\{\,(t,u)\mid |t-t_0|\le T, |u-u_0|\le L\,\}$. Let f be continuous on R, so there exists M such that $|f(t,u)|\le M$ on R.*

Time of existence *Let $\delta = \min\{T, L/M\}$. Then the solution $u(t)$ exists and stays in R for $|t-t_0|\le \delta$, i.e. $|u(t)-u_0|\le L$ for all $t\in[t_0-\delta, t_0+\delta]$.*

Uniqueness *If f is Lipshitz in u and uniformly continuous for $t\in R$, then the solution is unique for $|t-t_0|\le \delta$.*

Proof. **Time of existence** Define $D = \{\,\xi\mid 0\le\xi\le\delta, |u(t)-u_0|\le L, |t-t_0|\le\xi\,\}$. Clearly $0\in D$, and also if $\xi\in D$ and $\xi'\in[0,\xi)$, then $\xi'\in D$. Therefore D is an interval. D is closed because u is continuous.) Also, D is relatively open in $[0,\delta]$. If $\xi<\delta$ and $\xi\in D$, we have $f(t,u(t))\le M$ for $|t-t_0|\le\delta$. Therefore

$$|u(t)-u(t_0)| \le \left|\int_{t_0}^t f(s,u(s))\ \mathrm{d}s\right| \le M\xi < M\delta \le L.$$

Therefore there exists ε such that $|u(t)-u_0| < L$ for $|t-t_0|\le\xi+\varepsilon$. Therefore there exists a neighborhood of $\xi\subset D$. Therefore $D=[0,\delta]$.

Uniqueness Consider two solutions u,v. Then we see that

$$|u(t)-v(t)| \le |u_0-v_0| + \int_{t_0}^t |f(s,u(s)-f(s,v(s)))|\ \mathrm{d}s \le |u_0-v_0| + K\int_0^t |u(s)-v(s)|\ \mathrm{d}s,$$

where K is the unoform Lipshitz constant. Therefore set $w(t)=|u(t)-v(t)|$. It satisfies the inequality $w(t)\le w_0 + K\int_{t_0}^t w(s)\ \mathrm{d}s$ for $|t-t_0|\le\delta$. This is a linear integral inequality. We will apply "Gronwall's Lemma."

---

[8]He has no idea why he said that.

Lemma 73 (Gronwall). *If $w(t) \geq 0$ and $\varphi(t) \geq 0$ are continuous and defined on $[0, T]$, with $w_0 \geq 0$ ans satisfy*

$$w(t) \leq w_0 + \int_0^t \varphi(s) w(s) \ ds,$$

then for $t \leq T$ we have $w(t) \leq w_0 e^{\int_0^t \varphi(s) \ ds}$.

Applying Gronwall here, $w(t) \leq w_0 e^{K(t - t_0)}$ for all $|t - t_0| \leq \delta$. In particular, if $w_0 = 0$, then $w(t) = 0$ for all t. $\qquad\square$

Note that we have proved continuous dependence on the initial conditions for short times. In particular, chaotic systems are predictable for a short time, after which they become unpredictable.

Proof of Gronwall. Set $u(t) = w_0 + \int_0^t \varphi(s) w(s) \ ds$. Then we know that $w(t) \leq u(t)$ for all t. We differentiate to obtain $\dot{u}(t) = \varphi(t) w(t) \leq \varphi(t) u(t)$. Also if $w_0 > 0$, then $u(0) = w_0 > 0$. We see that $\dot{u}(t) \geq 0$, so $u(t) \geq 0$. We can now take logs to obtain for all t,

$$\frac{1}{u(t)} \dot{u}(t) \leq \varphi(t),$$

or $(\log u(t))^{\cdot} \leq \varphi(t)$. Therefore we integrate to find that $\log u(t) - \log u(0) \leq \int_0^t \varphi(s) \ ds$. Therefore $u(t) \leq u(0) e^{\int_0^t \varphi(s) \ ds}$. If $w_0 = 0$, then we can't do this. Instead replace w_0 with ε and let $\varepsilon \to 0$. $\quad\square$

Alternative existence proof for differential equations. This proof uses Picard iteration. Remember the initial equation was $\dot{u}(t) = f(t, u(t)), u(t_0) = u_0$. We integrate to obtain

$$u(t) = u_0 + \int_0^t f(s, u(s)) \ ds.$$

This is an implicit nonlinear integral equation, but there is a natural iteration, which is Picard iteration.

Set $u_{k+1}(t) = u_0 + \int_0^t f(s, u_k(s)) \ ds$. This is explicit and so convergence of the iteration to $u(t) \in C(K)$ yields a solution. Then abstractly: Let X be a complete metric space (think $X = C([t_0 - \delta, t_0 + \delta])$). Then define the operator $T : X \to X$ given by $Tu(t) = u_0 + \int_0^t f(s, u(s)) \ ds$. Then $u(t)$ solves the ODE if and only if it is a fixed point of T. We look for an easy condition on a map which yields a fixed point.

Definition 74 (Strict Contraction). A nonlinear map $T : X \to X$ is a strict contraction if there exists $\rho < 1$ such that for all $x, y \in X$, $d(Tx, Ty) \leq \rho d(x, y)$. Then we use the following theorem:

Theorem 75 (Contraction Mapping Theorem). *A strict contraction on a complete metric space has a unique fixed point.*

We show that the map T given above is a strict contraction. Then we see that

$$\|(Tu)(t) - (Tv)(t)\| \sup \left| \int_{t_0}^t K|u(s) - v(s)| \ ds \right| \leq K\delta \|u - v\|,$$

so if $K\delta < 1$, this is a contraction. $\qquad\square$

Page 18 of 43

Proof of contraction mapping theorem. For uniqueness, note that if there are two fixed points, then that contradicts the definition of a strict contraction.

For existence, just iterate. For any x_0, set $x_{k+1} = Tx_k$. We show that this sequence is Cauchy. After that, because X is complete, the sequence converges, and its limit is a fixed point by continuity. To see that the sequence is Cauchy, this is simply because geometric series converge (this is left as an exercise because I don't feel like typing). $\qquad\square$

6   BANACH SPACES

6.1   Lecture 10 (Feb 26) We began the proof using Picard iteration last time, and Robin is leaving that as an exercise. We begin the discussion of Banach spaces, which are covered in chapters 5 and 12 of the textbook. Robin talked about proving estimates. We will add arithmetic structure to our metric space.

Recall the definition of a normed vector space X:

Definition 76 (Normed Vector Space)**.** A normed vector space X is a vector space satisfying:

1. $\|u\| \geq 0$, $\|u\| = 0 \Rightarrow u = 0$;
2. $\|\alpha u\| = |\alpha|\|n\|$;
3. $\|u + v\| \leq \|u\| + \|v\|$.

The extra structure here is linearity, which allows us to translate and scale, so the entire topology is determined by the unit ball.

Definition 77 (Banach Space)**.** A Banach space is a complete normed vector space.

Example 78. The first example is $\mathbb{R}^n$ with any p-norm:

$$\|(x_1, \ldots, x_n)\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

The power of $1/p$ is forced on us by the second axiom, and the triangle inequality is not at all obvious. Then for $p = \infty$, define $\|x\|_\infty = \max|x_i|$, which is the limit of the p-norms.

We examine $B_1(0)$ for different p-norms, which has boundary $|x|^p + |y|^p = 1$. For $p = 1$ we get the cross-polytope, $p = 2$ the circle, and for $p = \infty$ we get the square. For $p < 1$ we lose convexity of the unit ball, but for $p > 1$ convexity holds, which is equivalent to the triangle inequality.

Note that the unit balls are nested for $1 \leq p \leq \infty$ but inclusion of unit balls reverses inequalities in sizes of norms.

Example 79. We have already seen that $C(K)$ with the sup norm is Banach. However, note that $C^p(K)$ is not closed with respect to the L^∞ norm. The reason is that the norm cannot control derivatives.

To make $C^p(K)$ Banach, we include the derivatives in the norm:

$$\|f\|_{C^p(K)} = \|f\|_\infty + \|f'\|_\infty + \cdots + \|f^{(p)}\|_\infty.$$

An alternative is to use $\|f\|_\infty + \|f^{(p)}\|_\infty$. The two norms are equivalent.

Definition 80. $(X, \|\cdot\|_1), (X, \|\cdot\|_2)$ are equivalent as Banach spaces if the norms are equivalent (they bound each other above and below up to some constant). Clearly equivalent norms generate identical topologies.

Example 81. Ww can extend norms on $\mathbb{R}^n$ to infinite dimensions: Let $\ell^p(\mathbb{N})$ be the set of sequences such that $\sum_i |x_i|^p$ is finite. This is Banach with the p-norm.

Example 82. If $K \subset \mathbb{R}^n$, $L^p(K) = \left\{ f \mid \int_K |f|^p < \infty \right\}$ is Banach with norm the integral to the $1/p$ power.

Example 83. If (X, μ) is a finite measure space, then we can define $L^p(X, \mu) = \{ f \mid \int |f|^p \ \mathrm{d}\mu < \infty \}$. Then this space with the obvious norm is a Banach space.

Note that the assumption that the space has finite measure is a convenience to ensure completeness. Otherwise, we must avoid issues of support "escaping to ∞."

Our goal is to show that $L^p(X, \mu)$ is a Banach space. Before we do this, we introduce Sobolev spaces:

Definition 84 (Sobolev Space). Define the Sobolev space to be:

$$W_{k,p}(X, \mu) = \left\{ f \ \middle| \ \left| \sum_{j=0}^{p} f^{(j)(x)} \right| \in L^p \right\}.$$

Set the norm to the only possible norm.

Note that here we are cheating because we don't know if the derivatives exist. We use a notion of weak derivative.

Theorem 85. $L^p(X, \mu)$ *is a Banach space.*

Proof. Non-negativity (up to sets of measure 0) and scaling are obvious. We need to prove the triangle inequality. Recall Jensen's inequality:

Theorem 86 (Jensen's Inequality). *Define the average* $\langle f \rangle_\mu$ *to be*

$$\langle f \rangle_\mu = \frac{1}{\mu(X)} \int_X f(x) \ \mathrm{d}\mu.$$

Then if $\varphi : \mathbb{R} \to \mathbb{R}$ *is convex,* $\varphi(\langle f \rangle_\mu) \leq \langle \varphi \circ f \rangle_\mu$.

Proof of Jensen was on the final exam last semester, so it is omitted.

Now recall Cauchy-Schwarz:

Theorem 87 (Cauchy-Schwarz). *If* f, g *are square-integrable, then*

$$\int |fg| \ \mathrm{d}\mu \leq \|f\|_2 \|g\|_2.$$

6.2   Lecture 11 (Feb 28)   Hölder's inequality generalizes this to p:

Theorem 88 (Hölder's Inequality)**.** *Let $1 \le p \le \infty$ and p' be the Hölder conjugate $p/(p-1)$. Then if $f \in L^p, g \in L^{p'}$, then $fg \in L^1$ and*

$$\int |fg| \ \mathrm{d}\mu \le \|f\|_p \|g\|_{p'}$$

Proof. If $p = 1$ or $p = \infty$ this is obvious. We can assume that $g \ne 0$ everywhere. The idea is to use Jensen. We will apply Jensen with $\varphi(z) = z^p$, which is convex. Set $\nu(A) = \int_A |g|^{p'} d\mu$, so $\nu(X) = \|g\|_{p'}^{p'}$. Note that by construction, $|g|^{p'}$ is the Radon-Nikodym derivative $\mathrm{d}\nu/\mathrm{d}\mu$. Therefore, by Jensen, we have

$$\left(\frac{1}{\nu(X)} \int_X h \ \mathrm{d}\nu \right)^p \le \frac{1}{\nu(X)} \int_X h^p \ \mathrm{d}\nu.$$

Taking roots and simplifying, we obtain

$$\int h|g|^{p'} \ \mathrm{d}\mu \le \left(\|g\|_{p'}^{p'} \right)^{1-1/p} \left(\int h^p |g|^{p'} \ \mathrm{d}\mu \right)^{1/p}$$

$$\int |fg| \ \mathrm{d}\mu \le \|g\|_{p'} \left(\int |f|^p |g|^{p(1-p')|g|^{p'}} \ \mathrm{d}\mu \right)^{1/p}$$

$$= \|f\|_p \|g\|_p$$

if we set $h = |f| |g|^{1-p'}$. $\square$

Corollary 89. *Assume $1 \le q \le p \le \infty$. Then $L^1(X, \mu) \supset L^q(X, \mu), \supset L^p(X, \mu) \supset L^\infty(X, \mu)$.*

Proof. Set $\frac{1}{p} + \frac{1}{r} = \frac{1}{q}$. Thus $\frac{1}{p/q} + \frac{1}{r/q} = 1$. Therefore if $f \in L^p$, then $f^q \in L^{p/q}$. Also, since $\mu(X) < \infty$, $1 \in L^\alpha$ for all α, and in particular $1 \in L^{r/q}$. Now we can write

$$\|f\|_q^q = \int |f|^q \cdot 1 \ \mathrm{d}\mu$$

$$\le \| |f|^q \|_{p/q} \|1\|_{r/q}$$

$$= \left(\int |f|^p \ \mathrm{d}\mu \right)^{q/p} \cdot \mu(X)^{q/r}$$

$$\le \|f\|_p^q \cdot \mu(X)^{q/r}.$$

Thus $f \in L^q$. $\square$

Now we can prove Minkowski's inequality:

Theorem 90 (Minkowski's inequality)**.** *If* $f, g \in L^p$, *then* $f + g \in L^p$ *and* $\|f + g\| \leq \|f\| + \|g\|$.

Proof. Firstly, note that $|f+g|^p \leq (|f|+|g|)^p \leq 2^p \max(|f|^p, |g|^p) \leq 2^p(|f|^p + |g|^p) < \infty$, so $f+g \in L^p$. This means that $|f + g|^p \in L^1$, so $|f + g|^{p-1} \in L^{p'}$. Therefore

$$\||f + g|^{p-1}\|_{p'} = (|f + g|^p \, \mathrm{d}\mu)^{1/p'} = \|f + g\|_p^{p/p'} = \|f + g\|_p^{p-1}.$$

Therefore,

$$\begin{aligned}
\|f + g\|_p^p &= \int |f + g|^p \, \mathrm{d}\mu \\
&= \int |f + g|^{p-1}|f + g| \, \mathrm{d}\mu \\
&\leq \int |f + g|^{p-1}|f| \, \mathrm{d}\mu + |f + g|^{p-1}|g| \, \mathrm{d}\mu \\
&\leq \||f + g|^{p-1}\|_{p'}(\|f\|_p + \|g\|_p) \\
&= \|f + g\|_p^{p-1}(\|f\|_p + \|g\|_p),
\end{aligned}$$

which gives the triangle inequality. $\qquad\square$

This implies that $L^p(X, \mu)$ is Banach. To prove L^p is Banach, use the completeness of L^1. If $\{f_k\}$ is Cauchy in L^p, then $|f_k|^p$ is Cauchy in L^1, so we take the $\frac{1}{p}$ power of the limit. $\qquad\square$

An alternative approach is to use Young's inequality $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$, which is not Young's convolution inequality.[9]

We return to an abstract Banach space X. Unless explicitly stated otherwise, assume that X is separable. Then we say X has a Schauder basis if:

Definition 91 (Schauder basis)**.** X has a Schauder basis if it has a countable set $x_1, \ldots, x_n$ such that for all $x \in X$, we can approximate X by $\sum c_n x_n$.

Example 92. The standard example is Fourier series, which can be uniquely approximated by $f(x) = \sum a_n \cos(nx) + \sum b_n \sin(nx)$.

Consider a linear operator T.

Definition 93 (Bounded Linear Operator)**.** T is bounded if there exists X such that for all $x \in X$, we have $\|Tx\|_Y \leq K\|x\|_X$.

The set of bounded linear operators is a normed vector space with the norm given by

$$\|T\| = \inf \{\, K \mid \|Tx\| \leq K\|x\| \,\}$$

. This is equivalent to $\sup_{\|x\|=1} \|Tx\|$.

Write $B(X, Y)$ to mean the space of bounded operators. If X, Y are Banach, so is $B(X, Y)$.

---

[9]Robin did not prove this inequality either.

Theorem 94. *T is bounded if and only if it is continuous at* 0.

Proof. It is easy to see that boundedness implies continuity. To prove the converse, suppose not. if T is unbounded, then for all N, there exists x_N such that $\|Tx_N\| > N\|x_N\|$. Take x_n to have length $1/N$. Then $\|Tx_N\| > 1$, so T is not continuous at 0. $\qquad\square$

We actually ended on time today, and Robin made a remark about me putting it in the notes.

6.3 Lecture 12 (Mar 5) Robin apologizes for not assigning enough homework. He will assign homework this week, which will be due after break. There will be no midterm, unlike last semester.[10]

Let $T : X \to Y$ be a linear map of normed vector spaces. We see that the kernel and image are clearly subspace, and if T is bounded, then $\ker T$ is closed (by continuity). X is the domain of T. Note that changing the domain changes the operator. In a finite dimensional space, all linear operators are bounded (the closed unit ball is compact).

Lemma 95. *The closed unit ball is compact if and only if* $\dim X < \infty$.

Proof. In finite dimensions, the closed unit ball is closed and bounded, so it is compact.

In infinite dimensions, this follows from Riesz's Lemma. We will find a sequence x_n in $\overline{B}$ which has no convergent subsequence. Take $r = \frac{1}{2}$. For each n, apply Riesz's Lemma to $X_{n-1} = \operatorname{Sp}\{x_1, \ldots, x_{n-1}\}$. X_{n-1} is not dense, so there exists x_n such that $\|x\|_n = 1$ and $d(x_n, X_{n-1}) \geq \frac{1}{2}$. Then no subsequence of $\{x_n\}$ can converge. $\qquad\square$

Lemma 96 (Riesz). *If* $X \subset Y$ *is a subspace, then if* X *is not dense, then for all* $0 < r < 1$*, there exists* $y \in Y$ *such that* $\|y\| = 1$ *and* $d(y, X) \geq r$.

Proof. First choose $y_1 \notin \overline{X}$. Let $R = \inf_{x \in X} \|x - y_1\| > 0$. We will scale and translate to get a unit vector. Choose $\varepsilon > 0$ and choose $x_1 \in X$ such that $\|x_1 - y_!\| < R + \varepsilon$. Now set $y = \frac{y_1 - x_1}{\|y_1 - x_1\|}$. Now we see that

$$
\begin{aligned}
\inf_{x \in X} \|x - y\| &= \inf_{x \in X} \left\| x - \frac{y_1}{y - x_1} + \frac{x_1}{y_1 - x_1} \right\| \\
&= \inf_{x \in X} \left\| \frac{x - y_1 + x_1}{\|y_1 - x_1\|} \right\| \\
&\geq \frac{R}{R + \varepsilon}.
\end{aligned}
$$

Now choose ε small enough such that $\frac{R}{R+\varepsilon} \geq r$. $\qquad\square$

Example 97. In $C_\infty[0,1]$ with any norm, the derivative is unbounded (take $e^{\lambda t}$ for example). Similarly, if we consider $\{\sin(nt), \cos(nt)\}$ as a Schauder basis on $C_{\operatorname{per}[0,2\pi]}$, $D\sin(nt) = n\cos(nt)$.

Example 98. On $C[0,1]$, there are two common integral operators:

---

[10]Robin went into how he doesn't see the point of mowing lawns.

Fredholm $f(x) \mapsto Kf = \int_0^1 f(y)k(x,y) \, \mathrm{d}y$, where $k(x,y)$ is some kernel. K is bounded in the sup norm if $\max_{0 \le x \le 1} \int_0^1 |k(x,k)| \, \mathrm{d}y$ is finite. We see that

$$\left\| \int_0^1 \right\| f(y)k(x,y) \, \mathrm{d}y = \sup_x \left| \int f(y)k(x,y) \, \mathrm{d}y \right| \le \|f\| \sup_x \int |k(x,y)| \, \mathrm{d}y.$$

Volterra $f(x) \mapsto Vf(x) = \int_0^x f(y)k(x,y) \, \mathrm{d}y$. Again $\|V\| = \sup_x \int_0^1 |k(x,y)| \, \mathrm{d}y$. Note that the range of V is not closed because Vf is smoothing $Vf \in C^1$.

Example 99. Let $\ell_\infty(\mathbb{N}) = \{(x_1, \ldots, x_n, \ldots)\}$ with the sup norm. The right shift S and left shift T are clearly bounded. S is not onto and the image of S is closed. Also, $\mathrm{Ker} T = \{(x, 0, \ldots)\}$. An example of this is integrals and derivatives of power series with basis $\left\{ \frac{x^n}{n!} \right\}$.

Theorem 100 (BLT). *If X is a normed vector space, Y Banach, and M a dense linear subspace of X, then any bounded operator $T : M \to Y$ has a unique extension $\overline{T} : X \to Y$ such that $\overline{T}x = Tx$ for all $x \in M$ and $\|\overline{T}\| = \overline{T}$.*

Proof. Clearly we define $\overline{T}$ to be the limit $\overline{T}x = \lim Tx_n$ where $x_n \to x$. Clearly this is well-defined (otherwise boundedness is contradicted). Also note that $\overline{T}$ is clearly bounded by boundedness of T and that $\overline{T}x = Tx$ for $x \in M$. By continuity, this is linear.

To prove uniqueness, let $\overline{T}, \widetilde{T}$ extend T. Then by continuity, they must be equal (M is dense). $\square$

Remark 101. The above theorem allows us to pass to the limit from, for example, C^∞.

Definition 102 (Linear Functional). A linear functional on X is a bounded linear map $x \to \mathbb{R}$.

Because $\mathbb{R}$ is complete, the set of such functionals is complete whether or not X is complete. We call this space X^*, the dual space of X.

6.4   Lecture 13 (Mar 7) Robin began class questioning the point of shoveling snow. He says that it is necessary so you can drive your car away from your house.

Let $\mathcal{B}(X, Y)$ be the space of bounded linear maps $X \to Y$. A natural question is: when can we invert T? Robin drew a diagram and mentioned something about us not taking his linear algebra class. In $\mathbb{R}^n$, we have the transpose $T^t : Y \to X$ (note that this is not possible unless we have an inner product).

Theorem 103 (Rank-Nullity). *For finite dimensional vector spaces $\mathbb{R}^n$, we have $\dim(\mathrm{im}\, T) = \dim(\mathrm{im}\, T^t)$, $\ker T \perp \mathrm{im}\, T^t$.*

The restriction of T is invertible as a map $T : (\ker T)^\perp \to \mathrm{im}\, T$. In finite dimensions, a necessary algebraic condition for solvability of $Tx = y$ is $\ker T = 0$. This is not enough. In our context, we want a bounded (continuous) inverse.

Theorem 104. *A bounded map $T : X \to Y$ of Banach spaces has a bounded inverse if and only if $\ker T = 0$ and $\mathrm{im}\, T$ is closed in Y.*[11]

---

[11] Proof of this is not given.

Proof. One of the directions is easy. To prove the converse, we use the open mapping theorem.

Assuming the open mapping theorem, we note that T open implies T^{-1} continuous. $\qquad\square$

Theorem 105 (Open mapping theorem)**.** *A bounded surjective map is open.*

Sketch of Proof. We show that $T(B_1(0))$ is open. Since T is surjective, write $Y = \bigcup_{n=1}^{\infty} T(B_n(0))$. Then apply the Baire Category Theorem. Therefore there exists n such that $\overline{T(B_n(0))}$ has an interior. Eventually, we prove that there exists ε such that $B_\varepsilon(0) \subset T(B_n(0))$. $\qquad\square$

Corollary 106. *If V is a vector space with norms $\|-\|_1$ and $\|-\|_2$ such that $\|v\|_1 \le K\|v\|_2$ and if V is complete with respect to both norms, then the norms are equivalent.*

Note that $\mathcal{B}(X,Y)$ has more structure, namely composition. In particular, $\mathcal{B}(X)$ is a Banach algebra. This allows us to consider nonlinear functions of $T \in \mathcal{B}(X)$.

Example 107. Some examples are exponentials and the Neumann series. To define the exponential, if $T \in \mathcal{B}(X)$, then we can define $e^{\alpha T}$ for any α by a power series. It is easy to see that the series is absolutely convergent. Some properties of the exponential are that $\frac{d}{d\alpha}e^{\alpha T} = Te^{\alpha T}$ and $e^T e^S = e^{T+S}$ is $TS = ST$.

The Neumann weries is defined as the inverse of the operator $I - T : X \to X$. This is given by the usual geometric series.

In general, in the Bamach algebra $\mathcal{B}(T)$, we can extend $f(T)$ for any given function $f : \mathbb{R} \to \mathbb{R}$ with convergent power series.

Definition 108. An operator $T : X \to Y$ is compact if for any bounded set $B \subset X$, the image TB is precompact in Y.

Sometimes, $\|T\|$ is too strong of a norm for our purposes. For $T \in \mathcal{B}(x,y)$, T_n converges uniformly if $\|T_n - T\| \to 0$. We will try to weaken this to broaden the class of convergent sequences. The first obvious weakening is pointwise convergence.

Example 109. Consider $\ell^p(\mathbb{N})$. Then let P_n be projection onto the first n components. We see that $\|P_n - P_m\| = 1$ for all $n \ne m$, so P_n is not Cauchy. However, $P_n \to \mathrm{id}$ pointwise.

Definition 110 (Weak convergence)**.** We weaken this notion further by declaring $T_n \to T$ weakly, written $T_n \rightharpoonup T$ if for each $x \in X$, $T_n x \rightharpoonup Tx$ in Y. Thus we use weak convergence in Y to define weak convergence in $\mathcal{B}(X,Y)$.

Definition 111 (Weak convergence in Y)**.** $y_k \rightharpoonup y$ weakly in Y if for all $f \in Y^*$, $f(y_k) \to f(y)$ in $\mathbb{R}$.[12]

6.5   Lecture 14 (Mar 19) We continue our discussion of dual spaces. Let X be a normed vector space and Y be Banach. Then $\mathcal{B}(X,Y)$, the space of bounded linear maps $X \to Y$ inherits completemenss from Y. Two of the most important Y are X and $\mathbb{R}$. Denote the algebraic dual space of X by X' and the topological dual by X^*. Note that $X^* \subseteq X'$ with the inclusion an equality when X is finite-dimensional and a strict inclusion when X is infinite-dimensional.

---

[12]This is analogous to convergence in measure.

Lemma 112. X^* *and* X *are isomorphic when* X *is finite dimensional.*

Proof. Choose a basis $e_1, \ldots, e_d$ of X. Then any $\varphi \in X^*$ satisfies $\varphi(x) = \varphi\left(\sum x_k e_k\right) = \sum x_k \varphi(e_k)$, so every φ is determined by its values on the basis. Choose a dual basis $w_1, \ldots, w_d$ given by $w_k(e_j) = \delta_{jk}$. Clearly the w_k are lienarly independent and any φ can be written as

$$\varphi = \sum \alpha_j w_j$$

by the expression for $\varphi(x)$ we computed above. The isomorphism is obvious now that they have the same dimension. $\qquad\square$

Remark 113. Note that this isomorphism is basis-dependent.

Example 114. If X is infinite dimensional, start with a countable independent set $\{e_1, \ldots, e_j \ldots\}$. Extend to a Hamel basis. Now given $f \in X$, write $f = \sum c_h h$ and note φ_j the usual way. Now form $\varphi = \sum n\varphi_n$. Then $\varphi \in X'$ but φ is unbounded.

Example 115 (Dual of $L^p(X, \mu)$). The answer is easy to find but hard to prove. Use Hölder's inequality. If $p'p'$ are conjugate and if $f \in L^p, g \in L^{p'}$, then $\|fg\|_1 \leq \|f\|_p \|g\|_{p'}$. Thus we expect that $(L^p)^* = L^{p'}$ for $1 \leq p < \infty$.

Sketch. The first step is to show the equality $\|g\|_{p'} = \sup_{\|f\|_p = 1} \|fg\|_1$. Take $f(x) = |g(x)|^{p'-1} \text{sgn}(g(x)) \in L^p$. Scale down with the appropriate powers and checking the desired equality is left as an exercise.

The next step is to show that $(L^p)^* \subset L^{p'}$. Start with $\varphi \in (L^p)^*$ and assume X has finite measure. Note that $\chi_E \in L^p$ for all p. For measurable E, define $\nu(E) = \varphi(\chi_E)$. Then $\|\chi_E\|_p = \mu(E)^{1/p}$. Also $\nu(E) \leq \|\varphi\|_{(L^p)^*} \|\chi_E\|_p = K\mu(E)^{1/p}$, where $K = \|\varphi\|_{(L^p)^*}$.

We show that ν is a signed measure and that $\nu \ll \mu$. It is clear that ν is finitely additive. Similarly, if $E = \cup_{j \geq 1} E_j$ a disjoint union, write $E = \cup_{j=1}^n E_j \cup F$. Then $\chi_E = \sum_1^N \chi_{E_j} + \chi_F$. Therefore

$$\nu(E) = \varphi(\chi_E) = \sum_{j=1}^N \varphi(\chi_{E_j}) + \varphi(\chi_F) = \sum_{j=1}^N \nu(E_j) + \nu(F).$$

From above, $\nu(\overline{E}) \leq K\mu(\overline{E})^{1/p}$, so if $p < \infty$, $\nu \ll \mu$. Also since $\mu(X) < \infty$, $\mu(F) \to 0$ as $N \to \infty$. Thus ν is countably additive.

Now we use Radon-Nikodym to obtain that there exists $g \in L^1(X)$ such that for all E measurable, $\nu(E) = \int_E g \, d\mu$. Thus $\nu(E) = \varphi(\chi_E) = \int_E g \, d\mu = \int g\chi_E \, d\mu$. Now φ extends to simple functions and then to measurable (and integrable) functions by taking limits. Therefore if $\varphi(f) = \int fg \, d\mu$ for $f \in L^1$. Then if $f \in L^p$,

$$\left| \int fg \, d\mu \right| = \varphi(f) \leq \|\varphi\|_{(L^p)^*} \|f\|_{L^p}.$$

Together with the first staement, this implies $g \in L^{p'}$ and that the norms are equal. $\qquad\square$

Remark 116. The assumption $p < \infty$ is essential, and in fact for X of finite measure, any finite measure ν on X with finite variation yields a bounded linear functional: $\langle \nu, f \rangle = \int f \, d\nu$. For general X, we require Radon measures.

So far, we have X, X^* and X^* is complete. Note that L^2 is self-dual and we know that L^p is reflexive for $1 < p < \infty$: $L^p \simeq (L^p)^{**}$. It is easy to show in general that $X \subset X^{**}$: $\langle x, \varphi \rangle = \varphi(x) = \langle \varphi, x \rangle$. Clearly this is bounded because $|\langle \varphi, x \rangle| \leq \|\varphi\|_{X^*} \|x\|_X$, so x is a bounded linear operator on X^*.

Linear functionals are very useful geometrically and analytically. We use them to define weak convergence and weak derivatives. A question is: can X^* separate a convex set from its exterior? if K is convex and $p \notin \overline{K}$, is there some $\ell \in X^*$ such that $\ell(x) \geq a$ for $x \in X$ and $\ell(p) < a$?

6.6   Lecture 15 (Mar 21)   Robin apologizes for not returning our homework. He's so organized.

We work in a vector space. Let K be bounded convex and $z \notin K$. We want to find a linear functional separating z from K such that $\ell(x) < a$ for $x \in K$ and $f(z) \geq a$. Note $\ell(x) = a$ defines a "hyperplane." Thus there is no loss of generality in taking $0 \in K$. We will build a "Minkowski gauge function" by scaling K around 0.

We will define $p_k(x) = \inf\{r \geq 0 \mid x \in rK\}$. Then $p_k(x) \leq 1$ if $x \in K$ and $p_k(x) \geq 1$ if $x \notin K$. Also $p_k(tx) = tp_k(x)$ for all $t \geq 0$ and $p_k(\alpha x + (1-\alpha)y) \leq \alpha p_k(x) + (1-\alpha)p_k(y)$ for all x, y and $\alpha \in [0,1]$.

A function p satisfying the last two conditions is called sublinear. The idea of Hahn-Banach is to extend a linear functional defined on the ray through z to a linear functinal on all of X, still dominated by p.

Theorem 117 (Hahn-Banach). *Let X be a vector space and $p : X \to \mathbb{R}$ sublinear. Then let $Y \subseteq X$ be a subspace and ℓ is a linear functional on Y such that $\ell(y) \leq p(y)$ for all $y \in Y$. Then ℓ extends to a linear functional φ on all of X, such that $\varphi(x) \leq p(x)$ for all $x \in X$ and $\varphi(y) \leq \ell(y)$ for all $y \in Y$.*

Proof. First we show that we can extend by one dimension, then we close the process using the axiom of choice. Take $z \notin Y$. Let $\widetilde{Y} = \mathrm{Sp}\{Y, z\}$. Then φ is defined on Y, so we need $\varphi(x)$. Note $\varphi(az + y) = a\varphi(z) + \varphi(y)$ for $y \in Y$. It is not enough to set $\varphi(z) = p(z)$ and extend by linearity. Choose two points $y_1, y_2 \in Y$ and $\alpha, \beta \in \mathbb{R}$. Consider

$$\beta\varphi(y_1) + \alpha y_1 = \varphi(\beta y_1 + \alpha y_2)$$
$$= (\alpha + \beta)\varphi\left(\frac{\beta}{\alpha + \beta}y_1 + \varphi(\alpha)\alpha + \beta y_2\right)$$
$$= (\alpha + \beta)p\left(\frac{\beta}{\alpha + \beta}(y_1 + \alpha z) + \frac{\alpha}{\alpha + \beta}(y_2 - \beta z)\right)$$
$$\leq \beta p(y_1 + \alpha z) + \alpha p(y_2 - \beta z).$$

Now divide by $\alpha\beta$ and separate:

$$\frac{1}{\alpha}(-p(y_1 + \alpha z) + \varphi(y_1)) \leq \frac{1}{\beta}(p(y_2 - \beta z) - \varphi(y_2)).$$

In conclusion, there exists $a \in \mathbb{R}$ such that

$$\sup_{\alpha, y} \frac{1}{\alpha}(-p(y_1 + \alpha z) + \varphi(y_1)) \leq a \leq \inf_{\beta, y_2} \frac{1}{\beta}(p(y_2 - \beta z) - \varphi(y_2)).$$

Now set $\varpi(z) = a$ and extend by linearity. Undoing this, we get $\varphi(x) \leq p(x)$ for all $x \in \widetilde{Y}$.

For the second step, use Zorn's lemma. Consider all extensions $(\widetilde{Y}, \varphi_{\widetilde{Y}})$. Partially order them in the obvious way (subspace and restriction agrees). Now consider a chain $\{(\widetilde{Y}_\alpha, \varphi_{\widetilde{Y}_\alpha})\}$. Then take $\widetilde{Y} = \cup_\alpha \widetilde{Y}_\alpha$. This is an upper bound for the chain. Therefore the set of extensions has a maximal element. Call it $(\hat{Y}, \varphi_{\hat{Y}})$. Clearly this must be all of X (otherwise apply step 1). $\qquad \square$

Remark 118. We can prove Hahn-Banach without requiring the axiom of choice.

Corollary 119. *Let X be a normed vector space and $Y \subset X$, and suppose $\lambda \in Y^*$. Then there exists $\Lambda \in X^*$ extending λ such that $\|\Lambda\|_{X^*} \le \|\lambda\|_{Y^*}$.*

Proof. Set $p(x) = \|\lambda\|_{Y^*} \|x\|$ and apply Hahn-Banach $\qquad \square$

Corollary 120. *For $y \in X$, there exists $\Lambda \in X^*$ such that $\lambda(y) = \|\Lambda\|_{X^*} \|y\|$.*

Proof. Set $Y = \text{Sp}(y), \lambda(y)\|y\|$. Apply the previous corollary. $\qquad \square$

Corollary 121. *Let $Z \subset X$ be a subspace. Let $y \in X$ with $d(y, z) = \delta > 0$. Then there exists a linear functional $\Lambda \in X^*$ satisfying $\|\Lambda\| \le 1$. $\Lambda(y) = \delta$, and $\lambda(z) = 0$ for all $z \in Z$.*

Theorem 122. *If X^* is separable, then X is separable.*

Proof. If $\{\lambda_n\}$ is dense in X^*, for each n, choose $x_n \in X$ such that $\|x_n\| = 1$ and $|\langle \lambda_n, x_n \rangle| \ge \|\lambda_n\|/2$. Consider

$$\mathcal{D} = \left\{ \sum_{n=1}^N r_n x_n \;\middle|\; r_n \in \mathbb{Z} \right\}.$$

We show $\overline{\mathcal{D}} = X$.

Suppose not. Then there exists $y \in X \smallsetminus \overline{\mathcal{D}}$ with $d(y, \mathcal{D}) = \delta > 0$. By Corollary 121, there exists Λ such that $\Lambda(x) = 0$ for all $x \in \mathcal{D}$. Because $\Lambda \in X^*$, there exists a subsequence $\lambda_{n_k} \to \Lambda$. Then

$$\|\Lambda - \lambda_{n_k}\| \ge \|\langle \Lambda - \lambda_{n_k}, x_{n_k} \rangle\| = \|\langle -\lambda_{n_k}, x_{n_k} \rangle\| \ge \|\lambda_{n_k}\|/2$$

. This implies that $\Lambda = 0$, which is a contradiction. $\qquad \square$

Remark 123. The converse of the above is not necessarily true. For example L^1 is separable but $(L^1)^* = L^\infty$ is not separable.

7   Hilbert Spaces

7.1   Lecture 16 (Mar 26) We will add more structure to our vector spaces. We revert to complex vector spaces. These are not conceptually harder, just computationally harder. Our inspiration comes from Pythagoras. The Euclidean distance comes from the dot product in $\mathbb{R}^n$.

Definition 124 (Inner Product)**.** An inner product in a complex vector space is a "sesquilinear form," i.e. a map $(\ ,\ ) : X \times X \to \mathbb{C}$ such that:

1. $(x, \alpha y + \beta z) = \alpha(x, y) + \beta(x, z)$ where $\alpha, \beta \in \mathbb{C}$;
2. $(x, y) = \overline{(y, x)}$ (Hermitian Symmetric);

3. $(x, x) \geq 0$. In particular, $(x, x) \in \mathbb{R}$;
4. $(x, x) = 0$ if and only if $x = 0$.

We observe that the inner product is conjugate linear in the first variable and that sesquilinearity is needed to ensure non-negativity. We refer to X equipped with $(\, , \,)$ as an inner product space.

Definition 125 (Hilbert Space)**.** A Hilbert space is a complete inner product space.

Example 126. Consider complex matrices $\mathbb{C}^{m \times n}$ equipped with the inner product $(A, B) = \mathrm{tr}(A^H B)$. This is called the "Hilbert-Schmidt norm."

Example 127. Consider $L^2(X, \mu)$ with functions to $\mathbb{C}$. Then set $(f, g) = \int_X \overline{f(x)} g(x) \, \mathrm{d}\mu$. This is an inner product space, but is not Hilbert in general. We simply take the completion.

Example 128. Now work with $C^{(k)}(X)$ complex valued, k-times differentiable f. Set

$$(f, g) = \sum_{j=0}^{k} \int_X \overline{f^{(j)}(x)} g^{(j)}(x) \, \mathrm{d}\mu;$$

and set $H^k(X)$ to be the completion of $C^k(X)$.

Lemma 129. *An inner product $(\, , \,)$ induces a norm via $\|x\| = \sqrt{(x, x)}$.*

Proof. Linearity is obvious ($\sqrt{\lambda \overline{\lambda}} = |\lambda|$) for all $\lambda \in \mathbb{C}$ and nonnegativity is also trivial. The triangle inequality follows from Cauchy-Schwarz. $\qquad \square$

Theorem 130 (Cauchy-Schwarz)**.** *For all $x, y \in X$, we have $|(x, y)| \leq \|x\| \|y\|$ with inequality if and only if x, y are linearly dependent.*

Proof. Set $\lambda = (x, y) / \|y\|^2$. Then

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

This implies the desired result. $\qquad \square$

Theorem 131 (Parallelogram Law)**.** *In any inner product space, $\|a + b\|^2 + \|a - b\|^2 = 2(\|a\|^2 + \|b\|^2)$.*

Remark 132. In fact, a norm on a vector space is induced from an inner product if and only if the parallelogram law holds.

Corollary 133. *L^p is not Hilbert unless $p = 2$.*

The law of cosines holds in any inner product space. More importantly, a, b are orthogonal if and only if $(a, b) = 0$.

Theorem 134. *Let $A \subset H$ be a subset of a Hilbert space. Define $A^\perp = \{x \mid x \perp a \text{ for all } a \in A\}$. Then $A^\perp$ is a closed linear subspace of H.*

Proof. It is clear that this is a subspace. Closedness is a consequence of continuity of the inner product. $\qquad\square$

A partial converse is:

Theorem 135. *If M is a closed linear subspace of H. Then:*

1. *For all $x \in X$, there exists a unique $y \in M$ such that $\|x - y\| = \min_{z \in M} \|x - z\|$;*
2. *y is the unique element such that $(x - y) \perp M$.*

Proof. Recall that in finite dimensions, we simply use a sequence of projections. Set $\delta = \inf_{z \in M} \|x - z\|$. There exists a sequence $z_n \in M$ such that $\|x - z_n\| \le \delta + 1/n$. We will show that the limit exists and that this element is unique. $\qquad\square$

7.2 Lecture 17 (Mar 28) We work in a Hilbert space H with M a closed subspace. We conclude the proof of Theorem 135 from last time.

Proof of Theorem 135, continued. We set $\delta = d(x, M)$. There exists a sequence $z_n \in M$ such that $\|x - z_n\| \le \delta + 1/n$. We will show that the limit exists and that this element is unique. Consider the triangle with vertices x, z_n, z_m. We will use the parallelogram law:

$$\begin{aligned}
\|z_n - z_m\|^2 &= 2\|z_n - x\|^2 + 2\|z_m - x\|^2 - 4\|x - \overline{z}\|^2 \\
&\le 2(\delta + 1/n)^2 + 2(\delta + 1/m)^2 - 4\delta^2 \\
&= 4\delta/n + 4\delta/m + 2/n^2 + 2/m^2.
\end{aligned}$$

Therefore the sequence is cauchy and converges, so y exists.

To show uniqueness, suppose there are two such points y_1, y_2. Then $\|y_1 - y_2\|^2 = 2\|x - y_1\|^2 + 2\|x - y_2\|^2 - 4\|x - \overline{y}\|^2 \le 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$.

We show that $(x - y, z) = 0$. For all $\lambda \in \mathbb{C}$, we have that $\|x - y\|^2 \le \|x - y - \lambda z\|^2 = \|x - y\|^2 + |\lambda|^2 \|z\|^2 - 2\mathrm{Re}\lambda(x - y, z)$. Let $(x - y, z) = re^{i\theta}$. Then setting $\lambda = te^{-i\theta}$, we get that $0 \le t^2 \|z\|^2 - 2tr$. Dividing by t^2, we have $\frac{2r}{t} \le \|z\|^2$ for all nonzero t, which is only possibleif $r = 0$.

To show uniqueness, suppose there is another such y'. We know that $\|x - y'\| \ge \|x - y\|$. Using the parallelogram law, we know $\|x - y\|^2 = \|x - y\|^2 + \|y - y'\|^2$ because $(x - y, y - y') = 0$. However, we see $\|x - y\|^2 = \|x - y'^2\| + \|y - y'\|^2$ and thus $y = y'$. $\qquad\square$

Remark 136. We will reserve the direct sum $\oplus$ for an orthogonal direct sum because if $M \subset X$ is closed with X Banach, there may not be a closed space N such that $v = m + n$, $m \in M, n \in N$. In a Hilbert space, for any set X, we can write $H = S^\perp \oplus S^{\perp\perp}$.

Definition 137 (Orthogonal Subset)**.** U is an orthogonal subset if $u \perp v$ for all $u, v \in U$.

Definition 138 (Orthonormal Subset)**.** U is orthonormal if U is orthogonal and $\|u\| = 1$ for all $u \in U$.

Example 139. Consider 2π-periodic functions, or functions on $\mathrm{T}^n = (S^1)^n$. Then $e^{in \cdot x}$ for $n \in \mathbb{Z}^d, x \in \mathrm{T}^d$ are orthogonal in $L^2(\mathrm{T}^d)$. This forms a Fourier basis.

Example 140. We can also consider quasi-periodic functions. f is quasi-periodic if it is the sum of finitely many periodic functions. Let X be the space of periodic functions of the form $\sum_{k=1}^{n} a_k e^{i\omega_k t}$. We can define an inner product on X by $(f, g) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \overline{f(t)} g(t) dt..$ We see X is not complete. The completion of X consists of L^2 almost periodic functions $f(t) = \sum_{k=1}^{\infty} a_k e^{i\omega_k t}$. Note that $e^{i\omega t}$ are orthogonal for $\omega \in \mathbb{R}$. Thus we have an uncountable orthogonal set and thus $\overline{X}$ is not separable.

In general we have an uncountable basis. Abstractly, we must make sense of $\sum_{\alpha \in I} X_\alpha$ where I could be uncountable.

Definition 141 (Unconditional Convergence)**.** In a Banach space, we'll say the unordered sum $\sum_{\alpha \in I}$ converges if the partial sums S_J converge where J is finite: there exists J_ε finite such that $\|S_j - x\| < \varepsilon$ for any $J \supset J_\varepsilon$

We define absolute convergence like we do for $\mathbb{R}$.

Lemma 142. *Absolute convergence implies unconditional convergence. Also, for all n there exists J_n finite such that $\|S_{J_n} - x\| \le 1/n$ and set $J = \cup J_n$, which is countable. We see that $x_\alpha = 0$ for all $\alpha \notin J$.*

Definition 143. An unordered sum is Cauchy if for all $\varepsilon > 0$ there exists J_ε such that $\|S_k\| < \varepsilon$ for all finite K such that $K \subset I \smallsetminus J_\varepsilon$.

Lemma 144. *In a Banach space, an unordered sum converges if and only if it is Cauchy.*

Proof. Suppose the sum converges. Then there exists J_ε such that $\|S_J - x\| < \varepsilon$ for any finite $J \supset J_\varepsilon$. IF $K \subset I \smallsetminus J_\varepsilon$ is finite, set $J = J_\varepsilon \cup K$. Then $\|S_k\| = \|S_J - S_{J_\varepsilon}\| \le \|S_J - x\| + \|S_{J_\varepsilon} - x\| \le 2\varepsilon$.

Now suppose the sum is Cauchy. Choose J_n increasing such that $\|S_K\| < 1/n$ whenever $K \subset I \smallsetminus J_n$ is finite. Then $\|S_{J_m} - S_{J_n}\| < 1/n$, so S_{J_n} is Cauchy. Thus it converges to some x. Now given ε, choose n such that $1/n < \varepsilon/2$. Then set $J_\varepsilon = J_n$, so if $J \supset J_\varepsilon$ is finite, then $\|S_J - x\| \le \|S_J - S_{J_n}\| + \|S_{J_n} - x\| \le 2/n < \varepsilon$. $\qquad\square$

7.3   Lecture 18 (Apr 02)   Recall the definition of unconditional convergence. Our goal is Bessel.

Lemma 145. *If $U = \{u_\alpha \mid \alpha \in J\}$ is an orthogonal set in H, then $\sum_{\alpha \in I} u_\alpha$ converges unconditionally if and only if $\sum_{\alpha \in I} \|u_\alpha\|^2 < \infty$, and in this case $\|\sum u_\alpha\|^2 = \sum \|u_\alpha\|^2$.*

Proof. For any finite set J, $\|\sum_J u_\alpha\|^2 = \sum_J \|u_\alpha\|^2$. Therefore unconditional convergence is equivalent to Cauchy inplies that $\sum u_\alpha$ is Cauchy if and only if $\sum \|u_\alpha\|^2$ is Cauchy in $\mathbb{R}$. $\qquad\square$

Corollary 146. *$\sum_I c_\alpha u_\alpha$ converges unconditionally in H if and onlf if $c_\alpha = 0$ for all but countable many α.*

Theorem 147 (Bessel)**.** *If $U = \{u_\alpha \mid \alpha \in I\}$ is an orthonormal set in H, then*

(a) $\sum_{\alpha} \in I|\langle u_{\alpha,x}\rangle|^2 \le \|x\|^2$ for all $x \in H$.
(b) $x_U := \sum_{\alpha \in I}\langle u_{\alpha}, x\rangle u_{\alpha}$ converges unconditionally for all $x \in Hl$
(c) $x - x_U \in U^{\perp}$.

Proof. We must understand the sum in (b).

$$\left\| x - \sum_J \langle u_{\alpha}, x\rangle u_{\alpha} \right\|^2 = \langle x - \sum_J \langle u_{\alpha}, x\rangle u_{\alpha}, x - \sum_J \langle u_{\alpha}, x\rangle u_{\alpha}\rangle$$

$$= \langle x, x\rangle - \sum_{\beta \in J} \langle u_{\beta}, x\rangle\langle x, u_{\beta}\rangle - \sum_{\alpha \in J} \overline{\langle u_{\alpha}, x\rangle}\langle u_{\alpha}, x\rangle + \sum_{\alpha, \beta} \overline{\langle u_{\alpha}, x\rangle}\langle u_{\beta}, x\rangle\langle u_{\alpha}, u_{\beta}\rangle$$

$$= \|x\|^2 - \sum_{\alpha \in J} |\langle u_{\alpha}, x\rangle|^2$$

Therefore, for any finite J, we have

$$\sum_J |\langle u_{\alpha}, x\rangle|^2 = \|x\|^2 - \left\| x - \sum_J \langle u_{\alpha}, x\rangle \right\|^2 \le \|x^2\|$$

This implies that $\sum_{\alpha \in I}|\langle u_{\alpha,x}\rangle|^2$ converges if and only if $\|\sum_{u_{\alpha},x} u_{\alpha}\|^2$ converges.

If $v \in U$, then

$$\langle c, x - x_U\rangle = \langle v, x - \sum\langle u_{\alpha}, x\rangle u_{\alpha}\rangle = \langle c, x\rangle - \sum_{\alpha \in I}\langle u_{\alpha}, x\rangle\langle v, u_{\alpha}\rangle = 0.$$

$\square$

We will write $[U]$ for the closed linear span of U for an orthonormal set $U \subset H$.

Theorem 148. $[U] = \left\{ \sum c_u u \mid \sum|c_u|^2 < \infty \right\}$.

Theorem 149. *Let U be an orthonormal subset of H. Then the following are equivalent:*

(a) $\langle u_{\alpha}, x\rangle = 0$ for all $\alpha \in I$ only if $x = 0$;
(b) $x = \sum_{\alpha \in I}\langle u_{\alpha}, x\rangle u_{\alpha}$ for all $x \in H$;
(c) $\|x\|^2 = \sum|\langle u_{\alpha}, x\rangle|^2$ for all $x \in H$;
(d) $[U] = H$;
(e) U is a maximal orthonormal set in H.

We call such a U an orthonormal basis for H.

Proof. **(a) implies (b)** Note that $U^{\perp} = \{0\}$. Bessel implies that $x = x_U$.
(b) implies (c) In the proof of Bessel.
(c) implies (d) (c) states that $x \in [U]$. Thus $[U]^{\perp} - 0$, so $H \subset [U]$.
(d) implies (e) If there exists v such that $v \perp U$, then since $v \in H$, $v = \sum c_{\alpha} u_{\alpha}$ for some c_{α} such that $\sum|c_{\alpha}|^2 < \infty$. Since $\langle v, u_{\alpha}\rangle = 0$ we must have $c_{\alpha} = 0$ and thus $v = 0$.
(e) implies (a) This is obvious.

$\square$

We conclude that if we have an orthonormal basis, any $\langle f, g\rangle$ can be componentwise.

Theorem 150. *Any Hilbert space has an orthonormal basis.*

Proof. Use Zorn's lemma. Consider the collection $\{U\}$ of orthonormal subsets ordered by inclusion. Given a chain, the union of all sets in the chain is an upper bound. Thus the collection has a maximal element. By Theorem 149, it is an orthonormal basis. $\qquad\square$

7.3.1 Adjoints We return to Banach spaces. Let $A : X \to Y$ be a bounded linear map. We want to define a transpose. Recall the dual $X^* = \{\ell : X \to \mathbb{R} \text{ bounded linear }\}$. Thus $\langle \ell, x \rangle$ is given by evaluation. Then given a map $A : X \to Y$ we define an adjoint $A^+ : Y^* \to X^*$ by $\langle A^+ \lambda, x \rangle = \langle \lambda, Ax \rangle$. We see that A^+ is linear and bounded and satisfies $\|A^+\| = \|A\|$.

Recall rank-nullity. Then analogous statements hold for A, A^+. We introduce some notation: Given a subset $V \subset X$, the annihilator $N(V)$ is the subspace of X^* such that $f \in N(V)$ if and onlf if $\langle f, x \rangle = 0$ for all $x \in V$. Then the range of A is closed in Y because A is bounded. We want that $R(A) = N(A^+)^\perp$. Thus $R(A) = \mathrm{Ker}(A^+)^\perp$.

If $y \in R(A) \subset Y$, then $y = Ax$ for some x. Under the identification $Y \subset Y^{**}$, we have $R(A) \subset \ker(A^+)^\perp$.

7.4 Lecture 19 (Apr 04) Last time we defined the adjoint on a Banach space. Robin is going to do a lot of handwaving as he talks about integration by parts. To motivate this discussion, consider L^p (our spaces have not yet been properly defined). We consider C_c^∞ the space of compactly supported C^∞ functions on $\mathbb{R}$. These are dense in all interesting spaces and vanish at the boundary.

Now define norms which include derivatives: $X = \overline{C_c^\infty}$ in the norm $\|u\|_p + \|u'\|_p$. Then $Y = L^p$ and our map is D, which is differentiation. We see that differentiation is linear and that $\|Du\|_Y = \|u'\|_{L^p} \leq \|u\|_X$. We attempt to find the adjoint using $(L^p)^* = L^q$. Then the adjoint of D satisfies $\langle D^* \psi, u \rangle = \int D^* \psi \cdot u \, \mathrm{d}x$, which must be

$$\langle \psi, Du \rangle = \int \psi \cdot Du \, \mathrm{d}x = -\int (D\psi)u \, \mathrm{d}x.$$

We conclude that D^* should be $-D$, whatever this means. Note the domain of D^* should be bigger than L^q.

Remark 151. Our couplings used to define adjoings are in some sense based on the L^2 inner product.

In a Hilbert space H, we have a natural pairing, which is given by the inner product. We would lise to define the adjoint of $A : H \to H$ by $(A^\dagger x, y) = (x, Ay)$.[13] This is consistent because any Hilbert space is self-dual.

Theorem 152 (Riesz Representation Theorem). *Let H be a Hilbert space. Then for all $\varphi \in H^*$ there exists a unique $y \in H$ such that $\varphi = (y, -)$.*

We develop some machinery to prove Riesz. Let X be a vector space. Then $P : X \to X$ is a projection is $P^2 = P$. We can enforce an orthogonal projection if $(x, Py) = (Px, y)$. In a Hilbert space, $\|P\| = 1$. For a proof, consider the inner product of Px with itself. Define the reflection $\mathcal{R}f$ by $\mathcal{R}f(x) = f(-x)$. Then $\frac{I \pm \mathcal{R}}{2}$ are projections onto even and odd parts of f, respectively.

---

[13]Note that Robin used + last time, $*$ this time, and $\dagger$ most recently.

Proof of Riesz Representation. If $\varphi = 0$ then we take $y = 0$. Thus assume $\varphi \neq 0$. Then there exists $z \in H$ such that $\varphi(z) = 0$ and $z \in (\ker P)^\perp$. Define $P : H \to H$ by $Px = \varphi x \cdot \frac{z}{\varphi(z)}$. It is clear that P is a projection. Also note that $P(y) = 0$ if and only if $\varphi(y) = 0$. Thus $H = \ker \varphi \oplus \operatorname{Ran} P$, which is an orthogonal direct sum.

Writing $x = w + \alpha z$, we see that $\varphi(x) = \varphi(w) + \alpha \varphi(z)$. Thus we want $y = \beta z$ such that $\varphi(x) = (y, x)$. We see that $\beta = \frac{\overline{\varphi(x)}}{\|z\|^2}$. Then for all x, we can calculate that $(y, x) = \varphi(x)$.

To show uniqueness, note that $(y - y_1, x) = 0$ for all x if and only if $y = y_1$. $\qquad \square$

7.5   Lecture 20 (Apr 09)

Last time we proved that a Hilbert space is self-dual. Define $I : H^* \to H$ by the mapping given by the Riesz Representation Theorem. Adjoints in Hilbert spaces can now be defined using the inner product and we can define the adjoint by $(x, Ay) = (A^*x, y)$. We want to generalize the rank-nullity theorem. Because $(z, Ax) = (A^*z, x)$, then $A^*z = 0$ implies $z \perp \operatorname{ran} A$.

Theorem 153. *Let $A : H \to H$ be linear. Then $\overline{\operatorname{ran}(A)} = (\ker(A^*))^\perp$ and $\ker A = (\operatorname{ran}(A^*))^\perp$.*

Proof. If $x \in \operatorname{ran} A$, then $x \perp \ker(A^*)$. If $x \in \operatorname{ran}(A)^\perp$, then $0 = (x, Ay) = (A^*x, y)$ for all y. Thus $x \in \ker A^*$. However we know that the closure of $\operatorname{ran}(A)$ contains $(\ker A^*)^\perp$. This gives the first part.

For the second part, apply the first part to $A^{**} = A$. $\qquad \square$

This theorem provides a necessary solvability condition for the abstract equation $Ax = y$, which is that y is orthogonal to the kernel of A^*. One context for this is partial differential equations. We take a PDS and look for perturbations of some known solution: $u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots$. Then we collect powers of ε, which typically leads to separate PDEs for $u_1, u_2, \ldots$. The goal is to approximate the solution by u_1. We need u_2 to have some bounded solution, which is where we write the solvability condition.

The second application is the relation between uniqueness and existence. Suppose that the adjoint problem $A^*x = y$ has a unique solution. Then $\ker A^* = 0$ Thus by the theorem, if A has closed range, then $Ax = y$ has a solution for any y.

Lemma 154. *If $A^*x = y$ has a unique solution and if $\|Ax\| \geq c\|x\|$ for some $c > 0$, then $Ax = y$ has a solution for all y.*

Definition 155. $A : H \to H$ has the Fredholm alternative if either:

1. $Ax = 0, A^*x = 0$ implies $x = 0$, and both $Ax = y$ and $A^*x = y$ have unique solutions;
2. $A^* = 0$ and $A^*x = 0$ have nontrivial finite-dimensional solution spaces of the same dimension, and $Ax = y$ has a solution iff $y \perp z$ for all z such that $A^*z = 0$, and the same holds for $A^*x = y$.

Definition 156. A is a Fredholm operator if $A : H \to H$ and

1. $\operatorname{ran} A$ is closed;
2. Both $\ker A$ and $\ker A^*$ are finite dimensional.

Define the index of A as $\dim \ker A - \dim \ker A^*$.

Proposition 157. *If A is Fredholm and K is compact, then $A + K$ is Fredholm.*

We will now discuss self-adjoint operators.

Definition 158. *A is self-adjoint if $A^* = A$.*

Self-adjoint operators satisfy the spectral theorem.

Lemma 159. *If A is self-adjoint, then:*

1. *$\|A\| = \sup_{\|x\|=1}(x, Ax)$;*
2. *$\|A^*A\| = \|A\|^2$, and if A is self-adjoint, then $\|A^2\| = \|A\|^2$;*

Theorem 160 (Spectral Theorem, compact case). *Let $A : H \to H$ be compact and self-adjoint on H. Then there is an orthonormal basis of eigenvectors: The non-zero eigenvalues form a countable set of reals, and we can write $A = \sum_k \lambda_k P_k$ where each P_k is an orthogonal projection onto the finite-dimensional eigenspace corresponding to λ_k. The series converges in operator norm.*

7.6   Lecture 21 (Apr 11) We consider the spectral theorem again. Recall the definition of a Fredholm operator. Note that $AA^*|_{\mathrm{ran}A}$ is invertible, and so is $A^*A|_{\mathrm{ran}A^*}$. If A is self-adjoint, then the three conditions coincidem and the only requirement is that the kernel is finite-dimensional. In particular, if A is invertible, it is Fredholm.

Recall the definitions of eigenvalue and eigenvector.

Proposition 161. *If A is self-adjoint, then*

1. *All eigenvalues are real.*
2. *Eigenvectors corresponding to different eigenvalues are orthogonal.*

Most of what we still will hold for normal operators, but we state them for self-adjoint operators.

Definition 162 (Normal Operator). *An operator is normal if it commutes with its adjoint.*

Example 163. Skew-adjoint operators $(A = -A^*)$ are normal.

Lemma 164. *Let $A : H \to H$ be self-adjoint. Then $\|A\| = \sup_{\|x\|=1}(x, Ax)$.*

Sketch of Proof. Let α be the supremum. Then $\alpha \leq \|A\|$ follows easily. Then note that $\|A\| = \sup|(y, Ax)|$, where y, x have norm 1. We can use the parallelogram law and choose y by a factor of $e^{i\varphi}$. $\qquad\qquad\square$

Definition 165 (Resolvent). Define the resolvent of A as $\rho(A) = \{z \in \mathbb{C} \mid A - zI \text{ is invertible }\}$. Note that because $A - zI$ is bounded, the inverse is also bounded. The spectrum of A is $\sigma(A) = \mathbb{C} \smallsetminus \rho(A)$.

There are two cases for $\lambda \in \sigma(A)$:

1. $A - \lambda I$ is Fredholm. Then there is a finite eigenspace and thus λ is an eigenvalue of finite multiplicity. The discrete spectrum is $\sigma_{dis}(A) = \{z \mid A - zI \text{ not invertible, but Fredholm }\}$. For each such z, there are a finite number of eigenvalues that span a finite dimensional space E_λ, and P_λ is the projection onto that space.
2. $A - \lambda I$ is not Fredholm. Define the essential spectrum $\sigma_{ess}(A) = \{z \in \mathbb{C} \mid A - zI \text{ is not Fredholm }\}$. We know that

(a) $\sigma_{ess}(A) \subset \sigma(A)$;

(b) σ_{ess} is closed in $\mathbb{C}$;

(c) If K is self-adjoint and compact, then $\sigma_{ess}(A) = \sigma_{ess}(A + K)$.

Theorem 166. *The following are equivalent:*

1. $\lambda \in \sigma_{ess}(A)$.
2. *Weyl's Criterion: there exists a sequence ψ_k in H with $\|\psi_k\| = 1$ such that $(a - \lambda I)\psi_k \to 0$, but $\{\psi_k\}$ has no convergent subsequence.*
3. $\dim(\ker(A - \lambda I)) = \infty$, *or there exist $\sigma_n \in \sigma(A)$ such that $\sigma_n \to \lambda$.*[14]

Remark 167. Robin believes that if A is normal, then there is no essential spectrum.

8   FOURIER SERIES

Consider $f : \mathbb{R} \to \mathbb{C}$ periodic with period 2π. We can regard $f : \mathbb{T} \to \mathbb{C}$. This becomes a Hilbert space with $(f, g) = \int_{\mathbb{T}} \overline{f(x)} g(x) dx$. The Fourier basis is the orthonormal basis $\{e_n(x)\}$ where $e_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}$. To get a hilbert space, we set $L^2(\mathbb{T}) = \overline{C(\mathbb{T})}$ in the corresponding L^2 norm.

Proposition 168. $\{e_n(x)\}$ *is a complete orthonormal basis for $L^2(\mathbb{T})$.*

We conclude that any function $f \in L^2(\mathbb{T})$ can be written

$$f = \sum_{n \in \mathbb{Z}} a_n e^{inx},$$

where $a_n = (e^{inx}, f)$. This sum converges provided $\sum |a_n|^2 < \infty$. Using Euler's identity, we can write $f(x)$ in the real Fourier sine/cosine series

$$f(x) = \sum_{n \geq 0} a_n \cos(nx) + \sum_{n \geq 0} b_n \sin(nx).$$

We have $a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(nx) dx$ and $b_n = \int_0^{2\pi} f(x) \sin(nx) dx$. Also $a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx$. The sine and cosine series are particularly nice for even and odd functions. A property of e_n plays very nicely with the derivative and is in fact an eigenfunction of the differentiation operator.

8.1   Lecture 22 (Apr 16)   We continue with our discussion of Fourier series. We will end lecture by solving a PDE. For simplicity, our period is 2π. We want to determine when and in what sense a Fourier series converges.

First, we see that $\sum a_n e^{inx}$ converges if $\sum |a_n| < \infty$, as in $a_n \in \ell_1$. Also, if $(a_n) \in \ell_2$, we get convergence.

Proposition 169. $\{e_n(x)\}$ *is a complete orthonormal basis on $C(\mathbb{T})$.*

The idea of the proof is to set $t = e^{ix}$ and follow the Weierstrass approximation theorem. For details, see the text.

---

[14]You should regard this morally as poles and essential singularities.

8.1.1 Distributions and Fourier Transform We will begin discussing Fourier transforms, which are more general than Fourier series.

Remark 170. Fourier transforms commute with convolution.

Theorem 171 (Carleson, 1966). *If $f \in L^2$, then $\sum a_n e^{ink} \to f$ pointwise almost everywhere, However, convergence fails at points of discontinuity of f.*[15]

Note that the power of Fourier series and the FFT is the speed with which this is calculated, which is $O(n \log n)$. Wavelets are also efficient, but they are discontinuous, so they avoid Gibbs phenomena and can be used in image processing.

We know that $D(e^{inx}) = ine^{inx}$, so $D(\sum a_n e^{inx}) = \sum in a_n e^{inx}$. Differentiating the Fourier series just scales the modes, so differentiating k times should lead to the Fourier series of k-th derivatives. Thus f is k-times differentiable if and only if $\sum |a_n|^2 n^{2k}$ converges. This is one way to define the Sobolev space $H^k(\mathbb{T})$.

Definition 172 (Fourier Transform). Let $u \in L^1(\mathbb{R}^d)$. Then the Fourier transform is

$$\widehat{u}(k) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-ik \cdot x} u(x) dx.$$

The inverse transform is given by

$$\overset{\vee}{v}(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{ik \cdot x} v(k) dk.$$

We think of x being physical space and k being frequency space.

Theorem 173 (Plancharel). *If $u \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, then $\widehat{u} \in L^2(\mathbb{R}^d)$ and $\|\widehat{u}\| = \|u\|$ in L^2.*

Idea of Proof. By our assumptions, the transform exists everywhere. Then we compute

$$\|\widehat{u}\|^2 = \int_k \overline{\widehat{u}(k)} \widehat{u}(k) dk$$

$$= \frac{1}{(2\pi)^d} \int \int \int e^{iky} \overline{u}(y) e^{-ikx} u(x) dx \, dy \, dk$$

$$= \frac{1}{(2\pi)^d} \int \int \int e^{ik(y-x)} dk \overline{u}(y) u(x) dx \, dy.$$

If $u, \widehat{u}$ have compact support, then we can integrate to get $\frac{1}{(2\pi)^{d/2}} \delta_{x,y}$. This all needs to be checked.

Taking this for granted, we get that the norm is $\int \overline{u}(x) u(x) dx = \|u\|^2$. $\square$

It follows that we can extend the Fourier transform to be defined on all of L^2. To do this, we approximate $u \in L^2$ be $u_n \in L^2 \cap L^1$ and take limits.

Proposition 174.   *1. $\int u(x) \overline{v}(x) dx = \int \widehat{u}(k) \overline{\widehat{v}}(k) dk$.*
     2. $\widehat{(u * v)} = (2\pi)^{d/2} \widehat{u} \widehat{v}$.

---

[15]This is known as the Gibbs phenomenon.

3. $(\widehat{u})^{\vee} = u$.
4. If $D^{\alpha}u \in L^2$, then $\widehat{D^{\alpha}u} = (ik)^{\alpha}\widehat{u}$ for $\alpha \in \mathbb{Z}^d$.[16]

8.2   Lecture 23 (Apr 18) We will define the notion of a distribution, following the note from Smoller on the webpage.

Definition 175 (Distribution)**.** A distribution is a continuous linear functional on C_0^{∞} the closure of C_C^{∞}, which is the space of compactly supported smooth function.

Here we have avoided the mention of a topology because C_0^{∞} is not metrizable. We define the strong topology:

Definition 176 (Strong Topology)**.** We declare $\phi_j \to 0$ in the strong topology if $\|D_{\phi_j}^{\alpha}\|_{\infty} \to 0$ for all nonzero multi-indices.

We call the $\phi \in C_0^{\infty}$ test dunctions. Then the disributions are continuous in the weak topology, where $T_j \to 0$ if $\langle T_j, \phi \rangle \to 0$ for all $\phi \in C_0^{\infty}$.

Example 177.   1. If $f \in L_{loc}^1$, then $\langle T_a, \phi \rangle = T_f(\phi) = \int_{\mathbb{R}^d} f(x)\phi(x)dx$.
   2. Evaluation at a, given by $\langle \delta_a, \phi \rangle = \delta_a(\phi) = \phi(a)$.
   3. Any Radon measure μ gives $\langle T_{\mu}, f \rangle = \int_{\mathbb{R}^d} \phi(x)d\mu$.
   4. Evaluation of derivatives: For $a \in \mathbb{R}^d, \alpha$ multiindex, $\langle T_{a,\alpha}, \phi \rangle = (-1)^{\alpha}D^{\alpha}\phi(a)$.

We define the support of a distribution T.

Definition 178. We say T vanishes on a set $U \subset \mathbb{R}^d$ if $\langle T, \phi \rangle = 0$ for all ϕ whose support is a subset of U. The support $\operatorname{supp} T = \{x \in \Omega \mid \text{ there exists } U \ni x, T = 0 \text{ on } U\}$.

8.2.1   Weak Derivatives These are defined by integration by parts. For example, if $u \in L_{loc}^1$ and α is a multi-index, V is the α-weak derivatives if integration by parts holds on all smooth test functions ϕ. That is, v is the α-weak derivative of u if for all test functions $\phi \in C_C^{\infty}$, $\int v\phi dx = (-1)^{|\alpha|}\int uD^{\alpha}\phi dx$.

We can define this for any distribution.

Definition 179 (Weak Derivative)**.** Let T be a distribution and α a multi-index. Then $D^{\alpha}T$ is defined by $\langle D^{\alpha}T, \phi \rangle = (-1)^{|\alpha|}\langle T, D^{\alpha}\phi \rangle$.

Note that we cannot multiply distributions, but we can multiply distributions by C^{∞} functions by $\langle fT, \phi \rangle = \langle T, f\phi \rangle$.

Example 180. If f is continuous at 0, we can define $\langle f\delta, \phi \rangle = f(0)\phi(0)$. However, if H is the Heaviside function, this does not make sense because $\delta = H'$, but $H = H^2$, so we don't know what to use.

An application of this is Burger's equation: $u_t + uu_x = 0$. This admits discontinuous solutions, so we can write it in the better form $u_t + \left(\frac{1}{2}u^2\right)_x = 0$.[17]

Recall that convolutions are like products. Then we can define

$$\langle f * g, \phi \rangle = \int f * g(x)\phi(x)dx = \int f(y)\left(g(z)\phi(z + y)dz\right)dy.$$

---

[16]Here we are using multi-index notation with attached multinomial coefficients.
[17]Thanks, physicists!

We can rewrite this as $Tf_y(Tg_z\phi(y+z))$. We can use this to define convolutions for distibutions.

8.3   Lecture 24 (Apr 23)   Recall the notion of a distribution. Today we will define the convolution of two distributions.

Definition 181 (Convolution of Distributions)**.** The convolution of two distributions S, T is defined by $\langle S * T, \phi \rangle = \langle S, \langle T, \phi(y+z) \rangle_z \rangle_y$. Here, we require at least one of S, T has compact support.

Lemma 182. *Suppose T is a distribution and S is a compactly supported distribution.*

1. *$T * S$ is a distribution.*
2. *$T * S = S * T$*
3. *If U is a compactly supported distribution then $(T * S) * U = T * (S * U)$*
4. *$\mathrm{supp}(T * S) \subset \mathrm{supp}\, T + \mathrm{supp}\, S$.*
5. *If $f \in C_0^\infty$ and T is a distribution, then regarding f as a distribution, $f * T(x) = \langle T, f(x-y) \rangle_y$.*

Remark 183. Note that the delta function is the unit with respect to convolution.

Lemma 184. *Convolution commutes with differentiation.*

Proof.

$$
\begin{aligned}
\langle D^\alpha(T * S), \phi \rangle &= (-1)^{|\alpha|} \langle T * S, D^\alpha \phi \rangle \\
&= (-1)^{|\alpha|} \langle T, \langle S, D^\alpha \phi(y+z) \rangle \rangle \\
&= \langle T, (-1^{|\alpha|}) \langle S, D^\alpha \phi(y+z) \rangle \rangle \\
&= \langle T, \langle D^\alpha S, \phi(y+z) \rangle \rangle \\
&= \langle T * D^\alpha S, \phi \rangle
\end{aligned}
$$

$\square$

We will use the same tricks to define the Fourier transform of some distributions. We have too many distributions, so we extend our class of test functions.

Definition 185. Schwartz space $\mathcal{S}$ is the set of those $\varphi \in C^\infty$ such that for all multi-indices $\alpha, \beta \in \mathbb{Z}_+^d$,
$$
\sup_{x \in \mathbb{R}^d} |x^\alpha D^\beta \varphi(x)| < \infty.
$$

Any $\varphi \in C^\infty$ such that $|\varphi| \le e^{-\varepsilon|x|^\delta}$ is in $\mathcal{S}$.

Note the following properties of multi-indices.

1. Taylor:
$$
f(x) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} D^\alpha f(x_0)(x-x_0)^\alpha + r_k(x),
$$
where $r_k(x) = o(|x-x_0|^k)$ and $\alpha! = \alpha_1! \cdots \alpha_d!$.
2. Leibniz:
$$
D^\alpha(fg) = \sum_{\beta + \gamma = \alpha} \frac{\alpha!}{\beta! \gamma!} (D^\beta f)(D^\gamma g).
$$

We consider the topology of $\mathcal{S}$. Set $P_{\alpha,\beta}(\varphi) = \sup_x |x^\alpha D^\beta \varphi(x)|$. This is not quite a norm but is a seminorm. Given a family of seminorms P_α on $\mathcal{Z}$, we define a topology on $\mathcal{Z}$ by $\mathcal{N}$ is an open neighborhood of z if $\mathcal{N} = \{y \mid p_\alpha(y - z) < \varepsilon$ for some finite set of $\alpha\}$. The p_α separate points if $p_\alpha(y) = 0$ implies $y = 0$ for all α. If the family separates points, the topology is Hausdorff. In this case $\mathcal{Z}$ is locally convex.

$\mathcal{Z}$ is also metrizable if A is countable. We need to take account of all the seminorms:

$$d(x,y) = \sum_n \frac{1}{2^n} \frac{p_n(x-y)}{1 + p_n(x-y)}$$

Theorem 186. *In this metric, $\mathcal{S}$ is complete.*

Proof. Consider a Cauchy sequence φ_n. Then $p_{0,0}(\varphi_n - \varphi_m) < \varepsilon$. Also by definition of the $p_{\alpha,\beta}$, the sequence is Cauchy in L^∞. Thus the limit $\widetilde{\varphi}$ is continuous. Then using $p_{0,\beta}$ we get a Cauchy sequence for the derivatives. We need to prove that the limits coincide with the derivarives of the limit $\widetilde{\varphi}$.

We do this by induction on $|\beta|$. Suppose $|\beta| = k+1$. Then there exists j such that $\beta = \alpha + e_j$ where $|\alpha| = k$. Now write

$$D^\alpha \varphi_n(x + te_j) - D^\alpha \varphi_n(x) = \int_0^t \frac{\partial}{\partial x_j} \left(D^\alpha \varphi_n(x + se_j) \right) ds = \int_0^t D^\beta \varphi_n(x + se_j) ds.$$

Taking the limit as $n \to \infty$, we see that

$$D^\alpha \widetilde{\varphi}(x + te_j) - D^\alpha \widetilde{\varphi(x)} = \int_0^t \psi_\beta(x + se_j) ds.$$

Dividing by t and taking $t \to 0$, we see that $\frac{\partial}{\partial x_j} D^\alpha \widetilde{\varphi(x)} = \psi_\beta(x)$ by continuity. This gives that $x^\alpha D^\beta \varphi_b$ is Cauchy and thus converges to $x^\alpha D^\beta \widetilde{\varphi}$. $\qquad \square$

8.4   Lecture 25 (Apr 25)   We continue our discussion of Schwarz space. We will apply the Fourier transform to $\mathcal{S}$ and see what happens. Recall that the Fourier transform morally decomposes a function into its pieces which look like $e^{ik\cdot x}$ with wavenumber k. We know $\mathcal{F} : L^1 \to L^\infty$ and $\|\widehat{f}(k)\| \le \|f\| \sup \frac{|e^{ikx}|}{(2\pi)^{d/2}}$. In fact, if $f \in L^1$, then the Fourier transform is continuous.

Not we introduce the translation operator $\tau_h f(x) = f(x - h)$. Then a simple calculation shows that $\widehat{\tau_h f} = e^{-ikh} \widehat{f}$. Assuming f is C^1, note that the Fourier transform of $\frac{1}{h}(\tau_{-h} f - f)$ is $\frac{e^{ikh} - 1}{h} \widehat{f}$. Taking the limit as $h \to 0$, we conclude that $\widehat{f'(x)} = ik\widehat{f}$.

Next we consider the Fourier transform of a convolution. This was on the homework last semester, so the calculation is omitted (just use Fubini):

$$\widehat{f * g} = (2\pi)^{d/2} \widehat{g}\widehat{f}.$$

Define $\mathcal{G}g = \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} e^{ikx}g(k)dk$. We will calculate

$$
\begin{aligned}
\mathcal{G}\widehat{f} &= \frac{1}{(2\pi)^{d/2}}\int e^{ikx}\frac{1}{2\pi}^{d/2} e^{-iky}f(y)dy\ dk \\
&= \frac{1}{(2\pi)^d}\int\int e^{ik(x-y)}f(y)dy\ dk \\
&= \frac{1}{(2\pi)^d}\int\left(\int e^{ik(x-y)}dk\right)f(y)dy
\end{aligned}
$$

We know that $\int e^{ik(x-y)}dx = c\delta(x-y)$. Then formally, $\widehat{\delta} = \frac{1}{(2\pi)^{d/2}}$. In order to justify this, we will need to consider tempered distributions. The tempered distributions are elements of $\mathcal{S}'$, the dual of $\mathcal{S}$. Note that $\widehat{D^\alpha f} = (ik)^\alpha\widehat{f}$. Thus $\mathcal{F}$ maps $\mathcal{S}$ to itself. Assuming f, g are nice enough, we can see that $\langle\widehat{f}, g\rangle = \langle\widehat{g}, f\rangle = \langle f, \widehat{g}\rangle$. Therefore, given $T \in \mathcal{S}'$, we define $\widehat{T}$ by $\langle\widehat{T}, \varphi\rangle = \langle T, \widehat{\varphi}\rangle$.

We conclude the calculation with

$$
\begin{aligned}
\mathcal{G}\widehat{f} &= \lim\frac{1}{(2\pi)^d}\int_{-M}^{M} e^{ikx}\int e^{-iky}f(y)dydk \\
&= \lim\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{-M}^{M} e^{ik(x-y)}dxf(y)dy \\
&= \lim\frac{1}{(2\pi)^d}\int \frac{1}{i(x-y)}e^{ik(x-y)}\Big|_{iM}^{M}f(y)dy \\
&= \lim\frac{1}{(2\pi)^d}\int \frac{e^{iM(x-y)} - e^{-iM(x-y)}}{i(x-y)}f(y)dy \\
&= \lim\frac{1}{(2\pi)^d}\int 2M\frac{\sin M(x-y)}{M(x-y)}f(y)dy.
\end{aligned}
$$

The sine term is a kernal approaching δ as $M \to \infty$, which will be confirmed later.

8.4.1 Yet another differential equation Consider a linear hyperbolic PDE $P(D)v = f$. Formally, we apply the Fourier transform. First we solve the easier problem $P(D)v = \delta$, where $P(D)$ is some element of the algebra of differential operators. We can apply the Fourier transform here, find the Fourier transform formally, and then invert. Because $\delta * f = f$, $v * f$ solves $P(D)v = f$.

8.5   Lecture 26 (Apr 30) We go back to the inversion formula.

Theorem 187. *If $f, g \in L^\infty \cap L^2 \cap L^1$ then $\mathcal{G} = \mathcal{F}^{-1}$.*

Proof. First restrict to $d = 1$. By Fubini, $d > 1$ follows immediately. Then for $d = 1$ we have

$$
\begin{aligned}
\mathcal{G}\widehat{f} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-iky} f(y) dy dk \\
&= \lim_{M\to\infty} \frac{1}{2\pi} \int_{-M}^{M} e^{ikx} e^{-iky} f(y) dy dk \\
&= \lim_{M\to\infty} \int_{\mathbb{R}} f(y) \left(\frac{1}{2\pi} \int_{-M}^{M} e^{ik(x-y)} dk \right) dy \\
&= \lim_{M\to\infty} S_M * f
\end{aligned}
$$

where $S_m(x) = \frac{1}{2\pi} \int_{-M}^{M} e^{ikx} dk = \frac{1}{\pi} \frac{\sin Mx}{x}$. Therefore $S_M(x) = \frac{M}{\pi} \operatorname{sinc} Mx$. We know that $\int_{-\infty}^{\infty} \operatorname{sinc} x = \pi$ (proved using complex analysis). Thus $\int S_M(x) = 1$. Then for M large, we have $S_m \to \delta$ as distributions. To justify this, we calculate:

$$
\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{\sin Mx}{x} \varphi(x) dx = \int_{|x|\geq 1} + \int_{|x|\leq 1}.
$$

Integrate by parts to get

$$
\begin{aligned}
\int_{1}^{R} \frac{1}{\pi} \frac{\sin Mx}{x} \varphi(x) dx &= \frac{1}{\pi} \left(\frac{-1}{M} \cos Mx \frac{\varphi(x)}{x} \right)\Big|_{1}^{R} - \int_{1}^{R} \frac{-1}{M} \cos Mx \frac{1}{\pi} \left(\frac{\varphi}{x} \right)' dx \\
&= O(1) \frac{1}{M} \to 0
\end{aligned}
$$

because $\varphi \in \mathcal{D}$. Now we see that

$$
\begin{aligned}
\int_{|x|\leq 1} \frac{1}{\pi} \frac{\sin Mx}{x} \varphi(x) dx &= \frac{1}{\pi} \int_{|x|\leq 1} \sin Mx \frac{\varphi(x) - \varphi(0)}{x} dx + \frac{\varphi(0)}{\pi} \int_{|x|\leq 1} \frac{\sin Mx}{x} dx \\
&= \frac{1}{\pi} \int_{|x|\leq 1} \sin(Mx) \psi(x) dx + \frac{\varphi(0)}{\pi} \int_{|x|\leq 1} \frac{\sin Mx}{x} dx.
\end{aligned}
$$

The first integral yields $1/M$ again, and the second integral becomes $\frac{\varphi(0)}{\pi} \int_{|y|\leq M} \operatorname{sinc}(y) dy \to 0$ as $M \to \infty$. $\qquad\square$

We consider more examples. For example, if $f \in L^1$, this determines a distribution. However, if $f \notin L^1$, can we still define a corresponding distribution? (As an example, take $1/x$.) Generally, it is not really possible. We lose "balance" by using φ weighted artificially to one side of the singularity at 0. On the other hand, if φ is even, we can use the "principal value" distribution:

The principal value $p.v.1/x$ is defined by $\langle p.v.1/x, \varphi \rangle = \lim_{\varepsilon\to 0} \int_{|x|\geq\varepsilon} \frac{\varphi(x)}{x} dx$. We show that this is a distribution. Write

$$
\int_{|x|\geq\varepsilon} \frac{\varphi(x)}{x} dx = \int_{\varepsilon}^{\infty} \frac{\varphi(x) - \varphi(-x)}{x} dx \to \int_{0}^{\infty} \frac{\varphi(x) - \varphi(-x)}{x} dx.
$$

Now observe that $\left| \frac{\varphi(x)-\varphi(-x)}{x} \right| = \left| \frac{1}{x} \int_{-x}^{x} \varphi'(t)dt \right| \le 2\|\varphi'\|_\infty$, wo the integrand is bounded. Now write

$$
\begin{aligned}
|\langle p.v.1/x, \varphi \rangle| &= \left| \int_0^\infty \frac{\varphi(x)-\varphi(-x)}{x} dx \right| \\
&\le \int_0^1 \left| \frac{\varphi(x)-\varphi(-x)}{x} dx \right| + \int_1^\infty \frac{|x(\varphi(x)-\varphi(-x))|}{x^2} dx \\
&\le 2\|\varphi'\| + 2\|x\varphi(x)\| \int_1^\infty \frac{1}{x^2} dx
\end{aligned}
$$

Now consider the fimal example, which is the PDE $P(D)u = f$ from last time. Given a polynomial $P(D)$, its symbol is the polynomial $p(ik)$. Observe that $\widehat{Df} = \int e^{-ikx} Df \; dx = -\int D(e^{-ikx})f \; dx = ik\widehat{f}$. Therefore a polynomial operator corresponds to a polynomial in Fourier space, and we can do this for more general functions. If a is a function of (ik), we can define an operator in physical space corresponding to a, which extends polynomials to power series and allows so-called "Fractional Laplacians."

For example, in $\mathbb{R}^1$, the Laplacian has $\widehat{\Delta u} = (ik)^u \widehat{u} = -k^2 \widehat{u}$. Thus we can define the operator $\sqrt{-\Delta}$ by $\sqrt{-\Delta}u = |k|\widehat{u}$. This is a radially symmetric derivative. Similarly, the operator $\wedge$ defined by $\widehat{\wedge} = (1+k^2)^{1/2}$ is very useful for defining Sobolev spaces. The fractional Sobolev space can be defined by $H^s = \{u \mid \wedge^s u \in L^2\}$.