MATH 175 NOTES

MOOR XU
NOTES FROM A COURSE BY LEON SIMON

Abstract. These notes were taken during Math 175 (Functional Analysis) taught by Leon Simon in Spring 2011 at Stanford University. They were live-TEXed during lectures in `vim` and compiled using `latexmk`. Each lecture gets its own section. The notes are not edited afterward, so there may be typos; please email corrections to `moorxu@stanford.edu`.

1. 3/28

You've probably got a vague notion of what functional analysis is about. It is the study of continuous linear operators on infinite dimensional spaces. This is interesting and has applications.

1.1. **Review of vector spaces.** Recall the basics of vector spaces.

Definition 1.1. A *linear space* or *vector space* X has two operations: addition and multiplication. For $u, v \in X$ and a scalar λ, we can define $u + v$ and λu. For this course, scalars will be in either $\mathbb{R}$ or $\mathbb{C}$, and X will be called either a *real* vector space or a *complex* vector space.

There are eight vector space axioms.

(1) $u + v = v + u$.
(2) $u + (v + w) = (u + v) + w$.
(3) There exists 0 such that $u + 0 = u$ for all $u \in X$.
(4) If $u \in X$ then there exists $-u \in X$ with $u + (-u) = 0$. (We usually write $u - v$ instead of the more cumbersome $u + (-v)$.
(5) $\lambda(\mu u) = (\lambda\mu)u$ for all $u \in X$ and scalars λ, μ.
(6) $\lambda(u + v) = \lambda u + \lambda v$.
(7) $(\lambda + \mu)u = \lambda u + \mu u$.
(8) $1u = u$ for all $u \in X$.

Recall the concepts of linear dependence, linear independence, and span.

Definition 1.2. $v_1, \dots, v_n \in X$ are *linearly dependent* if there exist scalars $c_1, \dots, c_n$ not all zero with $c_1v_1 + \cdots + c_nv_n = 0$. *Linearly independent* means "not linearly dependent."

Definition 1.3. If $A \subseteq X$ is a nonempty set,

$$\operatorname{span} A = \{c_1v_1 + \cdots + c_nv_n \mid n \geq 1, c_1, \dots, c_n \text{ scalars}\}.$$

Definition 1.4. $W \subset X$ is a *subspace* of X if $0 \in W$ and W is closed under addition and multiplication by scalars.

Definition 1.5. X is *finite dimensional* if there exist finitely many vectors $v_1, \ldots, v_n \in X$ with $X = \operatorname{span}\{v_1, \ldots, v_n\}$. Here, $v_1, \ldots, v_n$ is called a *basis* of X.

If X is not finite dimensional, we say that X is *infinite dimensional*.

Example 1.6. $\mathbb{R}^n = \{x = (x_1, \ldots, x_n) \mid x_i \in \mathbb{R}\}$ with addition defined as $x + y = (x_1 + y_1, \ldots, x_n + y_n)$ and multiplication by scalars defined as $\lambda x = (\lambda x_1, \ldots, \lambda x_n)$ for $\lambda \in \mathbb{R}$. This is a real vector space with the standard basis vectors.

Example 1.7. In the same way, we can define $\mathbb{C}^n$ as a complex vector space. This can also be viewed as a real vector space.

Viewed as a complex vector space, we clearly have $\dim \mathbb{C}^n = n$. Viewed as a real vector space, we have $\dim \mathbb{C}^n = 2n$.

We consider an example that is more relevant to functional analysis.

Example 1.8. Define

$$\ell^2(\mathbb{R}) = \left\{ x = \{x_j\}_{j=1,2,\ldots} \mid x_i \in \mathbb{R}, \sum_{j=1}^{\infty} x_j^2 < \infty \right\}$$

with addition and multiplication by scalars defined componentwise. Similarly, define

$$\ell^2(\mathbb{C}) = \left\{ z = \{z_j\}_{j=1,2,\ldots} \mid z_j \in \mathbb{C}, \sum_{j=1}^{\infty} |z_j|^2 < \infty \right\}$$

with the same operations. We should also check that they are vector spaces; this is easy to do.

Example 1.9. Define $C(\mathbb{R}) = \{\text{continuous functions } \mathbb{R} \to \mathbb{R}\}$ with the natural operations $(f + g)(x) = f(x) + g(x)$ and $(\lambda f)(x) = \lambda f(x)$. This is a real vector space. Similarly, we can define $C(\mathbb{C}) = \{\text{continuous functions } \mathbb{C} \to \mathbb{C}\}$ with the same operations; this is a complex vector space.

Notice that the set of real polynomials $\{p(x) = a_0 + a_1 x + \cdots + a_n x^n\}$ is a subspace of $C(\mathbb{R})$.

1.2. **Inner product spaces.**

Definition 1.10. A *complex inner product space* is a complex vector space with an *inner product*, denoted $(u, v) \in \mathbb{C}$. The inner product is a map $(\cdot, \cdot) : X \times X \to \mathbb{C}$ with the properties

(1) $(u, v) = \overline{(v, u)}$
(2) $(\lambda u + \mu v, w) = \lambda(u, w) + \mu(v, w)$ (linear in the first component)
(3) (u, u) is real and positive for $u \neq 0$.

A real inner product is defined analogously, with $\mathbb{C}$ replaced by $\mathbb{R}$.

Remark. Note that we have $(\lambda u, v) = \lambda(u, v)$, but $(u, \lambda v) = \overline{\lambda}(u, v)$. Also, check that $(u, v + w) = (u, v) + (u, w)$.

Example 1.11. $X = \mathbb{R}^n$ with (real) inner product $(x, y) = x \cdot y$ defined as the dot product.

Example 1.12. $X = \mathbb{C}^n$ with (complex) inner product $(z, w) = \sum_{j=1}^{n} z_j \overline{w}_j$.

Example 1.13. For $X = \ell^2(\mathbb{R})$, we can define the inner product as $(x, y) = \sum_{j=1}^{\infty} x_j y_j$.
Similarly, for $X = \ell^2(\mathbb{C})$, we can define the inner product as $(v, w) = \sum_{j=1}^{\infty} z_j \overline{w}_j$.
We should check that these series converge absolutely. This is an easy exercise.

Definition 1.14. Define the *inner product norm* or *length* of u to be

$$\|u\| = \sqrt{(u, u)}.$$

We can now derive some properties. We begin with a basic identity.

Proposition 1.15. $(u + v, u + v) = (u, u) + (u, v) + (v, u) + (v, v)$. *Therefore,* $\|u + v\|^2 = \|u\|^2 + (u, v) + \overline{(u, v)} + \|v\|^2$*, so hence*

$$\|u + v\|^2 = \|u\|^2 + 2 \operatorname{Re}(u, v) + \|v\|^2 .$$

Similarly, we have

$$\|u - v\|^2 = \|u\|^2 - 2 \operatorname{Re}(u, v) + \|v\|^2 .$$

Adding these gives the "parallelogram identity"

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

2. 3/30

Proposition 2.1. *If* $\sum_{j=1}^{\infty} x_j^2$ *and* $\sum_{j=1}^{\infty} y_j^2$ *are convergent, then* $\sum_{j=1}^{\infty} x_j y_j$ *is absolutely convergent.*

Proof. Recall the Cauchy-Schwarz inequality says that $|x_j y_j| \le \frac{1}{2}(x_j^2 + y_j^2)$. Then

$$\sum_{j=1}^{N} |x_j y_j| \le \frac{1}{2} \sum_{j=1}^{N} (x_j^2 + y_j^2) \le C.$$ □

2.1. Inner product spaces.

Recall that we were talking about inner product spaces. We have the inner product norm $\|u\| = \sqrt{(u, u)}$. This has a number of properties.

(1) $\|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2 \operatorname{Re}(u, v)$.
(2) $\|\lambda u\| = |\lambda| \, \|u\|$.

Proof. $\|\lambda u\| = \sqrt{(\lambda u, \lambda u)} = \sqrt{\lambda \overline{\lambda} (u, u)} = |\lambda| \, \|u\|$. □

(3) $|(u, v)| \le \|u\| \, \|v\|$ for all $u, v \in X$. This is the *Cauchy-Schwarz inequality.*

Proof. Exercise. □

(4) Triangle inequality: $\|u + v\| \le \|u\| + \|v\|$

Proof. Properties (1) and (3) imply that

$$\|u + v\|^2 \le \|u\|^2 + \|v\|^2 + 2|(u, v)| \le \|u\|^2 + \|v\|^2 + 2 \|u\| \, \|v\| = (\|u\| + \|v\|)^2.$$ □

2.2. **Norms.** Let X be any linear space.

Definition 2.2. $\|\cdot\|$ is a norm on X if
(a) $\|\lambda u\| = |\lambda| \|u\|$ for all scalars λ and for all $u \in X$.
(b) $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in X$.
(c) $\|u\| > 0$ if $u \neq 0$.

Example 2.3. The above properties are true if X is an inner product space and the $\|u\| = \sqrt{(u,u)}$. Every inner product gives a norm, but not all norms come from inner products.

2.2.1. *Relation to metric spaces.*

Definition 2.4. Define $d(u, v) = \|u - v\|$. This is called the *norm metric*.

Recall the three properties of a metric:

(1) $d(u, v) = d(v, u)$
(2) $d(u, v) \leq d(u, w) + d(w, v)$ for all u, v.
(3) $d(u, v) \geq 0$ and $d(u, v) = 0$ if and only if $u = v$

We can check that the norm metric is indeed a metric. For example, the triangle inequality follows from the triangle inequality for norms: $d(u, v) = \|u - v\| = \|u - w + w - v\| \leq \|u - w\| + \|v - w\| = d(u, w) + d(v, w)$.

Now that we have a metric, we can review some metric space terminology.

Definition 2.5. $B_\rho(u) = \{v \in X : d(v, u) < \rho\} = \{v \in X : \|v - u\| < \rho\}$ is the *open ball* of radius ρ and center u.

Similarly, $\overline{B}_\rho(u) = \{v \in X : d(v, u) \leq \rho\} = \{v \in X : \|v - u\| \leq \rho\}$ is the *closed ball* of radius ρ and center u.

Definition 2.6. $U \subset X$ is *open* if $u \in U$ implies that there exists $\rho > 0$ such that $B_\rho(u) \subset U$.

$C \subset X$ is *closed* if C contains all of its limit points, i.e. $y = \lim u_k$ with $u_k \in C$ for all k implies that $y \in C$.

Definition 2.7. A set $K \subset X$ is *compact* if for every sequence $\{x_k\}_{k=1,\ldots} \subset K$ there exists a convergent subsequence $\left\{x_{k_j}\right\}_{j=1,\ldots}$ with $\lim x_{k_j} \in K$.

Definition 2.8. X is *complete* if every Cauchy sequence in X converges in X.

Here's a diagram of what we've considered so far:

$$\text{inner product spaces} \subset \text{normed spaces} \subset \text{general metric space.}$$

Definition 2.9. A complete normed linear space is called a *Banach space*.

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

Remark. This theory might seem abstract, but functional analysis was developed to solve real problems, in areas such as partial differential equations. We'll respect that history and we will discuss applications of functional analysis to ODEs. All of this theory was created to attack concrete problems.

Before we can talk about more general spaces, we should first understand finite dimensional normed linear spaces. These have some special properties.

We can pick any basis $e_1, \ldots, e_n$. Then for any $x \in X$, we can write x as $x = \sum_{j=1}^n x_j e_j$.

Remark. We will call x_j as the *coordinates* of x with respect to the basis, and we will define $\underline{x} = (x_1, \dots, x_n)$ to be a point in $\mathbb{C}^n$ or $\mathbb{R}^n$.

Then (using the Cauchy-Schwarz inequality), we have

$$\|x\| = \left\|\sum_{j=1}^{n} x_j e_j\right\| \le \sum_{j=1}^{n} \|x_j e_j\| = \sum_{j=1}^{n} |x_j| \, \|e_j\| \le \sqrt{\sum_{j=1}^{n} |x_j|^2} \sqrt{\sum_{j=1}^{n} \|e_j\|^2} = M \, \|\underline{x}\|_{\mathbb{R}^n} .$$

We have therefore shown that $\|x\| \le M \|\underline{x}\|_{\mathbb{R}^n}$ for every $x \in X$.

Proposition 2.11. *Define $S = \{x \in X : \|\underline{x}\|_{\mathbb{R}^n} = 1\}$. Then S is a compact subset of X.*

Proof. Let $\{x^{(k)}\}_{k=1,\dots} \in S$. Then $\|\underline{x}^{(k)}\|_{\mathbb{R}^n} = 1$ for all k. This means that $\underline{x}^{(k)}$ are in the unit sphere of $\mathbb{R}^n$, which is compact. Hence, there exists a convergent subsequence $\{\underline{x}^{(k_j)}\}_{j=1,\dots}$ with $\lim \underline{x}^{(k_j)} = \underline{y}$. Here, $\underline{y}$ is on the unit sphere. Let $y = \sum_{j=1}^{n} y_j e_j$.

Note that $\|x^{(k_j)} - y\| \le M \|x^{(k_j)} - y\|_{\mathbb{R}^n} \to 0$ as $j \to \infty$. We also get the inequality

$$1 - \|\underline{y} - \underline{x}^{(k_j)}\|_{\mathbb{R}^n} \le \|\underline{y}\|_{\mathbb{R}^n} = \|\underline{y} - \underline{x}^{k_j} + \underline{x}^{k_j}\|_{\mathbb{R}^n} \le \|\underline{y} - \underline{x}^{(k_j)}\|_{\mathbb{R}^n} + 1,$$

which completes the proof. □

Proposition 2.12. *We also have that $\|x\|$ is continuous on X (and hence on S).*

Proof.

$$|f(x) - f(y)| = |\|x\| - \|y\|| \le \|x - y\| = d(x, y).$$

□

We know that continuous functions on compact sets attain their minima. Let $m = \min_{x \in S} \|x\|$. Then $x \ne 0$ implies that

$$\frac{x}{\|\underline{x}\|_{\mathbb{R}^n}} \in S \implies \left\|\frac{x}{\|\underline{x}\|_{\mathbb{R}^n}}\right\| \ge m,$$

so that $\|x\| \ge m \|x\|_{\mathbb{R}^n}$. Together with what we did before, we have proved that

$$m \|x\|_{\mathbb{R}^n} \le \|x\| \le M \|x\|_{\mathbb{R}^n}$$

for all $x \in X$. This norm $\|\cdot\|$ is therefore equivalent to the Euclidean norm.

3. 4/1

3.1. **Finite dimensional spaces.** Recall that we are working with a finite dimensional normed linear space X with norm $\|\cdot\|$. Last time, we showed that for any given basis $e_1, \dots, e_n$, we proved that there exist $M, m > 0$ with $m \|\underline{x}\|_{\mathbb{R}^n} \le \|x\| \le M \|\underline{x}\|_{\mathbb{R}^n}$ for all $x \in X$.

Proposition 3.1. *Note that this guarantees that all norms are equivalent in a finite dimensional space. This means that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms in a finite dimensional space X, then there exists a constant C with $C^{-1} \|x\|_2 \le \|x\|_1 \le C \|x\|_2$.*

Proof. We check this fact. We have

$$m_1 \|\underline{x}\|_{\mathbb{R}^n} \le \|x\|_1 \le M_1 \|\underline{x}\|_{\mathbb{R}^n}$$
$$m_2 \|\underline{x}\|_{\mathbb{R}^n} \le \|x\|_2 \le M_2 \|\underline{x}\|_{\mathbb{R}^n}$$

This shows that

$$\frac{m_1}{M_2}\left\|x\right\|_2 \leq m_1\left\|\underline{x}\right\|_{\mathbb{R}^n} \leq \left\|x\right\|_1 \leq M_1\left\|\underline{x}\right\|_{\mathbb{R}^n} \leq \frac{M_1}{m_2}\left\|x\right\|_2,$$

which proves our statement. □

Proposition 3.2. *All norms in a finite dimensional space gives the same open sets.*

Proof. Define

$$B_\rho^{\|\cdot\|_1}(y) = \left\{x \in X : \left\|x - y\right\|_1 < \rho\right\}.$$

Note that $\left\|x-y\right\|_1 < \rho$ implies that $\left\|x-y\right\|_2 < \frac{M_2}{m_1}\rho$. Therefore, $B_\rho^{\|\cdot\|_1}(y) \subseteq B_{\frac{M_2}{m_1}\rho}^{\|\cdot\|_2}(y)$. The opposite fact can be shown similarly. □

Proposition 3.3. *In a finite dimensional normed space X, the closed unit ball is compact.*

Proof. Let $\left\{x_k\right\}_{k=1,\ldots}$ be a sequence in $\overline{B_1(0)} = \{x \in X : \|x\| \leq 1\}$.

We have that $\left\|\underline{x}_k\right\|_{\mathbb{R}^n} \leq \frac{1}{m} \cdot 1$. Therefore, in $\mathbb{R}^n$, this sequence is bounded. Hence, there exists a subsequence $\left\{\underline{x}_{k_j}\right\}$ which converges to $\underline{y} \in \mathbb{R}^n$. In X, there is a corresponding point $y = \sum_{j=1}^n y_j e_j$. Then $\left\|x_{k_j} - y\right\| \leq M\left\|\underline{x}_{k_j} - \underline{y}\right\|_{\mathbb{R}^n} \to 0$. This proves compactness. □

Lemma 3.4. *If X is any infinite dimensional normed linear space, the closed unit ball is* not *compact.*

Proof. Take some $e_1 \in X$, say $\|e_1\| = 1$.

Take $e_2 \in X \setminus \operatorname{span}\{e_1\}$.

Take $e_3 \in X \setminus \operatorname{span}\{e_1, e_2\}$.

Inductively, take $e_n \in X \setminus \operatorname{span}\{e_1, \ldots, e_{n-1}\}$. These must exist because otherwise the space would be finite dimensional.

Homework 1, problem 8 said that there exists $w_n \in \operatorname{span}\{e_1, \ldots, e_{n-1}\}$ with $0 < \lambda_n = \|e_n - w_n\| = \min\{\|e_n - y\| : y \in \operatorname{span}\{e_1, \ldots, e_n\}\}$.

Define

$$\tilde{e}_n = \frac{e_n - w_n}{\|e_n - w_n\|} = \frac{e_n - w_n}{\lambda_n}.$$

For $n > l$, we have

$$\|\tilde{e}_n - \tilde{e}_l\| = \left\|\frac{e_n - w_n}{\lambda_n} - \tilde{e}_l\right\|.$$

We claim that $\|\tilde{e}_n - \tilde{e}_l\| \geq 1$. Otherwise,

$$\left\|\frac{e_n - w_n}{\lambda_n} - \tilde{e}_l\right\| < 1,$$

which means that

$$\|e_n - (w_n + \lambda_n \tilde{e}_l)\| < \lambda_n.$$

Note that $w_n + \lambda_n \tilde{e}_l \in \operatorname{span}\{e_1, \ldots, e_{n-1}\}$, which contradicts the definition of λ_n as the minimal distance. Therefore, we indeed have $\|\tilde{e}_n - \tilde{e}_l\| \geq 1$ and hence there is no convergent subsequence. □

Remark. Everything we said also works in complex spaces; just change $\mathbb{R}$ to $\mathbb{C}$.

This concludes the discussion of finite dimensional spaces vs infinite dimensional spaces.

3.2. **More about completeness.** We haven't yet proven that any infinite space is complete. We claim that this is true for the space $\ell^2_{\mathbb{R}} = \left\{x = (x_1, x_2, \dots) : \sum_{j=1}^{\infty} x_j^2 < \infty\right\}$.

Definition 3.5. Define $(x, y) = \sum_{j=1}^{\infty} x_j y_j$. Check that this is indeed an inner product. We also get the inner product norm $\|x\| = \sqrt{(x, x)}$.

Proposition 3.6. *$\ell^2_{\mathbb{R}}$ is complete with respect to this norm. This means that it is a Hilbert space.*

Proof. Let $\left\{x^{(k)}\right\}_{k=1,\dots}$ be a Cauchy sequence. Let $\varepsilon > 0$. Then there exists N such that $\left\|x^{(k)} - x^{(l)}\right\| < \varepsilon$ for every $k > l \geq N$. This is

$$\sqrt{\sum_{j=1}^{\infty}(x_j^{(k)} - x_j^{(l)})^2} = \left\|x^{(k)} - x^{(l)}\right\| < \varepsilon.$$

Therefore, $|x_j^{(k)} - x_j^{(l)}| < \varepsilon$ for all $k > l \geq N$ and for all $j = 1, 2, \dots$. Hence, $\left\{x_j^{(k)}\right\}_{k=1,2,\dots}$ is a Cauchy sequence in $\mathbb{R}$ for all $j = 1, 2, \dots$. There is therefore some $y_j \in \mathbb{R}$ with $\lim_{k\to\infty} x_j^{(k)} = y_j$.

We have the inequality

$$\sqrt{\sum_{j=1}^{M}(x_j^{(k)} - x_j^{(l)})^2} \leq \sqrt{\sum_{j=1}^{\infty}(x_j^{(k)} - x_j^{(l)})^2} < \varepsilon.$$

In this inequality, take the limit as $k \to \infty$. Then

$$\sqrt{\sum_{j=1}^{M}(y_j - x_j^{(l)})^2} \leq \varepsilon$$

for all $l \geq N$ and for all M. Therefore, we see that

$$\left\|y - x^{(l)}\right\| = \sqrt{\sum_{j=1}^{\infty}(y_j - x_j^{(l)})^2} \leq \varepsilon$$

for all $l \geq N$. This shows that $y - x^{(l)} \in \ell^2_{\mathbb{R}}$. We also know that $x^{(l)} \in \ell^2_{\mathbb{R}}$, so that $y \in \ell^2_{\mathbb{R}}$. Furthermore, for every $\varepsilon > 0$, there exists N with $\left\|y - x^{(l)}\right\| \leq \varepsilon$ for all $l \geq N$. Hence, $\lim x^{(l)} = y$ in $\ell^2_{\mathbb{R}}$. This proves that every Cauchy sequence converges, which proves completeness.

Note that the crucial step was to use the completeness of $\mathbb{R}$. $\square$

Definition 3.7. A set A is convex means that for all $x, y \in A$, the line segment joining them is in A, i.e. $tx + (1 - t)y = y + t(x - y) \in A$ for all $t \in [0, 1]$.

Theorem 3.8. *Let X be any Hilbert space, and let A be any nonempty closed convex subset of X. Let $x \in X \setminus A$. Then there exists a unique nearest point of A to x. More precisely, there exists $a \in A$ such that $\|x - a\| < \|x - y\|$ for every $y \in A \setminus \{a\}$.*

4. 4/4

Last time, we stated Theorem 3.8 that we can find a unique closest point. We can now prove it. The proof is based on the parallelogram inequality. We don't yet know that the minimum exists, but we can consider the infimum.

Proof of Theorem 3.8. Let $\alpha = \inf\{\|x-y\| : y \in A\}$. For all $k = 1, 2, \ldots$, there exists $y_k \in A$ with $\|x - y_k\| < \sqrt{\alpha^2 + \frac{1}{k}}$, since otherwise we would have $\|x-y\| \geq \sqrt{\alpha^2 + \frac{1}{k}}$ for all $y \in A$, contradicting the definition of α.

We can now apply the parallelogram identity. This states that $\|z-w\|^2 + \|z+w\|^2 = 2(\|z\|^2 + \|w\|^2)$. We plug in $z = x - y_k$ and $w = x - y_l$. Then

$$\|y_k - y_l\|^2 + \|(x-y_k) + (x-y_l)\|^2 = 2(\|x-y_k\|^2 + \|x-y_l\|^2) \leq 2\left(\alpha^2 + \frac{1}{k} + \alpha^2 + \frac{1}{l}\right)$$

Note that $\|(x-y_k)+(x-y_l)\|^2 = 4\|x - (y_k+y_l)/2\|^2$. By the convexity of A, we know that $(y_k+y_l)/2 \in A$. Therefore, $\|x - (y_k+y_l)/2\| > \alpha$. Hence, we have

$$\|y_k - y_l\|^2 + 4\alpha^2 \leq 2\left(\alpha^2 + \frac{1}{k} + \alpha^2 + \frac{1}{l}\right),$$

so

$$\|y_k - y_l\|^2 \leq \frac{2}{k} + \frac{2}{l}$$

for all $k, l = 1, 2, \ldots$. Therefore, $\{y_k\}$ is a Cauchy sequence. X is a Hilbert space, so it is complete. Hence, $\{y_k\}$ is convergent, so there exists $a \in X$ with $a = \lim y_k$, i.e. $\lim \|a - y_k\| = 0$. Hence, $a \in A$ because A is closed.

We now claim that $\|x-a\|$ is the minimum distance. We have

$$\alpha \leq \|x-a\| = \|x - y_k + y_k - a\| \leq \|x-y_k\| + \|y_k - a\| \leq \sqrt{\alpha^2 + \frac{1}{k}} + \|y_k - a\| \to \alpha + 0.$$

Therefore, $\|x-a\| = \alpha$.

We still need to show uniqueness. Suppose that $\tilde{a} \in A$ also has the minimum distance $\|x - \tilde{a}\| = \alpha$. We want to show that $a = \tilde{a}$. We can plug a and $\tilde{a}$ in the parallelogram law in place of y_k and y_l. This gives

$$\|a-\tilde{a}\|^2 + 4\alpha^2 \leq \|a-\tilde{a}\|^2 + 4\left\|x - \frac{a+\tilde{a}}{2}\right\|^2 = 2(\|x-a\|^2 + \|x-\tilde{a}\|^2) = 4\alpha^2,$$

so hence $\|a-\tilde{a}\|^2 = 0$ and therefore $a = \tilde{a}$. □

4.1. **Orthogonality.**

Definition 4.1. In a Hilbert space X, vectors $x_1, \ldots, x_N \in X$ are *orthogonal* means that $(x_i, x_j) = 0$, for all $i \neq j$, $i, j = 1, \ldots, N$.

$x_1, \ldots, x_N \in X$ are *orthonormal* if

$$(x_i, x_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases}$$

There is a fundamental identity for orthonormal vectors.

Proposition 4.2. *Let $x \in X$, and suppose that $e_1, \dots, e_N$ are orthonormal. Let $\lambda_1, \dots, \lambda_N$ be scalars, let $c_j = (x, e_j)$ for $j = 1, \dots, N$. Then*

$$\left\| x - \sum_{j=1}^{N} \lambda_j e_j \right\|^2 = \|x\|^2 + \sum_{i=1}^{N} |c_i - \lambda_i|^2 - \sum_{i=1}^{N} |c_i|^2.$$

Proof.

$$\begin{aligned}
\left\| x - \sum_{j=1}^{N} \lambda_j e_j \right\|^2 &= \left(x - \sum_{i=1}^{N} \lambda_i e_i, x - \sum_{j=1}^{N} \lambda_j e_j \right) \\
&= (x, x) - \left(\sum_{i=1}^{N} \lambda_i e_i, x \right) - \left(x, \sum_{j=1}^{N} \lambda_j e_j \right) + \left(\sum_{i=1}^{N} \lambda_i e_i, \sum_{j=1}^{N} \lambda_j e_j \right) \\
&= \|x\|^2 - \sum_{i=1}^{N} \lambda_i \bar{c}_i - \sum_{j=1}^{N} \overline{\lambda}_j c_j + \sum_{i=1}^{N} |\lambda_i|^2 \\
&= \|x\|^2 + \sum_{i=1}^{N} |c_i - \lambda_i|^2 - \sum_{i=1}^{N} |c_i|^2.
\end{aligned}$$

□

This identity is nice. Staring at it for a bit, we can read off the point that satisfies the closest point property.

Theorem 4.3. *The point of* span $\{e_1, \dots, e_N\}$ *which has minimum distance from x is exactly $\sum_{i=1}^{N} c_i e_i$ and it is the unique such point.*

Intuitively, we expect that this minimum point on span $\{e_i\}$ is orthogonal. We can check this:

$$\left(x - \sum_{i=1}^{N} c_i e_i, \sum_{j=1}^{N} \lambda_j e_j \right) = \sum_{j=1}^{N} \overline{\lambda}_j (x, e_j) - \sum_{i=1}^{N} c_i \overline{\lambda}_i = \sum_{j=1}^{N} c_j \overline{\lambda}_j - \sum_{i=1}^{N} c_i \overline{\lambda}_i = 0.$$

As we expected, this means that

$$\left(x - \sum_{i=1}^{N} c_i e_i \right) \in \left(\operatorname{span} \{e_1, \dots, e_N\} \right)^{\perp}.$$

Theorem 4.4. *Let $e_1, e_2, \dots$ be an (infinite) sequence of orthonormal vectors in the Hilbert space X. Then for every $x \in X$,*

(i) $\sum_{i=1}^{\infty} c_i e_i$ is convergent, where $c_i = (x, e_i)$ for all i, that is, there exists $y \in X$ with

$$\lim_{N \to \infty} \left\| \sum_{i=1}^{N} c_i e_i - y \right\| = 0.$$

(ii) This is Bessel's inequality and it is very important.

$$\sum_{i=1}^{\infty} |c_i|^2 \leq \|x\|^2$$

(iii) Equality holds in (ii) if and only if $x = \sum_{i=1}^{\infty} c_i e_i$.

The proof will use the identity 4.2.

Proof. We will first prove (ii). By identity 4.2 with $\lambda_i = c_i$ for all i, we have

$$0 \leq \left\| x - \sum_{i=1}^{N} c_i e_i \right\|^2 = \|x\| - \sum_{i=1}^{N} |c_i|^2,$$

so therefore $\sum_{i=1}^{N} |c_i|^2 \leq \|x\|^2$ for all N. This proves (ii).

Now we prove (iii). We have

$$\sum_{i=1}^{\infty} |c_i|^2 = \|x\|^2 \Leftrightarrow \lim_{N\to\infty} \sum_{i=1}^{N} |c_i|^2 = \|x\|^2 \Leftrightarrow \lim_{N\to\infty} \left\| x - \sum_{i=1}^{N} c_i e_i \right\| \Leftrightarrow x = \sum_{i=1}^{\infty} c_i e_i.$$

Finally, we prove (i). let $S_N = \sum_{i=1}^{N} c_i e_i$. Then $N > M$ implies that

$$\|S_N - S_M\|^2 = \left\| \sum_{i=M+1}^{N} c_i e_i \right\|^2 = \sum_{i=M+1}^{N} |c_i|^2 \leq \sum_{i=M+1}^{\infty} |c_i|^2 \to 0$$

as $M \to \infty$. This proves that S_N is a Cauchy sequence, and hence S_N is convergent. That concludes the proof. $\square$

5. 4/6

X is a Hilbert space and $e_1, e_2, \ldots$ is an orthonormal sequence.

Definition 5.1. $e_1, e_2, \ldots$ is a *complete* orthonormal sequence if $x = \sum_{i=1}^{\infty} (x, e_i) e_i$ for all $x \in X$.

Theorem 5.2. *The statement (C) that $e_1, e_2, \ldots$ is a complete orthonormal sequence (i.e. $x = \sum_{i=1}^{\infty} (x, e_i) e_i$ for all $x \in X$) is equivalent to each of the following:*

(i) Equality holds in Bessel's inequality for every $x \in X$ (i.e. $\|x\|^2 = \sum_{i=1}^{\infty} |c_i|^2$)
(ii) There does not exist an $x \in X \setminus \{0\}$ with $(x, e_i) = 0$ for every i.
(iii) $\operatorname{span}\{e_1, e_2, \ldots\}$ *is a dense subset of X.*

Proof. We already proved (i) last time in Theorem 4.4.

(ii) $\implies$ (C). Consider $\sum_{i=1}^{\infty} (x, e_i) e_i$. We know that this converges. Then

$$\left(x - \sum_{i=1}^{\infty} (x, e_i) e_i, e_j \right)$$
$$= (x, e_j) - \left(\sum_{i=1}^{\infty} (x, e_i) e_i, e_j \right) = (x, e_j) - \sum_{i=1}^{\infty} (x, e_i)(e_i, e_j) = (x, e_j) - (x, e_j) = 0.$$

By (ii), this means that $x - \sum_{i=1}^{\infty} (x, e_i) e_i = 0$, which is (C).

(C) $\implies$ (ii). Suppose $x \in X$, then $x = \sum_{i=1}^{\infty} (x, e_i) e_i = \sum_{i=1}^{\infty} c_i e_i$, so $x = 0$ if $(x, e_i) = 0$ for every i.

(C) $\implies$ (iii). Take $x \in X$. Then (C) implies that

$$x = \sum_{i=1}^{\infty} c_i e_i = \lim_{N\to\infty} \sum_{i=1}^{N} c_i e_i,$$

which is a limit of sums in span $\{e_1, e_2, \dots\}$. Therefore, x is a limit point of span $\{e_1, e_2, \dots\}$. Since x was arbitrary, this means that span $\{e_1, e_2, \dots\}$ is dense.

(iii) $\implies$ (C). Take any point $x \in X$. Then (iii) implies that $x = \lim_{N\to\infty} y_N$ where $y_N = \sum_{j=1}^{Q_N} \lambda_j^N e_j \in \operatorname{span}\{e_1, \dots, e_n\}$. This means that

$$\left\| x - \sum_{j=1}^{Q_N} \lambda_j^N e_j \right\|^2 \to 0$$

as $N \to \infty$. By the fundamental inequality that we proved last time, we have

$$\left\| x - \sum_{j=1}^{Q_N} c_j e_j \right\|^2 \le \left\| x - \sum_{j=1}^{Q_N} \lambda_j^N e_j \right\|^2 \to 0.$$

Note that this is also true with sums up to M, with $M \ge Q_N$ and $\lambda_j^L = 0$ for all $L > N$.

Let $\varepsilon > 0$. Then there exists N_0 such that

$$\left\| x - \sum_{j=1}^{M} c_j e_j \right\| < \varepsilon$$

for all $M \ge N_0$. This is the definition of convergence, so $x = \lim_{M\to\infty} \sum_{j=1}^{M} c_j e_j = \sum_{j=1}^{\infty} c_j e_j$. □

Example 5.3. Define

$$L^2_{\mathbb{C}}[-\pi, \pi] = \left\{ f : \int_{-\pi}^{\pi} |f|^2 \text{ exists and is finite} \right\}.$$

This has an inner product

$$(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}\, dx.$$

We can easily check the inner product properties.

This is not as simple as it looks. What type of integration are we using? The Riemann integral is not good enough. Using the Riemann integral, this space lacks completeness. There are a huge class of counterexamples.

Therefore, we will use the Lebesgue integral. We will spend a couple of lectures on this, but for now, don't worry about it. All Riemann-integrable functions are also Lebesgue-integrable with the same integral; however, Lebesgue integration allows us to handle a much larger class of functions. In particular, Lebesgue integration gives us completeness.

There is an extremely important application of the abstract theory that we have just developed. There is a simple orthonormal sequence in this space. This is

$$\left\{1, e^{ix}, e^{-ix}, e^{2ix}, e^{-2ix}, \dots, e^{nix}, e^{-nix}, \dots\right\}.$$

Proposition 5.4. *This is an orthonormal sequence.*

Proof.

$$(e^{inx}, e^{imx}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{inx} e^{-imx}\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)x}\, dx = \begin{cases} \frac{2\pi}{2\pi} = 1 & n = m \\ \frac{1}{2\pi}\left[\frac{e^{i(n-m)x}}{i(n-m)}\right] = 0 & n \ne m. \end{cases}$$ □

In particular, all the previous theory applies to this case. In fact, we will show that this is a *complete* orthonormal sequence.

6. 4/8

We are in the process of talking about an extremely important application. We are considering the following example:

Example 6.1.

$$X = L^2_{\mathbb{C}}([-\pi, \pi]) = \left\{ f = f_1 + if_2 : \int_{-\pi}^{\pi} |f|^2 \text{ exists and is finite} \right\}$$

The inner product $(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}\, dx$ is clear. For now, we will believe that this is a complete space (with the Lebesgue integral). Historically, this is the main reason why people use the Lebesgue integral. This will be proved soon.

The big claim is

Proposition 6.2.

$$1, e^{ix}, e^{-ix}, e^{2ix}, e^{-2ix}, \dots, e^{inx}, e^{-inx}, \dots$$

is a complete orthonormal sequence.

Proof. Recall that for an orthonormal sequence to be complete, we need

$$x = \sum_{n=1}^{\infty} (x, e_n) e_n$$

for all x. In this case, we have the *Fourier series of f*

$$\sum_{n=1}^{\infty} (x, e_n) e_n = \sum_{n=-\infty}^{\infty} (f, e^{inx}) e^{-inx} = \lim_{n \to \infty} \sum_{n=-N}^{N} (f, e^{inx}) e^{inx}.$$

(We don't have to worry too much about the limit because we already proved earlier that everything converges.) Here,

$$(f, e^{inx}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt.$$

Therefore,

$$S_N(f)(x) = \sum_{n=-N}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-int}\, dt e^{inx} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \sum_{n=-N}^{N} e^{in(x-t)}\, dt.$$

We have the *Dirichlet kernel*

$$\begin{aligned} D_N(s) = \sum_{n=-N}^{N} e^{ins} &= e^{-iNs} + e^{-i(N-1)s} + \cdots + 1 + \cdots + e^{iNs} \\ &= e^{-iNs}(1 + e^{is} + e^{2is} + \cdots + e^{2iNS}) = e^{-iNS} \frac{1 - e^{(2N+1)is}}{1 - e^{is}} = \frac{e^{-iNs} - e^{(N+1)is}}{1 - e^{is}} \\ &= \frac{e^{is/2}(e^{-i(N+1/2)s} - e^{i(N+1/2)s})}{e^{is/2}(e^{-is/2} - e^{is/2})} = \frac{\sin(N + \frac{1}{2})s}{\sin \frac{1}{2} s} \end{aligned}$$

for $0 < |s| \leq \pi$. Note also that $D_N(0) = 2N + 1$.

Here, convergence does *not* mean pointwise convergence. We actually mean L^2 convergence, i.e.

$$\left\| f - \sum_{n=-N}^{N} (f, e^{inx}) e^{inx} \right\| \to 0$$

as $N \to \infty$ with the L^2 inner product norm. This is what we need to show.

Remark. Bessel's inequality in this case states that

$$\sum_{n=-\infty}^{\infty} |c_n|^2 \leq \|f\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^2 \, dt.$$

This is nice, but we don't need it here.

We will now show that span $\{1, e^{ix}, e^{-ix}, \dots\}$ is dense in $L^2_{\mathbb{C}}([-\pi, \pi])$, i.e. if $f \in L^2_{\mathbb{C}}([-\pi, \pi])$ then for every $\varepsilon > 0$, there are λ_j for $j = -N, \dots, N$ such that

$$\left\| f - \sum_{j=-N}^{N} \lambda_j e^{ijx} \right\| < \varepsilon.$$

First, we check this for when f is a continuous function with $f(-\pi) = f(\pi) = 0$. We claim that we can use

$$\frac{1}{m+1} \sum_{N=0}^{m} S_N(f)(x) = \frac{1}{2\pi} \frac{1}{m+1} \int_{-\pi}^{\pi} f(t) \sum_{N=0}^{m} D_N(x-t) \, dt.$$

Here,

$$\begin{aligned}
K_m(x) &= \frac{1}{m+1} \sum_{N=0}^{m} D_N(s) = \frac{1}{m+1} \sum_{N=0}^{m} \frac{e^{-iNs} - e^{i(N+1)s}}{1 - e^{is}} \\
&= \frac{\frac{1}{m+1} \left(\sum_{N=0}^{m} e^{-iNs} - e^{is} \sum_{N=0}^{m} e^{iNs} \right)}{1 - e^{is}} \\
&= \frac{1}{m+1} \cdot \frac{1}{1 - e^{is}} \left(\frac{1 - e^{-i(m+1)s}}{1 - e^{-is}} - e^{is} \frac{(1 - e^{i(m+1)s})}{1 - e^{is}} \right) \\
&= \cdots \\
&= \frac{1}{m+1} \frac{\sin^2(\frac{m+1}{2} s)}{\sin^2(\frac{s}{2})}
\end{aligned}$$

where the calculation will be finished next time. This is the *Fejer kernel.* This is a nonnegative function. □

7. 4/11

We want to show completeness of the orthonormal sequence $\{e^{inx} : n = 0, \pm 1, \dots\}$ in $L^2_{\mathbb{C}}[-\pi, \pi]$.

Last time, we showed that

$$S_N(f)(x) = \sum_{n=-N}^{N} c_n e^{inx}$$

with

$$c_n = (f, e^{int}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt.$$

Then

$$S_N(f)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) D_N(x-t)\, dt$$

with

$$D_N(s) = \sum_{n=-N}^{N} e^{ins} = \frac{\sin(N+\frac{1}{2})s}{\sin\frac{s}{2}}.$$

Note that $D_N(0) = 2N+1$. Consider

$$\frac{1}{m+1} \sum_{N=0}^{m} S_N(f)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{1}{m+1} \sum_{N=0}^{m} D_N(x-t)\, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) K_n(x-t)\, dt,$$

where we define

$$\begin{aligned}
K_m(s) = \frac{1}{m+1} \sum_{N=0}^{m} D_N(s) &= \frac{1}{m+1} \frac{1}{1-e^{is}} \left(\frac{1-e^{-i(m+1)s}}{1-e^{-is}} - e^{is} \frac{1-e^{i(m+1)s}}{1-e^{is}} \right) \\
&= \frac{1}{m+1} \frac{1}{e^{is/2}(e^{-is/2}-e^{is/2})} \left(\frac{1-e^{-i(m+1)s}}{e^{-is/2}(e^{is/2}-e^{-is/2})} - \frac{e^{is}(1-e^{i(m+1)s})}{e^{is/2}(e^{-is/2}-e^{is/2})} \right) \\
&= \frac{1}{m+1} \frac{1}{e^{-is/2}-e^{is/2}} \left(\frac{1-e^{-i(m+1)s}}{e^{is/2}-e^{-is/2}} - \frac{1-e^{i(m+1)s}}{e^{-is/2}-e^{is/2}} \right) \\
&= \frac{1}{m+1} \left(\frac{e^{i(m+1)s}+e^{-i(m+1)s}-2}{(e^{is/2}-e^{-is/2})^2} \right) = \frac{1}{m+1} \frac{(e^{i(m+1)s/2}-e^{-i(m+1)s/2})^2}{(e^{is/2}-e^{-is/2})^2} \\
&= \frac{1}{m+1} \frac{\sin^2(\frac{m+1}{2}s)}{\sin^2\frac{s}{2}}
\end{aligned}$$

for $0 < |s| \leq \pi$. Additionally, $K_m(0) = \frac{1}{m+1}\sum_{N=0}^{m}(2N+1) = m+1$.

This was not an obvious thing to do. It's not clear that the average of the S_N should be nice, but it does come out nicely.

Let's consider some properties of K_m. Notice that it is even and positive. It is zero whenever $\frac{m+1}{2} = k\pi$. There is a sharp peak at zero, with height $m+1$. There are a lot of zeros, and between the zeros, there are violent oscillations. It is a good idea to draw a picture of this.

What is the area under the graph? We have

$$A = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_m(s)\, ds = \frac{1}{2\pi} \frac{1}{m+1} \sum_{N=0}^{m} \int_{-\pi}^{\pi} D_N(s)\, ds = 1.$$

This is nice. It is like an oscillatory version of the Dirac delta "function". In the distribution sense, these converge to the Dirac delta.

Consider this function on the intervals $[-\pi, -\delta) \cup (\delta, \pi]$, where $\delta < |s| \leq \pi$. Then

$$0 \leq K_m(s) \leq \frac{1}{m+1} \frac{1}{\sin^2\frac{\delta}{2}} \to 0$$

as $m \to \infty$ for fixed δ.

Take $f = g$, where g is continuous and $g(-\pi) = g(\pi) = 0$. Extend g to be 2π-periodic. Then

$$\frac{1}{m+1}\sum_{N=0}^{m} S_N(g)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(t)K_n(x-t)\,dt.$$

This is an average of the Fourier partial sums of g, and these functions are in span$\{e^{inx}\}$. We claim that this is a really good approximation to g. To see this, consider

$$\begin{aligned}\left|\frac{1}{m+1}\sum_{N=0}^{m} S_N(g)(x) - g(x)\right| &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} g(t)K_m(x-t)\,dt - g(x)\right| \\ &= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} (g(t)-g(x))K_m(x-t)\,dt\right|\end{aligned}$$

because $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_m(x-t)\,dt = 1$. Make a change of variable $s = x - t$, and observe that all functions under consideration are 2π-periodic. Then

$$= \left|\frac{1}{2\pi}\int_{x-\pi}^{x+\pi} (g(x-s)-g(x))K_m(s)\,ds\right| = \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} (g(x-s)-g(x))K_m(s)\,ds\right|.$$

For any $\delta \in (0,\pi)$, we can break this integral up into

$$= \left|\frac{1}{2\pi}\int_{-\delta}^{\delta} (g(x-s)-g(x))K_m(s)\,ds + \frac{1}{2\pi}\int_{\delta\le|s|\le\pi} (g(x-s)-g(x))K_m(s)\,ds\right|.$$

Note that g is uniformly continuous because it is continuous on a compact set. Let $\varepsilon > 0$. Then there exists $\delta > 0$ such that $|x-y| < \delta$ implies that $|g(x)-g(y)| < \varepsilon$. Therefore, picking the appropriate δ, we have

$$\begin{aligned}&\le \frac{1}{2\pi}\left|\int_{-\delta}^{\delta} (g(x-s)-g(x))K_m(s)\,ds\right| + \frac{1}{2\pi}\left|\int_{\delta\le|s|\le\pi} (g(x-s)-g(x))K_m(s)\,ds\right| \\ &\le \frac{1}{2\pi}\int_{-\delta}^{\delta} |g(x-s)-g(x)|K_m(s)\,ds + \frac{1}{2\pi}\int_{\delta\le|s|\le\pi} |g(x-s)-g(x)|K_m(s)\,ds.\end{aligned}$$

Let g is continuous on a compact set, so it attains its maximum. Let $L = \max|g|$. Then we have

$$\begin{aligned}&\le \frac{\varepsilon}{2\pi}\int_{-\delta}^{\delta} K_m(s)\,ds + \frac{2L}{2\pi}\int_{\delta\le|s|\le\pi} K_m(s)\,ds \\ &\le \frac{\varepsilon}{2\pi}\int_{-\pi}^{\pi} K_m(s)\,ds + \frac{L}{\pi}\frac{1}{m+1}\frac{1}{\sin^2\frac{\delta}{2}} \\ &= \varepsilon + \frac{L}{\pi}\frac{1}{m+1}\frac{1}{\sin^2\frac{\delta}{2}} < \varepsilon + \varepsilon = 2\varepsilon\end{aligned}$$

provided that $m > m_0$, where $\frac{L}{\pi}\frac{1}{m_0+1}\frac{1}{\sin^2\frac{\delta}{2}} < \varepsilon$. Therefore,

$$\max_{[-\pi,\pi]}\left|\frac{1}{m+1}\sum_{N=0}^{m} S_N(g)(x) - g(x)\right| < 2\varepsilon$$

for all $m > m_0$. We're almost done. We have proved this for when g is continuous. We want to show that this is true for all L^2 functions, and we'll finish this next time.

8. 4/13

We've almost finished the proof that our orthonormal sequence $\{e^{inx} : n = 0, \pm 1, \dots\}$ is dense in $L^2_{\mathbb{C}}[-\pi, \pi]$.

We proved that if $g : [-\pi, \pi] \to \mathbb{R}$ is continuous with $g(-\pi) = g(\pi) = 0$, then for any $\varepsilon > 0$, there exists m_0 with

$$\max_{x \in [-\pi,\pi]} \left| g(x) - \frac{1}{m+1} \sum_{N=0}^{m} S_N(g)(x) \right| < \varepsilon$$

for all $m \geq m_0$. Then,

$$\left\| g - \frac{1}{m+1} \sum_{N=0}^{m} S_N(g) \right\|_{L^2} = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| g(x) - \frac{1}{m+1} \sum_{N=0}^{m} S_N(g)(x) \right|^2 dx} \leq \varepsilon.$$

Now, take any $f \in L^2_{\mathbb{C}}[-\pi, \pi]$. One of the properties of the Lebesgue integral (that we will prove later) is that the continuous functions are dense in L^2.

Proposition 8.1. *There exists a continuous g with $g(-\pi) = g(\pi) = 0$ and $\|f - g\|_{L^2} < \varepsilon$.*

Using this claim, we're almost done. Then

$$\begin{aligned} \left\| f - \frac{1}{m+1} \sum_{N=0}^{m} S_N(g)(x) \right\|_{L^2} &= \left\| f - g + g - \frac{1}{m+1} \sum_{N=0}^{m} S_N(g)(x) \right\|_{L^2} \\ &\leq \|f - g\|_{L^2} + \left\| g - \frac{1}{m+1} \sum_{N=0}^{m} S_N(g)(x) \right\|_{L^2} < \varepsilon + \varepsilon \end{aligned}$$

for all $m > m_0$.

We have now shown that $\{e^{inx}\}$ is complete in L^2 space.

8.1. Lebesgue integral.

Let's talk about the Lebesgue integral.

We should first consider the idea of "measure zero." Let's remind ourselves of what this means in the Riemann theory.

Definition 8.2. A set $S \subset [a, b]$ has *content zero* if for every $\varepsilon > 0$ there exists finitely many open intervals $I_1, \dots, I_N$ with $S \subset \cup_{j=1}^{N} I_j$ and $\sum_{j=1}^{N} |I_j| < \varepsilon$, where $|I_j|$ is the length of I_j.

In the Lebesgue theory, we make what appears to be an innocent change, by allowing infinitely many intervals.

Definition 8.3. A set $S \subset [a, b]$ has *Lebesgue measure zero* if for every $\varepsilon > 0$ there exists open intervals $I_1, I_2, \dots$ such that $S \subset \cup_{j=1}^{\infty} I_j$ and $\sum_{j=1}^{\infty} |I_j| < \varepsilon$.

This small change doesn't seem profound, but it makes a huge difference to the theory.

Lemma 8.4. *If $S_1, S_2, \dots$ each has measure zero, then $\cup_{j=1}^{\infty} S_j$ also has measure zero.*

Remark. This is hopelessly false in the case of content zero.

A special case of this lemma is when each S_j contains precisely one point. Then every countable set of points also has measure zero. For example, the set of all rationals has Lebesgue measure zero.

Proof. Let $\varepsilon > 0$ be given. Since S_j has measure zero, there are intervals where $S_j \subset I_1^j \cup I_2^j \cup \cdots = \cup_{i=1}^\infty I_i^j$ with $\sum_{i=1}^\infty |I_i^j| < \frac{\varepsilon}{2^j}$.

Then $\cup_{j=1}^\infty S_j \subset \cup_{i,j} I_i^j$ and $\sum_{i=1}^\infty \sum_{j=1}^\infty |I_i^j| < \sum_{i=1}^\infty \frac{\varepsilon}{2^i} = \varepsilon$. □

Definition 8.5. Any property is true *almost everywhere* on $[a,b]$ if it holds except for a set of measure zero.

Definition 8.6. φ is a *step function* on an interval $[a,b]$ if there exists a partition $a = x_0 < x_1 < \cdots < x_N = b$ such that $\varphi|_{(x_{i-1},x_i)} = a_i$ is constant.

Note that the sum, difference, or product of step functions is a step function. This requires picking a common refinement of the two partitions. The integral of a step function is just

$$\int_{[a,b]} \varphi = \sum_{i=1}^N a_i(x_i - x_{i-1}).$$

We can now state the main technical lemma that allows us to define the Lebesgue integral for a huge class of functions. We defer the proof until later.

Lemma 8.7.

(1) *Suppose that $\{\varphi_k\}$ is an increasing sequence of step functions on $[a,b]$, (i.e. $\varphi_k(x) \le \varphi_{k+1}(x)$ for all $x \in [a,b]$ and for all k) such that $\left\{\int_{[a,b]} \varphi_k\right\}_{k=1,2,\ldots}$ is bounded. Then $\{\varphi_k(x)\}_{k=1,2,\ldots}$ is a bounded (hence convergent) sequence for almost all $x \in [a,b]$. (i.e. there exists $S \subset [a,b]$ of measure zero with $\{\varphi_k(x)\}$ bounded for all $x \in [a,b] \setminus S$)*

(2) *If $\{\psi_k\}$ is any other increasing sequence of step functions with $\int_{[a,b]} \psi_k$ bounded, and with $\lim \psi_k(x) = \lim \varphi_k(x)$ for almost every $x \in [a,b]$, then $\lim \int_{[a,b]} \varphi_k = \lim \int_{[a,b]} \psi_k$.*

Definition 8.8. Let $\mathcal{L}_0[a,b]$ be the set of all real valued functions $f : [a,b] \to \mathbb{R}$ such that there exists an increasing sequence of step functions $\{\varphi_k\}$ with $\int_{[a,b]} \varphi_k$ bounded and $\lim \varphi_k(x) = f(x)$ for almost every $x \in [a,b]$.

Now, define

$$\int_{[a,b]} f = \lim \int_{[a,b]} \varphi_k.$$

The first thing we do when we meet a mathematical definition is to ask: Does it make sense? It is well-defined by the second part of our lemma.

We now look at some properties of this integral. We will prove these next time.

Properties 8.9.

(1) If $\alpha, \beta \ge 0$, and $f, g \in \mathcal{L}_0$, then $\alpha f + \beta g \in \mathcal{L}_0$ and $\int_{[a,b]}(\alpha f + \beta g) = \alpha \int_{[a,b]} f + \beta \int_{[a,b]} g$.

(2) If $f, g \in \mathcal{L}_0$ then $\max\{f,g\}, \min\{f,g\} \in \mathcal{L}_0$.

(3) If $f, g \in \mathcal{L}_0$ with $f \le g$ then $\int_{[a,b]} f \le \int_{[a,b]} g$.

Warning: $f \in \mathcal{L}_0$ does *not* imply that $-f \in \mathcal{L}_0$.

9. 4/15

Recall that $\mathcal{L}_0$ is the set of $f : [a,b] \to \mathbb{R}$ such that $f(x) = \lim \varphi_k(x)$ for almost every $x \in [a,b]$, with $\varphi_{k+1} \ge \varphi_k$, where φ_k are step functions with $\int \varphi_k$ bounded.

We have some properties:

Properties 9.1.

(1) If $f \in \mathcal{L}_0$ and if $\tilde{f} : [a,b] \to \mathbb{R}$ with $\tilde{f}(x) = f(x)$ for almost every x, then $\tilde{f} \in \mathcal{L}_0$ and $\int_{[a,b]} \tilde{f} = \int_{[a,b]} f$.
(2) $\alpha, \beta \geq 0$ and $f, g \in \mathcal{L}_0$ imply that $\alpha f + \beta g \in \mathcal{L}_0$ and $\int(\alpha f + \beta g) = \alpha \int f + \beta \int g$.
(3) $f, g \in \mathcal{L}_0$ and $f \leq g$ implies that $\int f \leq \int g$.

Proof. We prove (3). Pick increasing sequences φ_k and ψ_k with $f = \lim \varphi_k$ and $g = \lim \psi_k$ almost everywhere.

Let $\tilde{\varphi}_k = \min\{\varphi_k, \psi_k\}$. Note that this is still a step function, and it still converges almost everywhere to f. To see this, notice that

$$\lim \tilde{\varphi}_k(x) = \min\{\lim \varphi_k(x), \lim \psi_k(x)\} = \min\{f(x), g(x)\} = f(x)$$

almost everywhere. Similarly, we define $\tilde{\psi}_k(x) = \max\{\varphi_k(x), \psi_k(x)\}$ to be a step function with $\lim \tilde{\psi}_k(x) = g(x)$. Therefore, we see that $\int f \leftarrow \int \tilde{\varphi}_k \leq \int \tilde{\psi}_k \to \int g$.

The point is that we should reduce everything to step functions. □

A big defect is that $f \in \mathcal{L}_0$ does not imply that $-f \in \mathcal{L}_0$. We enlarge this space to get around this problem.

Definition 9.2. Let $\mathcal{L}^1 = \{g - h : g, h \in \mathcal{L}_0\}$. This is the space of *Lebesgue integrable* functions.

Observe that $\mathcal{L}^1$ is indeed a linear space.

We should check this. For example, for multiplication by scalars, if $g - h \in \mathcal{L}^1$, we should show that $\lambda(g-h) \in \mathcal{L}^1$. To do this, we have two cases: $\lambda > 0$ and $\lambda < 0$. In each case, either $\lambda g, \lambda h \in \mathcal{L}_0$ or $-\lambda g, -\lambda h \in \mathcal{L}_0$, so the result follows simply.

Definition 9.3. For $f = g - h \in \mathcal{L}^1$, we define $\int f = \int g - \int h$.

Such a definition might plausibly be nonsense. There might be lots of ways of writing f as a difference of two functions in $\mathcal{L}_0$. Suppose that $g_1 - h_1 = g_2 - h_2$. Then $g_1 + h_2 = g_2 + h_1$. Each is a sum of functions in $\mathcal{L}_0$, so we can use the linearity in $\mathcal{L}_0$ to see that $\int g_1 + \int h_2 = \int g_2 + \int h_1$. Therefore, $\int g_1 - \int h_1 = \int g_2 - \int h_2$. Therefore, our definition works.

It seems like we got something for nothing, by extending $\mathcal{L}_0$ so simply to $\mathcal{L}^1$. But we can't get something for nothing. It turns out that it will be hard to show some really simply facts.

We have some properties:

Properties 9.4.

(1) If $f \in \mathcal{L}^1$ and $\tilde{f} : [a,b] \to \mathbb{R}$ with $\tilde{f}(x) = f(x)$ almost everywhere then $\tilde{f} \in \mathcal{L}^1$ and $\int \tilde{f} = \int f$.
(2) If $f_1, f_2 \in \mathcal{L}^1$ with $f_1 \leq f_2$, then $\int f_1 \leq \int f_2$.
(3) If $f \in \mathcal{L}^1$ then $|f| \in \mathcal{L}^1$ and $|\int f| \leq \int |f|$.
(4) If $f \in \mathcal{L}^1$ then there exists a decreasing sequence $\{g_k\} \subset \mathcal{L}_0$ with $g_k(x) \to f(x)$ for almost every $x \in [a,b]$ and $\int g_k \to \int f$.
(5) If $f \in \mathcal{L}^1$ and $\int |f| = 0$, then $f(x) = 0$ for almost every $x \in [a,b]$.
(6) If we have a sequence $\{f_k\} \subset \mathcal{L}^1$ of nonnegative functions $f_k \geq 0$ with $\int f_k \to 0$, then there exist a subsequence $\{f_{k_j}\}_{j=1,2,\ldots}$ with $f_{k_j}(x) \to 0$ for almost every $x \in [a,b]$.

Proof. (1) follows trivially from the corresponding fact about $\mathcal{L}_0$.

We check (2). $f_1 = g_1 - h_1$ and $f_2 = g_2 - h_2$ for $g_1, g_2, h_1, h_2 \in \mathcal{L}_0$. We know that $g_1 - h_1 \le g_2 - h_2$, so $g_1 + h_2 \le g_2 + h_1$, and so $\int (g_1 + h_2) \le \int (g_2 + h_1)$, and hence $\int f_1 \le \int f_2$.

We check (3). $f \in \mathcal{L}^1$, so $f = g - h$, where $g, h \in \mathcal{L}_0$. Here, we use a trick to express $|g - h|$ as a difference of functions in $\mathcal{L}_0$, that is, $|g - h| = \max\{g, h\} - \min\{g, h\} \in \mathcal{L}^1$. Now, $f \le |f|$, so that $\int f \le \int |f|$.

We check (4). $f \in \mathcal{L}^1$, so that $f = g - h$ for $g, h \in \mathcal{L}_0$. Then $h = \lim \varphi_k$ with $\varphi_{k+1} \ge \varphi_k$ and $\int \varphi_k$ bounded and $\int \varphi_k \to \int h$; a similar statement is true for $g = \lim \psi_l$. Then $g - h = \lim(g - \varphi_k)$ almost everywhere, and $g - \varphi_k = \lim_{l \to \infty}(\psi_l - \varphi_k)$ is the limit is a sequence of step functions, so $g - \varphi_k \in \mathcal{L}_0$, so we are done.

Note that (6) implies (5); just take $f_k = |f|$ for all k.

We now need to prove (6). This is a bit trickier than the previous properties. We pick f_{k_j} such that $\int f_{k_j} < \frac{1}{2^j}$ and $k_{j+1} > k_j$. We can certainly do this because $\int f_k \to 0$. By (4), there exist $g_j \in \mathcal{L}_0$ with $g_j \ge f_{k_j}$ and $\int g_j \le \int f_{k_j} + \frac{1}{2^j}$. Pick an increasing sequence $\{\psi_{j,i}\}_{i=1,2,\ldots}$ of step functions with $\lim_{i \to \infty} \psi_{j,i}(x) = g_j(x)$ for almost every x and $\lim_{i \to \infty} \int \psi_{j,i} = \int g_j$ for $j = 1, 2, \ldots$. We can take these to be nonnegative, as otherwise we could just redefine $\psi_{j,i} = \max\{\psi_{j,i}, 0\}$.

Define $\psi_i = \sum_{j=1}^{i} \psi_{j,i}$, where ψ_i is a nonnegative step function. Then for $i > N$,

$$\begin{aligned} \sum_{j=1}^{N} \int \psi_{j,i} &\le \int \sum_{j=1}^{N} \psi_{j,i} \le \int \psi_i = \sum_{j=1}^{i} \int \psi_{j,i} \\ &\le \sum_{j=1}^{i} \int g_j \le \sum_{j=1}^{i} \int f_{k_j} + \frac{1}{2^j} \le \sum_{j=1}^{i} \frac{1}{2^{j-1}} = 2 \end{aligned}$$

for every i. Then $\int \psi_i$ is bounded and $\psi_{i+1} \ge \psi_i$. By the main technical lemma 8.7, we see that $\psi_i(x)$ are bounded almost everywhere, i.e. for almost every x, there exists M_x with

$$\sum_{j=1}^{N} \psi_{j,i}(x) \le \psi_i(x) \le M_x$$

for all $i > N$. Therefore, $\sum_{j=1}^{N} g_j(x) \le M_x$ for all N. Hence, $\sum_{j=1}^{\infty} g_j(x) \le M_x$. This series converges, so $g_j(x) \to 0$ as $j \to \infty$ for almost every x. Hence, we also have $f_{k_j}(x) \to 0$ for almost every $x \in [a, b]$, which concludes the proof. ☐

10. 4/18

Recall that we defined $\mathcal{L}^1([a, b]) = \{f : f = g - h, g, h \in \mathcal{L}_0[a, b]\}$.

Definition 10.1. Define

$$\|f\|_1 = \int_{[a,b]} |f|.$$

Note that this is almost a norm. It has two of the three norm properties. We have

$$\|f_1 + f_2\|_1 = \int_{[a,b]} |f_1 + f_2| \le \int_{[a,b]} (|f_1| + |f_2|) = \|f_1\|_1 + \|f_2\|_1,$$

and $\|\lambda f\|_1 = |\lambda| \|f\|_1$. However, for the third property, $\|f\|_1 = 0$ implies that $f(x) = 0$ for almost every $x \in [a, b]$. It formally fails the third property, but it does not fail badly. We say that $\|\cdot\|_1$ is a *seminorm*.

Remark. If $f \in \mathcal{L}^1[a, b]$, there exists a sequence $\{\zeta_k\}$ of step functions such that $\|f - \zeta_k\|_1 \to 0$ and $\zeta_k(x) \to f(x)$ for almost every $x \in [a, b]$. This sequence converges in the seminorm and converges almost everywhere pointwise.

Proof. We have $f = g - h$ with $g, h \in \mathcal{L}_0$. This means that $g(x) = \lim \varphi_k(x)$ for almost every x and $\lim \int \varphi_k = \int g$, with $\varphi_{k+1} \geq \varphi_k$. Likewise, $h(x) = \lim \psi_k(x)$ for almost every x and $\lim \int \psi_k = \int h$ with $\psi_{k+1} \geq \psi_k$.

Take $\zeta_k = \varphi_k - \psi_k$. Then $\psi_k(x) = \varphi_k(x) - \psi_k(x) \to g(x) - h(x)$ for almost every x. Also, since $g - \varphi_k \geq 0$ and $h - \psi_k \geq 0$ almost everywhere, we have

$$\int_{[a,b]} |f - \zeta_k| \leq \int_{[a,b]} |g - \varphi_k| + |h - \psi_k| \leq \int_{[a,b]} (g - \varphi_k) + \int_{[a,b]} (h - \psi_k) \to 0. \qquad \square$$

Theorem 10.2 (Completeness of $\mathcal{L}^1$). *Suppose that $\{f_k\} \subset \mathcal{L}^1[a, b]$ is Cauchy with respect to $\|\cdot\|_1$. (This means that for every $\varepsilon > 0$ there exists N such that $\|f_k - f_l\|_1 < \varepsilon$ for all $l > k \geq N$.) Then there exists $f \in \mathcal{L}^1[a, b]$ such that $\|f_k - f\|_1 \to 0$ (i.e. $f_k \to f$ with respect to $\|\cdot\|_1$).*

Proof. Consider applying the Cauchy property with $\varepsilon = 2^{-j}$. Then there exists k_j such that $\|f_l - f_{k_j}\| < 2^{-j}$ for all $l \geq k_j$, and we can choose $k_{j+1} > k_j$ for every j. Hence, $\|f_{k_{j+1}} - f_{k_j}\|_1 < 2^{-j}$.

By the above remark, there exists a step function ζ_j with $\|f_{k_j} - \zeta_j\|_1 < 2^{-j}$. Then we can see that

$$\begin{aligned}\|\zeta_{j+1} - \zeta_j\|_1 &= \|\zeta_{j+1} - f_{k_{j+1}} + f_{k_{j+1}} - f_{k_j} + f_{k_j} - \zeta_j\|_1 \\ &\leq \|\zeta_{j+1} - f_{k_{j+1}}\|_1 + \|f_{k_{j+1}} - f_{k_j}\|_1 + \|f_{k_j} - \zeta_j\|_1 < 2^{-j} + 2^{-j} + 2^{-j} \leq 2^{-j+2}.\end{aligned}$$

Take $\zeta_0 \equiv 0$. Observe that we can always write $a_+ = \max\{a, 0\} \geq 0$ and $a_- = \max\{-a, 0\} \geq 0$, so that $a = a_+ - a_-$; then

$$\zeta_l = \sum_{j=1}^{l} (\zeta_j - \zeta_{j-1}) = \sum_{j=1}^{l} (\zeta_j - \zeta_{j-1})_+ - \sum_{j=1}^{l} (\zeta_j - \zeta_{j-1})_- = \Phi_l - \Psi_l,$$

where Φ_l and Ψ_l are increasing sequences of step functions. Then

$$\int_{[a,b]} (\Phi_l + \Psi_l) = \int_{[a,b]} \sum_{j=1}^{l} |\zeta_j - \zeta_{j-1}| = \sum_{j=1}^{l} \|\zeta_j - \zeta_{j-1}\|_1 \leq 4.$$

Then the first part of the main technical lemma 8.7 tells us that $\Phi_l(x) \to g(x)$ and $\Psi_l(x) \to h(x)$ almost everywhere, where $g, h \in \mathcal{L}_0$, and $\int_{[a,b]} g = \lim \int_{[a,b]} \Phi_l$ and $\int_{[a,b]} h = \lim \int_{[a,b]} \Psi_l$. Therefore, $\zeta_l \to g(x) - h(x)$ almost everywhere. Define $f = g - h \in \mathcal{L}^1$. We know that this is the pointwise limit of ζ_l. In addition,

$$\begin{aligned}\|f - \zeta_l\|_1 &= \int |f - \zeta_l| = \int |(g - \Phi_l) - (h - \Psi_l)| \\ &\leq \int |g - \Phi_l| + \int |h - \Psi_l| = \int (g - \Phi_l) + \int (h - \Psi_l) \to 0.\end{aligned}$$

Now,

$$\int |f - f_{k_l}| = \int |f - \zeta_l + \zeta_l - f_{k_l}| \leq \|f - \zeta_l\|_1 + \|\zeta_l - f_{k_l}\|_1 \to 0.$$

On the other hand, we know that $\int |f_l - f_{k_l}| = \|f_l - f_{k_l}\|_1 \to 0$ as $l \to \infty$ by the Cauchy property. Hence,

$$\|f - f_l\|_1 \leq \|f - f_{k_l}\|_1 + \|f_{k_l} - f_l\|_1 \to 0.$$

We cooked up a function f, and this took some work to do. We needed to find an increasing sequence of step functions. Then we checked that f actually satisfied the properties that we wanted. This concludes the proof. □

There is an important corollary, called the Monotone Convergence Theorem.

Theorem 10.3 (Monotone Convergence Theorem)**.** *Suppose that we have an increasing sequence of functions $\{f_k\} \subset \mathcal{L}^1$, i.e. $f_{k+1} \geq f_k$. Further suppose that $\int_{[a,b]} f_k$ is bounded. Then there exists an $f \in \mathcal{L}^1$ such that $f_k(x) \to f(x)$ almost everywhere and $\int_{[a,b]} f_k \to \int_{[a,b]} f$.*

Proof. Take $l > k$. Then

$$\|f_k - f_l\|_1 = \int_{[a,b]} (f_l - f_k) = \int_{[a,b]} f_l - \int_{[a,b]} f_k.$$

Also, $\int_{[a,b]} f_k$ is a bounded increasing sequence, so $\lim \int_{[a,b]} f_k$ exists. This implies that $\{f_k\}$ is Cauchy. By completeness, we have therefore shown that there exists $f \in \mathcal{L}^1$ such that $\int_{[a,b]} f_k \to \int_{[a,b]} f$.

Now, $\|f_k - f\|_1 \to 0$, which means that $\int_{[a,b]} |f_k - f| \to 0$. By Property 9.4, there exists a subsequence $\left\{f_{k_j}\right\}$ with $f_{k_j}(x) \to f(x)$ almost everywhere, which therefore means that $f_k(x) \to f(x)$ almost everywhere. □

Corollary 10.4. *Let $\{f_k\} \subset \mathcal{L}^1$ with $\int_{[a,b]} |f_k|$ is bounded, and assume also that there exists $f : [a,b] \to \mathbb{R}$ such that $f(x) = \lim f_k(x)$ for almost every $x \in [a,b]$. Then $f \in \mathcal{L}^1$.*

Remark. Caution: It may be false in general that $\int f_k \to \int f$.

This is a very powerful fact that the pointwise limits of $\mathcal{L}^1$ functions converge to an $\mathcal{L}^1$ function. Nothing of this sort is true for the Riemann integral. This fact is extremely general.

11. 4/20

Today we will discuss linear functionals, and go back to the Lebesgue integral next time. We first make some final remarks about orthogonality.

11.1. **Final remarks on orthogonality.** We work in a Hilbert space X. Let E be any non-empty set.

Definition 11.1. The *orthogonal complement* of E is

$$E^\perp = \{x \in X : (x,e) = 0 \text{ for all } e \in E\}.$$

We claim that $E^{\perp}$ is a closed linear subspace of X. This should be an easy exercise. We check that it is closed. Suppose that $y_k \to y$ and $y_k \in E^{\perp}$ for all k, and take any $e \in E$. Then

$$|(y,e)| = |(y-y_k,e)+(y_k,e)| = |(y-y_k,e)| \le \|y-y_k\| \, \|e\| \to 0,$$

so therefore $y \in E^{\perp}$.

Lemma 11.2. *If M is a closed linear subspace of X then*

(i) $M^{\perp} \cap M = \{0\}$.

(ii) $x \in X$ *implies that* $x = y + z$ *with* $y \in M$ *and* $z \in M^{\perp}$*, and this representation is unique.*

(iii) $(M^{\perp})^{\perp} = M$.

Proof. We prove (i). If $x \in M \cap M^{\perp}$ then $(x,x) = 0$, so that $\|x\|^2 = 0$ and hence $x = 0$.

We prove (ii). Recall that there exists $y \in M$ with $\|x-y\| = \min_{z\in M}\|x-z\|$. Notice that $\|x-y-t\lambda z\|^2$ has minimum at $t = 0$, for all scalar $\lambda \neq 0$ and nonzero $z \in M$. Then

$$\|(x-y)-t\lambda z\|^2 = \|x-y\|^2 + t^2\|\lambda z\|^2 - 2t\,\mathrm{Re}(x-y,\lambda z)$$

has minimum at $t = 0$. This implies that $\mathrm{Re}(\overline{\lambda}(x-y,z)) = 0$. Take $\lambda = (x-y,z)$, so we get $(x-y,z) = 0$, and hence $x - y \in M^{\perp}$.

We still need to check uniqueness. Assume that $x = y + z = \tilde{y} + \tilde{z}$ with $y, \tilde{y} \in M$ and $z, \tilde{z} \in M^{\perp}$. Subtracting, we see that $y - \tilde{y} = \tilde{z} - z$ are in M and $M^{\perp}$, so by part (i), we see that $y - \tilde{y} = \tilde{z} - z = 0$, which implies uniqueness.

We prove (iii). Take $x \in M$. Then $(x,w) = 0$ for every $w \in M^{\perp}$. This says that $x \in (M^{\perp})^{\perp}$. Hence, $M \subset (M^{\perp})^{\perp}$. We should prove the reverse inclusion.

Take any $x \in (M^{\perp})^{\perp}$. By part (ii), we can write $x = y + z$ with $y \in M$ and $z \in M^{\perp}$. This tells us that $0 = (x,z) = (y,z) + (z,z) = (z,z) = \|z\|^2$. This means that $z = 0$, so $x = y \in M$. Hence, we've shown that $(M^{\perp})^{\perp} \subset M$. Therefore, $(M^{\perp})^{\perp} = M$. ☐

The main content of this lemma was part (ii).

11.2. **Linear functionals.** Let X be any normed space.

Definition 11.3. f is a *linear functional* on X if $f : X \to \mathbb{R}$ (if X is a real vector space) or if $f : X \to \mathbb{C}$ (if X is a complex vector space).

Lemma 11.4. *Let f be such a linear functional on X. The following are equivalent:*

1. *f is continuous (at each point of X)*
2. *f is continuous at 0*
3. *there exists M with $|f(x)| \le M\|x\|$ for every $x \in X$.*

When the final condition holds, we say that f is a bounded linear function*. The final condition is called* Lipschitz continuity.

Proof. (1) $\implies$ (2) is trivial.

We want to prove that (2) $\implies$ (3). We use the ε-δ definition of continuity at 0 with $\varepsilon = 1$. This means that there exists $\delta > 0$ such that $|f(x)| = |f(x) - f(0)| < 1$ whenever $\|x\| < \delta$. Take $x \in X \setminus \{0\}$. Then

$$\left| f\left(\frac{\delta}{2}\frac{x}{\|x\|}\right)\right| < 1,$$

so hence $|f(x)| \leq \frac{2}{\delta} \|x\|$ for every $x \in X$. This proves (2) $\implies$ (3).

Finally, we will show that (3) $\implies$ (1). Take $x, y \in X$, and $|f(x) - f(y)| = |f(x-y)| \leq M \|x - y\|$, which implies continuity at y. □

From now on, we will usually refer to bounded linear functionals instead of continuous linear functionals; this is more convenient for our purposes.

Definition 11.5. If f is a bounded linear functional on a nontrivial space X, we define

$$\|f\| = \sup\left\{\frac{|f(x)|}{\|x\|} : x \in X \setminus \{0\}\right\}$$

This makes sense because f is a bounded linear functional, so that the set above is nonempty and bounded.

Definition 11.6. Let X^* be the set of all bounded linear functionals on X.

Note that X^* is a linear space.

Proposition 11.7. *$\|f\|$ is a norm.*

Proof. We check each of the three properties:

(1) $\|\lambda f\| = |\lambda| \|f\|$.
(2) $\|f + g\| \leq \|f\| + \|g\|$.
(3) $\|f\| \geq 0$, and $\|f\| = 0$ if and only if $f = 0$. □

Proposition 11.8. *X^* is a Banach space.*

We will prove this proposition later, but it is fairly elementary, so try to do this on your own.

Theorem 11.9 (Riesz(-Fréchet) representation theorem)**.** *Suppose that X is any Hilbert space, any f is any bounded linear functional on X. Then there exists $y \in X$ such that $f(x) = (x, y)$ for all $x \in X$, and such y is unique.*

Proof. First, we consider uniqueness. We have $(x, y) = (x, \tilde{y})$ for all $x \in X$, Then $(x, y - \tilde{y}) = 0$ for every $x \in X$. In particular, choose $x = y - \tilde{y}$, so that $\|y - \tilde{y}\|^2 = (y - \tilde{y}, y - \tilde{y}) = 0$, so that $y = \tilde{y}$.

Now, let $K = \ker f = \{x \in X : f(x) = 0\}$. This is clearly a linear subspace of X. We claim that it is closed. Suppose that $y_k \to y$ with $y_k \in K$ for every K. Then $f(y) = \lim f(y_k)$ by continuity of f at y, but $f(y_k) = 0$ for every k. Hence $f(y) = 0$ and hence $y \in K$.

There are two possibilities. If $f = 0$ then $K = X$. This implies that we can take $y = 0$. We can therefore safely assume that f is not the zero functional. This means that $K \neq X$, so there exists a point $z \in X \setminus K$, and we can choose $z \in K^\perp \setminus \{0\}$. Observe that $f(x) - \lambda f(z) = f(x - \lambda z) = 0$ when $\lambda = \frac{f(x)}{f(z)}$, which means that $x - \frac{f(x)}{f(z)} z \in K$, so it is orthogonal to z, i.e.

$$(x, z) - \frac{f(x)}{f(z)} \|z\|^2 = \left(x - \frac{f(x)}{f(z)} z, z\right) = 0,$$

so hence

$$(x, z) = \frac{f(x)}{f(z)} \|z\|^2,$$

and hence

$$f(x) = \left(x, \frac{\overline{f(z)}}{\|z\|^2} z \right),$$

so we can simply choose $y = \frac{\overline{f(z)}}{\|z\|^2} z$. □

12. 4/22

Let's finish our discussion of the Lebesgue integral. Recall that we proved the Monotone Convergence Theorem 10.3.

This has a very useful corollary:

Corollary 12.1. *We have a given function $f : [a,b] \to \mathbb{R}$ and $\{f_k\} \subset \mathcal{L}^1([a,b])$ with $f_k \to f$ almost everywhere, and $\left\{ \int_{[a,b]} |f_k| \right\}_{k=1,2,\dots}$ is bounded, then $f \in \mathcal{L}^1$.*

This has not many hypothesis.

Proof. The proof is two applications of the Monotone Convergence Theorem. Without loss of generality, we may assume $f_k \geq 0$ for all k because otherwise we can write $f_k = f_{k+} - f_{k-}$, with f_{k+} and f_{k-} as $\mathcal{L}^1$ functions converging to f_+ and f_- respectively.

Suppose that $l \geq 1$, and define the increasing sequence

$$F_{k,l} = -\min\{f_k, \dots, f_{k+l}\},$$

for $l = 1, 2, \dots$. Each of these is bounded above by zero, so the Monotone Convergence Theorem applies. There is therefore some $F_k = \lim_{l\to\infty} F_{k,l} \in \mathcal{L}^1([a,b])$. Here,

$$F_k(x) = -\inf\{f_k(x), \dots, f_{k+l}(x)\},$$

so that

$$-F_k = \inf\{f_k(x), \dots, f_{k+l}(x)\} \leq f_k.$$

is an increasing sequence bounded above by f_k. We can again apply the Monotone Convergence Theorem. Then $\lim -F_k = f$ almost everywhere, so $f \in \mathcal{L}^1([a,b])$. □

Definition 12.2. Define $\mathcal{L}^2([a,b])$ to be the set of functions $f : [a,b] \to \mathbb{R}$ such that both $f, f^2 \in \mathcal{L}^1$.

Proposition 12.3. $\mathcal{L}^2([a,b])$ *is a linear space.*

Proof. It is clear that if $f \in \mathcal{L}^2$ and $\lambda \in \mathbb{R}$, then trivially $\lambda f \in \mathcal{L}^2$.

What's less obvious is that if $f, g \in \mathcal{L}^2$ then $f + g \in \mathcal{L}^2$. We should check this. We are given $f, f^2, g, g^2 \in \mathcal{L}^1$. There exists a sequence φ_k of step functions with $\varphi_k \to f$ almost everywhere. Likewise, there exists a sequence ψ_k of step functions with $\psi_k \to f^2$ almost everywhere. We can arrange to have $\lim \int |f - \varphi_k| \to 0$ and $\lim \int |f^2 - \psi_k| \to 0$. Without loss of generality, we can take $\psi_k \geq 0$ because otherwise, replace ψ_k by $\max\{\psi_k, 0\}$.

Recall that

$$\operatorname{sgn} h(x) = \begin{cases} 1 & h(x) > 0 \\ -1 & h(x) < 0 \\ 0 & h(x) = 0. \end{cases}$$

Let $\Psi_k = \operatorname{sgn}\{\varphi_k\} \sqrt{\psi_k}$. Then $\Psi_k(x) \to f(x)$ almost everywhere, and $\Psi_k^2 \leq \psi_k$, which implies that $\int_{[a,b]} \Psi_k^2$ is bounded (in fact, $\int \Psi_k^2 \leq \int \psi_k$).

Similarly, there exists a sequence Φ_k of step functions with $\Phi_k(x) \to g(x)$ almost everywhere and $\int_{[a,b]} \Phi_k^2$ bounded. Therefore, $(\Phi_k + \Psi_k)^2 \to (f+g)^2$ almost everywhere. Also, $(\Phi_k + \Psi_k)^2 \leq 2(\Phi_k^2 + \Psi_k^2)$, so $\int (\Psi_k + \Phi_k)^2$ is bounded.

Finally, clearly $f + g \in \mathcal{L}^1$, and by the Corollary 12.1 implies that $(f+g)^2 \in \mathcal{L}^1$. $\square$

Suppose that $f, g \in \mathcal{L}^2[a,b]$. Then $fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right) \in \mathcal{L}^1$. We can therefore define an semi-inner product:

Definition 12.4.

$$(f,g)_{\mathcal{L}^2} = \int_{[a,b]} fg$$

for any $f, g \in \mathcal{L}^2$.

We should check the properties. Trivially, this is linear in f, and $(f,g) = (g,f)$. It mildly fails the last property: it is true that $(f,f) \geq 0$, and $(f,f) = 0$ if and only if $f = 0$ almost everywhere. We can define the $\mathcal{L}^2$-seminorm:

Definition 12.5. $\|f\|_2 = \|f\|_{\mathcal{L}^2} = \sqrt{(f,f)}$.

The triangle inequality still holds: $\|f+g\|_{\mathcal{L}^2} \leq \|f\|_{\mathcal{L}^2} + \|g\|_{\mathcal{L}^2}$.

Happily, that's enough to show the Cauchy-Schwarz inequality:

Proposition 12.6. $|(f,g)_{\mathcal{L}^2}| \leq \|f\|_{\mathcal{L}^2} \|g\|_{\mathcal{L}^2}$. *Equivalently,*

$$\int_{[a,b]} fg \leq \sqrt{\int_{[a,b]} f^2 \int_{[a,b]} g^2}.$$

We proved that $\mathcal{L}^1$ is complete with respect to the $\mathcal{L}^1$-seminorm. We claim that this is also true in the $\mathcal{L}^2$ case.

Proposition 12.7. $\mathcal{L}^2$ *is complete with respect to the seminorm* $\|\cdot\|_2$*, that is, if* $\{f_k\} \subset \mathcal{L}^2[a,b]$ *is a Cauchy sequence with respect to this seminorm* $\|\cdot\|_2$ *(i.e. for every* $\varepsilon > 0$ *there exists* N *such that* $\|f_k - f_k\|_2 < \varepsilon$ *for every* $k > l \geq N$*), then there exists* $f \in \mathcal{L}^2[a,b]$ *with* $\|f_k - f\|_2 \to 0$.

Proof. We can always write $f_k = f_{k_+} - f_{k_-}$, so hence without loss of generality we can take $f_k \geq 0$. Then by the Cauchy-Schwarz inequality,

$$\|f_k - f_l\|_1 = \int_{[a,b]} |f_k^2 - f_l^2| = \int_{[a,b]} |f_k - f_l|(f_k + f_l) \leq \|f_k - f_l\|_2 \|f_k + f_l\|_2 .$$

If we use the Cauchy sequence property with $\varepsilon = 1$, then there exists some N_1 such that $\|f_k - f_l\|_2 < 1$ for every $k > l \geq N_1$. In particular, this means that $\|f_k - f_{N_1}\|_2 \leq 1$ for every $k \geq N_1$. By the triangle inequality, we can then write

$$\|f_k\|_2 = \|f_k - f_{N_1} + f_{N_1}\|_2 \leq \|f_k - f_{N_1}\|_2 + \|f_{N_1}\|_2 \leq 1 + \|f_{N_1}\|_2$$

for all $k \geq N_1$, so $\|f_k\|_2$ is bounded. Hence, $\|f_k + f_l\|_2 \leq \|f_k\|_2 + \|f_l\|_2 \leq C$, so therefore

$$\left\|f_k^2 - f_l^2\right\|_1 \leq C\varepsilon$$

for all $k > l > N$. Therefore, $\{f_k^2\}$ is Cauchy with respect to the $\mathcal{L}^1$-seminorm $\|\cdot\|_1$. Therefore, there exists $h \in \mathcal{L}^1([a,b])$ with $\int |f_l^2 - h| \to 0$. Therefore, there exists a subsequence $\{f_{k_j}\}$ with $f_{k_j}^2 \to h$ almost everywhere.

Also, by Cauchy-Schwarz,

$$\|f_k - f_l\|_1 = \int_{[a,b]} |f_k - f_l| \cdot 1 \leq \|f_k - f_l\|_2 \cdot \sqrt{b-a},$$

i.e. $\{f_k\}$ is a Cauchy sequence in $\mathcal{L}^1$, and so hence there exists $f \in \mathcal{L}^1$ with $\|f_k - f\|_1 \to 0$. There is again a subsequence $\{f_{l_j}\}$ with $f_{l_j}(x) \to f(x)$ for almost every x. We can take f_{l_j} to be a subsequence of the f_{k_j} because f_{k_j} is also Cauchy in $\mathcal{L}^1$. Therefore, we have a subsequence $\{f_{l_j}\}$ such that $f_{l_j} \to f$ almost everywhere (with $f \in \mathcal{L}^1$) and $f_{l_j}^2 \to h$ with $h \in \mathcal{L}^1$. Hence, $h = f^2$ almost everywhere, and hence $f^2 \in \mathcal{L}^1$. Therefore, $f \in \mathcal{L}^2$.

We need to check convergence in the $\mathcal{L}^2$ norm. Then by Cauchy-Schwarz,

$$\begin{aligned}\|f_k - f\|_2^2 &= \int_{[a,b]} (f_k - f)^2 = \int_{[a,b]} |f_k - f||f_k - f| \\ &\leq \int_{[a,b]} |f_k - f|(f_k + f) = \left\|f_k^2 - f^2\right\|_1 \to 0,\end{aligned}$$

so hence $\mathcal{L}^2$ is a complete space. □

There are still a few points to clean up, such as the proof of the technical Lemma 8.7. Also, we'll turn the seminorm to an actual norm through an underhanded trick.

13. 4/25

We will finish the discussion of the Lebesgue integral. We want to get around the difficulty that our supposed "norm" is only a seminorm.

Recall that for $f \in \mathcal{L}^1[a,b]$, we have the seminorm $\|f\|_1 = \int_{[a,b]} |f|$. Similarly, for $f \in \mathcal{L}^2[a,b]$, we have a semi-inner product $(f,g) = \int_{[a,b]} fg$.

The problem is that $\|f\|_1 = 0$ or $\|f\|_2 = 0$ implies only that $f(x) = 0$ almost everywhere.

To get around this, instead of considering functions, we consider classes of functions.

Definition 13.1. The $\mathcal{L}^1$ class of f is defined as

$$\underline{f} = \left\{g \in \mathcal{L}^1[a,b] : g(x) = f(x) \text{ for almost every } x \in [a,b]\right\}.$$

Note that $g \in \underline{f}$ if and only if $f \in \underline{g}$ if and only if $\underline{f} = \underline{g}$. Treat these as elements in a new space instead of thinking of them as classes.

We can now *define* $\underline{f} + \underline{g} = \underline{f+g}$. We should check that this makes sense. Indeed, if $f_1 \in \underline{f}$ and $g_1 \in \underline{g}$, then certainly $f_1 + g_1 \in \underline{f+g}$. Similarly, $\lambda\underline{f} = \underline{\lambda f}$. Then we can define

Definition 13.2. $L^1[a,b] = \left\{\underline{f} : f \in \mathcal{L}^1[a,b]\right\}$ is a linear space.

We can now define:

Definition 13.3. $\int_{[a,b]} \underline{f} = \int_{[a,b]} f$ and $\left\|\underline{f}\right\|_1 = \|f\|_1 = \int_{[a,b]} |f|$.

Proposition 13.4. *$L^1[a,b]$ is a normed (Banach) space.*

Proof. We just need to check the properties: $\left\|\lambda\underline{f}\right\|_1 = |\lambda| \left\|\underline{f}\right\|$ and $\left\|\underline{f}+\underline{g}\right\|_1 \leq \left\|\underline{f}\right\|_1 + \left\|\underline{g}\right\|_1$. Also, $\left\|\underline{f}\right\| = 0$ implies that $\|f\|_1 = 0$, so that $f(x) = 0$ almost everywhere, and so $f \in \underline{0} = \{f \in \mathcal{L}^1[a,b] : f(x) = 0 \text{ a.e. } x \in [a,b]\}$. □

Similarly, we can define

Definition 13.5. $L^2[a,b] = \left\{\underline{f} : f \in \mathcal{L}^2[a,b]\right\}$, and we can define $(\underline{f}, \underline{g}) = (f,g) = \int_{[a,b]} fg$ with a norm $\left\|\underline{f}\right\|_2 = \sqrt{(\underline{f}, \underline{f})}$. This actually is now an inner product, so this is now a genuine Hilbert space.

To finish off the Lebesgue theory, we still need to finish the proof of our main technical lemma 8.7.

Lemma 13.6 (Lemma 8.7)**.** *If $\{\psi_k\}$ is an increasing sequence of step functions on $[a,b]$ with $\left\{\int_{[a,b]} \psi_k\right\}$ bounded, then $\{\psi_k(x)\}$ is bounded (and hence convergent) for almost every $x \in [a,b]$.*

Proof of Lemma 8.7. Let $S = \{x \in [a,b] : \{\psi_k(x)\}$ is not bounded$\}$. Therefore, for $x \in S$, we see that $\lim_{k\to\infty} \psi_k(x) = \infty$. We want to prove that S has measure zero.

Let's agree to use the notation $\tilde{\psi}_k(x) = \psi_k(x) - \psi_1(x)$. Notice that this is a nonnegative function, and $\tilde{\psi}_1(x) \equiv 0$. Equivalently, we see that $S = \{x \in [a,b] : \{\tilde{\psi}_k(x)\}$ is not bounded$\}$.

Take an arbitrary $\alpha > 0$. Let $S_k = \{x \in [a,b] : \tilde{\psi}_k(x) > \alpha\}$. Then $S_{k+1} \supset S_k$ for all k, and $S_1 = \emptyset$. We can therefore write

$$\begin{aligned} S_N &= (S_N \setminus S_{N-1}) \cup (S_{N-1} \setminus S_{N-2}) \cup \cdots \cup (S_2 \setminus S_1) = \bigcup_{k=1}^{N-1} (S_{k+1} \setminus S_k) \\ &= \bigcup_{k=1}^{N-1} \left\{x : \tilde{\psi}_k(x) \le \alpha < \tilde{\psi}_{k+1}(x)\right\}. \end{aligned}$$

Recall that $\tilde{\psi}_k$ are step functions. For each $k = 1, 2, \ldots$, consider the partition $a = x_0^{(k)} < x_1^{(k)} < \cdots < x_{N_k}^{(k)} = b$ so that $\tilde{\psi}_k$ is constant on each interval of this partition. Choose these partitions to be compatible with both $\tilde{\psi}_k$ and $\tilde{\psi}_{k+1}$ (e.g. by taking a common refinement).

Define $\mathcal{P}_k = \{(x_{i-1}^{(k)}, x_i^{(k)}) : i = 1, \ldots, N_k\}$. Then $S_{k+1} \setminus S_k$ is a subcollection of these. Then we have

$$\mathcal{Q}_k \subset S_{k+1} \setminus S_k \subset \mathcal{Q}_k \cup \{x_0^{(k)}, \ldots, x_{N_k}^{(k)}\}$$

where $\mathcal{Q}_k$ is some subcollection of $\mathcal{P}_k$.

Therefore,

$$\bigcup_{k=1}^{N-1} \mathcal{Q}_k \subset S_N \subset \bigcup_{k=1}^{N-1} \left(\mathcal{Q}_k \cup \{x_0^{(k)}, \ldots, x_{N_k}^{(k)}\}\right)$$

Thus $\tilde{\psi}_N(x) > \alpha$ on $\cup_{k=1}^{N-1} \mathcal{Q}_k$, i.e.

$$\tilde{\psi}_N \ge \alpha \sum_{k=1}^{N-1} \sum_{I \in \mathcal{Q}_k} \chi_I.$$

We can now integrate this. This implies that

$$C \ge \int_{[a,b]} \tilde{\psi}_N \ge \alpha \sum_{k=1}^{N-1} \sum_{I \in \mathcal{Q}_k} |I|$$

for some constant upper bound C independent of N. Hence

$$\alpha \sum_{k=1}^{\infty} \sum_{I \in \mathcal{Q}_k} |I| \leq C = \lim \int_{[a,b]} \tilde{\psi}_k.$$

Notice that by our preceding inclusions, we have

$$S_N \subset \bigcup_{k=1}^{N-1} (\mathcal{Q}_k) \cup E_N$$

where E_N is finite. Then

$$\bigcup_{N=1}^{\infty} S_N \subset \left(\bigcup_{N=1}^{\infty} \bigcup_{k=1}^{N-1} \mathcal{Q}_k \right) \cup \left(\bigcup_{N=1}^{\infty} E_N \right).$$

Note that $\cup_{N=1}^{\infty} E_N$ is countable and hence has measure zero, and $\cup_{N=1}^{\infty} \cup_{k=1}^{N-1} \mathcal{Q}_K$ has sum of lengths of intervals $\leq C/\alpha$.

Observe that if we take a point $y \in S$ then $y \in S_k$ for all sufficiently large k, so $S \subset \cup_{N=1}^{\infty} S_N$.

We can now finish the proof. Let $\varepsilon > 0$ be given. Choose $\alpha > C/\varepsilon$. Then select $J_1, J_2, \ldots$ with $\cup_{N=1} E_N \subset \bigcup_{j=1}^{\infty} J_j$ and $\sum |J_j| < \varepsilon$.

$$S \subset \bigcup_{N=1}^{\infty} S_N \subset \left(\bigcup_{N=1}^{\infty} \bigcup_{k=1}^{N-1} \mathcal{Q}_k \right) \cup \left(\bigcup_{N=1}^{\infty} E_N \right) \subset \bigcup_{N=1}^{\infty} \left(\bigcup_{k=1}^{N-1} \mathcal{Q}_k \right) \cup \left(\bigcup_{j=1}^{\infty} J_j \right),$$

which now has length $< 2\varepsilon$, so we are done. □

This was the hardest proof of the Lebesgue theory, which shouldn't be surprising since everything followed from it.

13.1. **Linear operators.** Now we will consider the next topic, which is linear operators.

Definition 13.7. Let E and F be normed spaces. Then $T : E \to F$ is a *linear operator* (i.e. linear map) if $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ for all $x, y \in E$ and for all scalars α, β.

Definition 13.8. We say that T is *bounded* if there exists M such that $\|T(x)\| \leq M \|x\|$ for all $x \in E$.

We should be somewhat careful here; $\|T(x)\|$ is a norm in F, while $\|x\|$ is a norm in E.

In the same way as for the linear functionals (as in Lemma 11.4), we get the following theorem:

Theorem 13.9. *T is continuous at each point of E if and only if T is continuous at 0 if and only if T is bounded.*

14. 4/29

14.1. **Bounded linear operators.** We started our discussion of bounded linear operators. Let E and F be normed spaces, and suppose that $T : E \to F$ is linear. Then T is *bounded* means that there exists M with $\|T(x)\| \leq M\|x\|$ for every $x \in E$. In this case, we can define the *operator norm*

$$\|T\| = \sup\left\{\frac{\|T(x)\|}{\|x\|} : x \in E \setminus \{0\}\right\} = \sup\left\{\|T(x)\| : \|x\| = 1\right\}.$$

Exercise 14.1. Check that $\|T\|$ is a norm on $\mathcal{L}(E,F)$.

A special case of this discussion is when $F = \mathbb{R}$ or $\mathbb{C}$, when T is just a linear functional. Note that T is continuous if and only if T is continuous at 0 if and only if T is bounded.

Definition 14.2. Let $\mathcal{L}(E,F)$ be the set of all bounded linear maps $E \to F$, equipped with the operator norm.

Lemma 14.3. *$\mathcal{L}(E,F)$ is a Banach space provided that F is a Banach space.*

The proof will be very similar to all of the other completeness proofs that we've seen.

Proof. $\mathcal{L}(E,F)$ is clearly a normed linear space, so we just have to check completeness. Let $\{T_k\}_{k=1,2,\ldots}$ be any Cauchy sequence in $\mathcal{L}(E,F)$. That is, let $\varepsilon > 0$; then there exists N such that $\|T_k - T_l\| < \varepsilon$ for every $l > k \geq N$. This implies that $\|T_k(x) - T_l(x)\| \leq \varepsilon\|x\|$ for every $x \in E$. Take any fixed x; this means that $\{T_k(x)\}_{k=1,2,\ldots}$ is Cauchy in F and hence convergent (since F is a Banach space) to some limit which we call $T(x)$.

Now,

$$T(\lambda x + \mu y) = \lim_{k\to\infty} T_k(\lambda x + \mu y) = \lim_{k\to\infty}(\lambda T_k(x) + \mu T_k(y)) = \lambda T(x) + \mu T(y).$$

Hence, $T : E \to F$ is linear. Now, $\lim_{l\to\infty}(T_k(x) - T_l(x)) = T_k(x) - T(x)$.

Recall that if $\lim z_k = z$ then $\lim \|z_k\| = \|z\|$ because $|\,\|z_k\| - \|z\|\,| \leq \|z_k - z\|$.

Therefore, $\lim_{k\to\infty}\|T_k(x) - T_l(x)\| = \|T_l(x) - T(x)\|$, and thus $\|T_k(x) - T(x)\| \leq \varepsilon\|x\|$ for each $x \in E$ and for all $k \geq N$. This says that $T_k - T \in \mathcal{L}(E,F)$, and it has $\|T_k - T\| < \varepsilon$ for all $k \geq N$. Since $\mathcal{L}(E,F)$ is linear, we see that $T = T_k - (T_k - T) \in \mathcal{L}(E,F)$, and furthermore $\|T_k - T\| < \varepsilon$ for all $k \geq N$ implies that $\lim_{k\to\infty}\|T_k - T\| = 0$. This concludes the proof. $\square$

Recall that E^* is the set of all bounded linear functionals on E, i.e. it is the set of all bounded linear maps $E \to \mathbb{R}$ if E is a real vector space, and it is the set of all bounded linear maps $E \to \mathbb{C}$ if E is a complex vector space. Since $\mathbb{R}$ and $\mathbb{C}$ are complete, we can apply the previous lemma to get the following corollary:

Corollary 14.4. *Let E be a normed space. Then E^* is complete.*

Proposition 14.5. *Let E, F, G be three normed spaces, and suppose $S \in \mathcal{L}(E,F)$ and $T \in \mathcal{L}(F,G)$. Then $T \circ S \in \mathcal{L}(E,G)$ and $\|T \circ S\| \leq \|T\|\,\|S\|$.*

Proof. We check this claim. Here, $\|T \circ S(x)\| = \|T(S(x))\| \leq \|T\|\,\|S(x)\| \leq \|T\|\,\|S\|\,\|x\|$ for all $x \in E$. The desired result now follows. $\square$

In particular, when $T \in \mathcal{L}(E,E)$, this result allows us to consider the nth power $T^n = T \circ \cdots \circ T$, and by induction, $\|T^n\| \leq \|T\|^n$. Notice that equality needn't hold. We might even get that $T^n = 0$. We use the convention that $T^0 = I$ is the identity.

14.2. **Inverses.** Now we discuss inverses.

Definition 14.6. $T \in \mathcal{L}(E, F)$ is *invertible* if there exists $S \in \mathcal{L}(F, E)$ with $T \circ S = I_F$ and $S \circ T = I_E$.

Remark. There exists a linear S (not necessarily bounded) with $T \circ S = I_F$ and $S \circ T = I_E$ if and only if T is both one-to-one and onto.

Theorem 14.7 (Bounded inverse theorem)**.** *Such an S is automatically bounded if E and F are Banach spaces.*

We will not prove this theorem in this course.

Remark. Suppose that $T \in \mathcal{L}(E, F)$. If there exists S such that $T \circ S = I_F$ then T is onto. If there exists S such that $S \circ T = I_E$ then T is one-to-one. These statements are not the same. Here's a nice example:

Example 14.8. Let $E = F = \ell^2$, and define S to be the *shift operator* $S(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$. This is one-to-one but *not* onto. There is also a reverse shift operator, so that $\tilde{S}(x_1, x_2, \dots) = (x_2, x_3, \dots)$ is onto but *not* one-to-one.

Note that $\tilde{S} \circ S = 1_{\ell^2}$ is the identity, but $S \circ \tilde{S} \neq 1_{\ell^2}$ is not the identity.

Here is a very useful theorem about finding inverses.

Theorem 14.9. *Suppose that E is Banach and $T \in \mathcal{L}(E, E)$ with $\|T\| < 1$. Then $I - T$ is invertible. In fact,*

$$(I - T)^{-1} = \sum_{n=0}^{\infty} T^n.$$

Here, we have $\sum_{n=0}^{\infty} T^n = \lim_{N \to \infty} \sum_{n=0}^{N} T^n$ with the limit taken in the operator norm.

Proof. Suppose that $M > N$. Then

$$\left\| \sum_{n=0}^{M} T^n - \sum_{n=0}^{N} T^n \right\| = \left\| \sum_{n=N+1}^{M} T^n \right\| \leq \sum_{n=N+1}^{M} \|T\|^n \leq \sum_{n=N+1}^{\infty} \|T\|^n = \|T\|^{N+1} \left(\frac{1}{1 - \|T\|} \right) \to 0$$

as $N \to \infty$, which means that $\left\{ \sum_{n=0}^{N} T^n \right\}_{N=1,2,\dots}$ is a Cauchy sequence with respect to the operator norm, and is hence convergent. Hence, our infinite sum actually makes sense.

Now

$$(I - T) \circ \sum_{n=0}^{\infty} T^n = \sum_{n=0}^{\infty} T^n - \sum_{n=1}^{\infty} T^n = I,$$

and similarly, $(\sum_{n=0}^{\infty} T^n) \circ (I - T) = I$. ∎

This is a very useful fact.

15. 5/2

Last time we proved that if E is a Banach space, $A \in \mathcal{L}(E, E)$, and $\|A\| < 1$, then $I - A$ is invertible (i.e. $(I - A)^{-1} \in \mathcal{L}(E, E)$). In fact, $(I - A)^{-1} = \sum_{n=0}^{\infty} A^n$ in the operator norm.

Corollary 15.1. *If E and F are Banach spaces, the set of all invertible $A \in \mathcal{L}(E, F)$ is an open subset of $\mathcal{L}(E, F)$.*

Proof. We claim that if $A \in \mathcal{L}(E,F)$ is invertible then $B \in \mathcal{L}(E,F)$ is invertible for every $B \in \mathcal{L}(E,F)$ such that $\|B-A\| < \frac{1}{\|A^{-1}\|}$. First, we check this claim.

Observe that if $\|B-A\| < \frac{1}{\|A^{-1}\|}$ then $\|(B-A)\circ A^{-1}\| \le \|B-A\|\,\|A^{-1}\| < 1$. This means by Theorem 14.9, $I-(B-A)\circ A^{-1}$ is invertible, which means that $(B\circ A^{-1})\circ A = B$ is invertible.

This proves the claim and hence the corollary. □

15.1. **Adjoints.** We briefly digress and talk about adjoints.

Definition 15.2. If H and K are Hilbert spaces and $A \in \mathcal{L}(H,K)$ is a bounded linear operator, then there exists an unique operator $A^* \in \mathcal{L}(K,H)$ with $(A(x),y) = (x, A^*(y))$ for every $x \in H$ and $y \in K$. Here, A^* is called the *adjoint operator*.

This is a direct consequence of the Riesz representation theorem.

Proof. Homework 5, problem 3 handled the real case. The complex case is almost exactly the same, except we need to check that linearity works out correctly. We can do this:

$$\begin{aligned}(x, A^*(\lambda y+\mu w)) &= (A(x), \lambda y + \mu w) = \overline{\lambda}(A(x),y) + \overline{\mu}(A(x),w) = \overline{\lambda}(x,A^*y)+\overline{\mu}(x,A^*w)\\ &= (x,\lambda A^*y) + (x,\mu A^* w)\end{aligned}$$

for every $x \in H$ and $y, w \in K$, so that $(x, A^*(\lambda y+\mu w) - (\lambda A^*y + \mu A^*w)) = 0$ for all $x \in H$, which means that $A^*(\lambda y + \mu w) = \lambda A^*(y) + \mu A^*(w)$.

We also want to say that A^* is unique. Suppose that $(x, A^*y - \tilde{A}^*y) = 0$ for all $x \in H$ and $y \in K$. Then $A^*y = \tilde{A}^*y$, and hence A^* is unique. □

Proposition 15.3. *A^* is bounded, $\|A^*\| = \|A\|$, and in fact, $(A^*)^* = A$.*

Proof. Take $x = A^*y$ in the definition of the adjoint, and use the Cauchy-Schwarz inequality to see that

$$\|A^*y\|^2 = (A(A^*(y)),y) \le \|A(A^*(y))\|\,\|y\| \le \|A\|\,\|A^*y\|\,\|y\|,$$

which means that $\|A^*y\| \le \|A\|\,\|y\|$ for every $y \in K$, which tells us that $\|A^*\|$ is bounded and $\|A^*\| \le \|A\|$.

In addition, we also have that $(A^*y,x) = (y,Ax)$. This shows that $(A^*)^* = A$, which means that $\|A\| \le \|A^*\|$. □

Let's consider what this looks like in finite dimensional space.

Example 15.4. Suppose that $H = \mathbb{C}^n$ and $K = \mathbb{C}^m$. Then $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^m)$ is the same as saying that A is given by an $m \times n$ matrix. Given $z = (z_1, \ldots, z_n)$, we want to have

$$Az = \sum_{i=1}^m \sum_{j=1}^n a_{ij} z_j e_i.$$

Suppose that $w = (w_1, \ldots, w_m)$. In this case, the adjoint property says that

$$(A(z),w) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} z_j \overline{w_i} = \sum_{i=1}^m \sum_{j=1}^n z_j \overline{(\overline{a_{ij}} w_i)}.$$

Here, A^* is given by

$$A^* w = \sum_{i=1}^{m} \overline{a_{ij}} w_i,$$

which means that $A^* = (\overline{a_{ij}})^T$.

Definition 15.5. $A \in \mathcal{L}(H, H)$ is *self-adjoint* (or in the complex case, *Hermitian*) if $A^* = A$.

In the finite dimensional case, $H = \mathbb{C}^n$. This means that $(\overline{a_{ij}})^T = (a_{ij})$, which means that $\overline{a_{ij}} = a_{ji}$ for all i and j. This also means that the diagonal entries are real.

Proposition 15.6. *If $A \in \mathcal{L}(H, H)$ is self-adjoint then (Ax, x) is always real.*

Proof. This is because $(Ax, x) = \overline{(x, Ax)} = \overline{(Ax, x)}$ by the self-adjointness property. ☐

Proposition 15.7. *If $A \in \mathcal{L}(H, H)$ is self-adjoint, then*

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x \neq 0} \frac{|(Ax, x)|}{\|x\|^2}.$$

Proof. Let $m = \sup_{x \neq 0} \frac{|(Ax,x)|}{\|x\|^2}$. Notice that Cauchy-Schwarz implies that $m \leq \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \|A\|$. We need to prove the reverse inequality.

Note that $(A(x+y), x+y) = (A(x), x) + (A(y), y) + (A(y), x) + (A(x), y)$. Here, notice that $(A(y), x) = \overline{(x, A(y))} = \overline{(A(x), y)}$. Then we have

$$\begin{aligned} (A(x+y), x+y) &= (A(x), x) + (A(y), y) + 2\operatorname{Re}(A(x), y) \\ (A(x-y), x-y) &= (A(x), x) + (A(y), y) - 2\operatorname{Re}(A(x), y). \end{aligned}$$

Subtracting, we get that

$$\begin{aligned} 4\operatorname{Re}(A(x), y) &= (A(x+y), x+y) - (A(x-y), x-y) \\ &\leq m(\|x+y\|^2 + \|x-y\|^2) \leq 2m(\|x\|^2 + \|y\|^2). \end{aligned}$$

Suppose that $A(x) \neq 0$. Then choose $y = \frac{\|x\|}{\|Ax\|} A(x)$, to get $4\|x\|\|A(x)\| \leq 4m\|x\|^2$. Therefore, $\|Ax\| \leq m\|x\|$ for every x, so therefore $\|x\| \leq m$, which means that $\|A\| = m$. ☐

16. 5/4

Today we will talk about the spectrum. Suppose that E is a complex Banach space and suppose that we have a bounded linear map $A \in \mathcal{L}(E, E)$.

16.1. **Spectrum.**

Definition 16.1. The *spectrum* of A is $\sigma(A) = \{\lambda \in \mathbb{C} : \lambda I - A \text{ is not invertible}\}$.

Recall that we say that $A \in \mathcal{L}(E, E)$ is *invertible* if A is 1-1, onto, and $A^{-1} \in \mathcal{L}(E, E)$.

Remark. $\{\lambda : \lambda I - A \text{ is not 1:1}\} \cup \{\lambda : \lambda I - A \text{ is not onto}\} \subset \sigma(A)$.

If $\lambda I - A$ is not 1:1, then there exists $x \neq 0$ with $(\lambda I - A)(x) = 0$, i.e. $Ax = \lambda x$. Here, λ is an *eigenvalue*, and this means that all eigenvalues are in $\sigma(A)$. Note that the spectrum might be much bigger, however.

Example 16.2. Let $E = \mathbb{C}^n$ with the usual inner product norm. $A \in \mathcal{L}(E, E)$ is given by an $n \times n$ matrix, so $\sigma(A)$ is exactly the set of eigenvalues.

Example 16.3. Let $E = \ell^2_{\mathbb{C}}$. Take any $z = (z_1, z_2, \dots) \in \ell^2_{\mathbb{C}}$. Let S be the shift operator $S(z) = (0, z_1, z_2, \dots)$. Clearly, $S \in \mathcal{L}(E, E)$; in fact, it is norm preserving. It has no eigenvalues. To see this, suppose that there were some z such that $Sz = \lambda z$. Then $(0, z_1, z_2, \dots) = (\lambda z_1, \lambda z_2, \dots)$. This means that $z_1 = 0$, which means that $z_2 = 0, \dots$. That is, $z = 0$, and hence no λ is an eigenvalue. However, the spectrum of this shift operator is actually very big: $\sigma(S) = \{\lambda \in \sigma : |\lambda| \le 1\}$.

Lemma 16.4. *$\sigma(A)$ is a closed subset of $\{\lambda \in \mathbb{C} : |\lambda| \le \|A\|\}$.*

Proof. Take any $\lambda \notin \sigma(A)$, which implies that $\lambda I - A$ is invertible. Recall that the set of invertible operators in $\mathcal{L}(E, E)$ is an open subset of $\mathcal{L}(E, E)$ with respect to the operator norm. Then there exists $\delta > 0$ such that $\|B - (\lambda I - A)\| < \delta$ implies that B is invertible. In particular, if $\mu \in \mathbb{C}$ then $|\mu - \lambda| < \delta$ implies that $\|(\mu I - A) - (\lambda I - A)\| < \delta$, and hence $\mu I - A$ is invertible. We have therefore shown that $\sigma(A)$ is closed.

Now, suppose that $|\lambda| > \|A\|$. Then $\lambda I - A = \lambda(I - \frac{1}{\lambda}A)$. Note that $\left\|\frac{A}{\lambda}\right\| = \frac{\|A\|}{|\lambda|} < 1$, so by our general theorem 14.9, this means that $\lambda I - A$ is invertible.

This means that no λ with $|\lambda| > \|A\|$ is in $\sigma(A)$, or in other words, we've proved that $\sigma(A) \subset \{\lambda : |\lambda| \le \|A\|\}$. □

16.2. **Compact operators.**

Definition 16.5. Suppose that E and F are normed linear spaces. A linear operator $T : E \to F$ is *compact* if $T(\{x \in E : \|x\| \le 1\})$ is contained in a compact subset of F.

Remark. Note that this is the same as saying that for all $R > 0$, we know that $T(\{x \in E : \|x\| < R\})$ is contained in a compact set.

T compact implies that T is bounded. Assume otherwise; then T is not bounded. Then there exists a sequence $\{x_k\} \subset E$ with $\|x_k\| = 1$ and $\|T(x_k)\| \ge k$, which means that there does not exist a convergent subsequence of $\{T(x_k)\}$. If there were such a convergent subsequence $T(x_{k_j}) \to y$ then $\left\|T(x_{k_j})\right\| \to \|y\|$.

Example 16.6. Suppose that E is an infinite dimensional normed space. Then $I_E \in \mathcal{L}(E, E)$ is certainly bounded, but it is not compact because the closed unit ball in E is not compact: there exists a sequence $e_1, e_2, \dots \in E$ with $\|e_j\| = 1$ for all j and $\|e_i - e_j\| \ge 1$ with $i \ne j$.

Example 16.7. Suppose that $T \in \mathcal{L}(E, F)$ with $H = T(E)$ as a finite dimensional subspace of F. We say that T is an operator of finite rank, and we claim that T is compact.

We should check this. Take any sequence $\{T(x_k)\}_{k=1,2,\dots} \subset H$, where $x_k \in E$ and $\|x_k\| \le 1$. Since H is finite-dimensional, this means that all bounded subsets of H have compact closure. Thus there exists a convergent subsequence $\{T_{y_{k_j}}\}$.

There is an extremely important theorem.

Theorem 16.8. *If E and F are Banach spaces and we take a sequence $\{T_k\}_{k=1,2,\dots} \subset \mathcal{L}(E, F)$ with each T_k compact and $\|T_k - T\| \to 0$ for some $T \in \mathcal{L}(E, F)$ (i.e. $T_k \to T$ in $\mathcal{L}(E, F)$). Then T is compact.*

Proof. Take any sequence $\{T(x_k)\}_{k=1,2,\dots}$ with $\|x_k\| \le 1$. Then $\{T_1(x_k)\}$ has a convergent subsequence $T_1(x_{1,k})$. Now, $\{T_2(x_{1,k})\}$ has a convergent subsequence $T_2(x_{2,k})$. The important thing is that $\{x_{2,k}\}$ is a subsequence of $\{x_{1,k}\}$ is a subsequence of $\{x_k\}$. We continue this process of finding subsequences of subsequences. Inductively, $T_q(x_{q-1,k})$ has a convergent subsequence $T_q(x_{q,k})$.

Now we use a standard trick in analysis by taking the diagonal sequence. Let $\varepsilon > 0$ be given. Consider $\{x_{q,q}\}$. This is a subsequence of $\{x_{q,k}\}_{k=q,q+1,\dots}$. Therefore, $\{T_p(x_{q,q})\}_{q=1,\dots}$ converges. That's because $\{x_{q,q}\}_{q=p,p+1,\dots}$ is also a subsequence of $\{x_{p,k}\}_{k=1,2,\dots}$.

The desired result should now follow easily, and we'll finish it next time. □

17. 5/6

We have Banach spaces E and F, and $T_k \to T$ in $\mathcal{L}(E,F)$. We want to show that if T_k is compact then T is compact. This is Theorem 16.8.

Finishing proof of theorem 16.8. We proved that for any sequence $\{x_k\}_{k=1,2,\dots}$ with $\|x_k\| \le 1$ for all k, there exists a subsequence $\{x_{q,q}\}_{q=1,2,\dots}$ of $\{x_k\}_{k=1,2,\dots}$ such that $\{T_p(x_{q,q})\}_{q=1,2,\dots}$ converges for each p. This was done with a diagonal argument. We claim that $\{T(x_{q,q})\}_{q=1,2,\dots}$ converges.

Let $\varepsilon > 0$ be given. Pick p such that $\|T_p - T\| < \varepsilon$. Using the Cauchy property, we can pick N such that $\|T_p(x_{q,q}) - T_p(x_{r,r})\| < \varepsilon$ for every $q > r \ge N$. Then

$$\begin{aligned}\|T(x_{q,q}) - T(x_{r,r})\| &= \|T(x_{q,q}) - T_p(x_{q,q}) + T_p(x_{q,q}) - T_p(x_{r,r}) + T_p(x_{r,r}) - T(x_{r,r})\| \\ &\le \|T(x_{q,q}) - T_p(x_{q,q})\| + \|T_p(x_{q,q}) - T_p(x_{r,r})\| + \|T_p(x_{r,r}) - T(x_{r,r})\| \\ &\le \|T - T_p\| + \varepsilon + \|T_p - T\| < 3\varepsilon\end{aligned}$$

for every $q > r \ge N$. Hence, $\{T(x_{q,q})\}_{q=1,2,\dots}$ is Cauchy and hence convergent. □

17.1. Hilbert-Schmidt operators.

Definition 17.1. Suppose that H is a Hilbert space and Y is a Banach space. An operator $T \in \mathcal{L}(H,Y)$ is said to be a *Hilbert-Schmidt* operator if there exists a complete orthonormal sequence $e_1, e_2, \dots$ for H such that $\sum_{j=1}^{\infty} \|T(e_j)\|^2 < \infty$.

Remark. It doesn't matter which complete orthonormal sequence we use. If $(f_1, f_2, \dots)$ is any other complete orthonormal sequence then $\sum_{j=1}^{\infty} \|T(f_j)\|^2 < \infty$ as well. This is something that will be checked on the homework.

Hilbert-Schmidt operators are important because a lot of common operators satisfy this property, and because of the following theorem.

Theorem 17.2. *Hilbert-Schmidt operators are automatically compact.*

Proof. Let $x_j = (x, e_j)$. Define $T_N(x) = T(\sum_{j=1}^{N} x_j e_j) = \sum_{j=1}^{N} x_j T(e_j)$. Then $T_N(H)$ is a finite-dimensional space $T_N(H) = \operatorname{span}\{T(e_1), \dots, T(e_N)\}$. That is, T_N is finite rank and hence compact by example 16.7. Then

$$(T - T_N)(x) = T\left(\sum_{j=1}^{\infty} x_j e_j\right) - T\left(\sum_{j=1}^{N} x_j e_j\right) = T\left(\sum_{j=N+1}^{\infty} x_j e_j\right) = \sum_{j=N+1}^{\infty} x_j T(e_j).$$

We should be careful and check that this final equality holds. We have

$$\sum_{j=N+1}^{\infty} x_j e_j = \lim_{M\to\infty} \sum_{j=N+1}^{M} x_j e_j \implies T\left(\sum_{j=N+1}^{\infty} x_j e_j\right) = \lim_{M\to\infty} \sum_{j=N+1}^{M} x_j T(e_j).$$

For $M_2 > M_1$, we can write

$$\left\| \sum_{j=N+1}^{M_2} x_j T(e_j) - \sum_{j=N+1}^{M_1} x_j T(e_j) \right\| = \left\| \sum_{j=M_1+1}^{M_2} x_j T(e_j) \right\| \leq \sum_{j=M_1+1}^{M_2} |x_j| \, \|T(e_j)\|$$
$$\leq \|x\| \sqrt{\sum_{j=M_1+1}^{M_2} \|Te_j\|^2},$$

which goes to zero because T is Hilbert-Schmidt, and $T(\sum_{j=1}^{\infty} x_j e_j)$ exists.

We now have

$$\|(T - T_N)(x)\| \leq \|x\| \sqrt{\sum_{j=N+1}^{\infty} \|T(e_j)\|^2} \implies \|T - T_N\| \leq \sqrt{\sum_{j=N+1}^{\infty} \|T(e_j)\|^2} \to 0$$

as $N \to \infty$. □

We should see that not all compact operators are Hilbert-Schmidt.

Example 17.3. Recall that we have the space $\ell^2 = \{x = (x_1, x_2, \dots) : x_j \in \mathbb{R}, \sum x_j^2 < \infty\}$. This has the usual inner product norm.

Consider $T : \ell^2 \to \ell^2$ defined by $T(x) = (x_1, \frac{x_2}{\sqrt{2}}, \frac{x_3}{\sqrt{3}}, \dots, \frac{x_n}{\sqrt{n}}, \dots)$. We claim that this is a compact operator that is not Hilbert-Schmidt.

Let $e_1 = (1, 0, \dots)$, $e_2 = (0, 1, 0, \dots)$ be the standard complete orthonormal basis. Then $T(x) = \sum_{j=1}^{\infty} \frac{x_j}{\sqrt{j}} e_j$. Now, $T(e_j) = \frac{1}{\sqrt{j}}$ for all $j = 1, 2, \dots$, which implies that $\sum \|T(e_j)\|^2 = \sum \frac{1}{j} = \infty$. Therefore, T is not Hilbert-Schmidt.

We still need to show that T is compact. Consider $T_N(x) = \sum_{j=1}^{N} \frac{x_j}{\sqrt{j}} e_j$. These have finite rank and are hence compact. Then

$$\|T(x) - T_N(x)\|^2 = \left\| \sum_{j=N+1}^{\infty} \frac{x_j}{\sqrt{j}} e_j \right\|^2 = \sum_{j=N+1}^{\infty} \frac{x_j^2}{j} \leq \frac{1}{N+1} \sum_{j=N+1}^{\infty} x_j^2 \leq \frac{1}{N+1} \|x\|^2 .$$

This means that $\|T - T_N\| \leq \frac{1}{\sqrt{N+1}} \to 0$, and hence T is compact.

17.2. **Spectral theorem.** This the simplest spectral theorem, for compact Hermitian operators. We might have seen the spectral theorem for symmetric finite dimensional matrices. This is the generalization to infinite dimensional complex space.

Theorem 17.4. *Suppose that H is a Hilbert space and T is a complex Hermitian operator, i.e. $(Tx, y) = (x, Ty)$ for all $x, y \in H$. Suppose $T \in \mathcal{L}(H, H)$. Then $(\ker T)^{\perp} = \overline{T(H)}$ (the closure of $T(H)$). Equivalently, we can write $\ker T = (\overline{T(H)})^{\perp}$.*

If $T(H)$ is infinite-dimensional, there exists a complete orthonormal sequence $e_1, e_2, \dots$ for $\overline{T(H)}$ such that $T(e_j) = \lambda_j e_j$ for every j, where λ_j are real numbers and $\lambda_j \to 0$ as $j \to \infty$. Also, for every $x = \sum_{j=1}^{\infty} x_j e_j \in (\ker T)^{\perp}$, so that $T(x) = \sum_{j=1}^{\infty} \lambda_j x_j e_j$.

If $T(H)$ is finite dimensional, then there exists an orthonormal basis $e_1, \dots, e_N$ with $T(e_j) = \lambda_j e_j$ for all $j = 1, \dots, N$.

We will prove the Spectral Theorem 17.4. This requires three lemmas. The first two only require Hermitian and do not need compactness.

Lemma 18.1. *Let $T : H \to H$ be Hermitian. Then all of its eigenvalues are real and eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof. Suppose that $Tx = \lambda x$ for $x \neq 0$. Then $(Tx, x) = \lambda(x, x) = \lambda \|x\|^2$. Note that $(Tx, x) = (x, Tx)$ so therefore this is real, and hence λ is real.

Now, suppose that $Tx = \lambda x$ and $Ty = \lambda y$ for $\lambda \neq \mu$. Then $(Tx, y) = \lambda(x, y)$ and $(x, Ty) = \mu(x, y)$, which means that $\lambda(x, y) = \mu(x, y)$, which means that $0 = (\lambda - \mu)(x, y)$, which means that $(x, y) = 0$. □

Lemma 18.2. *Suppose that $T : H \to H$ is Hermitian, and suppose that L is any subspace of H with $T(L) \subset L$. Then $T(L^\perp) \subset L^\perp$.*

Proof. Take $y \in T(L^\perp)$, that is $y = T(x)$ with $x \in L^\perp$. We need to check that y is orthogonal to L. Suppose that $z \in L$. Then by the Hermitian property, we have $(z, T(x)) = (T(z), x) = 0$ because $T(z) \in L$ and $x \in L^\perp$. This means that $y = T(x) \in L^\perp$. □

These two lemmas were straightforward and just required following the definitions. The next lemma is slightly more complicated and is the main ingredient of the proof of the Spectral Theorem.

Lemma 18.3. *Suppose that $T : H \to H$ is nonzero, compact, and Hermitian. Then there exists a vector e_1 with $\|e_1\| = 1$ and $T(e_1) = \lambda_1 e_1$ where either $\lambda_1 = \|T\|$ or $\lambda_1 = -\|T\|$. Therefore, either $\|T\|$ or $-\|T\|$ is an eigenvalue of T.*

Proof. Recall from Proposition 15.7 that for Hermitian operators, $\|T\| = \sup_{\|x\|=1} |(Tx, x)|$. Then there exists a sequence $\{x_k\}_{k=1,2,\dots}$ with $\|x_k\| = 1$ for all k and $(T(x_k), x_k) \to \lambda_1$, where either $\lambda_1 = \|T\|$ or $\lambda_1 = -\|T\|$. We need to show that x_k converges, because then the limit could be taken as our e_1.

Note that (using the Hermitian property),

$$\|T(x_k) - \lambda_1 x_k\|^2 = \|T(x_k)\|^2 + \lambda_1^2 - 2\lambda_1(T(x_k), x_k) \leq 2\lambda_1^2 - 2\lambda_1(T(x_k), x_k) \to 0$$

as $k \to \infty$. This means that $T(x_k) - \lambda_1 x_k \to 0$ in H. Since T is compact, $T(x_k)$ must have a convergent subsequence. Then pick $\{x_{k_j}\}_{j=1,2,\dots}$ such that $y = \lim T(x_{k_j})$ exists. Then $T(x_{k_j}) - \lambda_1 x_{k_j} \to 0$ in H, which means that $\lambda_1 x_{k_j} \to y$ and hence $x_{k_j} \to \frac{y}{\lambda_1}$. Therefore $T(x_{k_j}) \to T(\frac{y}{\lambda_1})$. This means that $y = T(\frac{y}{\lambda_1})$, or $Ty = \lambda_1 y$. We can now pick $e_1 = y/\lambda_1$, and this concludes the proof. □

This is a powerful result, and this is the main result and the inductive step of the proof of the Spectral Theorem.

We are now ready to prove the Spectral Theorem.

Proof of the Spectral Theorem 17.4. First, Lemma 18.3 implies that we have $T(e_1) = \lambda_1 e_1$ with $\lambda_1 \neq 0$ and $\|e_1\| = 1$. Then either $T|_{(\mathrm{span}\{e_1\})^\perp} = 0$ or we can repeat this step with $(\mathrm{span}\{e_1\})^\perp$ in place of H and $T|_{(\mathrm{span}\{e_1\})^\perp}$ in place of T. This works because Lemma 18.2 implies that $T((\mathrm{span}\{e_1\})^\perp) \subset (\mathrm{span}\{e_1\})^\perp$

In the first case, we have $(\operatorname{span}\{e_1\})^\perp \subset \ker T$. Also, Lemma 18.1 implies that $\operatorname{span}\{e_1\} \subset (\ker T)^\perp$, which implies that $\ker T \subset (\operatorname{span}\{e_1\})^\perp$. This means that $(\operatorname{span}\{e_1\})^\perp = \ker T$, so that $\operatorname{span}\{e_1\} = (\ker T)^\perp$. In this case, we are done with the proof. This is the one-dimensional case of the Spectral Theorem.

Now, we consider the inductive step. This is more or less the same as the preceding argument. Assume that $k \geq 2$ and we have orthonormal $e_1, \dots, e_{k-1}$ with $\operatorname{span}\{e_1, \dots, e_{k-1}\}^\perp$ and with $T(e_j) = \lambda_j e_j$ for $\lambda_j \neq 0$. Then by Lemma 18.3, either $T|_{\operatorname{span}\{e_1,\dots,e_{k-1}\}^\perp} = 0$ or there exists $e_k \in (\operatorname{span}\{e_1, \dots, e_{k-1}\})^\perp$ with $T(e_k) = \lambda_k e_k$ for $\lambda_k \neq 0$. By Lemma 18.2, we then have $T(\operatorname{span}\{e_1, \dots, e_{k-1}\}^\perp) \subset \operatorname{span}\{e_1, \dots, e_{k-1}\}^\perp$.

In the first case, we have $(\operatorname{span}\{e_1, \dots, e_k\})^\perp \subset \ker T$, and by Lemma 18.1, we have $\operatorname{span}\{e_1, \dots, e_n\} \subset (\ker T)^\perp$. Therefore, just as before, $\ker T = (\operatorname{span}\{e_1, \dots, e_{k-1}\})^\perp$ and hence $(\ker T)^\perp = \operatorname{span}\{e_1, \dots, e_{k-1}\}$. In this case, we terminate and are done. This is the finite dimensional case.

Therefore, we've proved that either the process terminates and we are in the finite dimensional case of the theorem (i.e. $(\ker T)^\perp$ is a finite dimensional subspace) or we get an orthonormal sequence $e_1, e_2, \dots$ with $T(e_j) = \lambda_j e_j$ with $\lambda_j \neq 0$ for all j and $e_j \in (\ker T)^\perp$.

We claim that in this case, $e_1, e_2, \dots$ is *complete* in $(\ker T)^\perp$. This is tricky to check.

Suppose that $n \geq 1$. Then $x \in H$ implies that $x = y_n + x_n$ where $y_n \in (\operatorname{span}\{e_1, \dots, e_n\})^\perp$ and $x_n \in \operatorname{span}\{e_1, \dots, e_n\}$. Note that in this case, Pythagoras's Theorem holds, so that $\|x\|^2 = \|y_n\|^2 + \|x_n\|^2 \geq \|y_n\|^2$. In addition, we know that $T(x) = T(y_n) + T(x_n)$.

In Homework 6, question 4, we showed that $\lambda_j \to 0$. In addition, we have

$$\|T(y_n)\| \leq \left\| T|_{\operatorname{span}\{e_1,\dots,e_n\}^\perp} \right\| \|y_n\|$$

where $|\lambda_{n+1}| = \left\| T|_{\operatorname{span}\{e_1,\dots,e_n\}^\perp} \right\|$. This means that $\|T(y_n)\| \leq |\lambda_{n+1}| \|x\| \to 0$, and hence $T(x_n) \to T(x)$.

Recall that we have $x_n = \sum_{j=1}^n (x, e_j) e_j$, so we have

$$\sum_{j=1}^n (x, e_j) \lambda_j e_j \to T(x),$$

and hence this series converges with respect to the inner product norm. Therefore $T(x) = \sum_{j=1}^\infty \lambda_j (x, e_j) e_j$.

It remains to check completeness. We know from Lemma 18.1 that $\operatorname{span}\{e_1, e_2, \dots\} \subset (\ker T)^\perp$. Also, $x \in (\operatorname{span}\{e_1, e_2, \dots\})^\perp$, which means that $T(x) = 0$ and hence that $x \in \ker T$. Therefore, $(\operatorname{span}\{e_1, e_2, \dots\})^\perp \subset \ker T$. Note that $\operatorname{span}\{e_1, e_2, \dots\}$ is not closed, so hence $\operatorname{span}\{e_1, e_2, \dots\}^{\perp\perp} = \overline{\operatorname{span}\{e_1, e_2, \dots\}}$. That's the end of the proof. □

19. 5/11

First, we need to finish the last few lines of the proof of the Spectral Theorem 17.4.

End of proof of Spectral Theorem 17.4. We have already showed that either there exists some $e_1, \dots, e_n$ orthonormal with $\operatorname{span}\{e_1, \dots, e_n\} = (\ker T)^\perp$ and $Te_j = \lambda_j e_j$ for all $j = 1, 2, \dots, n$ or there exists an orthonormal sequence $e_1, e_2, \dots$ with $\operatorname{span}\{e_1, e_2, \dots\} \subset (\ker T)^\perp$ and $T(e_j) = \lambda_j e_j$ for all j and $T(x) = \sum_{j=1}^\infty \lambda_j (x, e_j) e_j$ for all x. Our inclusion shows that

$\ker T \subset (\overline{\operatorname{span}\{e_1, e_2, \dots\}})^\perp$. Also, $x \in (\overline{\operatorname{span}\{e_1, e_2, \dots\}})^\perp$, which implies that $T(x) = 0$ and hence $x \in \ker T$. This means that $(\overline{\operatorname{span}\{e_1, e_2, \dots\}})^\perp \subset \ker T$.

Therefore, $\ker T = (\overline{\operatorname{span}\{e_1, e_2, \dots\}})^\perp$ and hence $(\ker T)^\perp = \overline{\operatorname{span}\{e_1, e_2, \dots\}}$. This means that $\operatorname{span}\{e_1, e_2, \dots\}$ is dense in $(\ker T)^\perp$. That is, $e_1, e_2, \dots$ is a *complete* orthonormal sequence in $(\ker T)^\perp$. This concludes the end of the proof of the Spectral Theorem. $\square$

19.1. **An application of the Spectral Theorem.** We will spend the next few lectures on an application of the spectral theorem. We will give an application to ordinary differential equations. There are equally important applications to partial differential equations, but those require more machinery than what we have available.

Consider an interval $[a, b]$ and two real valued function $p : [a, b] \to \mathbb{R}$ is C^1 and $q : [a, b] \to \mathbb{R}$ is C^0. Assume that $p > 0$ on $[a, b]$.

For $u \in C^2([a, b])$, consider the differential operator $Lu = (pu')' + qu = pu'' + p'u' + qu$.

There is an existence and uniqueness theorem for ordinary differential equations, which we will not prove in this class.

Theorem 19.1 (Existence and uniqueness theorem)**.** *Let $g : [a, b] \to \mathbb{R}$ be a given C^0 function, and suppose we are given that $u(t_0) = c_1$ and $u'(t_0) = c_2$ at some given point $t_0 \in [a, b]$. There exists a unique $C^2([a, b])$ function u with $Lu = g$ on $[a, b]$ and $u(t_0) = c_1$ and $u(t_0) = c_2$.*

We will be particularly interested in what is called an "eigenvalue problem", and we'll see that this relates to a certain Hermitian operator.

19.2. **Sturm-Liouville.**

Problem 19.2. We are given the *Sturm-Liouville eigenvalue problem*:

$$\begin{cases} -Lu = \lambda u \text{ on } [a, b] \\ \alpha u(a) + \beta u'(a) = 0 \text{ for given } \alpha, \beta \text{ not both zero} \\ \gamma u(b) + \delta u'(b) = 0 \text{ for given } \gamma, \delta \text{ not both zero.} \end{cases}$$

Example 19.3. We do a very simple example. Consider the interval $[a, b] = [0, \pi]$ with $p \equiv 1$ and $q \equiv 0$ on the interval. Take $(\alpha, \beta) = (1, 0)$ and $(\gamma, \delta) = (1, 0)$. In this case, our problem is

$$\begin{cases} -u'' = \lambda u \text{ on } [a, b] \\ u(0) = u(\pi) = 0. \end{cases}$$

In the case $\lambda > 0$, the general solution is $u = A \cos \sqrt{\lambda} t + B \sin \sqrt{\lambda} t$. Which of these satisfy the boundary conditions? Checking $u(0) = 0$, we see that $A = 0$. Then $u(\pi) = 0$ forces either $B = 0$ or $\sqrt{\lambda} = j$ for $j = 1, 2, \dots$. In particular, if u is nonzero then $\lambda = j^2$ and the corresponding solution is $u_j = B \sin jt$.

In the case $\lambda = 0$, we get $u'' = 0$, which means that $u = At + B$. The boundary conditions force $A = B = 0$, so there are no nonzero solutions.

In the case $\lambda < 0$, we have $u = Ae^{\sqrt{-\lambda} t} + Be^{-\sqrt{-\lambda} t}$, and again, the boundary conditions imply that $A = B = 0$.

The point is that there are some special values of λ that work. In addition, the functions u_j form a complete orthonormal sequence for $L^2_{\mathbb{R}}[0, \pi]$.

We want to show that the same type of behavior occurs for the general Sturm-Liouville problem. We will end up showing that there exist a complete orthonormal sequence of eigenfunctions.

Proposition 19.4. *Suppose that $u, v \in C^2([a, b])$ are not identically zero, and suppose that $Lu = 0$ and $Lv = 0$. Then either $u \equiv cv$ for some constant c or the vectors $\binom{u(t)}{u'(t)}$ and $\binom{v(t)}{v'(t)}$ are linearly independent vectors in $\mathbb{R}^2$ for every $t \in [a, b]$.*

Proof. We can check this. If $\binom{u(t_0)}{u'(t_0)}$ and $\binom{v(t_0)}{v'(t_0)}$ are linearly dependent then either $\binom{u(t_0)}{u'(t_0)} = c_1\binom{v(t_0)}{v'(t_0)}$ or $\binom{v(t_0)}{v'(t_0)} = c_2\binom{u(t_0)}{u'(t_0)}$ (these cases are not the same in the case when one of the vectors is 0).

We can just consider the first case; the second case is the same. We then have $u(t_0) - c_1 v(t_0) = 0$ and $u'(t_0) - c_1 v'(t_0) = 0$, which means that $w(t_0) = w'(t_0) = 0$ for $w = u - c_1 v$. By the uniqueness theorem, this means that $w \equiv 0$, which means that $u \equiv cv$. This proves the proposition. □

Corollary 19.5. *In the case $Lu = 0$ and $Lv = 0$, we have either $u = cv$ or $uv' - vu' \neq 0$ for all $t \in [a, b]$. Actually, $p(uv' - vu')$ is constant.*

Proof. The first statement is simply a restatement of the preceding proposition. Now, we have

$$\frac{d}{dt}(p(uv' - vu')) = \frac{d}{dt}((pv')u - (pu')v) = (pv')'u + pv'u' - p(u')'vv = -qvu + quv = 0,$$

so indeed $p(uv' - vu')$ is constant. □

Here, the quantity $p(uv' - vu')$ is called the *Wronskian*.

Proposition 19.6. *As before, suppose that $u, v \in C^2([a, b])$ with $Lu = 0$ and $Lv = 0$ and u, v not identically zero. Assume that 0 is* not *an eigenvalue of the Sturm-Liouville problem, and we have the boundary conditions*

$$\begin{cases} \alpha u(a) + \beta u'(a) = 0 \\ \gamma v(b) + \delta v'(b) = 0, \end{cases}$$

where each of u and v satisfies the boundary conditions of the Sturm-Liouville problem on one endpoint. Then $u \neq cv$, and hence, $p(uv' - vu') \equiv c \neq 0$ on $[a, b]$.

Proof. Otherwise, we would have that both u and v would satisfy the same Sturm-Liouville problem with $\lambda = 0$, which is a contradiction.

Therefore, we see that $p(uv' - vu') \equiv c \neq 0$ on $[a, b]$. □

20. 5/13

Recall that we are talking about the Sturm-Liouville eigenvalue problem 19.2. We have $Lu = (pu')' + qu$ for $p \in C^1([a, b])$ and $q \in C^0([a, b])$. Assume for the moment that $\lambda = 0$ is not an eigenvalue.

Recall that we showed last time that if $Lu = 0$ and $Lv = 0$, and each of u and v satisfy one of the two boundary conditions, then we have $uv' - vu' \neq 0$. Also $p(uv' - vu') = c$ is a constant called the Wronskian. Note that we can always get such u and v due to the existence and uniqueness theorem. That is, we can solve $Lu = 0$ with $u(a) = c_1$, $u'(a) = c_2$ with $c_1\alpha + c_2\beta = 0$, and we can solve $Lv = 0$ with $v(b) = d_1$, $v'(b) = d_2$ with $d_1\gamma + d_2\delta = 0$.

Problem 20.1. For a given $g \in C^0([a,b])$, we want to solve

$$\begin{cases} Lw = g \\ \alpha w(a) + \beta w'(a) = 0 \\ \gamma w(b) + \delta w'(b) = 0 \end{cases}$$

To do this, we use the *method of variation of parameters*. The goal is to find w in the form $w = \varphi u + \psi v$ with some φ and ψ to be chosen. We want to find φ and ψ.

Note that $w' = \varphi u' + \psi v' + u\varphi' + v\psi'$. Suppose that we stipulate that $\varphi' u + \psi' v = 0$. Then $(pw')' = (\varphi(pu'))' + (\psi(pv'))' = -qu\varphi + pu'\varphi' - qv\psi + p\psi' v' = -qw + p(u'\varphi' + \psi' v')$. Hence $Lw = (pw')' + qw = p(u'\varphi' + \psi' v') = g$ if and only if $u'\varphi' + v'\psi' = \frac{g}{p}$. Therefore we have two equations:

$$\begin{cases} \varphi' u + \psi' v = 0 \\ u'\varphi' + v'\psi' = \frac{g}{p}. \end{cases}$$

This can be written in matrix form as

$$\begin{pmatrix} u & v \\ u' & v' \end{pmatrix} \begin{pmatrix} \varphi' \\ \psi' \end{pmatrix} = \frac{g}{p} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This can be solved by brute force, or by computing the inverse of the 2×2 matrix, yielding solutions

$$\begin{cases} \varphi' = -\frac{g}{c} v \\ \psi' = \frac{g}{c} u. \end{cases}$$

Hence, we have

$$\varphi(t) = \int_t^b \frac{g(s)v(s)}{c}\, ds + A$$
$$\psi(t) = \int_a^t \frac{g(s)u(s)}{c}\, ds + B.$$

Thus, supposing that $A = B = 0$,

$$w(t) = \int_t^b \frac{g(s)v(s)}{c} u(t)\, ds + \int_a^t \frac{g(s)u(s)}{c} v(t)\, ds.$$

We check the boundary conditions: $w(a) = u(a)\int_a^b \frac{g(s)v(s)}{c}\, ds$ and $w'(a) = u'(a)\int_a^b \frac{g(s)v(s)}{c}\, ds$. Then

$$\alpha w(a) + \beta w'(a) = (\alpha u(a) + \beta u'(a)) \int_a^b \frac{g(s)v(s)}{c}\, ds = 0.$$

Similarly, we have $w(b) = v(b)\int_a^b \frac{g(s)u(s)}{c}\, ds$ and $w'(b) = v'(b)\int_a^b \frac{g(s)u(s)}{c}\, ds$. Then again,

$$\gamma w(b) + \delta w'(b) = (\gamma v(b) + \delta v'(b)) \int_a^b \frac{g(s)u(s)}{c}\, ds = 0,$$

so our proposal solution $w(t)$ is actually the solution to this problem. That is,

$$w(t) = \int_a^b \frac{g(s)}{c} \left[H(s-t)u(t)v(s) - H(t-s)u(s)v(t)\right] ds = \int_a^b k(s,t)g(s)\, ds$$

where $k(s,t) = \frac{H(s-t)u(t)v(s)+H(t-s)u(s)v(t)}{c}$. This is a symmetric function and $H(s)$ is the Heaviside function

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

Remark. We have

$$k(s,t) = \begin{cases} \frac{u(t)v(s)}{c} & t \leq s \\ \frac{u(s)v(t)}{c} & t > s. \end{cases}$$

is continuous on the closed square $[a,b] \times [a,b]$.

Define $T : C^0([a,b]) \to C^0([a,b])$ by $T(g)(t) = \int_a^b k(s,t)g(s)\, ds$. Observe that T is linear and this is well-defined for all $g \in L^2([a,b])$. Also, we claim that $T(g)$ is $C^0([a,b])$ even when $g \in L^2([a,b])$. We can check this last statement. Let $\varepsilon > 0$ be given. By continuity of k, there exists $\delta > 0$ such that $|k(s,t_1) - k(s,t_2)| < \varepsilon$ whenever $|t_1 - t_2| < \delta$ for $t_1, t_2 \in [a,b]$ and $s \in [a,b]$. Therefore,

$$\begin{aligned} |T(g)(t_1) - T(g)(t_2)| &= \left| \int_a^b (k(s,t_1) - k(s,t_2))g(s)\, ds \right| \leq \int_a^b |k(s,t_1) - k(s,t_2)||g(s)|\, ds \\ &\leq \varepsilon \int_a^b 1 \cdot |g(s)|\, ds \leq \varepsilon\sqrt{b-a}\, \|g\|_{L^2([a,b])} = C\varepsilon \end{aligned}$$

for some fixed constant C whenever $t_1, t_2 \in [a,b]$ with $|t_1 - t_2| < \delta$. This means that $T : L^2([a,b]) \to C^0([a,b])$.

Now, we have (by Fubini's Theorem):

$$\begin{aligned} (h, T(g)) &= \int_a^b T(g)(t)h(t)\, dt = \int_a^b h(t) \int_a^b k(s,t)g(s)\, ds\, dt \\ &= \int_a^b \int_a^b k(s,t)h(t)\, dt\, g(s)\, ds = (T(h), g). \end{aligned}$$

Therefore, $(T(h), g) = (T(g), h)$ for all $h, g \in C^0([a,b])$. Next time, we'll show that this is also true for $h, g \in L^2$.

21. 5/16

We continue the discussion of the Sturm-Liouville eigenvalue problem 19.2.

We are assuming for the moment that zero is not an eigenvalue. We used this to get particular solutions u, v with $Lu = 0$, $Lv = 0$ each satisfying half of the boundary conditions: $\alpha u(a) + \beta u'(a) = 0$ and $\gamma v(b) + \delta v'(b) = 0$, with $u \neq cv$ and $v \neq cu$. This means that $p(uv' - vu') = k \neq 0$. We'll remove this assumption later.

Recall that we can now solve the inhomogeneous problem 20.1. Our solution was

$$w(t) = \int_a^b k(s,t)g(s)\, ds,$$

where $k(s,t) = k(t,s)$ and k is continuous on the square $[a,b] \times [a,b]$. In fact, we have

$$k(s,t) = \begin{cases} \frac{v(s)u(t)}{c} & s \geq t \\ \frac{u(s)v(t)}{c} & s \leq t. \end{cases}$$

We also showed that this is $\varphi(t)u(t) + \psi(t)v(t)$, where

$$\varphi(t) = \frac{1}{k}\int_t^b v(s)g(s)\,ds \text{ and } \psi(t) = \frac{1}{k}\int_a^t u(s)g(s)\,ds.$$

This allowed us to define an operator $T : C^0([a,b]) \to C^2([a,b])$ and $T : L^2[a,b] \to C^0([a,b])$. Acting with $g,h \in C^0([a,b])$, we saw that T is self-adjoint, giving $(T(g),h)_{L^2[a,b]} = (g,T(h))_{L^2[a,b]}$. We claim that this is also true for $g,h \in L^2[a,b]$.

Proposition 21.1. *T is self-adjoint as an operator from L^2 to L^2.*

Proof. Recall that the continuous functions are dense in L^2, so for any $g,h \in L^2[a,b]$, there exists sequences $g_k, h_k \in C^0([a,b])$ of continuous functions with $\|g_k - g\|_{L^2} \to 0$ and $\|h_k - h\|_{L^2} \to 0$.

According to the self-adjointness property that we showed for continuous functions, we have $(T(g_k),h_k)_{L^2[a,b]} = (g_k,T(h_k))_{L^2[a,b]}$ for all k. Recall from last time that we also have $T(g)(t) = \int_a^b k(s,t)g(s)\,ds = (k(s,t),g(s))_{L^2[a,b]}$. Hence,

$$|T(g)(t)| \leq \int_a^b |k(s,t)||g(s)|\,ds \leq \sqrt{\int_a^b k(s,t)^2\,ds}\,\|g\|_{L^2[a,b]},$$

so therefore $\max_{[a,b]} |T(g)| \leq \max|k|\sqrt{b-a}\,\|g\|_{L^2[a,b]}$. Note that for any $f \in L^2[a,b]$, we have $\|f\|_{L^2} = \sqrt{\int_a^b f(s)^2\,ds} \leq \max|f|\sqrt{b-a}$. Therefore, we've shown that $\|T(g)\|_{L^2[a,b]} \leq \max|k|(b-a)\,\|g\|_{L^2[a,b]}$.

We now claim that as $k \to \infty$, we have $(T(g_k),h_k)_{L^2[a,b]} \to (T(g),h)_{L^2[a,b]}$ as $k \to \infty$. We check this:

$$(T(g_k),h_k)_{L^2[a,b]} = (T(g_k) - T(g),h_k) + (T(g),h) + (T(g),h_k - h)$$

Here, the first term is

$$\begin{aligned}(T(g_k) - T(g),h_k) &\leq \|T(g_k) - T(g)\|\,\|h_k\| = \|T(g_k - g)\|\,\|h_k\| \\ &\leq \max|k|(b-a)\,\|g_k - g\|\,\|h_k\| \to 0\end{aligned}$$

and the last term is $(T(g),h_k - h) \leq \|T(g)\|\,\|h_k - h\| \to 0$. Therefore, we see that indeed $(T(g_k),h_k) \to (T(g),h)$ and similarly $(g_k,T(h_k)) \to (g,T(h))$.

Therefore, we see that $T : L^2([a,b]) \to C^0([a,b]) \subset L^2([a,b])$ with $(T(g),h)_{L^2[a,b]} = (g,T(h))_{L^2[a,b]}$, so hence T is self-adjoint as an operator $L^2 \to L^2$. $\square$

In addition, we also have the following result:

Proposition 21.2. *$T : L^2[a,b] \to L^2[a,b]$ is compact. In fact, it is Hilbert-Schmidt, i.e. there exists a complete orthonormal sequence $f_1, f_2, \ldots$ for $L^2[a,b]$ with $\sum_{j=1}^\infty \|T(f_j)\|^2_{L^2[a,b]} < \infty$.*

Proof. Recall that for $L^2([0,2\pi])$, with the inner product $(f,g) = \frac{1}{\pi}\int_0^{2\pi} f(x)g(x)\,dx$, we have already found an orthonormal sequence $\{\frac{1}{\sqrt{2}}, \cos x, \sin x, \cos 2x, \sin 2x, \dots\}$. To do this, we just need to rescale them. Define a new variable t, where $x = \frac{2\pi(t-a)}{b-a}$ and $dx = \frac{2\pi}{b-a}\,dt$. This yields the new inner product $(f,g) = \frac{2}{b-a}\int_a^b \tilde{f}(t)\tilde{g}(t)\,dt$. Our new orthonormal sequence is then

$$\frac{1}{\sqrt{2}}, \cos\frac{2\pi(t-a)n}{b-a}, \sin\frac{2\pi(t-a)n}{b-a}, n = 1, 2, \dots.$$

Note that this is a complete orthonormal sequence, and all functions are $C^0[a,b]$.

We now have that

$$\frac{2}{b-a}T(f_j)(t) = \frac{2}{b-a}\int_a^b k(s,t)f_j(s)\,ds = (k(s,t), f_j(s))_{L^2[a,b]}.$$

Bessel's identity then tells us that

$$\left(\frac{2}{b-a}\right)^2 \sum_{j=1}^{\infty}(T(f_j)(t))^2 = \frac{2}{b-a}\int_a^b k(s,t)^2\,ds$$

for all $t \in [a,b]$. This means that

$$\frac{2}{b-a}\sum_{j=1}^{N}(T(f_j)(t))^2 \leq \int_a^b k(s,t)^2\,ds,$$

which implies that

$$\sum_{j=1}^{N}\|T(f_j)\|_{L^2[a,b]} \leq \int_a^b\int_a^b k(s,t)^2\,ds\,dt,$$

which is a fixed constant. This shows that $\sum_{j=1}^{\infty}\|T(f_j)\|^2_{L^2[a,b]} < \infty$, so T is Hilbert-Schmidt. We already showed that T is bounded (in fact, $\|T(g)\|_{L^2[a,b]} \leq \max|k|\sqrt{b-a}\,\|g\|_{L^2[a,b]}$), and hence T is compact. (Note that Hilbert-Schmidt does not imply compact; boundedness is necessary!) $\square$

Therefore, the Spectral Theorem 17.4 applies, and hence there exists a complete orthonormal sequence $h_1, h_2, \dots$ for $(\ker T)^\perp$ with $T(h_j) = \lambda_j h_j$, $\lambda_j \neq 0$, and $\lambda_j \to 0$ as $j \to \infty$.

We claim that $\ker T = 0$ (this will be checked next time). Assuming this for now, we see that $h_1, h_2, \dots$ is a complete orthonormal sequence for $L^2[a,b]$ and $T(h_j) = \lambda_j h_j$ with $\lambda_j \neq 0$ for all j. How does this relate to the original Sturm-Liouville problem 20.1?

Recall that $T : L^2[a,b] \to C^0[a,b] \subset L^2[a,b]$ maps into the continuous functions, so therefore each $h_j \in C^0[a,b]$ is continuous. However, we also know that $T : C^0[a,b] \to C^2[a,b]$, which means that in fact each $h_j \in C^2[a,b]$. We also assumed that this operator solves the Sturm-Liouville problem, so that $L(T(g)) = g$, so hence $\lambda_j L(h_j) = h_j$, so we've transformed the problem 20.1 into

$$\begin{cases} L(h_j) = \frac{1}{\lambda_j}h_j \\ \alpha h_j(a) + \beta h_j'(a) = 0 \\ \gamma h_j(b) + \delta h_j'(b) = 0. \end{cases}$$

In summary, we have found a complete orthonormal sequence of functions for $L^2[a,b]$, each of which is in $C^2[a,b]$, and that solve the Sturm-Liouville problem with our boundary conditions.

This completes the solution of the Sturm-Liouville eigenvalue problem, except that we still need to check that $\ker T = 0$.

This was a highly nontrivial result, so it wasn't surprising that we had to work so hard for it.

22. 5/18

Recall from last time that we still need to check that $\ker T = \{0\}$.

Proposition 22.1. $\ker T = \{0\}$.

Proof. Recall that for $g \in L^2[a,b]$, we have

$$T(g)(t) = \int_a^b k(s,t)g(s)\,ds = u(t)\varphi_g(t) + v(t)\psi_g(t),$$

where $\beta = p(uv' - vu') \neq 0$ and

$$\varphi_g(t) = \frac{1}{\beta}\int_t^b v(s)g(s)\,ds \text{ and } \psi_g(t) = \frac{1}{\beta}\int_a^t u(s)g(s)\,ds.$$

Suppose that $g \in L^2[a,b]$. Recall that there exists a sequence $\{g_k\} \subset C^0([a,b])$ with $\|g_k - g\|_{L^2} \to 0$. Then

$$T(g_k) = u(t)\varphi_{g_k}(t) + v(t)\psi_{g_k}(t).$$

Note that

$$\begin{aligned}|\varphi_{g_k}(t) - \varphi_g(t)| &= \frac{1}{|\beta|}\left|\int_t^b v(s)(g_k(s) - g(s))\,ds\right| \leq \frac{1}{|\beta|}\int_a^b |v(s)||g_k(s) - g(s)|\,ds \\ &\leq \frac{\max_{[a,b]}|v|}{|\beta|}\int_a^b 1\cdot|g_k(s) - g(s)|\,ds \leq \frac{\max_{[a,b]}|v|}{|\beta|}\sqrt{b-a}\,\|g_k - g\|_{L^2[a,b]} \to 0.\end{aligned}$$

The same argument is true for $\psi_{g_k}(t)$, so therefore $T(g_k)(t) \to u(t)\varphi_g(t) + v(t)\psi_g(t)$ uniformly. This means that $T(g_k)(t) \in C^0[a,b]$.

Last time, we also checked that $T(g_k)(t) \to T(g)(t)$. Thus, $T(g)(t) = u(t)\varphi_g(t) + v(t)\psi_g(t)$. We knew this to be true for $g \in C^0([a,b])$, and now we've checked it for $g \in L^2[a,b]$.

Also, by construction, $(T(g_k))'(t) = u'(t)\varphi_{g_k}(t) + v'(t)\psi_{g_k}(t)$; we chose φ and ψ to make this happen, and in fact $u(t)\varphi'_{g_k}(t) + v(t)\psi'_{g_k}(t) \equiv 0$. We use the same uniform convergence as before to see that $(T(g_k))'(t) \to u'(t)\varphi_g(t) + v'(t)\psi_g(t)$ uniformly.

We integrate this identity over $[a,t]$ to get that

$$T(g_k)(t) = T(g_k)(a) + \int_a^t (u'(s)\varphi_{g_k}(s) + v'(s)\psi_{g_k}(s))\,ds.$$

By uniform convergence, we can take limits under the integral, and we see that

$$T(g)(t) = T(g)(a) + \int_a^t (u'(s)\varphi_g(s) + v'(s)\psi_g(s))\,ds.$$

Here, as we checked earlier, the integrand is a continuous function. This means that $T(g)(t)$ is $C^1[a,b]$ and $T(g)'(t) = u'(t)\varphi_g(t) + v'(t)\psi_g(t)$.

To summarize, we have that

$$T(g)(t) = u(t)\varphi_g(t) + v(t)\psi_g(t)$$
$$T(g)'(t) = u'(t)\varphi_g(t) + v'(t)\psi_g(t).$$

Therefore, if $T(g) = 0$ then

$$u(t)\varphi_g(t) + v(t)\psi_g(t) = 0$$
$$u'(t)\varphi_g(t) + v'(t)\psi_g(t) = 0.$$

This yields

$$\begin{pmatrix} u(t) & v(t) \\ u'(t) & v'(t) \end{pmatrix} \begin{pmatrix} \varphi_g(t) \\ \psi_g(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The determinant of the matrix is the Wronskian, which we know is nonzero. Hence, $\varphi_g(t) \equiv 0$ and $\psi_g(t) \equiv 0$. Hence, we've derived that

$$\int_t^b v(s)g(s)\,ds \equiv 0 \text{ and } \int_a^t u(s)g(s)\,ds \equiv 0$$

for all t. We claim that $g(s) = 0$. If g were continuous, we could differentiate and use the fundamental theorem of calculus; however, $g \in L^2[a,b]$ and we have to be a bit more careful. This is the content of the following lemma:

Lemma 22.2. *If $f \in L^2[a,b]$ with $\int_a^t f(s)\,ds = 0$ for all t then $f(x) = 0$ almost everywhere.*

Proof. There exists a sequence of step functions φ_k with $\|\varphi_k - f\|_{L^2} \to 0$. Then for each φ_k, there is an associated partition $x_1^{(k)}, x_2^{(k)}, \ldots, x_{N_k}^{(k)}$. Then for each k, we have that

$$\varphi_k = \sum_{j=1}^{N_k} a_j^{(k)} \chi_{[x_{j-1}^{(k)}, x_j^{(k)}]}.$$

Then

$$\|\varphi_k - f\|_{L^2}^2 = (\varphi_k - f, \varphi_k - f)_{L^2[a,b]} = \|\varphi_k\|^2 + \|f\|^2 - 2(\varphi_k, f).$$

Now,

$$(\varphi_k, f) = \int_a^b \varphi_k(s)f(s)\,ds = \sum_{j=1}^{N_k} a_j^{(k)} \int_{x^{(k)}}^{x_j^{(k)}} f(s)\,ds = 0$$

because

$$\int_{x_{j-1}^{(k)}}^{x_j^{(k)}} f(s)\,ds = \int_a^{x_j^{(k)}} f(s)\,ds - \int_a^{x_{j-1}^{(k)}} f(s)\,ds = 0.$$

Thus, we see that $\|f_k - f\|_{L^2}^2 \to 0$ and hence $f = 0$ almost everywhere. □

Thus, we've seen that $v(s)g(s) = 0$ and $v(s)g(s) = 0$ almost everywhere. We never have $u(s) = v(s) = 0$ because the Wronskian is nonzero, so hence $g(s) = 0$ almost everywhere. This shows that $\ker T = \{0\}$. □

Therefore, if 0 is not an eigenvalue, we know basically everything about the solutions of the Sturm-Liouville problem 19.2. That is, there exists a complete orthonormal sequence $h_1, h_2, \ldots$ for $L^2[a,b]$ with $h_i \in C^2[a,b]$ and $T(h_j) = \lambda_j h_j$ where $\lambda_j \neq 0$ and $\lambda_j \to 0$. Here, $-Lh_j = \frac{1}{\lambda_j} h_j$ and the h_j satisfy the boundary conditions.

So far, we've assumed that 0 is not an eigenvalue of the Sturm-Liouville eigenvalue problem 19.2:

$$\begin{cases} -Lu = \lambda u \text{ on [a, b]} \\ \alpha u(a) + \beta u'(a) = 0 \text{ for given } \alpha, \beta \text{ not both zero} \\ \gamma u(b) + \delta u'(b) = 0 \text{ for given } \gamma, \delta \text{ not both zero.} \end{cases}$$

Currently, if 0 is an eigenvalue, we know nothing. We want to remove this assumption, and this takes some work.

Proposition 22.3. *There exists $\mu_0 > 0$ such that no $\lambda \leq -\mu_0$ is an eigenvalue. (Hence, $\frac{1}{\lambda_j} \to +\infty$.)*

Proof. Suppose that $-Lu = \lambda u$ for $u \neq 0$, and suppose that we have the boundary condition $\alpha u(a) + \beta u'(a) = 0$. If $\beta = 0$, we have $u(a) = 0$; otherwise, $\beta \neq 0$ and $u'(a) = -\frac{\alpha}{\beta} u(a)$.

Consider $u^2(t)$. This is a continuous function on a closed interval, so it attains its minimum. Let $y \in [a,b]$ be the point where $u^2(t)$ attains a minimum. Then

$$\int_a^y (u^2)' = u^2(y) - u^2(a),$$

which implies that

$$u^2(a) = u^2(y) - 2\int_a^y u(s)u'(s)\,ds \leq \frac{1}{b-a}\int_a^b u^2(s)\,ds - 2\int_a^y u(s)u'(s)\,ds.$$

Now, integration by parts yields

$$\begin{aligned} \lambda \|u\|^2 = (-Lu, u)_{L^2} = \int_a^b -((pu')'u + qu^2) &= -pu'u|_a^b + \int_a^b (p(u')^2 - qu^2) \\ &= p(a)u'(a)u(a) - p(b)u'(b)u(b) + \int_a^b (p(u')^2 - qu^2) \\ &= cu(a)^2 - du(b)^2 + \int_a^b (p(u')^2 - qu^2) \end{aligned}$$

for constants c and d. Thus,

$$\begin{aligned} &-\lambda \|u\|^2 = -cu^2(a) - du^2(b) - \int_a^b (p(u')^2 - qu^2) \\ &\leq |c|\left(\frac{1}{b-a}\int_a^b u^2 - 2\int_a^y uu'\right) + |d|\left(\frac{1}{b-a}\int_a^b u^2 - 2\int_y^b |u||u'|\right) - \int_a^b (p(u')^2 - qu^2). \end{aligned}$$

We are now almost done, and we'll finish this next time. $\square$

23. 5/20

We're trying to prove that there is a fixed lower bound on all of the eigenvalues.

Finishing the proof of 22.3. Recall that we showed that if $-Lu = \lambda u$ (and $u \neq 0$) then

$$-\lambda \|u\|^2 < -\int_a^b (p(u')^2 - qu^2) + \frac{|c|+|d|}{b-a}\int_a^b u^2 + 2(|c|+|d|)\int_a^b |u||u'|.$$

Let $\delta = \min_{[a,b]} p$; we know that $p > 0$. Then this is:

$$\leq -\delta \int_a^b (u')^2 + \left(\max|q| + \frac{|c|+|d|}{b-a}\right)\int_a^b u^2 + 2(|c|+|d|)\int_a^b |u||u'|.$$

We will now use a tricky little inequality $|ab| \leq \frac{1}{2}(a^2+b^2)$; this is equivalent to $(|a|-|b|)^2 \geq 0$. For any $\varepsilon > 0$, we can write this as

$$|ab| = \left|(\sqrt{2\varepsilon}a)\left(\frac{b}{\sqrt{2\varepsilon}}\right)\right| \leq \varepsilon a^2 + \frac{1}{4\varepsilon}b^2.$$

Plugging in $|u| = b$ and $|u'| = a$, and setting $C = \max|q| + \frac{|c|+|d|}{b-a}$, we have

$$\leq -\delta \int_a^b (u')^2 + C\int_a^b u^2 + \varepsilon\int_a^b (u')^2 + \frac{(|c|+|d|)^2}{\varepsilon}\int_a^b u^2.$$

Summarizing, this yields:

$$-\lambda \int_a^b u^2 \leq -\delta\int_a^b (u')^2 + \varepsilon\int_a^b (u')^2 + \left(C + \frac{(|c|+|d|)^2}{\varepsilon}\right)\int_a^b u^2$$

for every $\varepsilon > 0$. Choose $\varepsilon = \delta$ to get

$$-\lambda\int_a^b u^2 \leq \left(\max|q| + \frac{|c|+|d|}{b-a} + \frac{(|c|+|d|)^2}{\delta}\right)\int_a^b u^2.$$

Let $\gamma = \max|q| + \frac{|c|+|d|}{b-a} + \frac{(|c|+|d|)^2}{\delta}$ be the constant in this expression. That is, if μ_0 is any constant with $\mu_0 > \gamma$, then any eigenvalue λ must satisfy $-\lambda < \mu_0$ and $\lambda > -\mu_0$. $\square$

Recall that if $\mu_1, \mu_2, \mu_3, \ldots$ are the eigenvalues then we showed that $|\mu_j| \to \infty$, and now that we've shown that they are all bounded below, we know that hence $\mu_j \to \infty$ and there are only finitely many negative eigenvalues μ_j.

Furthermore, we can now get rid of that annoying assumption that 0 is not an eigenvalue. Let $L_0 u = (pu')' + (q - \mu_0)u = Lu - \mu_0 u$. This means that $-L_0 u = \lambda u$ if and only if $-Lu = (\lambda - \mu_0)u$, and the boundary conditions for u are the same in each case of L and L_0.

Thus, 0 is *not* an eigenvalue of L_0. All of our preceding results therefore apply. This means that there exists a complete orthonormal sequence $h_1, h_2, \ldots$ for all of $L^2[a,b]$ with corresponding eigenvalues $\mu_1^0, \mu_2^0, \ldots$. Then $-L_0 h_j = \mu_j^0 h_j$, and hence $-Lh_j = \mu_j h_j$ for $\mu_j = \mu_j^0 - \mu_0$. The same result in fact holds; we have a complete orthonormal sequence of eigenvectors, and we know that the eigenvalues satisfy $\mu_j \to \infty$ as $j \to \infty$.

23.1. **Application: Heat Flow.** We now consider an application of this discussion to partial differential equations. This application is heat flow.

Suppose you have a region of $\mathbb{R}^n$. (The physically useful situations are $n = 1, 2, 3$.) Say this region is made of some homogeneous and isotropic material. Isotropic means that the material is not made of crystals, so heat has no preference for flowing in a particular direction. Let $u(x, t)$ be the temperature at position x and time t.

There is a very accurate model for heat flow. Here are the physical assumptions:

(1) The quantity of heat in any ball $B_\rho(y)$ is $\lambda \int_{B_\rho(y)} u(x, t)\, dx$ where λ is some constant dependent on the material.

(2) Heat should flow in the direction $-\nabla_x u(x, t)$, and the rate of flow should be proportional to $|\nabla u|$. The rate of flow of heat across the boundary $\partial B_\rho(y)$ should be $\mu \int_{\partial B_\rho(y)} \eta \cdot \nabla u$ where η is the unit normal of $\partial B_\rho(y)$.

Hence, we should have that

$$\lambda \frac{d}{dt} \int_{B_\rho(y)} u(x, t)\, dx = \mu \int_{\partial B_\rho(y)} \eta \cdot \nabla u.$$

We assume that u is a C^1 function so we put the derivative under the integral sign. This is just a model, and it is close enough to a smooth function so assuming this should not be problematic. Then we can apply the Gauss' Theorem to see that

$$\lambda \int_{B_\rho(y)} \frac{\partial u}{\partial t} = \mu \int_{\partial B_\rho(y)} \eta \cdot \nabla u = \mu \int_{B_\rho(y)} \operatorname{div}(\nabla u) = \mu \int_{B_\rho(y)} \Delta u.$$

Here, we used the fact that $\operatorname{div}(\nabla u) = \Delta u$, where Δ is the Laplacian.

Summarizing, we see that for any ball $B_\rho(y)$ of the material, we see that

$$\int_{B_\rho(y)} \frac{\partial u}{\partial t} = \frac{\mu}{\lambda} \int_{B_\rho(y)} \Delta u.$$

This means that

$$\frac{\int_{B_\rho(y)} \frac{\partial u}{\partial t}}{B_\rho(y)} = \frac{\frac{\mu}{\lambda} \int_{B_\rho(y)} \Delta u}{B_\rho(y)},$$

and letting $\rho \to 0$, we have that

$$\boxed{\frac{\partial u}{\partial t} = \frac{\mu}{\lambda} \Delta u}.$$

23.1.1. *Example:* $n = 1$. Let's now apply this in the case of $n = 1$. This is the case of a metal bar, and there are various boundary conditions. We could fix $u(b) \equiv 0$ to keep the ends have fixed temperature. Alternatively, we could insulate the ends to prevent heat flow, yielding $\frac{\partial u}{\partial x}(b) \equiv 0$.

Problem 23.1. In this case, our problem has boundary conditions and initial conditions.

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, t > 0, a < x < b \\ \alpha u(a, t) + \beta \frac{\partial u}{\partial x}(a, t) = 0, t > 0 \\ \gamma u(b, t) + \delta \frac{\partial u}{\partial x}(b, t) = 0, t > 0 \\ u(x, 0) = \varphi(x), a \le x \le b. \end{cases}$$

Remark. There is no loss of generality to drop the constant $\frac{\mu}{\lambda}$ simply by rescaling via $t \mapsto \frac{\lambda}{\mu} t$.

First, we look for separated variable solutions, i.e. $u(x,t) = f(x)g(t)$. This won't solve the problem completely, but it's a good start.

We have

$$\frac{\partial u}{\partial t} = f(x)g'(t)$$
$$\frac{\partial^2 u}{\partial x^2} = f''(x)g(t),$$

so our PDE becomes $f(x)g'(t) = f''(x)g(t)$. This is equivalent to

$$\frac{g'(t)}{g(t)} = \frac{f''(x)}{f(x)}$$

at points where $g(t) \neq 0$ and $f(x) \neq 0$. How could that be? One side is a function of t and the other is a function of x, so they can be equal only when both are constant. We can therefore get a family of many solutions as follows: Take $\lambda \in \mathbb{R}$, and solve

$$\begin{cases} g'(t) = -\lambda g(t) & t \geq 0 \\ f''(x) = -\lambda f(x) & a \leq x \leq b. \end{cases}$$

We can solve the first equation easily; this is just $g(x) = Ce^{-\lambda t}$ for some constant C. We know how to solve second order equations, so we can also do the second equation. We're interested in taking $u = f(t)g(x)$ while satisfying as many of our conditions as possible, so we'll make u also satisfy the boundary conditions. Observe that the boundary conditions for u translate directly into boundary conditions of f, so we want to solve

$$\begin{cases} -f'' = \lambda f \\ \alpha f(a) + \beta f'(a) = 0 \\ \gamma f(b) + \delta f'(b) = 0. \end{cases}$$

This is a Sturm-Liouville problem that we know how to solve. There exists a complete orthonormal sequence of eigenfunctions $h_1, h_2, \ldots$ with corresponding $\lambda_1, \lambda_2, \ldots$ with $\lambda_j \to \infty$.

That is, we have solutions $c_j h_j(x) e^{-\lambda_j t}$ for $j = 1, 2, \ldots$. These satisfy the partial differential equation and the boundary conditions. Now we have to try to satisfy the initial conditions. The eigenfunctions form a complete orthonormal sequence, so $\varphi \in L^2[a,b]$ implies that we can pick $c_j = (\varphi, h_j)$ to get that

$$\sum_{j=1}^{\infty} c_j h_j(x) e^{-\lambda_j t} = \varphi$$

at $t = 0$ in the L^2 norm.

So it looks like we're done! There's still something left to check; we need to make sure that this series remains smooth and still satisfies the PDE. There's still some checking and a little way to go, but that's the general idea.

Recall that we were consider the heat equation problem 23.1.

We showed that $u(x,t) = \sum_{j=1}^{\infty} c_j e^{-\lambda_j t} h_j(x)$, where λ_j is the j-th eigenvalue of a Sturm-Liouville problem, h_j is the j-th eigenfunction from Sturm-Liouville, and $c_j = (\varphi, h_j)_{L^2[a,b]}$ is a candidate solution.

By construction, each partial sum is a solution of the heat equation satisfying boundary conditions. Consider the initial conditions; at $t = 0$, we have $\sum c_j h_j = \varphi$. There are a number of things to check: does this sum converge? Is it continuous?

To check these properties, we need to use the Weierstrass M-test.

Proposition 24.1 (Weierstrass M-test)**.** *Suppose we have a sequence of given real-valued functions $f_n : X \to \mathbb{R}$, and suppose we have numbers $M_n \geq 0$ such that $\sup_X |f_n| \leq M_n$ and $\sum M_n$ converges. Then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on X.*

Proof. First, the comparison test guarantees convergence: $|f_n(x)| \leq M_n$ and $\sum M_n$ converges implies that $\sum f_n(x)$ is absolutely convergent.

Let $S_N(x) = \sum_{n=1}^{N} f_n(x)$ and $S(x) = \sum_{n=1}^{\infty} f_n(x)$.

Let $\varepsilon > 0$ be given. Pick J such that $\sum_{n=J+1}^{\infty} M_n \leq \varepsilon$. Then take any $N > J$. Then

$$|S_N(x) - S(x)| = \left| \sum_{n=N+1}^{\infty} f_n(x) \right| \leq \sum_{n=N+1}^{\infty} |f_n(x)| \leq \sum_{n=N+1}^{\infty} M_n < \varepsilon$$

for every x. This means that $\sup_{x \in X} |S_N(x) - S(x)| < \varepsilon$ for every $N > J$, which gives us uniform convergence. □

We only have finite time, so let's only do this under the simplest boundary conditions. Consider the special case of the interval $[0, \pi]$ with boundary conditions $u(a,t) = 0$ and $u(b,t) = 0$. In the notation of problem 23.1, this corresponds to $(\alpha, \beta) = (\gamma, \delta) = (1, 0)$. In this case, we know the eigenvalues of the Sturm-Liouville problem. This is $\lambda_j = j^2$ for $j = 1, 2, \ldots$ and $h_j(x) = \sin jx$. Here, we use the L^2 inner product with the appropriate scaling to make h_j have norm one: $(f, g) = \frac{2}{\pi} \int_0^{\pi} f(x) g(x)\, dx$. Then we have

$$u(x,t) = \sum_{j=1}^{\infty} c_j e^{-j^2 t} \sin jx$$

where

$$c_j = \frac{2}{\pi} \int_0^{\pi} \varphi(x) \sin jx \, dx.$$

Proposition 24.2. *This $u(x,t)$ converges uniformly.*

Proof. We will check that this converges using the Weierstrass M-test. To do this, we will also assume that $f \in C^2([0,\pi])$ satisfying the boundary conditions, so that $\varphi(0) = \varphi(\pi) = 0$. Note that here, $u(0,t) = u(\pi,t) = 0$.

We can integrate by parts twice, observing that all boundary terms are zero, to get that

$$\begin{aligned} c_j &= \frac{2}{\pi}\int_0^\pi \varphi(x)\sin jx\,dx = -\frac{2}{j\pi}\varphi(x)\cos jx\bigg|_0^\pi + \frac{2}{j\pi}\int_0^\pi \varphi'(x)\cos jx\,dx \\ &= \frac{2}{j^2\pi}\varphi'(x)\sin jx\bigg|_0^\pi - \frac{2}{j^2\pi}\int_0^\pi \varphi''(x)\sin jx\,dx = -\frac{2}{j^2\pi}\int_0^\pi \varphi''(x)\sin jx\,dx, \end{aligned}$$

and hence

$$|c_j| \le \frac{2\max_{[0,\pi]}|\varphi''|}{j^2}$$

for $j = 1, 2, \dots$. We claim that $\sum_{j=1}^\infty c_j e^{-j^2 t}\sin jx$ converges uniformly on $[0,\pi]\times[0,\infty)$. We can now check this:

$$\left|c_j e^{-j^2 t}\sin jx\right| \le 2\frac{\max_{[0,\pi]}|\varphi''|}{j^2},$$

which is the jth term of a convergent sequence $\sum \frac{1}{j^2}$. This is what we call M_j in the Weierstrass M-test. The Weierstrass M-test therefore shows uniform convergence. ☐

The uniform limit of continuous functions is continuous, which shows the following corollary:

Corollary 24.3. $\sum_{j=1}^\infty c_j e^{-j^2 t}\sin jx$ *is continuous in* $[0,\pi]\times[0,\infty)$. *Hence the boundary conditions and initial conditions are satisfied.*

Now, we just need to show that $u(x,t)$ satisfies the heat equation. To do this, we need to be able to differentiate; this will come from the following general result:

Theorem 24.4 (Differentiation of series)**.** *Suppose we have a sequence of* C^1 *functions* $f_n : [c,d] \to \mathbb{R}$ *for* $n = 1, 2, \dots$, *suppose that* $\sum_{n=1}^\infty f_n(x)$ *is convergent for all* $x \in [c,d]$, *and suppose that* $\sum_{n=1}^\infty f_n'(x)$ *is uniformly convergent on* $[c,d]$. *Then* $f(x) = \sum_{n=1}^\infty f_n(x)$ *is* C^1 *on* $[c,d]$ *and* $f'(x) = \sum_{n=1}^\infty f_n'(x)$.

Remark. We have to be careful, and there are counterexamples where we cannot differentiate inside an infinite series.

To check that $\sum f_n'(x)$ is uniformly convergent, the Weierstrass M-test gives a sufficient condition: $|f_n'(x)| \le M_n$ where $\sum M_n$ is convergent.

Proof. We can differentiate a finite sum without any problems, so that

$$\frac{d}{dx}\left(\sum_{n=1}^N f_n(x)\right) = \sum_{n=1}^N f_n'(x),$$

which implies that

$$\int_c^x \sum_{n=1}^N f_n'(t)\,dt = \sum_{n=1}^N f_n(x) - \sum_{n=1}^N f_n(c).$$

Since $\sum f_n'(x)$ is converging *uniformly*, we can take the limit as $N \to \infty$ under the integral sign. This gives

$$\int_c^x \sum_{n=1}^\infty f_n'(t)\,dt = \sum_{n=1}^\infty (f_n(x) - f_n(c)) = \sum_{n=1}^\infty f_n(x) - \sum_{n=1}^\infty f_n(c).$$

Note that $\sum_{n=1}^{\infty} f_n'(x)$ is a continuous function on $[c,d]$ because it is the uniform limit of continuous functions. Then

$$\int_c^x \sum_{n=1}^{\infty} f_n'(t)\,dt = \sum_{n=1}^{\infty} f_n(x) - \sum_{n=1}^{\infty} f_n(c).$$

By the fundamental theorem of calculus, $\sum_{n=1}^{\infty} f_n(x)$ is $C^1[c,d]$ and $\frac{d}{dx}(\sum_{n=1}^{\infty} f_n(x)) = \sum_{n=1}^{\infty} f_n'(x)$, which is what we wanted. $\square$

We can now use the theorem to see our next claim:

Proposition 24.5. *$u(x,t) = \sum_{j=1}^{\infty} c_j e^{-j^2 t}\sin jx$ is $C^2((0,\pi)\times(0,\infty))$ and its partial derivatives can be computed termwise, and hence the equation $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$ does hold.*

Proof. We will check that we can apply our general differentiation theorem. That is, we need to see that the derivative is uniformly convergent.

Consider any fixed $x \in (0,1)$ and suppose that $t \in [c,d]$ for some $0 < c < d < \infty$. Then the termwise differentiated series with respect to t

$$\sum_{j=1}^{\infty} c_j(-j^2)e^{-j^2 t}\sin jx$$

is clearly uniformly convergent on the interval $[c,d]$ by the Weierstrass M-test. Indeed,

$$|c_j j^2 e^{-j^2 t}\sin jx| \le 2\max_{[0,\pi]}|\varphi''|e^{-j^2 c} \le 2\max_{[0,\pi]}|\varphi''|e^{-jc},$$

which is again the jth term of a convergent series. Therefore the termwise differentiation theorem 24.4 holds and tells us that

$$\frac{\partial u}{\partial t} = -\sum_{j=1}^{\infty} j^2 c_j e^{-j^2 t}\sin jx.$$

Similarly, in an almost identical argument, if we take some $0 < c < d < \pi$ and some fixed $t > 0$, we see that the termwise differentiated series with respect to x

$$\sum_{j=1}^{\infty} c_j j e^{-j^2 t}\cos jx$$

is uniformly convergent by the Weierstrass M-test; indeed, $|c_j j e^{-j^2 t}\cos jt| \le 2\max|\varphi''|e^{-jt}$ is the jth term of a convergent series. Also, differentiating a second time, we see that

$$-\sum_{j=1}^{\infty} j^2 c_j e^{-j^2 t}\sin jx$$

is uniformly convergent on $[c,d]$. This shows that

$$\frac{\partial^2 u}{\partial x^2} = -\sum_{j=1}^{\infty} c_j j^2 e^{-j^2 t}\sin jx = \frac{\partial u}{\partial t}.$$

This concludes the argument, and we can conclude that $u(x,t)$ is indeed a solution to the heat equation 23.1. Note that if we had continued in this way, we could have shown that $u(x,t)$ is actually C^∞ in the interior. $\square$

25. 5/25

Let's briefly revisit the spectrum. Suppose that X is a Banach space and $A \in \mathcal{L}(X, X)$. Recall that the spectrum of A is $\sigma(A) = \{z \in \mathbb{C} : A - zI \text{ is not invertible}\}$. Recall that $\sigma(A)$ is a closed subset of the closed disc $\{z \in \mathbb{C} : |z| \leq \|A\|\}$. We also showed that the set of eigenvalues is trivially contained in $\sigma(A)$, though the set of eigenvalues might be empty (e.g. in the case of the shift operator).

There was one thing that neglected to address: Is the spectrum nonempty?

Theorem 25.1. *The spectrum $\sigma(A)$ is always nonempty.*

We need two preliminaries before we can prove this claim.

Definition 25.2. A function $h(z)$ is *holomorphic* if for every z_0 with $|z_0| \leq R$, we can write $h(z) = \sum_{j=0}^{\infty} a_j (z - z_0)^j$ for sufficiently small $|z - z_0|$.

Theorem 25.3 (Maximum modulus principle)**.** *If $h(z)$ is holomorphic on $\{z : |z| \leq R\}$ then* $\max_{|z|\leq R} |h(z)| = \max_{|z|=R} |h(z)|$.

We will also need another very important result, which is one of the basic results of functional analysis.

Theorem 25.4 (Hahn-Banach theorem)**.** *Let X be any normed linear space. Then there exists a nontrivial bounded linear functional on X.*

Example 25.5. For example, if X is an inner product space, $f(x) = (x, y)$ for some fixed $y \in X \setminus \{0\}$ would be a bounded linear functional. This is less obvious in normed spaces.

We will assume these two preliminaries for now, and defer their proofs to later. First, we will prove that the spectrum is nonempty.

Proof of theorem 25.1. Suppose to the contrary that $\sigma(A) = \emptyset$, i.e. $A - zI$ is invertible for every $z \in \mathbb{C}$. In particular, if $z_0 \in \mathbb{C}$ then we can write

$$A - zI = (A - z_0 I) - (z - z_0)I = (A - z_0 I)(I - (z - z_0)(A - z_0 I)^{-1})$$

Let $B = (A - z_0 I)^{-1}$. Recall that by theorem 14.9, $I - (z - z_0)(A - z_0 I)^{-1}$ is invertible if $|z - z_0| \, \|(A - z_0 I)^{-1}\| < 1$, i.e. if

$$|z - z_0| < \frac{1}{\|(A - z_0 I)^{-1}\|} = \frac{1}{\|B\|},$$

and in this case, we have

$$(A - zI)^{-1} = \left(\sum_{j=1}^{\infty} (z - z_0)^j B^j \right) B = \sum_{j=0}^{\infty} (z - z_0)^j B^{j+1},$$

so

$$(A - zI)^{-1}(y) = \sum_{j=0}^{\infty} (z - z_0)^j B^{j+1}(y)$$

for any $y \in X$ as long as $|z - z_0| < \frac{1}{\|B\|}$. By linearity, we get that for any nontrivial linear functional (which exists by the Hahn-Banach theorem 25.4),

$$f((A - zI)^{-1}(y)) = \sum_{j=0}^{\infty} (z - z_0)^j f(B^{j+1}(y))$$

is in $\mathbb{C}$, so then $h(z) = f((a - zI)^{-1}(y))$ is holomorphic for $|z - z_0| < \frac{1}{\|B\|}$. Since we picked z_0 arbitrarily, we've shown that $h(z)$ is holomorphic on the entire plane $\mathbb{C}$.

Suppose that $|z| > 2\|A\|$. Then we can write $A - zI = -z(I - \frac{1}{z}A)$, where $\left\|\frac{1}{z}A\right\| < \frac{1}{2} < 1$. Applying theorem 14.9 again, we see that

$$(A - zI)^{-1} = -\frac{1}{z}\sum_{j=0}^{\infty}\frac{1}{z^j}A^j.$$

This time, we see that

$$h(z) = f((A - zI)^{-1}y) = f\left(-\frac{1}{z}\sum_{j=0}^{\infty}\frac{1}{z^j}A^j y\right) = -\frac{1}{z}\sum_{j=0}^{\infty} f\left(\frac{1}{z^j}A^j y\right)$$

for $|z| > 2\|A\|$. Therefore,

$$|h(z)| \le \frac{1}{|z|}\sum_{j=0}^{\infty}\left|f\left(\frac{1}{z^j}A^j y\right)\right| \le \frac{1}{|z|}\sum_{j=0}^{\infty}\|f\|\frac{1}{|z|^j}\left\|A^j y\right\| \le \frac{\|f\|}{|z|}\sum_{j=0}^{\infty}\left(\frac{\|A\|}{|z|}\right)^j\|y\| \le \frac{2\|f\|\|y\|}{|z|}.$$

Therefore, we see that $\max_{|z|=R}|h(z)| \to 0$ as $R \to \infty$. By the maximum modulus principle 25.3, this shows that $\max_{|z|=R}|h(z)| \to 0$ as $R \to \infty$, which shows that $h \equiv 0$.

We have now shown that $0 \equiv h(z) = f((A - zI)^{-1}(y))$ for every y. Since the inverse maps the Banach space onto itself, this means that $f \equiv 0$, which is a contradiction because we chose f to be a nontrivial linear functional. $\square$

Now, we need to prove our preliminaries. First, we'll prove the maximum principle for subharmonic functions.

Proposition 25.6. *Let $u \in C^2(B_R) \cap C^0(\overline{B}_R)$, where $B^R = \{x \in \mathbb{R}^n : \|x\| < R\}$, and suppose that $\Delta u \ge 0$ on B_R. This means that the function is twice differentiable on the open ball and is continuous on the closed ball. Then $\max_{\overline{B}_R} u = \max_{\partial B_R} u$.*

Here, $\Delta u = \sum_{j=1}^{n}\frac{\partial^2}{\partial x_j^2}u$ is the Laplacian.

Proof. Let $\varepsilon > 0$, and let $v(x) = u(x) + \varepsilon\|x\|^2$. Here, $\Delta v = \Delta u + 2n\varepsilon > 0$ in B_R.

First, suppose there some $y \in B_R$ such that $v(y) = \max_{B_R} v$. Then

$$\frac{\partial v}{\partial x_j}(y) = 0, \frac{\partial v^2}{\partial x_j^2}(y) \le 0,$$

which implies that $\Delta v(y) \le 0$; this is a contradiction to the statement that $\Delta v > 0$. We defined v in order to have strictly positive Laplacian precisely to make this work. Hence

$$\max_{\overline{B}_R} u \le \max_{\overline{B}_R} v = \max_{\partial B_R} v \le \max_{\partial B_R} u + \varepsilon R^2$$

for all $\varepsilon > 0$, so therefore $\max_{\overline{B}_R} u \le \max_{\partial B_R} u$. The reverse inequality is trivial, so we've proven the maximum principle. $\square$

We are now ready to prove theorem 25.3.

Proof of theorem 25.3. Suppose that $h(z)$ is holomorphic on $\overline{D}_R = \{z \in \mathbb{C} : |z| \le R\}$. We can write $h(z) = h(x+iy) = u(x,y) + iv(x,y)$ where u and v are real-valued functions on $\{(x,y) : \sqrt{x^2+y^2} \le R\}$. Then recall the Cauchy-Riemann equations:

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \end{cases}$$

Differentiating the first with respect to x and the second with respect to y yields

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial v^2}{\partial x \partial y} = \frac{\partial^2 v}{\partial y \partial x}$$
$$\frac{\partial^2 u}{\partial y^2} = -\frac{\partial v^2}{\partial y \partial x},$$

so hence

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

Similarly, $\Delta v = 0$ as well; this means that u and v are harmonic. Now,

$$|h(z)|^2 = u^2 + v^2,$$

so hence

$$\begin{aligned}\Delta |h(z)|^2 &= \frac{\partial}{\partial x}\left(2u\frac{\partial u}{\partial x} + 2v\frac{\partial v}{\partial x}\right) + \frac{\partial}{\partial y}\left(2u\frac{\partial u}{\partial y} + 2v\frac{\partial v}{\partial y}\right) \\ &= 2\left(\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2\right) \ge 0,\end{aligned}$$

and therefore the maximum principle 25.6 holds, and we are done. □

26. 5/27

Today we will prove the Hahn-Banach theorem. To reiterate, here is the statement of the theorem:

Theorem 26.1 (Hahn-Banach theorem)**.** *Let X be any real vector space, and suppose that $p : X \to \mathbb{R}$ such that p is positively homogeneous (i.e. $p(tx) = tp(x)$ for all $t > 0$ and all $x \in X$) and subadditive (i.e. $p(x+y) \le p(x) + p(y)$ for all $x, y \in X$). Let S_0 be any subspace of X and suppose that $f_0 : S_0 \to \mathbb{R}$ is linear with $f_0(x) \le p(x)$ for all $x \in S_0$.*

Then there exists an extension $f : X \to \mathbb{R}$ that is linear with $f|_{S_0} = f_0$ and $f(x) \le p(x)$ for all $x \in X$.

As we discussed last time, we are interested in a corollary:

Corollary 26.2. *Let X be any normed space, and take $p(x) = \|x\|$. Consider some arbitrary $y \in X \setminus \{0\}$ and $S_0 = \operatorname{span}\{y\}$. Then $f_0(ty) = t\|y\|$ is linear on S_0. The Hahn-Banach theorem 26.1 then implies that there exists $f : X \to \mathbb{R}$ with $f|_{S_0} = f_0$ and $f(x) \le \|x\|$ for every $x \in X$. That implies that f is a nontrivial bounded linear functional.*

We are now ready to attack the proof of the Hahn-Banach theorem.

Proof of Hahn-Banach theorem 26.1. Suppose that S_0 is a subspace of X which contains S_0 such that there exists an $f_1 : S_1 \to \mathbb{R}$ which is linear and $f_1|_{S_0} = f_0$ and $f_1(x) \le p(x)$ for every $x \in S_1$.

If $S_1 \neq X$, then we can pick a vector $a \in X \setminus S_1$, and we can define a new subspace $S_2 = \{x + ta : x \in S_1, t \in \mathbb{R}\}$. We can find an extension $f_2 : S_2 \to \mathbb{R}$ defined by $f_2(x + ta) = f_1(x) + t\lambda$ where $\lambda = f_2(a)$ and $x \in S_1$. This is clearly linear, and it is an extension of f_1 because $f_2|_{S_1} = f_1$ because this is the case where $t = 0$.

Here is the main step of the proof:

Claim. There exists a choice of λ such that $f_2(x) \le p(x)$ for all $x \in S_2$.

Proof. In the case $t = +1$, we have $f_2(x + a) = f_1(x) + \lambda$ for all $x \in S_1$, and in the case $t = -1$, we have $f_2(y - a) = f_1(y) - \lambda$ for all $y \in S_1$. Adding, we have $f_1(x) + f_1(y) = f_1(x + y) \le p(x + y)$ for all $x, y \in S_1$.

We use subadditivity in a somewhat tricky way to see that

$$f_1(x) + f_1(y) \le p(x + y) = p((x + a) + (y - a)) \le p(x + a) + p(y - a),$$

which means that

$$f_1(y) - p(y - a) \le p(x + a) - f_1(x)$$

for all $x, y \in S_1$. By taking a fixed x, we see that

$$\sup_{y \in S_1} (f_1(y) - p(y - a)) \le p(x + a) - f_1(x)$$

for all x, which in turn implies that

$$\sup_{y \in S_1} (f_1(y) - p(y - a)) \le \inf_{x \in S_1} (p(x + a) - f_1(x)).$$

That's great, because there's a number that fits in between these two (possibly equal to both). Choose λ to be this number. That is, there exists $\lambda \in \mathbb{R}$ with $f_1(y) - p(y - a) \le \lambda \le p(x + a) - f_1(x)$ for every $x, y \in S_1$. These gives us two inequalities:

$$\begin{aligned} f_1(x) + \lambda &\le p(x + a) \quad \text{for every } x \in S_1 \\ f_1(y) - \lambda &\le p(y - a) \quad \text{for every } y \in S_1. \end{aligned}$$

If $t > 0$, the first inequality gives

$$f_2(x + ta) = f_1(x) + t\lambda = t\left(f_1\left(\frac{x}{t}\right) + \lambda\right) \le t\left(p\left(\frac{x}{t} + a\right)\right) = p(x + ta).$$

Similarly, if $t < 0$, the second inequality gives

$$f_2(x + ta) = f_1(x) + t\lambda = |t|\left(f_1\left(\frac{x}{|t|}\right) - \lambda\right) \le |t|\left(p\left(\frac{x}{|t|} - a\right)\right) = p(x + ta).$$

That was the clever part of the proof. □

It's starting to look like the theorem is true. If we can extend it at all, we've shown that we can extend it a bit more. It's tempting that say that we're done, but that's not true because we are working with infinite dimensional spaces. We need to do something a bit more sophisticated.

Let $\mathcal{S}$ be the set of ordered pairs (S, f_S) such that S is a subspace of X which contains S_0 and $f_S : S \to \mathbb{R}$ is linear with $f_S(x) \le p(x)$ for all $x \in S$, and $f_S|_{S_0} = f_0$. Here, we've collected all extensions of f_0. We can define a *partial ordering* on $\mathcal{S}$.

Definition 26.3. Recall that a partial ordering or any set Q means that $x \preceq x$ for all $x \in Q$, and that $x \preceq y$ and $y \preceq x$ implies that $x = y$, and also that $x \preceq y$ and $y \preceq z$ implies that $x \preceq z$. This differs total ordering because two elements of Q do not have to be related in a partial order.

Q is totally ordered (or a "chain") if $x, y \in Q$ implies that either $x \preceq y$ or $y \preceq x$.

Example 26.4. The real numbers are totally ordered (and hence partially ordered). The inclusion of sets is an example of a partial order.

We can define our partial ordering on $\mathcal{S}$ by $(S, f_S) \preceq (T, f_T)$ means that $S \subset T$ and $f_T|_S = f_S$. Suppose that $\mathcal{T} \subset \mathcal{S}$ is a chain. Then let $\overline{T} = \cup_{(T,f_T)\in\mathcal{T}} T$.

Claim. We claim that this is a subspace.

Normally, this is nonsense; the union of two subspaces is not in general a subspace. However, we do have a total ordering, and we will use this to check the claim.

Proof. Take any $x, y \in \overline{T}$ and $\alpha, \beta \in \mathbb{R}$. Then $x \in T_1, y \in T_2$ for some $(T_1, f_{T_1}), (T_2, f_{T_2}) \in \mathcal{T}$. Since $\mathcal{T}$ is totally ordered, we know that either $T_1 \subset T_2$ or $T_2 \subset T_1$. Assume without loss of generality that $T_1 \subset T_2$. Then $x, y \in T_2$, so therefore $\alpha x + \beta y \in T_2$ as well, and we are done. □

We can now define a function $\overline{f} : \overline{T} \to \mathbb{R}$. For $x \in \overline{T}$, define $\overline{f}(x) = f_T(x)$ for any T with $(T, f_T) \in \mathcal{T}$ and $x \in T$. This is unambiguous due to our total ordering.

Note that $\overline{f} : \overline{T} \to \mathbb{R}$ is linear. To see this, consider $\alpha, \beta \in \mathbb{R}$ and $x, y \in \overline{T}$ as above. Then we know that $x, y \in T_2$ for some T_2, so that $\overline{f}(\alpha x+\beta y) = f_{T_2}(\alpha x+\beta y) = \alpha f_{T_2}(x)+\beta f_{T_2}(y) = \alpha\overline{f}(x) + \beta\overline{f}(x)$. Clearly, we also have $\overline{f}(x) \leq p(x)$. In particular, this shows that $(\overline{T}, \overline{f}) \in \mathcal{S}$. By construction, $(T, f_T) \preceq (\overline{T}, \overline{f})$, i.e. $(\overline{T}, \overline{f})$ is an upper bound for $\mathcal{T}$ relative to this partial order.

These are all of the hypothesis that we need for Zorn's Lemma:

Lemma 26.5 (Zorn's Lemma)**.** *If $\mathcal{S}$ is any partially ordered set such that every chain $\mathcal{T}$ has an upper bound in $\mathcal{S}$, then $\mathcal{S}$ has at least one maximal element.*

That is, there exists $(S, f_S) \in \mathcal{S}$ such that if $(S, f_S) \preceq (T, f_T)$ for some $(T, f_T) \in \mathcal{S}$ then $(S, f_S) = (T, f_T)$.

We now claim that this maximal element is now the whole space: $X = S$. Indeed, if $S \neq X$ then we can find $S_1 \supsetneq S$ and an extension $\overline{f}$ of f_S, which contradicts maximality of (S, f_S). This proves the Hahn-Banach theorem. □

27. 6/1

It is impossible to review everything, so we'll give a brief and sketchy overview of most (but not all) of the main topics.

We had *inner product spaces* and *normed spaces*, and we usually denote inner products as (x, y). We checked that $\sqrt{(x, x)}$ is a norm, called the inner product norm. In particular, this means that inner product spaces are contained in normed spaces. To do this, we needed *Cauchy-Schwarz*: $|(x, y)| \leq \|x\| \, \|y\|$, and we had the *triangle inequality*: $\|x + y\| \leq \|x\| + \|y\|$.

We had some special results about finite-dimensional normed spaces.

(1) We showed that in such spaces, all norms are equivalent. This means that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are any two norms then there exists a constant C such that $C^{-1}\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1$ for every $x \in X$.
(2) Any closed bounded subset is compact. In particular, the closed unit ball is compact.

Of course, these fail miserably in infinite dimensional spaces. In fact, in infinite dimensional space, the closed unit ball is *never* compact. The proof was that we showed that there exists a sequence $e_1, e_2, \ldots$ with $\|e_j\| = 1$ for every j such that $\|e_i - e_j\| \geq 1$ for every $i \neq j$, which violates sequential compactness. This provided a major contrast between finite and infinite dimensional space.

We then talked about *complete* spaces. A complete inner product space is called a *Hilbert space* and a complete normed space is called a *Banach space*.

Example 27.1. For example, $\mathbb{R}^n, \mathbb{C}^n$ are standard examples of real and complex finite-dimensional Hilbert spaces, with the inner products $(x, y) = x \cdot y$ and $(z, w) = z \cdot \overline{w}$. Other Hilbert spaces included $\ell^2_{\mathbb{R}}$ and $\ell^2_{\mathbb{C}}$, and most importantly, $L^2[a, b]$ and $L^2_{\mathbb{C}}[a, b]$ with inner products $(f, g) = \int_{[a,b]} f\overline{g}$.

In a Hilbert space, we could discuss the *parallelogram identity* with the inner product norm:

$$\|x - y\|^2 + \|x + y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Using this, we proved the *nearest point property.* That is, if A is a closed convex subset of H and $x \in H$ then there exists a unique $a \in A$ with $\|x - a\| < \|x - y\|$ for all $y \in A \setminus \{a\}$. A special case of this is when A is a closed linear subspace, in which case the nearest point $a \in A$ has the additional property that $(x - a) \perp A$.

This brings us to orthogonality. If $E \subset H$ is any nonempty subset, then the *orthogonal complement* of E is $E^\perp = \{x \in H : (x, e) = 0 \text{ for all } e \in E\}$. We checked that $E^\perp$ is a closed linear subspace. We also had a theorem: If M is any closed linear subspace then

(1) $M^\perp \cap M = \{0\}$
(2) $x \in H$ implies that $x = y + z$, $y \in M$ and $z \in M^\perp$
(3) $(M^\perp)^\perp = M$

The proof of (2) uses the nearest point property, and the proof of (3) uses (2) to show one of two inclusions.

We then discussed orthonormal sequences in H. Suppose that $e_1, \ldots, e_N$ is a finite orthonormal sequence. We had a basic identity from which many results followed:

$$\left\| x - \sum_{j=1}^{N} \lambda_j e_j \right\|^2 = \|x\|^2 + \sum_{j=1}^{N} |c_i - \lambda_i|^2 - \sum_{i=1}^{N} |c_i|^2.$$

In particular, the nearest point of span $\{e_1, \ldots, e_N\}$ to x is $\sum_{j=1}^{N} c_j e_j$. In this case, our basic identity reduces to

$$\left\| x - \sum_{j=1}^{N} c_j e_j \right\|^2 = \|x\|^2 - \sum_{i=1}^{N} |c_i|^2,$$

which is *Bessel's identity.*

We can also take an infinitely orthonormal sequence $e_1, e_2, \ldots$. This allows several conclusions:

(1) $\sum_{j=1}^{\infty} c_j e_j$ always converges. We proved this by checking that the partial sums formed a Cauchy sequence. That is, there exists y such that

$$\left\| y - \sum_{j=1}^{N} c_j e_j \right\| \to 0$$

as $N \to \infty$.

(2) $\sum_{i=1}^{\infty} |c_i|^2 \leq \|x\|^2$. This is Bessel's inequality.

(3) $x = \sum_{j=1}^{\infty} c_j e_j$ if and only if equality holds in Bessel's inequality. In this case, this is called Bessel's identity.

This leads us to our next definition. The orthonormal sequence $e_1, e_2, \ldots$ is *complete* if $x = \sum_{j=1}^{\infty} c_j e_j$ for every x. We showed that the following are equivalent:

(1) $e_1, e_2, \ldots$ is complete
(2) Equality holds in Bessel's inequality for every $x \in H$
(3) No $x \neq 0$ satisfies $(x, e_j) = 0$ for all j
(4) $\operatorname{span}\{e_1, e_2, \ldots\}$ is dense in H.

Our key example of a complete orthonormal sequence is in the case $H = L^2_{\mathbb{C}}[-\pi, \pi]$. We showed that $\{e^{inx}\}_{n=0,\pm 1,\pm 2,\ldots}$ is a complete orthonormal sequence with the inner product $(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\overline{g}(t)\, dt$. This was a long story. The proof involved the *Fejér kernel*, and it's hard to overestimate the importance of this.

We digressed a little bit because we didn't fully understand $L^2_{\mathbb{C}}$. To understand this, we discussed the *Lebesgue integral*. We discussed:

(1) Definition of measure zero.
(2) Step functions.
(3) Main technical theorem.
   (a) If φ_k is an increasing sequence of step functions and $\int_a^b \varphi_k$ is bounded then we have $\{\varphi_k(x)\}_{k=1,2,\ldots}$ is bounded for almost every $x \in [a, b]$.
   (b) If ψ_k is an increasing sequence and $\lim \varphi_k = \lim \psi_k$ almost everywhere, then we have $\lim \int_a^b \varphi_k = \lim \int_a^b \psi_k$.
(4) Definition of $\int_{[a,b]} f$ for $f \in \mathcal{L}_0$
(5) $\mathcal{L}^1 = \{g - h : g, h \in \mathcal{L}_0\}$ is a linear space
(6) Properties of the integral. For example, if $f_k \geq 0$ and $\int_{[a,b]} f_k \to 0$ then there exists a subsequence f_{k_j} with $f_{k_j}(x) \to 0$ for almost every $x \in [a, b]$.
(7) $\mathcal{L}^2 = \{f : f \in \mathcal{L}^1, f^2 \in \mathcal{L}^1\}$ is a linear space
(8) Completeness.

We then discussed linear operators. Here, X is a normed space, and we let X^* be the set of bounded *linear functionals*, known as the *dual space*. This is a Banach space (even if X is not complete). Furthermore, if X and Y are both normed spaces then we defined $\mathcal{L}(X, Y) = \{\text{bounded linear operators } X \to Y\}$. This had an *operator norm*

$$\|T\| = \sup_{x \neq 0} \frac{\|T(x)\|}{\|x\|} = \sup_{\|x\|=1} \|T(x)\|.$$

We showed that if Y is complete then $\mathcal{L}(X, Y)$ is Banach.

We discussed *adjoint operators*. If H and K are Hilbert spaces and if $T \in \mathcal{L}(H, K)$ then there exists $T^* \in \mathcal{L}(K, H)$ such that $(T(x), y) = (x, T^*(y))$ for all $x \in H$ and $y \in K$. This needed the *Riesz representation theorem*. That is, in a Hilbert space H, any $f \in H^*$ can be written as $f(x) = (x, z)$ for some fixed $z \in H$.

Next, we discussed *compact operators*. If X and Y are Banach spaces, $T \in \mathcal{L}(X, Y)$ means that $T(\{x : \|x\| \leq 1\})$ is contained in a compact subset of Y. We proved that if $\{T_k\}$ is a sequence of compact operators and $T_k \to T$ in the operator norm then T is also compact. The proof of this required a diagonal process.

If X is a Hilbert space, we checked that any *Hilbert-Schmidt* operator is automatically compact, and we also gave an example of an operator that was compact but not Hilbert-Schmidt. Here, $H = \ell^2_{\mathbb{R}}$ and we had the operator $T \in \mathcal{L}(\ell^2_{\mathbb{R}}, \ell^2_{\mathbb{R}})$ with $T(x) = (x_1, \frac{x_2}{\sqrt{2}}, \frac{x_3}{\sqrt{3}}, \dots)$.

We also had various theorems about the *spectrum*. If X is a Banach space and $T \in \mathcal{L}(X, X)$ then we showed that $\sigma(T)$ is a closed subset of the closed disc in the complex plane of radius $\|T\|$. One of the last things that we did was to show that the spectrum is nonempty. One of the main tools for discussing the spectrum was that if $T \in \mathcal{L}(X, Y)$ with X a Banach space then $\|T\| < 1$ implies that $I - T$ is invertible.

The last part of the course was the *spectral theorem*, and the main application of that theory was the *Sturm-Liouville* theory.

E-mail address: `moorxu@stanford.edu`