

MATH602
Real Analysis II

Pingbang Hu

December 16, 2022

Abstract

This is a graduate level functional analysis taught by [Joseph Conlon](#). The prerequisites include linear algebra, complex analysis and also [real analysis](#). We'll use Peter Lax[[Lax02](#)] and Reed-Simon[[RS80](#)] as textbooks.

The focus of this course is rather standard, including Banach and Hilbert Spaces Theory, Bounded Linear, Compact, and Self-Adjoint Operators Theorem, Representation, Hahn-Banach, Open Mapping Theorem, and Spectral Theory. We also covered some point-set topology along the way.

Contents

1	Banach and Hilbert Spaces	2
1.1	Linear Spaces	2
1.2	Quotient Spaces	3
1.3	Normed Spaces	3
1.4	Banach Spaces	7
1.5	Inner Product Spaces	10
1.6	Hilbert Spaces	12
1.7	Fourier Series	14
2	Bounded Linear Operators	20
2.1	Bounded Linear Functionals	20
2.2	Representation Theorems	22
2.3	Hahn-Banach Theorem	26
2.4	Separation of Convex Sets	29
2.5	Bounded Linear Operators	32
3	Main Principles of Functional Analysis	39
3.1	Open Mapping Theorem	39
3.2	Closed Graph Theorem	41
3.3	Principle of Uniform Boundedness	43
3.4	Compact Sets in Banach Spaces	46
3.5	Weak Topology	49
3.6	Weak* Topology	51
4	Compact Operators and Spectral Theory	54
4.1	Compact Operators	54
4.2	Fredholm Theory	58
4.3	Spectrum of Bounded Linear Operators	60
4.4	Spectrum of Compact Operators	65
5	Self-Adjoint Operators on Hilbert Spaces	68
5.1	Spectrum of Self-Adjoint Operators	68
5.2	Spectral Theorem for Compact Self-Adjoint Operators	71
5.3	Continuous Functional Calculus	72
5.4	Borel Functional Calculus	78
6	Epilogue	81
6.1	Proof of Riesz Representation for $C(K)$	81
A	Review	90
A.1	Midterm Review	90

Chapter 1

Banach and Hilbert Spaces

Lecture 1: Introduction

We first briefly review different kinds of vector spaces.

30 Aug. 14:30

1.1 Linear Spaces

Let's first see the simplest (i.e., without structures) vector space called [linear vector space](#).

Definition 1.1.1 (Linear vector space). A *linear vector space* E over a field \mathbb{F} is a set with operations of addition and multiplication (by a scalar) such that it's closed under operations, and also the addition and scalar multiplication obey

- (a) $u + v = v + u$ for $u, v \in E$
- (b) $u + (v + w) = (u + v) + w$ for $u, v, w \in E$
- (c) $\exists 0 \in E$ such that $0 + u = u + 0 = u$ for $u \in E$
- (d) $\forall u \in E, \exists -u \in E$ such that $u + (-u) = 0$
- (e) $\lambda(u + v) = \lambda u + \lambda v$ for $u, v \in E, \lambda \in \mathbb{F}$
- (f) $(\lambda + \mu)u = \lambda u + \mu u$ for $u \in E, \lambda, \mu \in \mathbb{F}$
- (g) $\lambda(\mu u) = (\lambda\mu)u$ for $u \in E, \lambda, \mu \in \mathbb{F}$

Remark. If $v, w \in E, \lambda, \mu \in \mathbb{R}$ (or \mathbb{C}), then $\lambda v + \mu w \in E$.

Notation (Real and complex vector space). If E is over $\mathbb{F} = \mathbb{C}$, we usually call E a *complex vector space*; if $\mathbb{F} = \mathbb{R}$, we say E is a *real vector space*.

Example. Given $n \in \mathbb{N}$, \mathbb{R}^n is an n dimensional real [linear vector space](#).

Example. Given $n \in \mathbb{N}$, \mathbb{C}^n is an n dimensional complex [linear vector space](#).

We concentrate on ∞ dimensional [linear vector space](#).

Example. Let K is a [compact Hausdorff space](#), then

$$E = \{f: K \rightarrow \mathbb{R} \mid f(\cdot) \text{ is continuous}\}$$

is a ∞ dimensional real [linear vector space](#).

Notation (Subspace). If E is a **linear vector space**, then we say $E_1 \subseteq E$ is a *subspace* if $E_1 \subseteq E$ and E_1 is itself a **linear vector space**. Moreover, if $E_1 \subsetneq E$, we say E_1 is a *proper subspace*.

Observe that a **linear vector space** can have many subspaces.

1.2 Quotient Spaces

Sometimes we don't care about vectors in some directions, suggesting the notion of **quotient space**.

Definition 1.2.1 (Quotient Space). The *quotient space* E / E_1 of two **linear vector spaces** E, E_1 such that $E_1 \subseteq E$ is the set of equivalence classes of vectors in E where equivalence is given by $x \sim y$ if $x - y \in E_1$. Additionally, denote $[x]$ as the equivalence class of $x \in E$, i.e., $[x] = x + E_1$.

One can see that **quotient space** E / E_1 is a **linear vector space** since if $x_1 + x_2 \in E$, $[x_1] + [x_2] = [x_1 + x_2]$, and also, $\lambda[x] = [\lambda x]$ for $\lambda \in \mathbb{R}$ or \mathbb{C} , i.e., $v, w \in E / E_1$, $\lambda, \mu \in \mathbb{R}$ or \mathbb{C} implies $\lambda v + \mu w \in E$. The dimension of a **quotient space** is defined as follows.

Definition 1.2.2 (Codimension). If E / E_1 has finite dimension, then the dimension of E / E_1 is called the *codimension* of E_1 in E , denoted as $\text{codim}(E_1)$.

Definition 1.2.2 is introduced since the way of defining dimensions for finite dimensional **vector spaces** doesn't work here. Recall **Theorem 1.2.1** in the finite dimension case.

Theorem 1.2.1. If E is finite dimensional, then $\text{codim}(E_1) + \dim(E_1) = \dim(E)$

We see that we may encounter something like $\infty - \infty$ if we define $\text{codim}(E_1) := \dim(E) - \dim(E_1)$, and indeed, **Definition 1.2.2** is well-defined in this sense.

Example. There exists the case that $\dim(E) = \infty$, $\dim(E_1) < \infty$ where $\dim(E / E_1) < \infty$.

Proof. Let $E = \{f: K \rightarrow \mathbb{R} \mid f(\cdot) \text{ continuous}\}$ and $E_1 = \{f \in E: f(k_1) = 0\}$ for a fixed $k_1 \in K$. We see that the dimension of E / E_1 is exactly 1 since E / E_1 is the set of constant functions. \circledast

Definition 1.2.3 (Linear operator). A map $T: E \rightarrow F$ between **linear spaces** E and F is a *linear operator* if it preserves the properties of addition and multiplication by a scalar, i.e., for $v, w \in E$ and $\lambda, \mu \in \mathbb{R}$ or \mathbb{C} ,

$$T(\lambda v + \mu w) = \lambda T(v) + \mu T(w).$$

Definition. Given a **linear operator** $T: E \rightarrow F$ we have the following.

Definition 1.2.4 (Kernel). The *kernel* of T is the subspace $\ker(T) = \{x \in E \mid Tx = 0\}$.

Definition 1.2.5 (Image). The *image* of T is the subspace $\text{Im}(T) = \{Tx \in F \mid x \in E\}$.

1.3 Normed Spaces

Given a vector, we want to measure the length of which. This suggests the following definitions.

Definition 1.3.1 (Norm). Let E be a **linear vector space**. A *norm* $\|\cdot\|: E \rightarrow \mathbb{R}$ on E is a function from E to \mathbb{R} with the properties:

- (a) $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$.

$$(b) \|\lambda x\| = |\lambda| \|x\|, \lambda \in \mathbb{R} \text{ or } \mathbb{C}.$$

$$(c) \|x + y\| \leq \|x\| + \|y\|.$$

Notation (Dilation). We say that the second condition is the *dilation* property.

Definition 1.3.2 (Normed vector space). A linear vector space E equipped with a norm $\|\cdot\|$ is called a *normed vector space*, denoted by $(E, \|\cdot\|)$.

A similar notion called *metric* is also widely used, though the structure is slightly coarser.

As previously seen (Metric). Given a vector space E , the metric $d(\cdot, \cdot): E \times E \rightarrow \mathbb{R}$ on E is a function from $E \times E$ to \mathbb{R} with the properties:

$$(a) d(x, y) \geq 0. \text{ Also, } d(x, x) = 0 \text{ and } d(x, y) \text{ implies } x = y.$$

$$(b) d(x, y) = d(y, x).$$

$$(c) d(x, z) \leq d(x, y) + d(y, z).$$

As one can imagine, if we can measure the length of a vector (by a *norm*), we can also measure the distance between vectors (by a *metric*).

Remark (Induced metric space). A normed vector space $(E, \|\cdot\|)$ induces a metric space (E, d) with the induced metric $d(x, y) = \|x - y\|$.

Now we give some well-known examples of *normed spaces*.

Example (Bounded sequences ℓ^∞). Let ℓ^∞ be the space of bounded sequences $x = (x_1, x_2, \dots)$ with $x_i \in \mathbb{R}$ for $i = 1, 2, \dots$. Then we define $\|x\| = \|x\|_\infty = \sup_{i \geq 1} |x_i|$.

Example (Absolutely summable sequences ℓ_1). Let ℓ_1 be the space of absolutely summable sequences $x = (x_1, x_2, \dots)$ and $\sum_{i=1}^\infty |x_i| < \infty$. Then we define $\|x\| = \|x\|_1 = \sum_{i=1}^\infty |x_i| < \infty$.

Example (Continuous functions $C(k)$). The space $C(k)$ of continuous functions $f: K \rightarrow \mathbb{R}$ where K is compact Hausdorff. Then we define $\|f\| = \|f\|_\infty = \sup_{x \in K} |f(x)|$.

1.3.1 Geometry of Normed Spaces

Now we can look into the structure of a *normed space* we're referring to without actually explaining what this really means previously. Intuitively, it's about the geometric properties of the spaces like how do *balls*, *spheres* and other shapes look like in that space when defining these shapes with *norms*.

Definition 1.3.3 (Ball). A (closed) *ball* centered at a point $x_0 \in E$ with radius $r > 0$ is the set

$$B(x_0, r) = \{x \in E \mid \|x - x_0\| \leq r\}.$$

Definition 1.3.4 (Sphere). The *sphere* centered at x_0 with radius $r > 0$ is the set

$$S(x_0, r) = \{x \in E \mid \|x - x_0\| = r\}.$$

Note. We see that $S(x_0, r)$ is the **boundary** of $B(x_0, r)$, i.e., $S(x_0, r) = \partial B(x_0, r)$.

Let's first look at *balls*. In finite dimensional, all *norms* are equivalent, which is not true for infinite dimensional *vector spaces*. This has something to do with the geometry of *balls*.

Explicitly, **balls** can have different geometries depending on the properties of the **norms**. We see that a $\|\cdot\|_\infty$ can have multiple supporting **hyperplane** at the corner, while for a $\|\cdot\|_2$ can have only one at each point.

Remark. The unit **balls** for $\|\cdot\|_1$ looks like **squares**, where we have

$$B(0, 1) = \{x = (x_1, x_2, \dots) \mid -1 < y_\epsilon < 1 \text{ for all } \epsilon\}$$

such that $y_\epsilon = \sum_{i=1}^{\infty} \epsilon_i x_i$, $\epsilon_i = \pm 1$ and $\epsilon = (\epsilon_1, \epsilon_2, \dots)$.

We see that different **norms** give different geometry, but they have important common features, most notably, **convexity** properties.

Definition 1.3.5 (Convex set). Given E a **linear vector space**, a set $K \subset E$ is *convex* if for $x, y \in K$ and $0 \leq \lambda \leq 1$,

$$\lambda x + (1 - \lambda)y \in K.$$

Definition 1.3.6 (Convex function). Given E a **linear vector space**, a function $f: E \rightarrow \mathbb{R}$ is called *convex* if for $x, y \in E$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Remark (Sublevel set). If $f: E \rightarrow \mathbb{R}$ is a **convex function**, then for any $M \in \mathbb{R}$ the *sublevel set* $\{x \in E \mid f(x) \leq M\}$ is **convex**.

The upshot is that **norms** are **convex**, and the unit **balls** are **convex** as well.

Lecture 2: Banach Spaces and Completion

Let's first see a proposition.

01 Sep. 14:30

Proposition 1.3.1. Let $(E, \|\cdot\|)$ be a **normed linear space**, then the norm is **convex** and continuous.

Proof. Let $f: E \rightarrow \mathbb{R}$ be $f(x) = \|x\|$. Then $f(x) - f(y) = \|x\| - \|y\| \leq \|x - y\|$, which implies $|f(x) - f(y)| \leq \|x - y\|$ for $x, y \in E$, i.e., f is Lipschitz continuous hence continuous. For **convexity**, let $0 \leq \lambda \leq 1$, we have

$$f(\lambda x + (1 - \lambda)y) = \|\lambda x + (1 - \lambda)y\| \leq \|\lambda x\| + \|(1 - \lambda)y\| = \lambda \|x\| + (1 - \lambda) \|y\| = \lambda f(x) + (1 - \lambda)f(y).$$

■

Note. Note that $f(\cdot) = \|\cdot\|$ is continuous implies the closed **ball**

$$B(x_0, r) = \{x \in E \mid \|x - x_0\| \leq r\} = \{x \in E \mid f(x - x_0) \leq r\}$$

is closed in topology of E . Also, $f(\cdot)$ is **convex** implies $B(x_0, r)$ is **convex**.

Remark. If $f: E \rightarrow \mathbb{R}$ is **convex**, then the sets $\{x \in E \mid f(x) \leq M\}$ is also **convex**. However, it's possible to have non-**convex functions** f such that all sets $\{x \in E \mid f(x) \leq M\}$ are **convex**.

Proof. Take $f(x) = |x|^p$ for $x \in \mathbb{R}$ and $p > 0$. We see that f is **convex** if $p > 1$, and non-**convex** if $p < 1$. However, the sets $\{x \in \mathbb{R} \mid f(x) \leq M\}$ are all **convex** since it's independent of p . \otimes

Lemma 1.3.1. Suppose $x \mapsto \|x\|$ satisfies

(a) $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$.

(b) $\|\lambda x\| = |\lambda| \|x\|$, $\lambda \in \mathbb{R}$ or \mathbb{C} .

(c) The unit ball $B(0, 1)$ is convex.

Then $f(x) = \|x\|$ satisfies the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$.

Proof. We see that if the third condition is true, then for $u, v \in B(0, 1)$ and $0 < \lambda < 1$, we have $\lambda u + (1 - \lambda)v \in B(0, 1)$. Let $x, y \in E$, and

$$\lambda = \frac{\|x\|}{\|x\| + \|y\|} \Rightarrow 1 - \lambda = \frac{\|y\|}{\|x\| + \|y\|}.$$

By letting $u = x / \|x\|$, $v = y / \|y\|$ we see that

$$\lambda u + (1 - \lambda)v = \frac{\|x\|}{\|x\| + \|y\|} \frac{x}{\|x\|} + \frac{\|y\|}{\|x\| + \|y\|} \frac{y}{\|y\|} \in B(0, 1) \Rightarrow \left\| \frac{x}{\|x\| + \|y\|} + \frac{y}{\|x\| + \|y\|} \right\| \leq 1.$$

From the second condition, it follows that $\|x + y\| \leq \|x\| + \|y\|$, which is the triangle inequality. ■

Remark. If $x \mapsto \|x\|$ satisfies the first two conditions and is convex, then it satisfies the triangle inequality.

Proof. Since

$$\frac{1}{2} \|x + y\| = \left\| \frac{x}{2} + \frac{y}{2} \right\| \leq \frac{1}{2} \|x\| + \frac{1}{2} \|y\|.$$

⊛

Now, given a quotient space E / E_1 , the question is can we try to define a norm?

Problem 1.3.1. On E / E_1 , is $\|[x]\| := \inf_{y \in E_1} \|x + y\|$ a norm?

Answer. No! If $x \in \overline{E_1} \setminus E_1$, then $\|[x]\| = 0$ but $0 \neq [x] \in E / E_1$.

⊛

We now see the difference from finite dimensional situation. All finite dimensional spaces E_1 are closed but not in general if E_1 has ∞ dimensions.

Example. Let $\ell_1(\mathbb{R})$ be the sequence of x_n for $n \geq 1$ in \mathbb{R} such that $\sum_{i=1}^{\infty} |x_i| \leq \infty$. Define

$$\|x\|_1 := \sum_{i=1}^{\infty} |x_i|,$$

and let E_1 be all sequences with finite number of the x_n are nonzero. We see that $\overline{E_1} = \ell_1(\mathbb{R})$ is infinite dimensional.

Proposition 1.3.2. Let $(E, \|\cdot\|)$ be a normed space and $E_1 \subseteq E$, E_1 is closed. Then

$$\|\cdot\| : E / E_1 \rightarrow \mathbb{R}, \quad \|[x]\| = \inf_{y \in E_1} \|x + y\|$$

is a norm on E / E_1 .

Proof. If $\|[x]\| = 0$, then $\inf_{y \in E_1} \|x - y\| = 0$, which implies $x \in E_1$ since E_1 is closed, so $[x] = 0$. Also, since

$$\|\lambda[x]\| = \inf_{y \in E_1} \|\lambda x + y\| = \inf_{z \in E_1} \|\lambda x + \lambda z\| = |\lambda| \inf_{z \in E_1} \|x + z\| = |\lambda| \|[x]\|,$$

the dilation property is satisfied. Finally, for triangle inequality, we have

$$\|[x] + [y]\| = \inf_{x_1, y_1 \in E} \|x + y + x_1 + y_1\| \leq \inf_{x_1 \in E_1} \|x + x_1\| + \inf_{y_1 \in E_1} \|y + y_1\| = \|[x]\| + \|[y]\|.$$

■

Remark. This shows that the only obstacle for this kind of **norm** being an actual **norm** is whether E_1 is closed.

1.4 Banach Spaces

Turns out that a **normed vector space** is not enough in general, hence we introduce the following.

Definition 1.4.1 (Banach space). A **linear normed space** is a *Banach space* if it's complete, i.e., every Cauchy sequence converges.

This implies that given a **Banach space** $(E, \|\cdot\|)$, if $\{x_n\}_{n \geq 1}$ is a sequence in E with the property such that $\lim_{m \rightarrow \infty} \sup_{n \geq m} \|x_n - x_m\| = 0$, then $\exists x_\infty \in E$ such that $\lim_{n \rightarrow \infty} \|x_n - x_\infty\| = 0$ as well.

Example. The spaces ℓ_1 , ℓ_∞ and $C(K)$ are **Banach spaces**.

1.4.1 Completion of Normed Space

We now show an important theorem which characterizes completeness in terms of convergence of series rather than sequences. We first see the definition.

Definition 1.4.2 (Absolutely summable). Let E be a **linear normed space** and a sequence $\{x_i\}_{i \geq 1}$ in E . Then $\{x_i\}_{i \geq 1}$ is *absolutely summable* if $\sum_{i=1}^{\infty} \|x_i\| < \infty$.

Then, we have the following.

Theorem 1.4.1 (Criterion for completeness). A **normed space** $(E, \|\cdot\|)$ is a **Banach space** if and only if every **absolutely summable** series in E converges.

Proof. We need to prove two directions.

(\Rightarrow) Suppose E is a **Banach space** and $\{x_k\}_{k \geq 1}$ an **absolutely summable** series. Set $s_n = \sum_{k=1}^n x_k$ for $n \geq 1$, we want to show s_n is Cauchy, and if this is the case, completeness of E implies $\exists s_\infty$ and $\lim_{n \rightarrow \infty} \|s_n - s_\infty\| = 0$. Let $n > m$, we see that

$$\|s_n - s_m\| = \left\| \sum_{k=m+1}^n x_k \right\| \leq \sum_{k=m+1}^n \|x_k\| \leq \sum_{k=m+1}^{\infty} \|x_k\|.$$

Observe that $\lim_{m \rightarrow \infty} \sum_{k=m+1}^{\infty} \|x_k\| = 0$, we see that the sequence $\{s_n\}$ is Cauchy, hence it converges.

(\Leftarrow) Conversely, suppose E is **not** complete. Then there exists a Cauchy sequence $\{x_n\}_{n \geq 1}$ which does not converge, implying no subsequence of $\{x_n\}_{n \geq 1}$ converges.^a We now construct an **absolutely summable** series which does not converge.

Define $n(1) \geq 1$ such that $\|x_n - x_{n(1)}\| \leq 1/2$ if $n \geq n(1)$, similarly, let $n(2) > n(1)$ be such that $\|x_n - x_{n(2)}\| \leq 1/2^2$ if $n > n(2)$. In all, we have $n(1) < n(2) < n(3) < \dots$ such that

$\|x_n - x_{n(k)}\| \leq 1/2^k$ if $n > n(k)$. Define $w_j := x_{n(j+1)} - x_{n(j)}$ for $j = 1, 2, \dots$. We see that

$$x_{n(m)} = x_{n(1)} + \sum_{j=1}^m w_j$$

for $m = 1, 2, \dots$, and $\{x_{n(m)}\}$ does not converge, hence so does the series $\sum_{j=1}^{\infty} w_j$. However, $\sum_{j=1}^{\infty} \|w_j\| \leq \sum_{j=1}^{\infty} 1/2^j = 1$, which implies $\{w_j\}$ is **absolutely summable**. ■

^aOtherwise, the whole sequence converges by the fact that it's Cauchy.

Theorem 1.4.2 (Completion). Suppose E is a **normed space**. Then there exists a **Banach space** \hat{E} called *the completion* of E with the following properties:

- (a) There exists a **linear map** $\iota: E \rightarrow \hat{E}$ such that $\|\iota x\| = \|x\|$.^a
- (b) $\text{Im}(\iota)$ is dense in \hat{E} , and \hat{E} is the smallest **Banach space** containing **image** of E .

^aThis is called an *isometric embedding* of E into \hat{E} .

Lecture 3: Banach, Inner Product Spaces

Notice that ℓ_1 and ℓ_∞ are **Banach**, and we want to generalize to ℓ_p with $1 < p < \infty$. For $x = \{x_n\}_{n \geq 1}$ in ℓ_p and if $\sum_{n=1}^{\infty} |x_n|^p < \infty$, for $\|x\|_p = (\sum_{n=1}^{\infty} |x_n|^p)^{1/p}$, we want to show that $x \mapsto \|x\|_p$ satisfies properties of a **norm**. The first two properties of a **norm** is easy check. As for triangle inequality, we have the following.

06 Sep. 14:30

Lemma 1.4.1 (Minkowski inequality). Let $1 \leq p < \infty$, for $x, y \in \ell_p$,

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

Proof. Recall that from **Lemma 1.3.1**, we only need to show that $B(0, 1)$ is **convex**, where

$$B(0, 1) = \left\{ x = \{x_n : n \geq 1\} \mid f(x) = \sum_{n=1}^{\infty} |x_n|^p \leq 1 \right\}.$$

But $f(x)$ is **convex** since $x \mapsto |x|^p$, $x \in \mathbb{R}$ is **convex** if $p \geq 1$, we're done. ■

Lemma 1.4.2 (Hölder's inequality). Let $1 < p < \infty$, for $x \in \ell_p$, $y \in \ell_q$, we have

$$\|x \cdot y\|_1 \leq \|x\|_p \|y\|_q$$

where $1/p + 1/q = 1$.

Proof. Note first that we can assume without loss of generality, $\sum_{j=1}^{\infty} |x_j|^p = \sum_{j=1}^{\infty} |y_j|^q = 1$. Then, result follows from the **Young's inequality**,

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

for $x, y > 0$, $x, y \in \mathbb{R}$. ■

Remark (Legendre transform and the inequality). **Young's inequality** is a special case of the inequality

$$xy \leq f(x) + \mathcal{L}f(y)$$

where $\mathcal{L}f(\cdot)$ is the **Legendre transform** of $f(\cdot)$, i.e., $\mathcal{L}f(y) = \sup_x [xy - f(x)]$.

If f is **convex**, then the function $xy \mapsto xy - f(x)$ is concave so has unique maximum. And $\mathcal{L}f(\cdot)$ always **convex** even if $f(\cdot)$ is not. In particular, if $f(x) = x^p/p$, then $\mathcal{L}f(y) = y^q/q$.

Note. **Minkowski inequality** is usually proved via the **Hölder's inequality**.

Proof. To show this, since

$$\sum_{j=1}^{\infty} |x_j + y_j|^p \leq \sum_{j=1}^{\infty} |x_j| |x_j + y_j|^{p-1} + \sum_{j=1}^{\infty} |y_j| |x_j + y_j|^{p-1}.$$

Then **Hölder's inequality** implies

$$\sum_{j=1}^{\infty} |x_j| |x_j + y_j|^{p-1} \leq \left(\sum_{j=1}^{\infty} |x_j|^p \right)^{1/p} \left(\sum_{j=1}^{\infty} |x_j + y_j|^{(p-1)q} \right)^{1/q},$$

and similarly,

$$\sum_{j=1}^{\infty} |y_j| |x_j + y_j|^{p-1} \leq \left(\sum_{j=1}^{\infty} |y_j|^p \right)^{1/p} \left(\sum_{j=1}^{\infty} |x_j + y_j|^{(p-1)q} \right)^{1/q}.$$

Note that $(p-1)q = p$, hence by combining both, we have

$$\sum_{j=1}^{\infty} |x_j + y_j|^p \leq \left[\left(\sum_{j=1}^{\infty} |x_j|^p \right)^{1/p} + \left(\sum_{j=1}^{\infty} |y_j|^p \right)^{1/p} \right] \left(\sum_{j=1}^{\infty} |x_j + y_j|^p \right)^{1/q},$$

i.e.,

$$\left(\sum_{j=1}^{\infty} |x_j + y_j|^p \right)^{1-1/q} = \left(\sum_{j=1}^{\infty} |x_j + y_j|^p \right)^{1/p} \leq \left(\sum_{j=1}^{\infty} |x_j|^p \right)^{1/p} + \left(\sum_{j=1}^{\infty} |y_j|^p \right)^{1/p},$$

proving the result. ⊗

Notice that **Minkowski inequality** and **Hölder's inequality** also hold for $1 \leq p \leq \infty$, or more generally, both hold for L^p spaces also. Let (Ω, Σ, μ) be a measure space and $L^p(\Omega, \Sigma, \mu)$ where all Σ measurable functions $f: \Omega \rightarrow \mathbb{R}$ (or \mathbb{C}) such that $\int_{\Omega} |f|^p d\mu < \infty$. Then, $L^p(\Omega, \Sigma, \mu)$ is a **normed space** with **norm**

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

It's more tricky to show that L^p is a **Banach space**, but it's indeed still the case.

Theorem 1.4.3 (Riesz-Fisher). The space $L^p(\Omega, \Sigma, \mu)$ is a **Banach space** for $1 \leq p < \infty$.

Proof. Toward using **Theorem 1.4.1**, let $\{f_n\}_{n \geq 1}$ be an **absolutely summable** sequence in L^p . Then the **norm** satisfies

$$\left\| \sum_{k=1}^N f_k \right\|_p \stackrel{!}{\leq} \sum_{k=1}^N \|f_k\|_p \leq C < \infty \Rightarrow \int_{\Omega} \left| \sum_{k=1}^N f_k \right|^p d\mu \leq C^p.$$

- Assume all f_k are non-negative. From **monotone convergence theorem**, we have

$$\lim_{N \rightarrow \infty} \int_{\Omega} \left(\sum_{k=1}^N f_k \right)^p d\mu = \int_{\Omega} \left(\sum_{k=1}^{\infty} f_k \right)^p d\mu \leq C^p.$$

Hence, $g = \sum_{k=1}^{\infty} f_k \in L^p$. We now want to show that $\sum_{k=1}^N f_k \rightarrow g$ in L^p . Set $r_n = \sum_{k=n+1}^{\infty} f_k$ where r_n is a decreasing sequence where $r_n \rightarrow 0$ a.e. and also

$$\int_{\Omega} r_1^p d\mu < \infty.$$

This means that $\lim_{n \rightarrow \infty} \|r_n\|_p = 0$ by **dominate convergence theorem**.

- For arbitrary $f_k: \Omega \rightarrow \mathbb{R}$, write $f_k = f_k^+ + f_k^-$ where $f_k^+ = \sup(f_k, 0)$ and $f_k^- = \inf(f_k, 0)$. The sequence $\{f_k^+\}_{k \geq 1}$ are **absolutely summable**, and we just proceed as before. Similarly, if $f_k: \Omega \rightarrow \mathbb{C}$, we get the same result. ■

1.5 Inner Product Spaces

Indeed, a slightly stronger structure than a **normed space** equipped is the so-called **inner product**, since it actually induces a **norm**.

Definition 1.5.1 (Inner product). Let E be a **linear space** over \mathbb{C} . An *inner product* $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{C}$ is a function which has the following properties:

- (a) $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- (b) $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$ for $a, b \in \mathbb{C}$.
- (c) $\langle x, y \rangle = \overline{\langle y, x \rangle}$.

Notation (Real inner product). We can also define **inner products** of spaces over \mathbb{R} with no extra conjugation in the last property.

Definition 1.5.2 (Inner product space). An *inner product space* is a **linear space** E with an **inner product** $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{C}$.

Definition 1.5.3 (Orthogonal). Given a **linear space** E , $x, y \in E$ are *orthogonal* if $\langle x, y \rangle = 0$, denote as $x \perp y$.

Theorem 1.5.1 (Cauchy-Schwarz inequality). Let $x, y \in E$ and an **inner product** $\langle \cdot, \cdot \rangle$, then

$$|\langle x, y \rangle| \leq \langle x, x \rangle^{\frac{1}{2}} \langle y, y \rangle^{\frac{1}{2}}.$$

Proof. Define $Q(t)$ by $Q(t) = \langle x + ty, x + ty \rangle = \langle y, y \rangle t^2 + 2t \operatorname{Re} \langle x, y \rangle + \langle x, x \rangle$ if $t \in \mathbb{R}$. Then we see that $Q(t) \geq 0$ with $t \in \mathbb{R}$, by looking at the discriminant, we have $(\operatorname{Re} \langle x, y \rangle)^2 \leq \langle x, x \rangle \langle y, y \rangle$. Finally, the result follows by choosing $\theta \in \mathbb{R}$ such that $\langle x, y \rangle = \operatorname{Re} \langle x e^{i\theta}, y \rangle$, we then see that

$$|\langle x, y \rangle| = |\operatorname{Re} \langle x e^{i\theta}, y \rangle| = |\operatorname{Re} \langle x, y \rangle| \leq \sqrt{\langle x, x \rangle \langle y, y \rangle},$$

proving the result. ■

Corollary 1.5.1. The function $x \mapsto \|x\| := \langle x, x \rangle^{\frac{1}{2}}$ is a **norm** on E .

Proof. The first two properties of a **norm** is easy to verify, and the triangle inequality is a conse-

quence of [Cauchy-Schwarz inequality](#) such that

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2 \operatorname{Re} \langle x, y \rangle + \|y\|^2 \stackrel{!}{\leq} \|x\|^2 + 2 \|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

Remark (Pythagorean theorem). The calculation in [Corollary 1.5.1](#) clearly implies *Pythagorean theorem*, which states that if $x \perp y$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Example (Canonical inner product in ℓ_2). Consider $x = (x_1, x_2, \dots), y = (y_1, y_2, \dots) \in \ell_2$, the space of square summable sequences, then

$$\langle x, y \rangle := \sum_{j=1}^{\infty} x_j \bar{y}_j$$

defines an [inner product](#).

Example (Canonical inner product in L^2). Consider $f, g \in L^2(\Omega, \Sigma, \mu)$, the space of square integrable functions, then

$$\langle f, g \rangle = \int_{\Omega} f(x) \bar{g}(x) \, d\mu(x)$$

defines an [inner product](#). Furthermore, $\|f\|_2 = \langle f, f \rangle^{1/2}$.

Proof. The only non-trivial fact to prove is that $\langle f, g \rangle$ is finite, i.e., $f\bar{g}$ is integrable. Firstly, f^2, \bar{f}^2 and $(f + g)^2$ are all integrable since f, \bar{g} and $f + \bar{g}$ are all in L^2 , hence $f\bar{g}$ is also integrable. \circledast

Example. Consider A, B in the space of $m \times n$ matrices $A = (a_{ij}), 1 \leq i \leq m, 1 \leq j \leq n$, then

$$\langle A, B \rangle = \operatorname{tr}(AB^*)$$

defines an [inner product](#), where B^* is the [Hermitian adjoint](#) of B , i.e., for $B = (b_{ij})$, then $B^* = (b_{ij}^*)$ for $b_{ij}^* = \bar{b}_{ji}$.

Remark (Hilbert-Schmidt (Frobenius) norm). Specifically, the [norm](#) corresponding to this [inner product](#) is

$$\|A\|_{\text{HS}} := \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2},$$

which is known as the [Hilbert-Schmidt](#) or [Frobenius norm](#).

1.5.1 Geometry of Inner Product Spaces

Indeed, the structure of an [inner product space](#) is much more interesting, since we can now consider the notion of angle between vectors.

As previously seen. Recall that in Euclidean space \mathbb{R}^n , the [inner product](#) can be computed by the formula

$$\langle x, y \rangle = \|x\| \|y\| \cos \theta(x, y)$$

where $\theta(x, y)$ denotes the angle between x and y .

Similarly, we can define the angle between x, y in an [inner product space](#) by

$$\cos \theta(x, y) := \frac{\langle x, y \rangle}{\|x\| \|y\|} \in [-1, 1]$$

where the range is ensured by [Cauchy-Schwarz inequality](#), so it's well-defined. Though this concept is rarely used anyway. Indeed, the only useful case is when $\cos \theta = 0$, namely when x and y are perpendicular, or [orthogonal](#).

But beyond [orthogonality](#), there are other geometric properties in an [inner product space](#) captured by [norms](#). Specifically, both [parallelogram law](#) and [polarization identity](#) hold, and the result is stated in terms of [norm](#) while they actually rely on the property of [inner product](#).

Lemma 1.5.1 (Parallelogram law). Given E an [inner product space](#), we have

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

Proof. Recall that $\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2\operatorname{Re} \langle x, y \rangle + \|y\|^2$ and similarly, $\|x - y\|^2 = \|x\|^2 - 2\operatorname{Re} \langle x, y \rangle + \|y\|^2$, hence the result follows. ■

Lemma 1.5.2 (Polarization identity). Given E an [inner product space](#), we have

$$\langle x, y \rangle = \frac{1}{4} \left\{ \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 \right\}$$

Proof. The proof is just to expand the right-hand side in terms of [inner product](#). ■

Remark. [Polarization identity](#) shows that the function $x \mapsto \|x\|^2$ determines the [inner product](#).

Lecture 4: Orthogonality and Projection

1.6 Hilbert Spaces

08 Sep. 14:30

Just like the case of [normed spaces](#), the [inner product spaces](#) are incomplete in general, hence we define the completed spaces of which, called [Hilbert spaces](#).

Definition 1.6.1 (Hilbert space). A complete [inner product space](#) is called a *Hilbert space*.

Example. Both ℓ_2 and $L^2(\Omega, \Sigma, \mu)$ are [normed spaces](#) and complete, hence are [Hilbert space](#).

1.6.1 Orthogonality

We'll soon see that the key notion in [Hilbert space](#) theory is [orthogonality](#).

Definition 1.6.2 (Orthogonal complement). Let $A \subseteq \mathcal{H}$ where \mathcal{H} is a [Hilbert space](#), then the *orthogonal complement* A^\perp of A is

$$A^\perp := \{x \in \mathcal{H} : \langle x, y \rangle = 0 \text{ for } y \in A\}.$$

Remark. A^\perp is also a [Hilbert space](#), in particular, closed and $A^\perp \cap A \subseteq \{0\}$.

Proof. A^\perp is closed [linear subspace](#) of \mathcal{H} where the closure follows from the continuity of the function $x \mapsto \langle x, y \rangle$ for $x \in \mathcal{H}$ by looking at the inverse [image](#) of $\{0\}$. Also, for $x \in A^\perp \cap A$, $\langle x, x \rangle = 0$ implies $x = 0$. The reverse inclusion is false since A can be empty. *

The fundamental theory of [Hilbert spaces](#) is [Theorem 1.6.1](#).

Theorem 1.6.1 (Orthogonality principle). Assume $E \subseteq \mathcal{H}$ is a closed [linear subspace](#) of the [Hilbert space](#) \mathcal{H} and $x \in \mathcal{H}$. Then we have the following.

- (a) There exists a unique closest point $y = P_E x \in E$ to x , i.e., $\|x - P_E x\| = \inf_{y' \in E} \|x - y'\|$.
- (b) The point $y = P_E x \in E$ is the unique vector such that $x - y \in E^\perp$.



Proof. Note that the function $y' \mapsto \|x - y'\|$ for $y' \in E$ is **convex**. We expect a minimizer y' .

- (a) Let $y_n \in E$ for $n = 1, 2, \dots$ be a minimizing sequence, i.e.,

$$\lim_{n \rightarrow \infty} \|x - y_n\| = \inf_{y' \in E} \|x - y'\| =: d.$$

From **parallelogram law**, we have

$$\|y_n - y_m\|^2 + 4\|x - (y_n + y_m)/2\|^2 = 2\|x - y_n\|^2 + 2\|x - y_m\|^2.$$

As $n, m \rightarrow \infty$, the right-hand side goes to $4d^2$. But since $\frac{1}{2}(y_n + y_m) \in E$, we have $\|x - \frac{1}{2}(y_n + y_m)\| \geq d$, so

$$\lim_{m \rightarrow \infty} \sup_{n \geq m} \|y_n - y_m\|^2 = 0,$$

which implies $\{y_n\}$ is a Cauchy sequence. As \mathcal{H} is complete, we see that $y_n \rightarrow y_\infty \in E$, with $\|x - y_\infty\| = d$.

Now, with the fact that E is closed, we set $y_\infty = P_E x$ where y_∞ is unique since if $\|x - y_\infty\| = \|x - y'_\infty\| = d$, again by the **parallelogram law** where we now plug in y_∞ and y'_∞ instead of y_n and y_m as above, we see that $\|y_\infty - y'_\infty\| = 0$, hence $y_\infty = P_E x \in E$ is well-defined.

- (b) We now show $P_E x$ is the unique vector $y \in E$ such that $x - y \perp E$, i.e., $x - y \in E^\perp$. Let $y' \in E$ and let $Q(t)$ be the quadratic

$$Q(t) := \langle x - P_E x + ty', x - P_E x + ty' \rangle = \|x - P_E x + ty'\|^2.$$

Since $t \mapsto Q(t)$ has a **strict** minimum at $t = 0$, which implies $Q'(0) = 0$, i.e., $\operatorname{Re} \langle x - P_E x, y' \rangle = 0$ for all $y' \in E$, which further implies $\langle x - P_E x, y' \rangle = 0$ for all $y' \in E$. This shows that $x - P_E x \in E^\perp$.

Finally, we need to show $P_E x \in E$ is the unique vector such $x - P_E x \in E^\perp$. This can be seen from $Q(t) = \|x - P_E x\|^2 + t^2 \|y'\|^2$ for any $y' \in E$.

■

We see that **orthogonality principle** is actually quite surprising, since to show existence of such a closest point, we typically need

- (a) **Compactness**,
- (b) Non-degeneracy for uniqueness.

But here by using **parallelogram law** and the completeness of \mathcal{H} , we don't need these.

Remark. **Orthogonality principle** shows that the minimizer for the function $y' \mapsto \|x - y'\|$ for $y' \in E$ is characterized by the orthogonality condition, i.e., $x - y \perp E$ for some $y \in E$.

This suggests the following definition.

Definition 1.6.3 (Orthogonal projection). Let \mathcal{H} be a **Hilbert space** and let $E \subseteq \mathcal{H}$ be a closed subspace. The *orthogonal projection operator* $P_E: \mathcal{H} \rightarrow E$ is given by $x \mapsto P_E x$ where $P_E x$ is defined uniquely via $x - P_E x \in E^\perp$.

The [orthogonal projection](#) is actually a so-called [bounded linear map](#) which is defined below.

Definition 1.6.4 (Bounded linear map). Given a mapping $A: E \rightarrow E$ on a [Banach space](#) E , we say it's a *bounded linear map* if it's [bounded](#) and [linear](#).

Definition 1.6.5 (Linear map). The operator A is *linear* if for $x, y \in E$, $a, b \in \mathbb{C}$,

$$A(ax + by) = aA(x) + bA(y).$$

Definition 1.6.6 (Bounded map). The operator A is *bounded* if

$$\|A\| := \sup_{\|x\|=1} \|Ax\| < \infty.$$

Remark. Note that $\|Ax\| \leq \|A\| \|x\|$ for $x \in E$.

We see that $P_E x$ is a [bounded linear map](#) $P_E: \mathcal{H} \rightarrow E \subseteq \mathcal{H}$ with the properties $P_E^2 = P_E$ and $\|x\|^2 = \|P_E x\|^2 + \|(I - P_E)x\|^2$ since $(I - P_E)x \perp P_E x$. The latter property shows that

$$\|P_E\| \leq 1, \quad \|I - P_E\| \leq 1,$$

and in fact, $\|P_E\| = \|I - P_E\| = 1$. Also, $I - P_E$ is also an [orthogonal projection](#) onto E^\perp .

1.7 Fourier Series

[Hilbert space](#) gives a geometric framework for studying [Fourier series](#). The classical Fourier analysis studies situations where a function $f: [-\pi, \pi] \rightarrow \mathbb{C}$ can be expanded as [Fourier series](#)

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k) \frac{1}{\sqrt{2\pi}} e^{ikt}$$

with the Fourier coefficients

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t) e^{-ikt} dt.$$

In order to make Fourier analysis rigorous, we have to understand what functions f can be written as [Fourier series](#), and in what sense the [Fourier series](#) converges. To do so, it's of great advantage to depart from this specific situation and carry out Fourier analysis in an abstract [Hilbert space](#). Let $f(t)$ be a vector in the function space $L^2([-\pi, \pi])$, and the exponential functions e^{-ikt} will form a set of [orthogonal](#) vectors in this space. Then, [Fourier series](#) will become an orthogonal decomposition of a vector f w.r.t. an [orthogonal system](#) of coordinates.

1.7.1 Orthogonal Systems

We first give the definition.

Definition 1.7.1 (Orthogonal system). A sequence $\{x_k\}_{k \geq 1}$ of non-zero vectors in a [Hilbert space](#) \mathcal{H} is *orthogonal* if $\langle x_k, x_\ell \rangle = 0$ for all $\ell \neq k$.

Definition 1.7.2 (Orthonormal system). An [orthogonal system](#) $\{x_k\}_{k \geq 1}$ is an *orthonormal system* if in addition, we have $\|x_k\| = 1$ for all k .

Write it in a more compact way, $\{x_k\}_{k \geq 1}$ is [orthonormal](#) if $\langle x_k, x_\ell \rangle = \delta_{k,\ell}$ where δ is the [Kronecker delta](#). Here is an immediate generation the given [remark](#).

Theorem 1.7.1 (Pythagorean theorem). Let $\{x_k\}_{k \geq 1}$ be an **orthogonal system** in a **Hilbert space** \mathcal{H} . Then for every $n \in \mathbb{N}$,

$$\left\| \sum_{k=1}^n x_k \right\|^2 = \sum_{k=1}^n \|x_k\|^2$$

Proof. From **orthogonality**, we have

$$\left\| \sum_{k=1}^n x_k \right\|^2 = \left\langle \sum_{k=1}^n x_k, \sum_{k=1}^n x_k \right\rangle = \sum_{k,j=1}^n \langle x_k, x_j \rangle = \sum_{k=1}^n \langle x_k, x_k \rangle = \sum_{k=1}^n \|x_k\|^2.$$

■

We now see some examples.

Example (Canonical basis of ℓ_2). In the space ℓ_2 , $x_k = (0, 0, \dots, 1, 0, \dots, 0) \in \ell_2$ for $k = 1, 2, \dots$ is an **orthonormal system** in ℓ_2 forming a basis.

Example (Fourier basis in L^2). In the space $L^2([-\pi, \pi])$, consider the exponential

$$e_k(t) = \frac{1}{\sqrt{2\pi}} e^{ikt}$$

for $t \in [-\pi, \pi]$. The set $\{e_k\}_{k=-\infty}^{\infty}$ is an **orthonormal system** in $L^2[-\pi, \pi]$, forming a basis.

1.7.2 Fourier Series

We can further generalize **Fourier series** to any **Hilbert space** by letting $\{x_k\}_{k \geq 1}$ be an **orthonormal** set in \mathcal{H} as follows.

Definition. Consider an **orthonormal system** $\{x_k\}_{k=1}^{\infty}$ in a **Hilbert space** \mathcal{H} and a vector $x \in \mathcal{H}$.

Definition 1.7.3 (Fourier series). The *Fourier series* of x w.r.t. $\{x_k\}_{k \geq 1}$ is the formal series

$$\sum_{k=1}^{\infty} \langle x, x_k \rangle x_k.$$

Definition 1.7.4 (Fourier coefficient). The coefficient $\langle x, x_k \rangle$ in the **Fourier series** are called *Fourier coefficients* of x .

To understand the convergence of **Fourier series**, we first focus on the finite case and study the partial sums of **Fourier series**. For $n = 1, 2, \dots$, we define $S_n: \mathcal{H} \rightarrow E_n$ such that

$$S_n(x) = \sum_{k=1}^n \langle x, x_k \rangle x_k$$

for $x \in \mathcal{H}$ where $E_n = \text{span}(\{x_1, \dots, x_n\})$. We see that S_n is a **linear operator** and $S_n = P_{E_n}$ is the **orthogonal projection** onto E_n since $\langle x - S_n(x), x_k \rangle = 0$ for $k = 1, \dots, n$, hence $S_n(x) \in E_n$ and $x - S_n(x) \perp E_n$.

Theorem 1.7.2 (Bessel's inequality). Let $\{x_k\}_k$ be an **orthogonal system** in a **Hilbert space** \mathcal{H} . Then for every $x \in \mathcal{H}$,

$$\sum_{k=1}^{\infty} |\langle x, x_k \rangle|^2 \leq \|x\|^2.$$

Proof. To estimate the size of $S_n(x)$, consider $x - S_n(x)$ and from [Pythagorean theorem](#),

$$\|S_n(x)\|^2 + \|x - S_n(x)\|^2 = \|x\|^2 \Rightarrow \|S_n(x)\|^2 \leq \|x\|^2.$$

On the other hand, again by [Pythagorean theorem](#) and [orthogonality](#),

$$\|S_n(x)\|^2 = \sum_{k=1}^n \|\langle x, x_k \rangle x_k\|^2 = \sum_{k=1}^n |\langle x, x_k \rangle|^2.$$

We see that by combining these two inequalities and let $n \rightarrow \infty$, we have the result. \blacksquare

Remark. In particular, we see that $\|S_n(x)\|^2 = \sum_{k=1}^n |\langle x, x_k \rangle|^2$, with $S_n = P_{E_n}$ we have $\|P_{E_n}x\|^2 \leq \|x\|^2$ for all $x \in \mathcal{H}$.

This implies the following.

Corollary 1.7.1. Let $\{x_k\}_{k \geq 1}$ be an [orthonormal system](#) in a [Hilbert space](#) \mathcal{H} . Then the [Fourier series](#) $\sum_k \langle x, x_k \rangle x_k$ for every $x \in \mathcal{H}$ converges in \mathcal{H} .

Proof. This follows directly from [Bessel's inequality](#) with the fact that the tail sum is Cauchy, i.e., we have

$$\left\| \sum_{k=n}^m x_k \right\|^2 = \sum_{k=n}^m \|x_k\|^2 \rightarrow 0$$

as $n, m \rightarrow \infty$ from [Pythagorean theorem](#). \blacksquare

[Corollary 1.7.1](#) tells us that Fourier series of x converge. However, it needs not converge to x , although we can still compute the point where it converges to by considering [Bessel's inequality](#), and the optimality is guaranteed by [orthogonality principle](#).

Theorem 1.7.3 (Optimality of Fourier series). Let $\{x_k\}_k$ be an [orthonormal system](#) in a [Hilbert space](#) \mathcal{H} . Then the corresponding [Fourier series](#) $S_n(x) = \sum_{k=1}^n \langle x, x_k \rangle x_k$ converges, i.e., $\lim_{n \rightarrow \infty} S_n(x) = S_\infty(x)$ exists for $x \in \mathcal{H}$. Furthermore, $S_n = P_{E_n}$ for every n where E_n is the space spanned by $\{x_i\}_{i=1}^n$.^a

^aThis includes $n = \infty$, where E_∞ is the **closure** of the space spanned by $\{x_k\}_{k \geq 1}$.

Proof. We show that the sequence $S_n(x)$ for $n = 1, 2, \dots$ is Cauchy. This is because

$$\|S_n(x) - S_m(x)\|^2 = \sum_{k=m+1}^n |\langle x, x_k \rangle|^2,$$

and [Bessel's inequality](#) implies $\sum_{k=1}^\infty |\langle x, x_k \rangle|^2 \leq \|x\|^2$. Hence, for any $\epsilon > 0$, there exists $m(\epsilon)$ such that

$$\sum_{k=m(\epsilon)+1}^\infty |\langle x, x_k \rangle|^2 < \epsilon,$$

which implies $\|S_n(x) - S_m(x)\|^2 < \epsilon$ if $n > m(\epsilon)$, hence $\{S_n(x)\}_{n \geq 1}$ is Cauchy, implying

$$\lim_{n \rightarrow \infty} S_n(x) = S_\infty(x) \in \mathcal{H}.$$

Also, $S_\infty = P_{E_\infty}$ where E_∞ is the closure of the [linear space](#) generated by the $\{x_k\}_{k \geq 1}$. \blacksquare

Remark. [Orthogonality principle](#) states that among all convergent series of the form $S = \sum_k a_k x_k$, the approximation error $\|x - S\|$ is minimized by the Fourier series of x since it's the [projection](#).

We finally note that the closedness of E_∞ makes sense since the self-dual of a set's **orthogonal complement** is itself if it's closed in the first place.

Lecture 5: Abstract Fourier Series

1.7.3 Orthonormal Bases

13 Sep. 14:30

It should be easy to identify an extra condition which makes the **Fourier series** of every vector x converges to x .

Definition 1.7.5 (Complete system). A system of vector $\{x_k\}_k$ in **Hilbert space** \mathcal{H} is *complete* if the space spanned by $\{x_k\}_k$ is dense in \mathcal{H} , i.e., $\overline{\text{span}(\{x_k\}_k)} = \mathcal{H}$.

Definition 1.7.6 (Orthonormal basis). A **complete orthonormal system** in a **Hilbert space** \mathcal{H} is called an *orthonormal basis* of \mathcal{H} .

Then, we have the so-called **Fourier expansion**, which is sometimes called **Fourier inversion formula**.

Theorem 1.7.4 (Fourier expansions). Let $\{x_k\}_k$ be an **orthonormal basis** of a **Hilbert space** \mathcal{H} . Then every vector $x \in \mathcal{H}$ can be expanded in its **Fourier series**

$$x = \sum_k \langle x, x_k \rangle x_k.$$

Proof. If an **orthogonal set** $\{x_k\}_k$ is **complete**, then $E_\infty = \mathcal{H}$, $P_{E_\infty} = I$. This implies $x = \sum_{k=1}^\infty \langle x, x_k \rangle x_k$ for $x \in \mathcal{H}$. ■

Corollary 1.7.2 (Parseval's identity). Let $\{x_k\}_k$ be an **orthonormal basis** of a **Hilbert space** \mathcal{H} . Then

$$\|x\|^2 = \sum_{k=1}^\infty |\langle x, x_k \rangle|^2.$$

Proof. From **Fourier expansion**, $\|x\|^2 = \|P_{E_n}x\|^2 + \|I - P_{E_n}\|^2$. By letting $n \rightarrow \infty$, we have

$$\|x\|^2 = \lim_{n \rightarrow \infty} \|P_{E_n}x\|^2 = \lim_{n \rightarrow \infty} \sum_{k=1}^n |\langle x, x_k \rangle|^2 = \sum_{k=1}^\infty |\langle x, x_k \rangle|^2.$$

■

1.7.4 Gram-Schmidt Orthogonalization

Suppose $x_1, x_2, \dots \in \mathcal{H}$ is a set of vectors and $E_n = \text{span}(\{x_1, \dots, x_n\})$. Then we can find an **orthonormal set** $\{y_k\}_{k \geq 1}$ in \mathcal{H} such that $E_n = \text{span}(\{y_1, y_2, \dots, y_{m(n)}\})$ where $m(n) \leq n$. Such a procedure is called *Gram-Schmidt orthogonalization*. To do this, firstly, set $y_1 = x_1 / \|x_1\|$, then for $n \geq 2$, we have

$$y_n = \frac{(I - P_{E_{n-1}})x_n}{\|(I - P_{E_{n-1}})x_n\|}$$

if $x_n \notin E_{n-1}$, i.e., E_{n-1} is properly contained in E_n .

Remark. Proving **completeness** of a set of vectors $\{x_k\}_{k \geq 1}$ in \mathcal{H} can be **non-trivial**.

Note that we can effectively compute the vectors $P_{E_n}(x_{n+1})$ since we know that $S_n(x)$ is the **orthogonal projection** of x onto $\text{span}(\{y_k\}_{k \geq 1})$, which is the partial sum of **Fourier series**

$$S_n(x) = \sum_{k=1}^n \langle x, y_k \rangle y_k.$$

As for $P_n(x)$, we see that it's the **orthogonal projection** onto the **orthogonal complement**, i.e.,

$$P_{E_n}(x) = x - S_n(x) = x - \sum_{k=1}^n \langle x, y_k \rangle y_k \Rightarrow P_{E_n}(x_{n+1}) = x_{n+1} - \sum_{k=1}^n \langle x_{n+1}, y_k \rangle y_k.$$

Let's now see some examples.

Example (Haar basis). We consider the *Haar basis* for $L^2([0, 1])$. Let $h: (0, 1) \rightarrow \mathbb{R}$ where

$$h(t) = \begin{cases} 1, & \text{if } 0 < t < 1/2; \\ -1, & \text{if } 1/2 < t < 1. \end{cases}$$

Extend $h(\cdot)$ by zero outside $(0, 1)$, we get $h: \mathbb{R} \rightarrow \mathbb{R}$, $h(t) = 0$ if $t \notin (0, 1)$, otherwise it's the same as above. The function $t \mapsto h(2^k t)$ has support in interval $0 < t < 2^{-k}$. Move the support to interval $\ell 2^{-k} < t < (\ell + 1)2^{-k}$ by translation. Set

$$h_{k,\ell}(t) = h(2^k t - \ell), \quad \ell = 0, 1, \dots, 2^k - 1.$$

The constant function plus functions $h_{k,\ell}$, $k = 0, 1, 2, \dots$, $0 \leq \ell \leq 2^k - 1$ are a **complete orthogonal set** for $\mathcal{H} = L^2([0, 1])$.

Proof. The span of the Haar functions includes characteristics functions χ_F for all dyadic intervals $[2^{-k}\ell, 2^{-k}(\ell + 1)]$ for $\ell = 0, 1, \dots, 2^k - 1$, $k = 0, 1, \dots$. If the set is **not complete**, then there exists $f \in L^2([0, 1])$ such that

$$\int_F f \, dt = 0$$

for all dyadic intervals F . Since we can approximate any measurable set $E \subseteq (0, 1)$ by a union of dyadic intervals.

Intuition. An easy way to see this is to consider

$$\left\{ F \in \mathcal{B}: \int_F f \, dt = 0 \right\},$$

which is the Borel subalgebra of \mathcal{B} , which indeed is a Borel algebra on $(0, 1)$. Then observe that dyadic intervals generate all open intervals.

Hence, we see that $\int_F f \, dt = 0$ for all measurable $F \subseteq (0, 1)$. Let $F = \{t \in (0, 1): f(t) > 0\}$, if $m(F) > 0$, then

$$\int_F f \, dt > 0.$$

Hence, a contradiction, so $m(F) = 0$. ⊗

Example (Fourier basis). Consider the Fourier basis $e_k(t) = \frac{1}{\sqrt{2\pi}} e^{ikt}$ for $k \in \mathbb{Z}$, $-\pi < t < \pi$. This is **complete** in $L^2([-\pi, \pi])$.

Proof. We use **Stone-Weierstrass theorem** and apply it to Fourier basis. All $e_k(\cdot)$ are in $C([-\pi, \pi])$, i.e., continuous functions $f: [-\pi, \pi] \rightarrow \mathbb{C}$. We know that $C([-\pi, \pi])$ is a **Banach space** with supremum norm $\|f\| := \sup_{t \in [-\pi, \pi]} |f(t)|$. **Stone-Weierstrass theorem** implies density of the space spanned by $e_k(\cdot)$, $k \in \mathbb{Z}$ in $C([-\pi, \pi])$, hence the completeness in $L^2([-\pi, \pi])$ follows from the density of continuous functions in $L^2([-\pi, \pi])$. ⊗

Proposition 1.7.1. Let $\{x_k\}_k$ be a linear independent system in a **Hilbert space** \mathcal{H} . Then the system $\{y_k\}_k$ obtained by Gram-Schmidt orthogonalization of $\{x_k\}_k$ is **orthonormal** in \mathcal{H} , and for all $n \in \mathbb{N}$,

$$\text{span}(\{y_k\}_{k=1}^n) = \text{span}(\{x_k\}_{k=1}^n).$$

Proof. The system $\{y_k\}_k$ is **orthonormal** by construction, and we obviously have the inclusion $\text{span}(\{y_k\}_k) \subseteq \text{span}(\{x_k\}_k)$. Furthermore, since the dimensions of these subspaces both equal n by construction, so they're indeed equal. ■

1.7.5 Existence of Orthogonal Bases

From [Proposition 1.7.1](#), every **Hilbert space** that is not *too large* has an **orthonormal basis**. We call this **Hilbert space separable**.

Definition 1.7.7 (Separable). A metric space is *separable* if it contains a countable dense subset.

For **Banach space**, **separability** follows from finding a countable set of vectors $\{x_k\}_k$ such that the span of $\{x_k\}_k$ is dense in E . Formally, we have the following.

Lemma 1.7.1 (Separable spaces). A **Banach space** E is **separable** if and only if it contains a system of vectors $\{x_k\}_{k \geq 1}$ whose linear span is dense in E , i.e., $\overline{\text{span}(\{x_k\}_{k \geq 1})} = E$.

Furthermore, we can prove the following.

Theorem 1.7.5. Every **separable Hilbert space** has an **orthonormal basis**.

Remark. We developed the theory for countable **orthogonal systems** and bases. One can generalize it for systems of arbitrary cardinality.

A final remark will be [Theorem 1.7.6](#), which states that all **Hilbert spaces** of the same cardinality have the same geometry.

Theorem 1.7.6. All infinite-dimensional **separable Hilbert spaces** are isometric to each other. Precisely, for every such spaces \mathcal{H}_1 and \mathcal{H}_2 , there is a **linear** bijective map $T: \mathcal{H}_1 \rightarrow \mathcal{H}_2$ preserving the **inner product**, i.e., for all $x, y \in \mathcal{H}_1$, $\langle Tx, Ty \rangle = \langle x, y \rangle$.

While leaving out the proof, we note that from [Theorem 1.7.6](#), $\|Tx\| = \|x\|$ for all $x \in \mathcal{H}_1$, hence

$$\|T(x - y)\| = \|Tx - Ty\| = \|x - y\|$$

for all $x, y \in \mathcal{H}_1$, i.e., T preserves all pairwise distances, hence the name *isometry*.

Remark. Every **separable Hilbert space** is isometric to ℓ_2 and $L^2([0, 1])$.

Proof. Since ℓ_2 and $L^2([0, 1])$ are **separable Hilbert spaces**, the result follows from [Theorem 1.7.6](#).

⊛

Chapter 2

Bounded Linear Operators

In this chapter we study certain transformations of [Banach spaces](#). Because these spaces are linear, the appropriate transformations to study will be [linear operators](#). Furthermore, since [Banach spaces](#) carry topology, it is most appropriate to study continuous transformations, i.e. continuous [linear operators](#). They are also called [bounded linear operators](#) for the reasons that will become clear shortly.

2.1 Bounded Linear Functionals

When the operators' range is \mathbb{R} or \mathbb{C} , it is interesting enough already, hence we study this case first.

2.1.1 Continuity and Boundedness

At this moment, the topology does not matter, so we define [linear functional](#) on general [linear vector spaces](#).

Definition. Let E be a [linear space](#) over \mathbb{R} or \mathbb{C} .

Definition 2.1.1 (Linear functional). A *linear functional* on E is a [linear operator](#) $f: E \rightarrow \mathbb{R}$ or \mathbb{C} such that for $x, y \in E$, $a, b \in \mathbb{R}$ or \mathbb{C} ,

$$f(ax + by) = af(x) + bf(y).$$

Definition 2.1.2 (Bounded linear functional). A [linear functional](#) $f(\cdot)$ is *bounded* if

$$\|f\| := \sup_{\|x\|=1} |f(x)| < \infty.$$

Clearly, the boundedness of $f(\cdot)$ implies $|f(x - y)| \leq \|f\| \|x - y\|$ for $x, y \in E$, hence, $f(\cdot)$ is continuous¹ if it's [bounded](#).

Remark. Conversely, if a [linear functional](#) is continuous, then it is [bounded](#).

Proof. Suppose $f(\cdot)$ is not bounded, then there exists a sequence $x_n \in E$ such that $|f(x_n)| \geq n \|x_n\|$ for $n = 1, 2, \dots$. By linearity,

$$\left| f\left(\frac{x_n}{n \|x_n\|}\right) \right| \geq 1, \quad n = 1, 2, \dots$$

But we know $\lim_{n \rightarrow \infty} \frac{x_n}{n \|x_n\|} = 0$ and $f(0) = 0$, hence $f(\cdot)$ is not continuous at 0. ⊛

¹In fact, it is Lipschitz continuous.

2.1.2 Dual Spaces and Hyperplanes

Indeed, we have a special name for the space of all **bounded linear functionals** called **dual spaces** due to its importance.

Definition 2.1.3 (Dual space). Let E be a **normed space**, then the space of all **bounded linear functionals** $f(\cdot)$ on E is called the *dual space* E^* of E .

The **dual space** is also a **normed space** with **norm**

$$\|f\| := \sup_{\|x\|=1} |f(x)|,$$

which is in fact a **Banach space**.

Remark. E^* is a **Banach space** even if the original E is not.

This definition of $\|\cdot\|_{E^*}$ implies $|f(x)| \leq \|f\|\|x\|$ for $x \in E$, $f \in E^*$, and $\|f\|$ is the smallest number in this inequality that makes it valid for all $x \in X$.

Definition 2.1.4 (Hyperplane). Given a **linear space** E , a subspace $H \subseteq E$ is a *hyperplane* if $\text{codim}(H) = 1$, i.e., $\dim(E/H) = 1$.

The following question arises.

Problem 2.1.1. Does there exist a **non-closed hyperplane**?

Answer. We know that this is not the case in finite dimension. And this question is analogous to *does there exist a subset $F \subseteq \mathbb{R}$ which is **not** Lebesgue measurable?* The answer to this is yes in both cases. However, construction uses **axiom of choice**. *

The goal is to make an equivalence between **bounded linear functionals** on E and **closed hyperplanes** in E . Turns out that there is a canonical correspondence between the **linear functionals** and the **hyperplanes** in E . This is clarified in the following.

Proposition 2.1.1 (Linear functionals and hyperplanes). Let E be a **linear space**.

- (a) For every **linear functional** f on E , $\ker(f)$ is a **hyperplane** in E . If E is a **Banach space**, and $f(\cdot)$ is **bounded**, then $\ker(f) = H$ is closed.
- (b) If $f, g \neq 0$ are **linear functionals** on E such that $\ker(f) = \ker(g)$, then $f = ag$ for some $a \neq 0$.
- (c) For every **hyperplane** $H \subseteq E$, there exists a **linear functional** $f \neq 0$ on E such that $\ker(f) = H$. If E is a **Banach space** and $\ker(f) = H$ is closed, then $f(\cdot)$ is **bounded**.

Lecture 6: Riesz Representation Theorem

Let's first see the proof of **Proposition 2.1.1**.

15 Sep. 14:30

Proof of Proposition 2.1.1. We prove them in order.

- (a) Let $x, y \notin \ker(f)$, then $f(x), f(y) \neq 0$, meaning that there exists a scalar $\lambda \neq 0$ such that $f(x) = \lambda f(y)$, i.e., $x - \lambda y \in \ker(f)$. Hence, if $[x], [y] \in E / \ker(f)$, $[x] = \lambda[y]$, implying $\dim(E / \ker(f)) = 1$.

Now, if f is **bounded**, then f is continuous, so $\ker(f) = f^{-1}(\{0\})$ is closed.

- (b) Consider the induced **functionals** $\tilde{f}, \tilde{g}: E/H \rightarrow \mathbb{R}$ or \mathbb{C} where $H = \ker(f) = \ker(g)$. This implies

$$\dim\left(E/H\right) = 1 \Rightarrow \tilde{f} = a\tilde{g} \text{ for some } a \neq 0 \Rightarrow f = ag.$$

- (c) Assume $\dim(E/H) = 1$, so $E/H = \{a[x_0] : a \in \mathbb{C} \text{ (or } \mathbb{R})\}$ for some $x_0 \in E$. Then, for any $x \in E$, $[x] = a(x)[x_0]$ for some $a(x) \in \mathbb{C}$ or \mathbb{R} . Define $f(x) := a(x)$, we see that f is linear and $\ker(f) = H$.

Now, if E is a Banach space and H is closed with $\dim(E/H) = 1$. Recall that E/H is also a Banach space with norm $\|[x]\| = \inf_{y \in H} \|x + y\|$ for $x \in E$.^a Let \tilde{f} be a linear functional on E/H . Since $\dim(E/H)$ is finite, \tilde{f} is continuous, implying $|\tilde{f}([x])| \leq A \|[x]\|$ for all $x \in E$ for some scalar A . Finally, we define $f(x) = \tilde{f}([x])$ for $x \in E$, then $\ker(f) = H$ and $|f(x)| \leq A \|[x]\| \leq A \|x\|$.

^aWe see now why we need the closure: otherwise we'll get a non-zero function with norm 0.

2.2 Representation Theorems

In concrete Banach spaces, the bounded linear functionals usually have a specific and useful form. Generally speaking, all linear functionals on function spaces (such as L^p and $C(K)$) act by integration of the function (with respect to some weight or measure). Similarly, all linear functionals on sequence spaces (such as ℓ_p) act by summation with weights.

2.2.1 Dual of Hilbert Spaces

We now start by characterizing bounded linear functionals on a Hilbert space \mathcal{H} .

Theorem 2.2.1 (Riesz representation theorem). Let \mathcal{H} be a Hilbert space. Then we have the following.

- (a) For every $y \in \mathcal{H}$, then function $f(x) = \langle x, y \rangle$ for $x \in \mathcal{H}$ is a bounded linear functional on \mathcal{H} .
- (b) If $f: \mathcal{H} \rightarrow \mathbb{C}$ or \mathbb{R} is a bounded linear functional on \mathcal{H} , then there exists $y \in \mathcal{H}$ such that $f(x) = \langle x, y \rangle$ for all $x \in \mathcal{H}$. Hence, the dual \mathcal{H}^* of \mathcal{H} is isometric to \mathcal{H} .

Proof. We prove this in order.

- (a) $f(x) = \langle x, y \rangle$ is clearly a linear functional. Boundedness follows from Cauchy-Schwarz inequality such that

$$|f(x)| = |\langle x, y \rangle| \leq \|x\| \|y\|,$$

and we can achieve $\|f\| = \|y\|$ by setting $x = y / \|y\|$.

Note. Note that there exists x_f such that $\|x_f\| = 1$ since $\|f\| = \sup_{\|x\|=1} |f(x)| = f(x_f)$, i.e., the supremum is achieved, although we're working on an infinite dimensional space. This property does not always hold for bounded linear functionals on Banach space since the unit ball can be not compact. But this holds for Hilbert space.

- (b) Let $f: \mathcal{H} \rightarrow \mathbb{C}$ or \mathbb{R} be a bounded linear functional on \mathcal{H} . Let $H = \ker(f)$, which is closed from Proposition 1.7.1. Let H^\perp be the orthogonal complement of H , i.e., $\mathcal{H} = H \oplus H^\perp$. Then $\dim(\mathcal{H}/H) = 1 \Rightarrow \dim(H^\perp) = 1$. Choose $y' \in H^\perp$ such that $g(x) = \langle x, y' \rangle$, which is in \mathcal{H}^* from (i). Furthermore, we see that $\ker(g) = \ker(f)$, so from Proposition 1.7.1, f and g are equal up to a constant $\lambda \in \mathbb{C}$ or \mathbb{R} , i.e., $f = \lambda g$. It follows that

$$f(x) = \lambda g(x) = \lambda \langle x, y' \rangle = \langle x, \lambda y' \rangle =: \langle x, y \rangle$$

for $y := \lambda y'$, hence we're done.^a

^aWe can even show that y here is unique.

In a concise form, [Riesz representation theorem](#) can be realized as $\mathcal{H}^* = \mathcal{H}$. Given a [Hilbert space](#) \mathcal{H} , [Riesz representation theorem](#) identifies the [dual space](#) \mathcal{H}^* , which can be used to show [Radon-Nikodym theorem](#).

2.2.2 Proof of Radon-Nikodym Theorem

[Riesz representation theorem](#) can be used to give a *soft* proof of [Radon-Nikodym theorem](#). Consider two measures μ, ν on the same σ -algebra.

As previously seen (Absolutely continuous). Recall that ν is called *absolutely continuous* w.r.t. μ , abbreviated as $\nu \ll \mu$, if for measurable sets A , $\mu(A) = 0$ implies $\nu(A) = 0$.

Theorem 2.2.2 (Radon-Nikodym theorem). Let μ, ν be two finite measures^a such that $\nu \ll \mu$, then there exists $g \geq 0$ such that g is μ -integrable and

$$\nu(A) = \int_A g \, d\mu$$

for A measurable.

^aThis can be extended to σ -finite measures by decomposition.

Proof. Consider the [linear functional](#) $F: L^2(\mu) \rightarrow \mathbb{R}$ or \mathbb{C} such that

$$F(f) = \int_{\Omega} f \, d\mu.$$

Then we have $\|F(f)\| \leq \|f\|_2 \sqrt{\mu(\Omega)}$, i.e., F is also a [bounded linear functional](#) on $L^2(\mu + \nu)$, hence by [Riesz representation theorem](#), there exists $h \in L^2(\mu + \nu)$ such that

$$F(f) = \int_{\Omega} f h \, d(\mu + \nu)$$

for $f \in L^2(\mu + \nu)$, i.e.,

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f h \, d\mu + \int_{\Omega} f h \, d\nu \quad (2.1)$$

if $f \in L^2(\mu + \nu)$. This further implies

$$\int_{\Omega} f h \, d\nu = \int_{\Omega} f [1 - h] \, d\mu \quad (2.2)$$

for $f \in L^2(\mu + \nu)$.

Claim. Such h satisfies $0 < h \leq 1$ μ -a.e., moreover, $(\mu + \nu)$ -a.e.

Proof. We first note that $\mu(A) = 0 \Leftrightarrow \mu(A) + \nu(A) = 0$. Let $A = \{h \leq 0\}$, $f = \mathbb{1}_A$ be the characteristic function on A . Then [Equation 2.1](#) implies

$$\int_A h \, (d\mu + d\nu) \leq 0 \Rightarrow \mu(A) = 0 \Rightarrow h > 0 \, \mu \text{ a.e.}$$

But since g is a positive function, so we also need $h \leq 1$. Again, set $B = \{h > 1\}$, $f = \mathbb{1}_B$. Then [Equation 2.1](#) implies

$$\mu(B) = \int_B h \, (d\mu + d\nu) > \mu(B)$$

unless $\mu(B) = 0$. ⊗

Now, by using **monotone convergence theorem**, we conclude^a that **Equation 2.2** holds for all $f \geq 0$, $f \in L^2(\mu + \nu)$.^b Finally, let $A \subseteq \Omega$ measurable and $hf = \chi_A$, from **Equation 2.2**,

$$\nu(A) = \int_A \frac{1-h}{h} d\mu.$$

By letting $g := 1 - h/h \Rightarrow g = d\nu/d\mu$, we're done. ■

^aConsider $f_n(t) := \min(f(t), n)$ and let $n \rightarrow \infty$.

^bBoth sides could be ∞ .

Notation (Radon-Nikodym derivative). g in **Radon-Nikodym theorem** is referred to as the *Radon-Nikodym derivative* where $g := d\nu/d\mu$.

Note (Uniqueness). The uniqueness of Radon-Nikodym derivatives can be shown via

$$\int_A g d\mu = 0$$

for all μ -measurable A , i.e., $g = 0$ μ -a.e.

2.2.3 Dual of L^p and ℓ_p

Another useful application of **Riesz representation theorem** is to characterize L^p and ℓ_p spaces and their dual L_p^* and ℓ_p^* . We first see the following.

Remark. Consider spaces $L^p(\Omega, \mu)$ for $1 \leq p \leq \infty$, then we have

$$L^q(\Omega, \Sigma, \mu) \subseteq (L^p(\Omega, \Sigma, \mu))^*$$

where $1/p + 1/q = 1$.

Proof. The easy part is that $g \in L^q$ induces a bounded linear functional on L^p by setting

$$F(f) = \int_{\Omega} fg d\mu.$$

By **Hölder's inequality**, $|F(f)| \leq \|f\|_p \|g\|_q$, hence $\|F\| \leq \|g\|_q$. To show the equality and $\sup_{\|f\|_p} |F(f)|$ is attained for $1 < p < \infty$, we choose $f = g^{q-1} \text{sgn}(g)$ since

$$F(f) = \int_{\Omega} |g|^q d\mu = \|g\|_q^q,$$

and from $1/p + 1/q = 1 \Rightarrow q - 1 = q/p$, we have

$$\|f\|_p^p = \int_{\Omega} |f|^p d\mu = \int_{\Omega} |g|^q d\mu = \|g\|_q^q \Rightarrow \|f\|_p = \|g\|_q^{q/p} = \|g\|_q^{q-1}.$$

This implies

$$F(f) = \int_{\Omega} |g|^q d\mu \Rightarrow \|g\|_q^q = \|g\|_q \|f\|_p.$$

Note. We see that $\sup_{\|f\|_p=1} |F(f)|$ is attained by taking $f = \text{sgn}(g)$.

⊛

In particular, we have the following.

Theorem 2.2.3 ($L^{p*} = L^q$). Consider the space $L^p = L^p(\Omega, \Sigma, \mu)$ with finite measure or σ -finite measure μ . Then for $1 \leq p < \infty$ and the conjugate exponent q of p .

- (a) For every weight function $g \in L^q$, integration with weight

$$G(f) := \int_{\Omega} fg \, d\mu$$

for $f \in L^p$ is a **bounded linear functional** on L^p , and its norm is $\|G\| = \|g\|_q$.

- (b) Conversely, every **bounded linear functional** $G \in L^{p*}$ can be represented as integration with weight for some unique weight function $g \in L^q$. Moreover, $\|G\| = \|g\|_q$.

Lecture 7: Hahn-Banach Theorem

Remark. When $p = 1$, the supremum is not attained necessarily. Consider $g \in L^\infty$, $F(f) := \int fg \, d\mu$ is **dual** of L^1 . If $g(\cdot)$ is continuous on \mathbb{R} with unique maximum, then the supremum $\sup_{\|f\|_1=1} |F(f)|$ is not attained. In all, for $1 \leq p \leq \infty$, L^q contained in the **dual** of L^p . If $1 < p \leq \infty$, then $\sup_{\|f\|_p=1} |F(f)|$ is attained. For $p = 1$, the supremum is not necessarily attained.

20 Sep. 14:30

Now, we're ready to prove **Theorem 2.2.3**.

Proof of Theorem 2.2.3. To show that the **dual** of L^p is L^q if $1 \leq p < \infty$ where $1/p + 1/q = 1$, we use **Radon-Nikodym theorem**. Suppose $E = L^p(\Omega, \Sigma, \mu)$ with $1 \leq p < \infty$ and $f \in E^*$. Just consider finite measure space, i.e., $\mu(\Omega) < \infty$. We define a measure ν on Σ by $\nu(A) := F(\chi_A)$ for $A \in \Sigma$, where χ_A is the characteristic function of A . We see that

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \Rightarrow \nu \ll \mu,$$

and **Radon-Nikodym theorem** implies

$$\nu(A) = \int_A g \, d\mu$$

for some $g =: \frac{d\nu}{d\mu} \in L^1(\Omega, \Sigma, \mu)$. Note that g may not be in L^q since $q > 1$. Hence, $F(f) = \int_{\Omega} fg \, d\mu$ for all simple function f assuming $g \geq 0$. Set $f = g^{q-1}$ with the fact that $|F(f)| \leq \|F\|_p \|f\|_p$. Recall that $q - 1 = q/p$, hence

$$\int g^q \, d\mu \leq \|F\|_p \left(\int g^q \, d\mu \right)^{1/p} \Rightarrow \|g\|_q^q \leq \|F\|_p \|g\|_q^{q/p} = \|F\|_p \|g\|_q^{q-1},$$

hence $\|g\|_q \leq \|F\|_p$.

Note. We assume $g \geq 0$ is because ν is a sign measure, then if we have a bounded variation function, we can just break it into $\nu^+ + \nu^-$.

■

Remark. L^1 is a subset of $L^{\infty*}$ but not equal to it. If $F: L^{\infty}(\mu) \rightarrow \mathbb{C}$ is a **bounded linear functional**, then if $\Omega = K$ is a compact **Hausdorff space** F induces a **bounded linear functional** on $C(K)$, i.e., the space of continuous functions on K . We see that $C(K) \subseteq L^{\infty}(K, \Sigma, \mu)$ where Σ is the Borel algebra on K .

2.2.4 Dual of $C(K)$

Finally, we state the following important characterization of **bounded linear functionals** on $C(K)$.

Theorem 2.2.4 (Riesz representation for $C(K)$). Let $E = C(K)$ be the space of continuous functions on compact Hausdorff space K . Then we have the following.

- (a) For every Borel regular signed measure on K , the functional $F(f) = \int_K f d\mu$ is a bounded linear functional on K .
- (b) Every bounded linear functional on $C(K)$ can be expressed as $F(f) = \int_K f d\mu$ for some measure μ , and $\|F\| = |\mu|(K)$, i.e., $TV(K)$.

Proof. In this case, the proof is much more difficult, and we put the proof in Section 6.1. ■

2.3 Hahn-Banach Theorem

Hahn-Banach theorem allows one to extend continuous linear functionals f from a subspace to the whole normed space, while preserving the continuity of f . **Hahn-Banach theorem** is a major tool in functional analysis. Together with its variants and consequences, this result has applications in various areas of mathematics, computer science, economics and engineering.

Theorem 2.3.1 (Hahn-Banach theorem). Let E_0 be a subspace of a Banach space E . Then every $f_0: E_0 \rightarrow \mathbb{R}$ or \mathbb{C} has a continuous extension $f: E \rightarrow \mathbb{R}$ or \mathbb{C} such that $\|f\| = \|f_0\|$.

Proof. We assume E is separable, otherwise we need transfinite induction. Let $\{x_n\}_{n \geq 1}$ have the property that its span is dense in E .

Intuition. Separability allows us to extend f_0 one dimension at a time. Now, if we can extend f_0 such that

$$E_0 \rightarrow E_0 + \text{span}(\{x_1\}) \rightarrow E_0 + \text{span}(\{x_1, x_2\}) \rightarrow \dots \rightarrow E_0 + \text{span}(\{x_n\}_{n \geq 1}),$$

then $\|f\| = \|f_0\|$, with the final space is dense in E , we can extend f to E by continuity.

Lecture 8: Proof of Hahn-Banach Theorem and Duality

Let's first proceed the proof of **Hahn-banach theorem**.

20 Sep. 14:30

Proof of Theorem 2.3.1 (Continued). To extend f by 1 dimension, i.e., $E \rightarrow E + \text{span}(\{x_1\})$, first note that extension is determined by a single number $\gamma = f(x_1)$ since f is a linear functional. We want that $\|f\| = \|f_0\|$ such that the linear functional $f_0: E_0 \rightarrow \mathbb{R}$ extends to $f: E_0 + \{x_1\} \rightarrow \mathbb{R}$, i.e., we want

$$|f_0(x_0) + \lambda\gamma| \leq \|x_0 + \lambda x_1\|$$

for $x_0 \in E$, $\lambda \in \mathbb{R}$. By dividing the inequality by $\lambda \neq 0$, it's sufficient to find γ such that $|f_0(x_0) + \gamma| \leq \|x_0 + x_1\|$, $x_0 \in E_0$.

- Suppose f_0 is a real-valued function, we need

$$- \|x_0 + x_1\| \leq f_0(x_0) + \gamma \leq \|x_0 + x_1\|$$

for all $x_0 \in E_0$. Such a γ exists, provides $\|x_0 + x_1\| - f_0(x_0) \geq -\|x'_0 + x_1\| - f_0(x'_0)$ for all $x_0, x'_0 \in E_0$. Furthermore, this is equivalent to write

$$f_0(x_0 - x'_0) \leq \|x_0 + x_1\| + \|x'_0 + x_1\|$$

for all $x_0, x'_0 \in E_0$, i.e., $f_0(x_0 - x'_0) \leq \|x_0 + x_1\| + \|-x_1 - x'_0\|$ for $x_0, x'_0 \in E_0$. Recall that $\|f_0\| = 1$, we have

$$f_0(x_0 - x'_0) \leq \|x_0 - x'_0\| \leq \|x_0 + x_1\| + \|-x_1 - x'_0\|.$$

- For complex valued f , consider $f: E \rightarrow \mathbb{C}$ be a **linear functional** over \mathbb{C} and let $g(x) = \operatorname{Re} f(x)$. Then $g: E \rightarrow \mathbb{R}$ is a real-valued **linear functional**. We see that $f(x) = g(x) - ig(ix)$ for all $x \in E$.^a Conversely, if $g: E \rightarrow \mathbb{R}$ is a real **linear functional** on **Banach space** E over \mathbb{C} , then $f: E \rightarrow \mathbb{C}$ defined by $f(x) = g(x) - ig(ix)$, $x \in E$ is a complex **linear functional** on E .

But we need to be a bit careful since when we extend $f_0: E_0 \rightarrow \mathbb{C}$, we're extending 2 real dimensions since for $g_0 = \operatorname{Re} f_0$, we need to do

$$E_0 \rightarrow E_0 + \operatorname{span}(\{x_1\}) \rightarrow E_0 \rightarrow \operatorname{span}(\{x_1, x_2\}).$$

Again, define $f(\cdot) = g(\cdot) - ig(i\cdot)$, we want to show $|f| = \|f_0\|$. We use the fact that for $x \in E_0 + \{\lambda x_0: \lambda \in \mathbb{C}\}$,

$$e^{i\theta} f(x) = f(xe^{i\theta})$$

for $\theta \in \mathbb{R}$. Choose θ such that $f(xe^{i\theta}) = g(xe^{i\theta})$, and since we already have $|g(xe^{i\theta})| \leq \|f_0\| \|xe^{i\theta}\|$, we see that $|f(x)| \leq \|f_0\| \|x\|$ for $x \in E_0 + \{\lambda x_1: \lambda \in \mathbb{C}\}$.

The above shows that we can indeed extend one dimension at a time, and the result follows from the fact that the space is **separable**. ■

^aSince $f(ix) = if(x)$, hence $g(ix) = -\operatorname{Im} f(x)$.

2.3.1 Supporting Functionals

Hahn-Banach theorem has a variety of analytic and geometric consequences. One of the basic tools guaranteed by **Hahn-Banach theorem** is the existence of a **supporting functional** $f \in X^*$ for every $x \in X$.

Theorem 2.3.2 (Supporting functional). Let E be a **Banach space**, then for every $x \in E$, there exists $f \in E^*$ such that $\|f\| = 1$, $f(x) = \|x\|$, i.e., $\sup_{\|y\|=1} |f(y)|$ attained at $y = x$.

Proof. Consider dimension 1 space $E_0 = \operatorname{span}(x) = \{tx, t \in \mathbb{R} \text{ or } \mathbb{C}\}$. Define $f_0: E_0 \rightarrow \mathbb{R}$ or \mathbb{C} such that $f_0(tx) = t\|x\|$, then $\|f_0\| = 1$, and **Hahn-Banach theorem** implies there exists $f \in E^*$ with $\|f\| = 1$. We see that $f(x) = \|x\|$ explicitly attain the **norm** and $\|\cdot\|$ is clearly a continuous extension of $\|\cdot\|_{E_0} = f_0$ as required. ■

Notice that we don't have uniqueness (as we don't have it in **Hahn-Banach theorem**) since a unit ball in L^∞ has corner, which will give multiple **hyperplanes**.

Remark (Geometric interpretation). Let B be a unit ball $\{x \in E: \|x\| \leq 1\}$ in a real **Banach space** E . Choose $x_0 \in \partial B$ such that $\|x_0\| = 1$. Then there exists $f \in E^*$, $\|f\| = 1$, $f(x) = \|x\|$. Let $H = \ker(f) + x_0$ where H intersects B at x_0 , we see that H divides E into 2 disjoint subsets, while B lies in one of which.

Proof. Since $x \in B$ and $\|x\| < 1$ implies $|f(x)| \leq \|x\| < 1$, we have $f(x) < 1$, i.e., $B \subseteq \{x: f(x) < 1\}$ and $E = \{x: f(x) < 1\} \cup H \cup \{x: f(x) > 1\}$. ⊛

supporting-functional theorem states that for every vector x , we indeed attain its **norm** on some **functional** $f \in E^*$, i.e., their **supporting functional**. But recall that the **norm** of a **functional** $f \in E^*$ is defined as

$$\|f\| := \sup_{x \neq 0} \frac{|f(x)|}{\|x\|},$$

and in general, f will not attain its **norm** on some vector x . This observation leads to the following.

Corollary 2.3.1. For every vector x in a **normed space** E ,

$$\|x\| = \max_{f \neq 0} \frac{|f(x)|}{\|f\|}$$

where the maximum is taken over all non-zero **linear functionals** $f \in E^*$.

Hahn-Banach theorem implies that there are enough **bounded linear functionals** $f \in E^*$ on every space E . One manifestation of this is the following.

Corollary 2.3.2 (Separation of points). For every two vectors $x_1 \neq x_2$ in a **normed space** E , there exists a **functional** $f \in E^*$ such that $f(x_1) \neq f(x_2)$.

Proof. The **supporting functional** $f \in E^*$ of the vector $x = x_1 - x_2$ must satisfy

$$f(x_1 - x_2) = \|x_1 - x_2\| \neq 0,$$

as required. ■

2.3.2 Second Dual Space

Let E be a **normed space**, then the **functionals** f^* are designed to act on vectors $x \in E$ via

$$f: x \mapsto f(x).$$

But indeed, we can instead say that *vectors* $x \in E$ *act on functionals* $f \in E^*$ via

$$x: f \mapsto f(x).$$

Thus, a vector $x \in E$ can itself be considered as a function from E^* to \mathbb{R} , i.e., a functional. Furthermore, this function x is clearly linear, so we may consider x as a **linear functional** on E^* . Also, the inequality

$$|f(x)| \leq \|x\| \|f\|$$

shows that this **functional** is bounded, so $x \in E^{**}$. We may instead write x as x^{**} for clarity. Note that the **norm** of x^{**} as a **functional** is $\|x^{**}\|_{E^{**}} \leq \|x\|$ since

$$\|x^{**}\| = \sup_{\substack{\|f\|=1 \\ f \in E^*}} |x^{**}(f)| = \sup_{\substack{\|f\|=1 \\ f \in E^*}} |f(x)| \leq \|x\|,$$

implying that $\|x^{**}\| \leq \|x\|$ for all $x \in E$. But from **supporting functional** $f \in E^*$ of x , we actually have

$$\|x^{**}\| = \|x\|,$$

i.e., we have a *canonical embedding* of E into E^{**} . The above discussion leads to the **second dual space theorem**.

Theorem 2.3.3 (Second dual space). Let E be a **normed space**. Then E can be considered as a **linear subspace** of E^{**} . For this, a vector $x \in E$ is considered as a **bounded linear functional** on E^* via the action

$$x: f \mapsto f(x), \quad f \in E^*.$$

To characterize the canonical embedding, we have the following definition.

Definition 2.3.1 (Reflexive space). A **normed space** E is called *reflexive space* if $E = E^{**}$ under the canonical embedding.

Example. L^p spaces for $1 < p < \infty$ are **reflexive spaces**.

Proof. We know that $L^{p*} = L^q$ where $1 \leq p < \infty$ for q being the conjugate index of p . ⊗

Example. L^p spaces for $p = 1$ or ∞ are not **reflexive spaces**

Proposition 2.3.1. Let E be a [reflexive space](#), then every [linear functional](#) $f \in E^*$ attains its [norm](#) on E .

Proof. By [reflexivity](#), the [supporting functional](#) of f is a vector $x \in E^{**} = E$, thus $\|x\| = 1$ and $f(x) = \|f\|$, as required. ■

Remark (James' theorem). The converse of [Proposition 2.3.1](#) is also true, i.e., if every [functional](#) $f \in E^*$ on a [Banach space](#) E attains its [norm](#), then E is [reflexive](#).

Lecture 9: Hahn-Banach Theorem for Sublinear Functions

From [Proposition 2.3.1](#), we see that to show a [Banach space](#) E is not [reflexive](#), it's sufficient to find $f \in E^*$ such that $\sup_{\|x\|=1} |f(x)|$ is not attained. 27 Sep. 14:30

Example. Let $C([0, 1])$ be the space of continuous functions $g: [0, 1] \rightarrow \mathbb{C}$ with $\|g\| := \sup_{0 \leq t \leq 1} |g(t)|$. Then for $f \in E^*$,

$$f(g) = \int_0^1 h(x)g(x) \, dx$$

for

$$h(x) = \begin{cases} -1, & \text{if } 0 < x < \frac{1}{2}; \\ 1, & \text{if } \frac{1}{2} < x < 1. \end{cases}$$

Then $\|f\| = 1 = \sup_{\|g\|=1} |f(g)|$, but the supremum is not attained since g needs to be continuous.

2.4 Separation of Convex Sets

In this section, we can extend [supporting functional theorem](#) such that we now have it for arbitrary [convex sets](#) other than the unit [ball](#). Since [supporting functional theorem](#) depends on [Hahn-Banach theorem](#), so we should first generalize [Hahn-Banach theorem](#).

2.4.1 Hahn-Banach Theorem for Sublinear Functions

By looking into the proof of [Hahn-Banach theorem](#), we see that we only used positive homogeneity and triangle inequality of the axiom of [norm](#), which suggests we define the following.

Definition 2.4.1 (Sublinear). Let E be a [linear space](#), a function $\|\cdot\| : E \rightarrow [0, \infty)$ is *sublinear* if

- (a) $\|\lambda x\| = \lambda \|x\|$ for $\lambda \in \mathbb{R}^+$, $x \in E$.
- (b) $\|x + y\| \leq \|x\| + \|y\|$, $x, y \in E$.

Remark (Differences from norm). Note that for a [sublinear](#) function to be a [norm](#), we need

- (a) $\|-x\| = \|x\|$, $x \in E$
- (b) $\|x\| = 0 \Rightarrow x = 0$.

Now, we can then generalize [Hahn-Banach theorem](#) to [sublinear functions](#).

Theorem 2.4.1 (Hahn-Banach theorem for sublinear functions). Let E_0 be a subspace of a [linear space](#) E over \mathbb{R} . Let $\|\cdot\|$ be a [sublinear functional](#) on E , and $f_0: E_0 \rightarrow \mathbb{R}$ be a [linear functional](#) on E_0 satisfying $f_0(x) \leq \|x\|$ for $x \in E_0$. Then f_0 admits an extension f to E such that $f(x) \leq \|x\|$ for $x \in E$.

Proof. The idea is the same from [Hahn-Banach theorem](#). ■

2.4.2 Geometric Properties of Sublinear Functions

We see that by considering [sublinear functionals](#) instead of [norms](#) offers us more flexibility in geometric applications. In particular, [sublinear functionals](#) arise as [Minkowski functionals](#) of [convex sets](#).

Definition 2.4.2 (Absorbing). A subset K of a [linear vector space](#) is *absorbing* if

$$E = \bigcup_{t \geq 0} tK$$

where $tK := \{tk : k \in K\}$.

Definition 2.4.3 (Minkowski functional). Let K be an [absorbing convex](#) subset of a [linear vector space](#) E such that $0 \in K$. Then the *Minkowski functional* $\|\cdot\|_K$ is defined as

$$\|x\|_K := \inf \{t > 0 : x/t \in K\}.$$

Proposition 2.4.1. Let K be an [absorbing convex](#) subset of a [linear vector space](#) E such that $0 \in K$. Then [Minkowski functional](#) $\|x\|_K$ is a [sublinear functional](#) on E . Conversely, let $\|\cdot\|$ be a [sublinear functional](#) on a [linear vector space](#) E , then the sub-level set

$$K = \{x \in E : \|x\| \leq 1\}$$

is an [absorbing convex set](#), and $0 \in K$.

Proof. To prove the forward direction, the main observation is that since $0 \in K$ and K is [convex](#), then $x \in K \Rightarrow tx \in K$ if $0 \leq t < 1$. To show dilation, for $\lambda > 0$,

$$\|\lambda x\| = \inf \left\{ t > 0 : x \in \frac{t}{\lambda} K \right\} = \lambda \inf \{s > 0 : x \in sK\} = \lambda \|x\|.$$

To show triangle inequality, suppose $x \in tK$, $y \in sK$, then $x = tk_1$, $y = sk_2$ for some $k_1, k_2 \in K$. We then have

$$x + y = (t + s) \left(\frac{t}{t + s} k_1 + \frac{s}{t + s} k_2 \right) = (t + s)k$$

for some $k \in K$ since K is [convex](#), hence $x + y \in (t + s)K$, we then have $\|x + y\| \leq \|x\| + \|y\|$.

Now, if $\|\cdot\|$ is [sublinear](#), then $K = \{x \in E : \|x\| \leq 1\}$ is [absorbing](#), [convex](#) and $0 \in K$.^a ■

^a $0 \in K$ since $\|0\| = 0$, while the [convexity](#) comes from the triangle inequality.

Remark. If $K \neq -K$, then $\exists x \in E$ with $\|x\| \neq \|-x\|$. If $K = E$, then $\|\cdot\| \equiv 0$.

2.4.3 Separation of Convex Sets

[Hahn-Banach theorem](#) has some remarkable geometric implications, which are grouped together under the name of *separation theorems*. Under mild topological requirements, these results guarantee that two [convex sets](#) A, B can always be separated by a [hyperplane](#).

Theorem 2.4.2 (Separation of a point from a convex set). Let K be an open [convex subset](#) of a [normed space](#) E and $x_0 \notin K$. Then there exists a continuous [linear functional](#) $f : E \rightarrow \mathbb{R}$ with $f \neq 0$ and $f(x) < f(x_0)$ for $x \in K$.

Proof. By translation, we can assume without loss of generality that $0 \in K$. Since K is open, it is

absorbing. Now, let $\|\cdot\|_K$ be the Minkowski functional, then

$$\|x\|_K \leq \frac{1}{r} \|x\|$$

for $x \in E$ if $B(0, r) \subseteq K$.



Proceed as in supporting functional theorem for unit ball, we define f_0 on $\text{span}(\{x_0\})$ by

$$f_0(tx_0) = t \|x_0\|_K$$

for $t \in \mathbb{R}$. Then if $E_0 = \{\lambda x_0 : \lambda \in \mathbb{R}\}$, $f_0(x) \leq \|x\|_K$ for $x \in E_0$ (i.e., $\|\cdot\|_K$ dominates f_0) since for $t \geq 0$,

$$f_0(tx_0) = t \|x_0\|_K = \|tx_0\|_K;$$

while for $t \leq 0$,

$$f_0(tx_0) = t \|x_0\|_K \leq 0 \leq \|tx_0\|_K.$$

Then from Hahn-Banach theorem, we can extend f_0 to $f: E \rightarrow \mathbb{R}$ such that

$$f(x) \leq \|x\|_K \leq \frac{1}{r} \|x\|$$

for $x \in E$, hence $f \in E^*$. For separation, we see that if $x \in K$ (hence in E),

$$f(x) \leq \|x\|_K \leq 1 \leq \|x_0\|_K = f_0(x_0) = f(x_0),$$

hence $f(x) \leq f(x_0)$. To get a strict separation, since K is open, so $x + tv \in K$ for $x \in K$ and some $t > 0$ and all v with $\|v\| = 1$. Hence, for all $t = t_x > 0$, we have

$$f(x + tv) \leq f(x_0) \Rightarrow f(x) + t \sup_{\|v\|=1} f(v) \leq f(x_0).$$

With the fact that $f \neq 0$, so $\|f\| = \sup_{\|v\|=1} f(v) \neq 0$, we conclude that $f(x) < f(x_0)$. ■

A more general version holds.

Theorem 2.4.3 (Separation of convex sets). Let A, B be disjoint convex subsets of a Banach space E .

- (a) If A is open, then there $\exists f: E \rightarrow \mathbb{R}$ such that $f(a) < f(b)$ for $a \in A, b \in B$.
- (b) If A, B are closed and B is compact, then there $\exists f: E \rightarrow \mathbb{R}$ such that $\sup_{a \in A} f(a) < \inf_{b \in B} f(b)$.

Proof. We have the following.

- (a) Let $K = A - B = \{a - b : a \in A, b \in B\}$, we then see that K is open, convex and $0 \notin K$. Since we can separate a point from a convex set, there exists $f \in E^*$ such that

$$f(a - b) < f(0) = 0$$

for $a \in A, b \in B$, hence $f(a) < f(b)$ for $a \in A, b \in B$.

- (b) Let A be closed, B be compact. Then we have

$$d(A, B) = \inf \{\|x - y\| : x \in A, y \in B\} = r > 0.$$

Define $A_\delta := \{x \in E: d(x, A) < \delta\}$ where A_δ is open. By setting $\delta := r/2$, we have $A_\delta \cap B = \emptyset$. From (a), we see that there exists $f \in E^*$ such that $f(x) < f(y)$ for $x \in A_\delta$, $y \in B$. Then $a \in A$ implies $a + \delta/2v \in A_\delta$ for some v such that $\|v\| = 1$, hence

$$f(a + \delta/2v) < f(b)$$

for $b \in B$. So

$$f(a) + \frac{\delta}{2}f(v) < f(b)$$

for $b \in B$, $\|v\| = 1$. Take the supremum over $\|v\| = 1$, we have $\sup_{\|v\|=1} |f(v)| = \delta > 0$, implying $f(a) < f(b) - \delta$, $a \in A$, $b \in B$. Finally, we have

$$\sup_{a \in A} f(a) < \inf_{b \in B} f(b).$$

■

Lecture 10: Adjoint Operators and Ergodic Theorem

Before ending this section, we have this final characterization of [convex sets](#): they're intersections of [half-spaces](#)! 29 Sep. 14:30

Definition 2.4.4 (Half-space). A *half-space* $H \subseteq E$ has the form of

$$H = \{x \in E: f(x) \leq \lambda\}$$

for $f \in E^*$ and $\lambda \in \mathbb{R}$, i.e., it is what lies on one side of a [hyperplane](#).

Corollary 2.4.1. Let $K \subseteq E$ be a closed [convex set](#), then K is the intersection of all [half-spaces](#) containing K .

Proof. Firstly, K is trivially contained in the intersection of the [half-spaces](#) that contain K . Denote such an intersection as S , then we have $K \subseteq S$. On the other hand, to show $K \supseteq S$, if $x_0 \notin K$, we show that there's a [half-space](#) contains K but not x_0 , hence $x_0 \notin S$ too, i.e., $S \subseteq K$.

From [separation of convex sets theorem](#) with $A = K$ and $B = \{x_0\}$, there exists $f \in E^*$ such that $\lambda := \sup_{k \in K} f(k) < f(x_0)$. We then see that the [half-space](#) $\{x \in E: f(x) \leq \lambda\}$ contains K but not x_0 . ■

2.5 Bounded Linear Operators

Turns out that we can generalize the notion of [linear functionals](#) $f: E \rightarrow \mathbb{R}$ or \mathbb{C} by further abstracting out the range by another [Banach space](#). As one can imagine, several results for [linear operators](#) will be generalizations of those we have already seen for [linear functionals](#), but there'll be important differences though. For example, a natural extension of [Hahn-Banach theorem](#) fails for [linear operators](#).

Firstly, the [operator norm](#) is defined as follows, which is a [norm](#) on [bounded linear operators](#).

Definition 2.5.1 (Operator norm). Given an operator $T: E \rightarrow F$ acting between [normed spaces](#) E and F , its *operator norm* is defined as

$$\|T\| := \sup_{\substack{\|x\|=1 \\ x \in E}} \|Tx\|.$$

2.5.1 Continuity and Boundedness

As for [Definition 2.1.2](#), we have the following.

Definition (Bounded linear operator). Let X, Y be two Banach spaces and let T be a linear operator between X and Y . Then we say T is a *bounded linear operator* if $\|T\| < \infty$.

Remark (Bounded operator). We can also talk about boundedness of a(n) (nonlinear) operator T just the same as requiring $\|T\| < \infty$.

As before, given Definition 2.5.1, we always have

$$\|Tx\| \leq \|T\|\|x\|$$

for a linear operator $T: X \rightarrow Y$, $x \in X$.

Definition 2.5.2 (Lipschitz). The operator T is called *Lipschitz* if for $x, y \in E$,

$$\|Tx - Ty\| \leq \|T\|\|x - y\|.$$

We see that for a bounded linear operator, it's Lipschitz as well.

Remark (Continuity and Boundedness). Same as linear functionals, the continuity and boundedness of linear operators are equivalent.

2.5.2 Space of Operators

Let X and Y be normed space, and let $\mathcal{L}(X, Y)$ be the space of bounded linear operators $T: X \rightarrow Y$, then $\mathcal{L}(X, Y)$ is a Banach space under the norm $T \mapsto \|T\|$.

Example. The dual space of E is just $E^* = \mathcal{L}(E, \mathbb{R})$.

Remark. In particular, we have

- (a) $\|T\| = 0 \Leftrightarrow T = 0$.
- (b) $\|\lambda T\| = |\lambda| \|T\|$ for $\lambda \in \mathbb{R}$ or \mathbb{C} , $T \in \mathcal{L}(X, Y)$.
- (c) $\|T + S\| \leq \|T\| + \|S\|$, $T, S \in \mathcal{L}(X, Y)$.
- (d) $\|TS\| \leq \|T\|\|S\|$, $T, S \in \mathcal{L}(X, Y)$.

2.5.3 Adjoint Operators

The concept of adjoint operators is a generalization of matrix transpose in linear algebra. Recall that if $A = (a_{ij})$ is an $n \times n$ matrix with complex entries, then the Hermitian transpose of A is an $n \times n$ matrix $A^* = (\overline{a_{ij}})$. The transpose thus satisfies the identity

$$\langle A^*x, y \rangle = \langle x, Ay \rangle$$

for $x, y \in \mathbb{C}^n$. We now extend this to linear operators.

Definition 2.5.3 (Adjoint operator). Let $T \in \mathcal{L}(X, Y)$, the *adjoint* $T^* \in \mathcal{L}(Y^*, X^*)$ of T is defined as

$$T^*f: X \rightarrow \mathbb{R} \text{ or } \mathbb{C}$$

for $f \in Y^*$, and $T^*f(x) = f(Tx)$ for $x \in X$.

We should note that T^* is indeed a bounded linear operator since

$$|T^*f(x)| = |f(Tx)| \leq \|f\|\|Tx\| \leq \|f\|\|T\|\|x\|$$

for $x \in X$, hence T^*f is a **linear functional** where

$$\|T^*f\| = \sup_{\|x\|=1} |T^*f(x)| \leq \sup_{\|x\|=1} \|f\| \|T\| \|x\| = \|f\| \|T\|,$$

so $T^*f \in X^*$ and $\|T^*f\| \leq \|T\| \|f\|$. This implies $T^*: Y^* \rightarrow X^*$ with T^* being a **linear operator** and T^* is **bounded** with

$$\|T^*\| \leq \|T\|.$$

In fact, we can achieve equality, which is shown in **Proposition 2.5.1**.

Proposition 2.5.1. For every $T \in \mathcal{L}(X, Y)$, the **adjoint** T^* is in $\mathcal{L}(Y^*, X^*)$ with $\|T^*\| = \|T\|$.

Proof. Since

$$\begin{aligned} \|T^*\| &= \sup_{\|f\|_{Y^*}=1} \|T^*f\|_{X^*} = \sup_{\|f\|_{Y^*}=1} \sup_{\|x\|_X=1} |T^*f(x)| \\ &= \sup_{\|f\|_{Y^*}=1} \sup_{\|x\|_X=1} |f(Tx)| = \sup_{\|x\|_X=1} \sup_{\|f\|_{Y^*}=1} |f(Tx)|. \end{aligned}$$

By choosing f to be a **supporting functional** of Tx , $\sup_{\|f\|_{Y^*}=1} |f(Tx)| = \|Tx\|_Y$, hence

$$\|T^*\| = \sup_{\|x\|_X=1} \|Tx\|_Y = \|T\|.$$

■

Let's look at some properties of **adjoint operators**. Let $T, S \in \mathcal{L}(X, Y)$ and $T^*, S^* \in \mathcal{L}(Y^*, X^*)$, then

- (a) $(aT + bS)^* = aT^* + bS^*$, $a, b \in \mathbb{R}$ or \mathbb{C} . Also, $(aT)^*f(x) = f(aTx) = af(Tx) = aT^*f(x)$.
- (b) $(ST)^* = T^*S^*$. This implies that if $T \in \mathcal{L}(X, X)$ is invertible, then $T^* \in \mathcal{L}(X^*, X^*)$ is invertible and $(T^*)^{-1} = (T^{-1})^*$.

Remark (Adjoint operators on Hilbert spaces). Specialize to **Hilbert space** \mathcal{H} , then by **Riesz representation theorem**, $\mathcal{H}^* \equiv \mathcal{H}$, i.e., $f \in \mathcal{H}^* \Leftrightarrow \exists y \in \mathcal{H}$ such that $f(x) = \langle x, y \rangle$ for $x \in \mathcal{H}$. Let $T \in \mathcal{L}(\mathcal{H}, \mathcal{H})$, and $T^* \in \mathcal{L}(\mathcal{H}^*, \mathcal{H}^*)$ with

$$T^*f(x) = f(Tx) = \langle Tx, y \rangle$$

for $x, y \in \mathcal{H}$, $f \in \mathcal{H}^*$. By writing $T^*f(x) = \langle x, T^*y \rangle$, which defined $T^*y: \mathcal{H} \rightarrow \mathcal{H}$, hence $\langle Tx, y \rangle = \langle x, T^*y \rangle$ for $x, y \in \mathcal{H}$. Clearly, T^* is a **bounded linear operator** on \mathcal{H} , i.e., $T^* \in \mathcal{L}(\mathcal{H}^*, \mathcal{H}^*)$ since

$$\|T^*\| = \sup_{\|y\|=1} \|T^*y\| = \sup_{\|y\|=\|x\|=1} \langle x, T^*y \rangle = \sup_{\|y\|=\|x\|=1} \langle Tx, y \rangle = \|T\|$$

just like **Proposition 2.5.1**. We see that $T^* \in \mathcal{L}(\mathcal{H}^*, \mathcal{H}^*) \Rightarrow T^* \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ via **Riesz representation**. Note that if $T^* \in \mathcal{L}(\mathcal{H}, \mathcal{H})$,

$$(aT)^* = \bar{a}T^*$$

for $a \in \mathbb{C}$, where for T defined on **Banach space**, $(aT)^* = aT^*$.

Just as with **Hilbert space**, we have a generalized notion of **orthogonality**, which we call **annihilator**.

Definition 2.5.4 (Annihilator). Let $A \subseteq X$ where X is a **Banach space**, then the *annihilator* A^\perp of A is a subset of X^* defined as

$$A^\perp := \{f \in X^*: f(x) = 0, x \in A\}.$$

Note. A^\perp is a closed linear subspace of X^* .

Proposition 2.5.2. Given two Banach spaces X and Y , let $T \in \mathcal{L}(X, Y)$ and $T^* \in \mathcal{L}(Y^*, X^*)$. Then $(\text{Im } T)^\perp, \ker(T^*) \subseteq Y^*$ satisfy

$$(\text{Im } T)^\perp = \ker(T^*).$$

Proof. Since $f \in (\text{Im } T)^\perp \Leftrightarrow f(Tx) = 0$ for all $x \in X$, i.e., $T^*f(x) = 0 \Leftrightarrow T^*f = 0 \Leftrightarrow f \in \ker(T^*)$, proving the result. ■

Corollary 2.5.1. Let \mathcal{H} be a Hilbert space, and $T \in \mathcal{L}(\mathcal{H}, \mathcal{H})$. Then the orthogonal decomposition holds, i.e.,

$$\mathcal{H} = \overline{\text{Im } T} \oplus \ker(T^*).$$

Proof. By Proposition 2.5.2, $\ker(T^*) = (\text{Im } T)^\perp$. And since \mathcal{H} is Hilbert space, $\overline{\text{Im } T} = \text{Im } T^a$ hence $(\overline{\text{Im } T})^\perp = \ker(T^*)$. Then by using orthogonality principle, the proof is complete. ■

^aSince if $E \subseteq \mathcal{H}$, $(E^\perp)^\perp = \overline{E}$.

2.5.4 Ergodic Theory

We now see an application on ergodic theorems. Ergodic theorems allow one to compute space averages as time averages. Given a probability space (Ω, \mathcal{F}, P) with $P(\Omega) = 1$, let $T: \Omega \rightarrow \Omega$ be a measurable map, i.e., $T^{-1}A \in \mathcal{F}$ if $A \in \mathcal{F}$. Then, we define the following.

Definition 2.5.5 (Measure-preserving). Let (Ω, \mathcal{F}, P) be a probability space. A transformation $T: \Omega \rightarrow \Omega$ is called *measure-preserving* if

$$P(T^{-1}A) = P(A)$$

for $A \in \mathcal{F}$, where $T^{-1}A = \{\omega \in \Omega: T\omega \in A\}$.

Let's first see some examples which illustrate the so-called *time and space averages*. We start with simple dynamical systems corresponding to rotation.

Example (Rotation). Let $\Omega = [0, 1]$, P be the Lebesgue measure and \mathcal{F} be Borel sets. Given $\lambda \in \mathbb{R}$, define

$$T\omega = \omega + \lambda \bmod 1.$$

This is equivalent to rotation on the unit circle through an angle $2\pi\lambda$, and we see that T is *measure-preserving* and one-to-one, and T^{-1} exists.

Example (Shift Operator). Let $\Omega = [0, 1]$, P be the Lebesgue measure and \mathcal{F} be Borel sets. Now, let

$$T\omega = 2\omega \bmod 1.$$

Then we see that T is just the shift operator on the binary representation, i.e., given $\omega = \sum_{j=1}^{\infty} \frac{a_j}{2^j}$ for $a_j = 0$ or 1 , then

$$T\omega = \sum_{j=1}^{\infty} \frac{a_{j+1}}{2^j}.$$

Now, let the *dyadic interval* $I_{n,k}$ be defined as

$$I_{n,k} := \left[\frac{k-1}{2^n}, \frac{k}{2^n} \right]$$

for $1 \leq k \leq 2^n$, we have $T^{-1}I_{n,k} = I_{n+1,k} \cup I_{n+1,k+2^n}$, hence $P(T^{-1}I_{n,k}) = P(I_{n,k})$ for all dyadic intervals $I_{n,k}$. This implies

$$P(T^{-1}O) = P(O)$$

for all $O \in \mathcal{F}$, hence T is **measure-preserving**, but not one-to-one. In fact, T is a two-to-one mapping. The action of T is $[0, 1/2] \xrightarrow{T} [0, 1]$, $[1/2, 1] \xrightarrow{T} [0, 1]$. We see that T doubles the length of a dyadic interval. To summarize,

- T is **measure-preserving** since it is two-to-one.
- T is an expanding map, which is called hyperbolic.

Lecture 11: Ergodic Theorem and Open Mapping

Now, we're ready to discuss ergodic theorem formally. Suppose $T: \Omega \rightarrow \Omega$ is **measure-preserving**, we can associate operator U on $L^2(\Omega)$ by defining $Uf(\omega) = f(T\omega)$ for $f \in L^2(\Omega)$ and $\omega \in \Omega$. Notice that

4 Oct. 14:30

$$\int_{\Omega} f(T\omega) d\mu(\omega) = \int_{\Omega} f(\omega) d\mu(\omega)$$

for all $f \in L^1(\Omega)$,² so for $\varphi \in L^2(\Omega)$, $U\varphi(\omega) = \varphi(T\omega)$ and since

$$\langle U\varphi, U\psi \rangle = \int_{\Omega} \varphi(T\omega)\psi(T\omega) d\mu(\omega) = \int_{\Omega} \varphi(\omega)\psi(\omega) d\mu(\omega) = \langle \varphi, \psi \rangle$$

for $\varphi, \psi \in L^2(\Omega)$, we see that U is a **bounded linear operator** on $\mathcal{H} = L^2(\Omega)$ with $\|U\| = 1$, $\|U\varphi\| = \|\varphi\|$, $\varphi \in \mathcal{H}$. In addition, for $\varphi, \psi \in \mathcal{H}$, $\langle U\varphi, U\psi \rangle = \langle \varphi, \psi \rangle$ implies $\langle U^*U\varphi, \psi \rangle = \langle \varphi, \psi \rangle$, which further implies $U^*U = I$, so U is one-to-one. Let's first see one more definition before we proceed.

Definition 2.5.6 (Unitary operator). A *unitary operator* is a **bounded linear operator** $U: \mathcal{H} \rightarrow \mathcal{H}$ on a **Hilbert space** \mathcal{H} such that U is surjective and for all $x, y \in \mathcal{H}$,

$$\langle Ux, Uy \rangle_{\mathcal{H}} = \langle x, y \rangle_{\mathcal{H}}.$$

Notice that U is not necessarily onto. However, if U is indeed onto, then $UU^* = U^*U = I$, implying that U is a **unitary operator** on \mathcal{H} and invertible.

Note. U is invertible if and only if T is one-to-one.

Proof. Since U just need to be onto for U being invertible, with $U^*\varphi(\omega) = \varphi(T^{-1}\omega)$ for $\omega \in \Omega$, if T is one-to-one then T^{-1} is onto, implying U^* is onto, so is U . *

Remark. $T: \Omega \rightarrow \Omega$ is one-to-one implies T is almost onto.

Proof. Let A be a set such that $T(\Omega) \subset A$, and hence $T^{-1}A = \Omega$ so $P(T^{-1}A) = P(\Omega) = 1$, implying that $P(A) = 1$, hence $P(\Omega \setminus A) = 0$. *

In the case T is not invertible (e.g. a 2-1 mapping), one might expect a similar formula for U^* . In the **shift operator** example, $T_1: [0, 1/2] \rightarrow [0, 1]$, $T_2: [1/2, 1] \rightarrow [0, 1]$, and T_1, T_2 are invertible, we have

$$U^*\varphi(\omega) = \frac{1}{2} (\varphi(T_1^{-1}\omega) + \varphi(T_2^{-1}\omega)).$$

Definition 2.5.7 (Ergodic transformation). A one-to-one, **measure-preserving** transformation T is *ergodic* if the only functions $f \in L^2(\Omega, \mathcal{F}, P)$ which satisfy $f(T\omega) = f(\omega)$ for almost all $\omega \in \Omega$ are the constant functions.

Remark (Eigenfunction). Phrasing differently, a **measure-preserving** mapping $T: \Omega \rightarrow \Omega$ is **ergodic** if and only if the only eigenfunction $\varphi \in L^2(\Omega)$ of the corresponding operator U is the constant

²This is true by letting $f = \mathbb{1}_A$ and then extend to $L^1(\Omega)$.

function, i.e. $U\varphi = \varphi$ implying φ is a constant.

Lemma 2.5.1. A **measure-preserving** mapping $T: \Omega \rightarrow \Omega$ is **ergodic** if and only if invariant sets of T have probability 0 or 1, i.e. if $A \in \mathcal{F}$ satisfies

$$P((A - T^{-1}A) \cup (T^{-1}A - A)) = 0,$$

then $P(A) = 0$ or $P(A) = 1$.

Proof. Assume T is not **ergodic**, then there exists $\varphi \in L^2(\Omega)$ such that $U\varphi = \varphi$. Hence, we can find $a, b \in \mathbb{R}$, $a < b$ such that $A = \{\omega \in \Omega: a < \varphi(\omega) < b\}$ has $0 < P(A) < 1$. However,

$$T^{-1}A = \{\omega: T\omega \in A\} = \{\omega: a < \varphi(T\omega) < b\} = \{\omega: a < \varphi(\omega) < b\} = A,$$

and thus A is invariant.

Conversely, suppose $A \in \mathcal{F}$, we have $A = T^{-1}A$ up to measure-zero sets and $0 < P(A) < 1$, then $\varphi = \mathbb{1}_A$ satisfies $U\varphi = \varphi \in L^2(\Omega)$ with the fact that φ is not constant., proving the result. ■

Proposition 2.5.3. Suppose $T: \Omega \rightarrow \Omega$ is **measure-preserving** and $\varphi \in L^2(\Omega)$, $\mathbb{E}[\varphi] = 0$, then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \varphi(T^n \cdot) \rightarrow 0$$

in $L^2(\Omega)$.

Proof. Note it suffices to assume $\mathbb{E}[\varphi] = 0$. We want to show

$$\lim_{N \rightarrow \infty} \frac{1}{N} [I + U + U^2 + \dots + U^{N-1}] \varphi(\cdot) = 0$$

in $L^2(\Omega)$. If φ is **orthogonal** to the constant function. Since $\mathbb{E}[\varphi] = 0$, then $\langle \varphi, 1 \rangle = 0$. Define a *derivative* operator on $L^2(\Omega)$ such that

$$D\varphi = (U - I)\varphi = \varphi(T \cdot) - \varphi(\cdot).$$

Using the **fundamental theorem of calculus** argument,

$$[I + U + U^2 + \dots + U^{N-1}]D\varphi = (U^N - I)\varphi.$$

Hence,

$$\left\| \frac{I + U + U^2 + \dots + U^{N-1}}{N} \varphi \right\| \leq \frac{2\|\psi\|}{N}$$

if $\varphi = D\psi$. In that case that as $N \rightarrow \infty$ is zero, i.e. if $\varphi \in \text{Im}(D) \subset \mathcal{H} = L^2(\Omega)$, then we're done. Note also that

$$\left\| \frac{I + U + U^2 + \dots + U^{N-1}}{N} \right\| \leq 1$$

since $\|U\| = 1$. Hence, converge to zero if $\varphi \in \overline{\text{Im}(D)}$, which implies that there exists $\varphi_\epsilon \in \text{Im}(D)$ such that $\|\varphi_\epsilon - \varphi\| < \epsilon$, i.e.,

$$\left\| \frac{I + U + \dots + U^{N-1}}{N} (\varphi_\epsilon - \varphi) \right\| < \epsilon.$$

Recall $\overline{\text{Im}(D)} \oplus \ker(D^*) = \mathcal{H} = L^2(\Omega)$. It suffices to show $\ker(D^*)$ is spanned by constant functions. Note T is **ergodic** implies $\ker(D)$ is spanned by constants, we have $D\varphi = 0 \Leftrightarrow U\varphi = \varphi$, and

$$(D^*\varphi = 0 \Leftrightarrow U^*\varphi = 0) \Rightarrow (\langle \varphi, U^*\varphi, \varphi \rangle = \langle \varphi, \varphi \rangle).$$

Therefore, we have $\langle U\varphi, \varphi \rangle = \langle \varphi, \varphi \rangle$, and also,

$$\int \varphi(T\omega)\varphi(\omega) \, dP(\omega) = \int \varphi(\omega)^2 \, d\omega = \int \varphi(T\omega)^2 \, d\omega,$$

which implies

$$\frac{1}{2} \int [\varphi(T\omega)^2 + \varphi(\omega)^2] \, dP(\omega) - \int \varphi(T\omega)\varphi(\omega) \, dP(\omega) = 0.$$

i.e. $\frac{1}{2} \int [\varphi(T\omega) - \varphi(\omega)]^2 \, dP(\omega) = 0$, which means

$$\varphi(T\omega) = \varphi(\omega), \quad \omega \in \Omega.$$

i.e. $\varphi \equiv \text{constant}$ by [ergodicity](#). ■

Theorem 2.5.1 (von Neumann ergodic theorem). Suppose $T: \Omega \rightarrow \Omega$ is [measure-preserving](#), then for any $\varphi \in L^2(\Omega)$, one has

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \varphi(T^n \cdot) = \int_{\Omega} \varphi(\omega) \, dP(\omega).$$

Remark. Convergence is in the $L^2(\Omega)$ sense, i.e. mean square.

Chapter 3

Main Principles of Functional Analysis

In this chapter, we'll study three of the fundamental theorems in functional analysis, which together with [Hahn-Banach theorem](#), form the main principles of functional analysis. Those are the [open mapping theorem](#), [closed graph theorem](#) and the [uniform boundedness principle](#).

3.1 Open Mapping Theorem

Suppose $T: X \rightarrow Y$ is a [bounded linear operator](#) on [Banach spaces](#), and T is injective and surjective, i.e. $T^{-1}: Y \rightarrow X$ exists. We'll soon see that the [open mapping theorem](#) implies T^{-1} is a [bounded operator](#), where the main argument relies on [Baire category theorem](#).

Definition 3.1.1 (Nowhere dense). A set S in a [metric space](#) M is *nowhere dense* if its closure \overline{S} has empty interior.

Example. The [Cantor set](#) is a [nowhere dense](#) set.

Lecture 12: Open Mapping Theorem

Let's start with [Baire category theorem](#), which essentially states that every complete [metric space](#) is a set of second category.¹ 6 Oct. 14:30

Proposition 3.1.1 (Baire category theorem). A complete [metric space](#) M is never the union of a countable number of [nowhere dense](#) sets.

Proof. We prove this by contradiction. Assume $M = \bigcup_{n=1}^{\infty} A_n$ with each A_n [nowhere dense](#). Since A_1 is [nowhere dense](#), so we can find $x_1 \in M - \overline{A_1}$. Furthermore, since $\overline{A_1}$ is closed, so we can find open [ball](#) B_1 centered at x_1 with radius less or equal to 1 such that $B_1 \cap A_1 = \emptyset$.

Similarly, A_2 is [nowhere dense](#), so there exists $x_2 \in B_1 - \overline{A_2}$, with $\overline{A_2}$ closed, we can still find [ball](#) B_2 centered at x_2 with radius less or equal to 1/2 such that

$$x_2 \in B_2 \subseteq \overline{B_2} \subseteq B_1$$

and $B_2 \cap A_2 = \emptyset$. By induction, we can find a sequence $\{x_n\}_{n=1}^{\infty}$ and open [balls](#) B_n such that

$$x_{n+1} \in B_{n+1} \subseteq \overline{B_{n+1}} \subseteq B_n$$

where B_n has radius smaller than $1/2^{n-1}$ and $B_n \cap A_n = \emptyset$.

Now, since the sequence $\{x_n\}$ is Cauchy and M is complete, we know that $x_n \rightarrow x_{\infty} \in M$, so

¹See [Meagre set](#). Notice that this is nothing to do with the category theory.

$x_\infty \in B_n$ for all n and hence $x_\infty \notin A_n$ for all n . This implies

$$M \neq \bigcup_{n=1}^{\infty} A_n,$$

which is a contradiction \nmid ■

We can now prove the central theorem in functional analysis, the [open mapping theorem](#).

Theorem 3.1.1 (Open mapping theorem). Let X, Y be [Banach spaces](#) and $T \in \mathcal{L}(X, Y)$. Assume T is surjective, i.e., $T(X) = Y$, then T maps open sets in X to open sets in Y .

Proof. Let $B_X := \{x \in X \mid \|x\| \leq 1\}$ be a unit [ball](#) in X , similarly B_Y be a unit [ball](#) in Y .

Claim. It's sufficient to show $T(B_X) \supseteq \epsilon B_Y$ for some $\epsilon > 0$.

Proof. To see this, let $U \subseteq X$ be an open set and $y \in TU$. We need to show TU contains a neighborhood of y . Let $x \in U$ such that $Tx = y$. Since U is open, so there exists $\delta > 0$ such that $U \supseteq x + \delta B_X$, so

$$TU \supseteq T(x + \delta B_X) = y + \delta T(B_X) \supseteq y + \delta' B_Y$$

for some $\delta' = \delta\epsilon > 0$, i.e., TU contains a neighborhood of y . ⊗

We now show $TB_X \supseteq \epsilon B_Y$ for some $\epsilon > 0$. Observe that $X = \bigcup_{n=1}^{\infty} nB_X$, hence

$$Y = TX = \bigcup_{n=1}^{\infty} nT(B_X).$$

From [Baire category theorem](#), we know that there exists $n \geq 1$ such that $\overline{nT(B_X)}$ has non-empty interior, i.e., $\overline{TB_X}$ has non-empty interior too. Hence, there exists $y \in Y$, $\delta > 0$ such that $y + \delta B_Y \subseteq \overline{TB_X}$. With $TX = Y$, there exists $x \in X$ such that $Tx = y$, hence $\delta B_Y \subseteq \overline{T(B_X - \{x\})}$. Since $B_X - \{x\} \subseteq nB_X$ for some $n \geq 1$, meaning that $\delta B_Y \subseteq \overline{nTB_X}$, implying $\overline{TB_X} \supseteq \epsilon B_Y$ for some $\epsilon > 0$. Finally, we show that $\overline{TB_X} \subseteq T(2B_X)$, which will imply

$$TB_X \supseteq \frac{1}{2} \overline{TB_X} \supseteq \frac{\epsilon}{2} B_Y,$$

completes the proof. To see this, we use a scaling argument. Let $y \in \overline{TB_X}$, then there exists $x_1 \in B_X$ such that

$$y - Tx_1 \in \frac{\epsilon}{2} B_Y \subseteq \overline{T\frac{1}{2}B_X}.$$

We can then choose $x_2 \in \frac{1}{2}B_X$ such that

$$y - Tx_1 - Tx_2 \in \frac{\epsilon}{4} B_Y \subseteq \overline{T\frac{1}{2^2}B_X}.$$

By induction, we can construct a sequence $\{x_n\}_{n \geq 1}$ such that

$$x_n \in \frac{1}{2^{n-1}}B_X, \quad y - \sum_{j=1}^n Tx_j \in \frac{\epsilon}{2^n}B_Y.$$

Then, $x = \sum_{j=1}^{\infty} x_n \in 2B_X$ where $Tx = y$. ■

3.1.1 Inverse Mapping Theorem

As an immediate consequence of the [open mapping theorem](#), we have the [inverse mapping theorem](#).

Theorem 3.1.2 (Inverse mapping theorem). Let $T: X \rightarrow Y$ be a **bounded linear operator** between **Banach spaces** X and Y which is both injective and surjective. Then T has a **bounded inverse** $T^{-1} \in \mathcal{L}(Y, X)$.

Proof. Since **open mapping theorem** states that the preimages of open sets under T^{-1} are open, hence T^{-1} is continuous. ■

Inverse mapping theorem is used to establish stability of solutions of linear equations. Consider a linear equation in x in a **Banach space**

$$Tx = b$$

for $T \in \mathcal{L}(X, Y)$ and $b \in Y$. Assume that a solution x exists and is unique for every b , then, from **inverse mapping theorem**, we see that the solution $x = x(b)$ is continuous w.r.t. b . In other words, the solution is stable under perturbations of b . In case T is not injective but is surjective, we can still apply **inverse mapping theorem** to the injectivization of T as follows.

Corollary 3.1.1 (Surjective operators are essentially quotient maps). Let X, Y be **Banach spaces**. Then every surjective **bounded linear operator** $T \in \mathcal{L}(X, Y)$ is a composition of a quotient map and an isomorphism. Specifically,

$$T = \tilde{T}q,$$

where $q: X \rightarrow X / \ker(T)$ is the quotient map, $\tilde{T}: X / \ker(T) \rightarrow Y$ is an isomorphism.

Proof. Let \tilde{T} be the injectivization of T then by construction, $T = \tilde{T}q$ and \tilde{T} is injective. Since T is surjective, \tilde{T} is also surjective. Hence, by **inverse mapping theorem**, \tilde{T} is an isomorphism. ■

3.1.2 Isomorphic Embeddings

Finally, as we know, the **kernel** of every **bounded linear operators** $T \in \mathcal{L}(X, Y)$ is always a closed subspace from **Proposition 2.1.1**, while the **image** of T may or may not be closed. We can also characterize this.

Proposition 3.1.2 (Isomorphic embedding). Given two **Banach spaces** X, Y and $T \in \mathcal{L}(X, Y)$, the following are equivalent.

- (a) T is injective and $\text{Im}(T)$ is closed.
- (b) T is **bounded below**, i.e., $\exists c > 0$, $\|Tx\| \geq c\|x\|$ for all $x \in X$.

Proof. To show that (a) implies (b), we see that $T^{-1}: \text{Im}(T) \rightarrow X$ is **bounded** since $\text{Im}(T)$ is **Banach space**, from **open mapping theorem**,

$$\|T^{-1}y\| \leq c^{-1}\|y\|$$

for $y \in \text{Im}(T)$, $c > 0$ being some constant. Set $y := Tx$, then $\|Tx\| \geq c\|x\|$ for $x \in X$, we're done.

To show another direction, suppose T is **bounded below**, then T is injective since $Tx = 0$ implies $x = 0$. To see $\text{Im}(T)$ is closed, let $x_n \in X$ for $n \geq 1$ be a sequence such that $\{Tx_n\}_{n \geq 1}$ is Cauchy such that $\|Tx_n - Tx_m\| \geq c\|x_n - x_m\|$ for all n, m , implying $\{x_n\}_{n \geq 1}$ is Cauchy, hence $x_n \rightarrow x_\infty \in X$, i.e., $Tx_n \rightarrow Tx_\infty \in \text{Im}(T)$, proving the result. ■

Remark. With a bit more work, one can show that T is an isomorphic embedding is equivalent to two conditions given in **Proposition 3.1.2**.

3.2 Closed Graph Theorem

We now study the second main theorem in functional analysis, which characterizes the property of the **graph** of a **bounded linear operator**.

Definition 3.2.1 (Graph). Let $T \in \mathcal{L}(X, Y)$ for X, Y being Banach spaces. Then the *graph* $\Gamma(T)$ of T is defined as

$$\Gamma(T) := \{(x, Tx) \in X \times Y \mid x \in X\}.$$

Clearly, $\Gamma(T)$ is a linear subspace of the normed space $X \oplus Y$.

Definition 3.2.2 (Closed graph). The graph $\Gamma(T)$ of T is *closed* if it is a closed subspace of $X \times Y$.

Comparing the continuity of T and closedness of $\Gamma(T)$, we see that T is continuous if and only if

$$x_n \rightarrow x \in X \text{ implies } Tx_n \rightarrow Tx;$$

while $\Gamma(T)$ is closed if and only if

$$x_n \rightarrow x \in X \text{ and } Tx_n \rightarrow y \in Y \text{ imply } y = Tx.$$

Hence, the continuity of T always implies the graph is closed, and indeed, the converse is also true.

Theorem 3.2.1 (Closed graph theorem). Let $T: X \rightarrow Y$ be a linear operator between Banach spaces X and Y . Then T is bounded (continuous) if and only if $\Gamma(T)$ is closed.

Proof. We have already shown that if T is bounded, then $\Gamma(T)$ is closed. Now assume $\Gamma(T)$ is closed, then we see that $\Gamma(T)$ is a Banach space, so we can now use open mapping theorem. Define a norm on $X \times Y$ by

$$\|(x, y)\| = \|x\| + \|y\|,$$

then $\Gamma(T)$ is a Banach space with this norm. Define $u: \Gamma(T) \rightarrow X$ by $u(x, Tx) = x$ for $x \in X$, then u is bounded since $\|u\| \leq 1$. From open mapping theorem, we know that u is surjective and injective, implying $u^{-1}: X \rightarrow \Gamma(T)$ is bounded from inverse mapping theorem. Hence, we have $\|u(x, Tx)\| \geq c\|(x, Tx)\|$ for all $x \in X$ and some $c > 0$, i.e.,

$$\|x\| \geq c(\|x\| + \|Tx\|) \Rightarrow \|Tx\| \leq \left(\frac{1}{c} - 1\right) \|x\|$$

for all $x \in X$, so T is bounded. ■

Remark. When trying to prove T is continuous by showing $x_n \rightarrow x \in X$ implying $Tx_n \rightarrow Tx$, we can always assume Tx_n converges in Y .

Proof. From closed graph theorem, checking $x_n \rightarrow x \in X$ implies $Tx_n \rightarrow Tx$ is equivalent to check $x_n \rightarrow x \in X$ and $Tx_n \rightarrow y \in Y$ implies $y = Tx$, so we may just assume the limit exists. ⊗

3.2.1 Symmetric Operators on Hilbert Spaces

One application of closed graph theorem to self-adjoint (symmetric) operator, i.e., $T^* = T$, on Hilbert space is the following.

Theorem 3.2.2 (Hellinger-Toeplitz theorem). Let $T: \mathcal{H} \rightarrow \mathcal{H}$ be a linear operator on a Hilbert space \mathcal{H} . If T is self-adjoint, i.e., $\langle Tx, y \rangle = \langle x, Ty \rangle$ for $x, y \in \mathcal{H}$, then T is bounded.

Proof. From closed graph theorem, it suffices to show that for a self-adjoint operator T , $\Gamma(T)$ is closed. Let $\{x_n\}_{n \geq 1}$ in \mathcal{H} such that $x_n \rightarrow x_\infty \in \mathcal{H}$ and $Tx_n \rightarrow y_\infty \in \mathcal{H}$, then we need to show $Tx_\infty \rightarrow y_\infty$. From the self-adjointness of T and the continuity of an inner product, for all $z \in \mathcal{H}$,

$$\langle z, y_\infty \rangle = \lim_{n \rightarrow \infty} \langle z, Tx_n \rangle = \lim_{n \rightarrow \infty} \langle Tz, x_n \rangle = \langle Tz, x_\infty \rangle = \langle z, Tx_\infty \rangle.$$

Since this holds for all $z \in \mathcal{H}$, we know that $Tx_\infty = y_\infty$, hence $\Gamma(T)$ is closed, so T is bounded. ■

Hellinger-Toeplitz theorem identifies the source of considerable difficulties in mathematical physics since many natural operators such as differential, though satisfy the symmetry condition, but are unbounded, and hence **Hellinger-Toeplitz theorem** declares that such operators *can not be defined everywhere* on the **Hilbert space**.

Example. There are no useful notions of differentiation that would make all $f \in L^2$ differentiable.

Lecture 13: Principle of Uniform Boundedness

3.3 Principle of Uniform Boundedness

11 Oct. 14:30

The final consequence of **open mapping theorem** is the following, which completes the whole picture of functional analysis. We first see some definitions.

Definition. Let X, Y be **Banach spaces** and let $\mathcal{T} \subseteq \mathcal{L}(X, Y)$ be a family of **bounded linear operators** from X to Y .

Definition 3.3.1 (Point-wise bounded). \mathcal{T} is *point-wise bounded* if $\sup_{T \in \mathcal{T}} \|Tx\| < \infty$ for all $x \in X$.

Definition 3.3.2 (Uniformly bounded). \mathcal{T} is *uniformly bounded* if $\sup_{T \in \mathcal{T}} \|T\| < \infty$.

Theorem 3.3.1 (Uniform boundedness theorem). Let X, Y be **Banach spaces** and let $\mathcal{T} \subseteq \mathcal{L}(X, Y)$ be a family of **bounded linear operators** from X to Y such that it's **point-wise bounded**, then it's **uniformly bounded**.

Proof. Define $M: X \rightarrow \mathbb{R}$ by $M(x) = \sup_{T \in \mathcal{T}} \|Tx\|$ for $x \in X$, also, let $X_n := \{x \in X: M(x) \leq n\}$, we can then write $X = \bigcup_{n=1}^{\infty} X_n$. From **Baire category theorem**, there exists $n \geq 1$ such that $\overline{X_n}$ has non-empty interior.

Claim. X_n is closed.

Proof. Note that the function $x \mapsto M(x)$ for $x \in X$ is **lower semi-continuous**, i.e., $M(x) \leq \liminf_{x_n \rightarrow x} M(x_n)$ since

$$\|Tx\| \leq \lim_{n \rightarrow \infty} \|Tx_n\| \leq \liminf_{n \rightarrow \infty} M(x_n),$$

and by taking supremum over x , we have $M(x) \leq \liminf_{n \rightarrow \infty} M(x_n)$. Hence, we see that X_n is closed, i.e., $\overline{X_n} = X_n$. ⊗

Hence, X_n itself has non-empty interior, so $X_n \supseteq x_0 + \epsilon B_X$ for some $\epsilon > 0$ and $B_X := \{x \in X: \|x\| \leq 1\}$.

Since $M(\cdot)$ is symmetric and **convex**, i.e., $M(-x) = M(x)$ for $x \in X$ and $M(\lambda x + (1 - \lambda)y) \leq \lambda M(x) + (1 - \lambda)M(y)$ for $x, y \in X$, $0 < \lambda < 1$, we see that $X_n \supseteq x_0 + \epsilon B_X$. From symmetric, we also have $X_n \supseteq -x_0 + \epsilon B_X$. Then by **convexity**, we together have $X_n \supseteq \epsilon B_X$, hence

$$\|x\| \leq \epsilon \Rightarrow \sup_{T \in \mathcal{T}} \|Tx\| \leq n \Rightarrow \sup_{T \in \mathcal{T}} \|T\| \leq n/\epsilon < \infty.$$

■

Remark (Completeness). **Uniform boundedness theorem** still holds if X is a **Banach space** while Y is only a **normed space**.

Proof. In the above proof, we only use the completeness of X , not Y . ⊗

Note (Principle of condensation of singularities). The [uniform Boundedness theorem](#) is called *principle of condensation of singularities* by Banach and Steinhaus initially.

Proof. Suppose a family $\mathcal{T} \subseteq \mathcal{L}(X, Y)$ is not [uniformly bounded](#), then the set of vectors

$$\{Tx : x \in B_X, T \in \mathcal{T}\}$$

is unbounded. Then from the [uniform boundedness theorem](#), \mathcal{T} is not even [point-wise bounded](#), so there exists *one* vector $x \in X$ with unbounded trajectory $\{Tx : T \in \mathcal{T}\}$. One can say that the unboundedness of the family \mathcal{T} is condensated in a single *singularity* vector x . \circledast

3.3.1 Weak and Strong Boundedness

[Principle of uniform boundedness](#) can be used to check whether a given set in a [Banach space](#) is bounded in the following way. Firstly, let's see some definitions.

Definition. Let $A \subseteq X$ where X is a [Banach space](#).

Definition 3.3.3 (Weakly bounded). A is *weakly bounded* if $\sup_{f \in X^*} |f(x)| < \infty$ for all $x \in A$.

Definition 3.3.4 (Strongly bounded). A is *strongly bounded* if $\sup_{x \in A} \|x\| < \infty$.

Clearly, [strong boundedness](#) implies [weak boundedness](#), moreover, the converse is also true.

Corollary 3.3.1 (Weak boundedness implies strong boundedness). Let $A \subseteq X$ for X being a [Banach space](#). If A is [weakly bounded](#), then A is [strongly bounded](#).

Proof. Firstly, embed A into $A^{**} \subseteq X^{**}$ by considering the conical embedding $X \rightarrow X^{**}$, then

$$\sup_{x^{**} \in A^{**}} |x^{**}(f)| < \infty$$

for all $f \in X^*$. From the [uniform boundedness theorem](#), we have $\sup_{x^{**} \in A^{**}} \|x^{**}\| < \infty$, and with [Hahn-Banach theorem](#), we have $\|x^{**}\| = \|x\|$ for all $x \in X$, proving the result. \blacksquare

We now review the midterm in [Appendix A.1](#).

Lecture 14: Midterm

Good luck!

13 Oct. 14:30

Lecture 15: Compactness in Banach Spaces

3.3.2 Schauder Bases

20 Oct. 14:30

Finally, we introduce a useful basis called [Schauder basis](#). Recall that some bases we have seen is uncountable, making them hard to work with in practice, on the other hand, [Schauder basis](#) is countable and with nice properties.

Definition 3.3.5 ([Schauder basis](#)). Let X be a [separable Banach space](#). A sequence $\{x_k\}_{k \geq 1}$ is a *Schauder basis* for X if every $x \in X$ can be uniquely represented as a convergent series

$$x = \sum_{k=1}^{\infty} a_k x_k$$

for $a_k \in \mathbb{R}$ or \mathbb{C} .

Remark. We can show that only [separable](#) spaces can have [Schauder basis](#), which is why we require it directly in [Definition 3.3.5](#).

It's clearly that if $\{x_k\}_{k \geq 1}$ is a [Schauder basis](#), it's linear independent and complete. Surprisingly, [Definition 3.3.5](#) is stronger than these.

Remark. For [completeness](#), given any $\epsilon > 0$, we can find $\{a_k\}_{k=1}^n$ such that

$$\left\| x - \sum_{k=1}^n a_k x_k \right\| \leq \epsilon.$$

However, it might be the case that a_k actually depends on ϵ , hence $\lim_{\epsilon \rightarrow 0} a_k(\epsilon)$ generally does not exist. But in fact, from the basis property, [Definition 3.3.5](#) guarantees that one can achieve higher accuracy by using more and more terms without changing the previous a_k .

A major property of [Schauder bases](#) is the uniform bound on the partial sums.

Theorem 3.3.2 (Partial sums of a Schauder basis). Let $\{x_k\}_{k \geq 1}$ be a [Schauder basis](#) for a [Banach space](#) X . Then there exists an $M \geq 0$ such that for all $n \geq 1$,

$$\left\| \sum_{k=1}^n a_k x_k \right\| \leq M \|x\| = M \left\| \sum_{k=1}^{\infty} a_k x_k \right\|$$

for $x \in X$.

Proof. Define a sequence space

$$E := \left\{ a = \{a_k\}_{k \geq 1} : \sum_{k=1}^{\infty} a_k x_k \text{ converges in } X \right\}$$

and for $a \in E$, define

$$\|a\| = \sup_{n \geq 1} \left\| \sum_{k=1}^n a_k x_k \right\| < \infty.$$

We see that $\|\cdot\|$ is a [norm](#) on E since $\|a\| = 0 \Rightarrow a = 0$ follows from the uniqueness property for [Schauder basis](#) and the fact that E is a [Banach space](#), so E is [complete](#).

Now, define a [linear operator](#) $T: E \rightarrow X$ by

$$Ta = \sum_{k=1}^{\infty} a_k x_k,$$

we have $\|Ta\| \leq \|a\|$, so T is also [bounded](#), injective and surjective. From [open mapping theorem](#), $T^{-1}: X \rightarrow E$ is [bounded](#) such that $\|T^{-1}\| \leq M < \infty$, i.e.,

$$\|Ta\| \geq \frac{1}{M} \|a\|$$

for all $a \in E$. This is equivalent to say

$$\sup_{n \geq 1} \left\| \sum_{k=1}^n a_k x_k \right\| \leq M \left\| \sum_{k=1}^{\infty} a_k x_k \right\|.$$

■

Notation (Basis constant). The $M \geq 0$ in [Theorem 3.3.2](#) is called the *basis constant*.

Definition 3.3.6 (Biorthogonal functional). Given a Schauder basis $\{x_k\}_{k \geq 1}$ and $x \in X$, we have the set of biorthogonal functionals $\{a_k(x)\}_{k \geq 1} =: \{x_k^*(x)\}_{k \geq 1}$ such that $x = \sum_{k=1}^{\infty} a_k x_k$.

The reason we call $\{a_k\}_{k \geq 1}$ for a specific x the set of biorthogonal functionals is as follows. We can define a partial sum operators S_n for $n = 1, 2, \dots$ such that

$$S_n: X \rightarrow X, \quad S_n(x) = \sum_{k=1}^n a_k x_k$$

for $x = \sum_{k=1}^{\infty} a_k x_k$, and we have shown that S_n is a bounded linear operator since $\sup_{n \geq 1} \|S_n\| = M < \infty$. Observe that $a_k = a_k(x)$ is a linear functional on X , so the basis expansion of $x \in X$ looks like

$$x = \sum_{k=1}^{\infty} x_k^*(x) x_k.$$

This resembles the Fourier series with respect to orthogonal bases in a Hilbert space, except now we discuss this in general Banach spaces.

Proposition 3.3.1. The biorthogonal functionals $\{x_k^*\}_{k \geq 1}$ of a Schauder basis $\{x_k\}_{k \geq 1}$ are uniformly bounded, i.e.,

$$\sup_{k \in \mathbb{N}} \|x_k^*\| \|x_k\| < \infty.$$

Proof. To do this, we write

$$x_n^*(x) x_n = S_n(x) - S_{n-1}(x)$$

for $n \geq 1$. From Theorem 3.3.2, we have

$$\|x_n^*(x) x_n\| \leq \|S_n(x)\| + \|S_{n-1}(x)\| \leq 2M \|x\|,$$

hence we conclude that $x_n^* \in X^*$ and $\sup_{n \geq 1} \|x_n^*\| \|x_n\| < \infty$. ■

3.4 Compact Sets in Banach Spaces

Compactness is a useful substitute of finite dimensionality as we'll see. Let's give a brief review.

3.4.1 Compactness

We first review some properties of compactness.

Definition 3.4.1 (Compact). A subset A of a topological space is compact if every open cover of A has a finite subcover.

This means, given a cover $A \subseteq \bigcup_{\alpha} U_{\alpha}$ for some collection of open sets U_{α} , then $A \subseteq \bigcup_{k=1}^n U_{\alpha_k}$ for some finite subcollection.

Remark. Properties of compact sets:

- (a) Compact sets of a Hausdorff space are closed (Corollary 6.1.1).
- (b) Closed subsets of compact sets are compact (Corollary 6.1.2).
- (c) The image of a compact set under a continuous function is compact (Theorem 6.1.5).
- (d) Continuous functions on compact sets are uniformly continuous and attain their maximum and minimum.

Definition 3.4.2 (Precompact). A set A is precompact if its closure \overline{A} is compact.

Definition 3.4.3 (ϵ -net). Let A be a subset of a metric space X . Then a subset $\Omega_\epsilon \subseteq X$ is an ϵ -net for A if A can be covered by balls of radius ϵ centered at points of Ω_ϵ , i.e.,

$$A \subseteq \{y : d(y, x) < \epsilon \text{ for some } x \in \Omega_\epsilon\}.$$

Theorem 3.4.1. Let A be a subset of a complete metric space X , the following are equivalent.

- (a) A is precompact.
- (b) Every sequence $\{x_n\}$ in A has a Cauchy subsequence which converges in X .
- (c) For every $\epsilon > 0$, there exists a finite ϵ -net for A .

Theorem 3.4.2 (Heine-Borel theorem). A subset A of a finite dimensional normed space X is precompact if and only if A is bounded.

3.4.2 Compactness in Infinite-Dimensional Normed Spaces

We can extend Heine-Borel theorem to infinite dimensional spaces.

Lemma 3.4.1 (Approximation by finite dimensional subspaces). A subspace A of a normed space X is precompact if and only if A is bounded, and for every $\epsilon > 0$, there exists a finite dimensional subspace Y_ϵ of X containing an ϵ -net for A .

Proof. We first prove the necessity. Let A be precompact and $\epsilon > 0$. Then there exists a finite ϵ -net Ω_ϵ for A . Now, take $Y_\epsilon = \text{span}(\Omega_\epsilon)$, which is finite-dimensional.

As for sufficiency, assume A is bounded, so $A \subseteq rB_X$ for some $r > 0$ where B_X is the unit ball $\{x \in X : \|x\| \leq 1\}$. Also, given ϵ , we have a finite-dimensional subspace Y_ϵ as an ϵ -net of A . Observe that we can restrict to points of Y_ϵ contained in $(r + \epsilon)B_{Y_\epsilon}$ since

$$A \subseteq \{x \in X : d(x, (r + \epsilon)B_{Y_\epsilon}) < \epsilon\},$$

i.e., $(r + \epsilon)B_{Y_\epsilon}$ is also an ϵ -net of A . Since Y is finite-dimensional, from Heine-Borel theorem, $(r + \epsilon)B_{Y_\epsilon}$ is precompact, i.e., we find a precompact ϵ -net of A , therefore A itself is precompact from Theorem 3.4.1. ■

On the other hand, Heine-Borel theorem states that the unit ball B_X of a finite-dimensional normed space X is compact, but this never holds in the infinite-dimensional case.

Theorem 3.4.3 (Riesz's theorem). The unit ball B_X of an infinite dimensional normed space X is never compact.

Proof. Suppose $B_X = \{x \in X : \|x\| \leq 1\}$ is compact. Then from Lemma 3.4.1, we can find a finite dimensional subspace Y containing an ϵ -net with $\epsilon = 1/2$ for B_X , i.e., $d(x, Y) \leq 1/2$ for all $x \in B_X$.

Recall that X is infinite dimensional, Y is finite dimensional, hence the quotient space X/Y is nontrivial. Note that Y is a closed subspace of X ,^a hence, the norm on X induces a norm on X/Y such that $\|[x]\| = \inf_{y \in Y} \|x - y\|$. We can then find an $x \in X$ and $[x] \in X/Y$ such that $\|[x]\| = 0.9$, i.e., there's an $\bar{y} \in Y$ such that $\|x - \bar{y}\| \leq 1$. In this case, $x - \bar{y} \in B_X$ and $d(x - \bar{y}, Y) = \|[x]\| = 0.9 > 1/2$, a contradiction. ■

^aSince it's finite-dimensional.

Lecture 16: Strong Convergence & Weak Topology

Lastly, we see that point-wise convergence of operators implies uniformly convergence on compact sets. 25 Oct. 14:30

Definition. Let X, Y be **normed spaces**, $\{T_n\}_{n \geq 1}$ be a sequence in $\mathcal{L}(X, Y)$ and $T \in \mathcal{L}(X, Y)$.

Definition 3.4.4 (Point-wise convergence). The sequence $\{T_n\}_{n \geq 1}$ *converges point-wise to T* if $\lim_{n \rightarrow \infty} T_n x = Tx$ for all $x \in X$.

Definition 3.4.5 (Uniformly convergence). The sequence $\{T_n\}_{n \geq 1}$ *converges uniformly to T on $A \subseteq X$* if $\lim_{n \rightarrow \infty} \|T_n x - Tx\| = 0$ for all $x \in A$.

Equivalently, $\{T_n\}_{n \geq 1}$ **converges uniformly** to T on A if

$$\lim_{n \rightarrow \infty} \sup_{x \in A} \|T_n x - Tx\| = 0.$$

Theorem 3.4.4 (Convergence on compact set). Let X, Y be **Banach spaces**, and $\{T_n\}_{n \geq 1}, T \in \mathcal{L}(X, Y)$. Suppose the sequence $\{T_n\}_{n \geq 1}$ **converges point-wise** to T , then $\{T_n\}_{n \geq 1}$ **converges uniformly** to T on all **precompact** subsets $A \subseteq X$.

Proof. Since $\{T_n\}$ **converges point-wise**, it's **point-wise bounded**. From **uniform boundedness theorem**, we know that $\sup_{n \geq 1} \|T_n\| < \infty$, i.e., $\exists M$ such that $\|T_n\| \leq M$ for all $n \geq 1$. Let $\epsilon > 0$ and choose a finite **ϵ -net** Ω_ϵ for A . Since Ω_ϵ is finite, there exists N_ϵ such that $\|T_n y - Ty\| \leq \epsilon$ for $n \geq N_\epsilon, y \in \Omega_\epsilon$, i.e., $T_n \rightarrow T$ **uniformly** on Ω_ϵ . Now, for an arbitrarily $x \in A$, there exists $y \in \Omega_\epsilon$ such that $\|x - y\| < \epsilon$, hence

$$\|T_n x - Tx\| \leq \|T_n y - Ty\| + \|(T_n - T)(x - y)\| \leq \epsilon + (\|T_n\| + \|T\|) \|x - y\| \leq \epsilon + 2M\epsilon$$

if $n \geq N_\epsilon$. This implies $\|T_n x - Tx\| \leq (2M + 1)\epsilon$ for $n \geq N_\epsilon$ for all $x \in A$, hence we have **uniform convergence** on A . ■

3.4.3 Compactness Criteria in Various Spaces

There is a useful criterion of **compactness** in spaces with **Schauder basis** (which covers all classical spaces).

Corollary 3.4.1. Let X be a **Banach space** with a **Schauder basis** $\{x_k\}_{k \geq 1}$. A subset $A \subseteq X$ is **precompact** if and only if A is bounded and the basis expansion of vectors $x \in A$ **converges uniformly**, i.e.,

$$\lim_{n \rightarrow \infty} \sup_{x \in A} \|x - S_n x\| = 0.$$

Proof. From **Theorem 3.4.4**, since $\{x_k\}_{k \geq 1}$ is a **Schauder basis**, the projection $S_n \rightarrow I$ **point-wise**, implying **uniform convergence** since A is **precompact**.

Conversely, for any $\epsilon > 0$, there exists n such that $\|x - S_n x\| < \epsilon$ for all $x \in A$, and $\text{Im}(S_n)$ is finite dimensional and $\text{Im}(S_n A)$ is bounded. Hence, there exists an **ϵ -net** Ω_ϵ for $\text{Im}(S_n A)$, so A is covered by a finite **2ϵ -net**, i.e., A is **precompact** from **Lemma 3.4.1**. ■

Finally, we state without proof one of the most important **compactness** criteria in $C([a, b])$. To start with, we introduce the notion of **equicontinuous**, specifically for real-valued function family.

Definition 3.4.6 (Equicontinuous). A real-valued function family \mathcal{F} is *equicontinuous* if for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(y)| \leq \epsilon$$

for all $f \in \mathcal{F}$ if $|x - y| \leq \delta$.

Remark (Uniformly Equicontinuous). **Definition 3.4.6** is often referred to *uniformly equicontinuous*. There's also a point-wise version of which, but we will not bother introducing it here.

Theorem 3.4.5 (Arzelà-Ascoli theorem). A subset $A \subseteq C([a, b])$ is **precompact** if and only if A is bounded and **equicontinuous**.

3.5 Weak Topology

Every **normed space** X is a **metric space** with the **metric** given by $d(x, y) = \|x - y\|$ for $x, y \in X$. This topology on X is called *strong topology*, i.e., a sequence $x_n \rightarrow x$ **converges strongly** in X if $\|x_n - x\| \rightarrow 0$. But actually, in addition to the strong topology, X carries a different topology called **weak topology**, as we're going to study it in this section.

3.5.1 Weak Convergence

Let's first formally introduce the definitions.

Definition. Let $\{x_n\}_{n \geq 1}$ be a sequence in a **normed space** X .

Definition 3.5.1 (Strongly convergence). The sequence $\{x_n\}_{n \geq 1}$ *converges strongly* to $x \in X$ if $\lim_{n \rightarrow \infty} \|x_n - x\| = 0$.

Definition 3.5.2 (Weakly convergence). The sequence $\{x_n\}_{n \geq 1}$ *converges weakly* to $x \in X$ if $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ for all $f \in X^*$.

Notation. If $\{x_n\}_{n \geq 1}$ is **weakly converging** to x , we write $x_n \xrightarrow{w} x$.

Remark (Strong and weak). As we have seen before (**Definition 3.3.4**, **Definition 3.3.3** and **Definition 3.5.1**, **Definition 3.5.2**), the convention is that *strong* is for **norm**, while *weak* is for **functional**.

We see that **strongly convergence** implies **weakly convergence**, while not as before, the converse is often not true. Even with this, there are several useful ties between **weak** and **strong** properties.

Proposition 3.5.1. If the sequence $\{x_n\}_{n \geq 1}$ **converges weakly** to $x \in X$, then we have the following.

- (a) $\sup_{n \geq 1} \|x_n\| < \infty$.
- (b) $\|x\| \leq \liminf_{n \rightarrow \infty} \|x_n\|$.
- (c) $x \in \overline{\text{conv}(\{x_n\}_{n \geq 1})}$.^a

^a $\text{conv}(A)$ is the **convex hull** of A , i.e., the smallest closed **convex set** containing the sequence.

Proof. Let's prove this one by one.

- (a) For $y \in X$, let $y^{**} \in X^{**}$ be from the embedding $X \rightarrow X^{**}$, $y \mapsto y^{**}$ such that $\|y^{**}\| = \|y\|$. Then for $n \geq 1$, $x_n \in X$, so $x_n^{**} \in X^{**}$, we have $\sup_{n \geq 1} |x_n^{**}(f)| < \infty$ since $f(x_n) \rightarrow f(x)$. Then, **uniform boundedness theorem** implies $\sup_{n \geq 1} \|x_n^{**}\| < \infty$. Now, since $\|x_n\| = \|x_n^{**}\|$, we conclude $\sup_{n \geq 1} \|x_n\| < \infty$.
- (b) If $x_n \xrightarrow{w} x$ in X , by **supporting functional theorem**, there exists $f \in X^*$ with $\|f\| = 1$ and $f(x) = \|x\|$. Since $\|f\| = 1$, $f(x_n) \leq \|x_n\|$ for $n \geq 1$. And since $x_n \xrightarrow{w} x$, we have $f(x_n) \rightarrow f(x) = \|x\|$, i.e., $\liminf_{n \rightarrow \infty} \|x_n\| \geq \|x\|$ as desired.
- (c) To show x lies in the closure of the **convex hull** of $\{x_n\}_{n \geq 1}$, denoted it by K , we first note that K is a closed **convex set**. If $x \notin K$, by **separating hyperplane theorem**, there exists $f \in X^*$ such that $\sup_{y \in K} f(y) < f(x)$, and hence $\sup_{n \geq 1} f(x_n) < f(x)$. Since $\lim_{n \rightarrow \infty} f(x_n) = f(x)$,

we have a contradiction. ■

There are some known criteria of [weak convergence](#) in classical [normed spaces](#), one of them is as follows.

Lemma 3.5.1 (Testing weak convergence on a dense set). Let X be a [normed space](#) and $A \subseteq X^*$ be a dense set. Then $x_n \xrightarrow{w} x$ if and only if $\{x_k\}_{k \geq 1}$ is bounded and $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ for every $f \in A$.

Proof. The necessity follows from [Proposition 3.5.1](#). To show the sufficiency, let $g \in X^*$, we need to show that $\lim_{n \rightarrow \infty} g(x_n) = g(x)$. Let $\epsilon > 0$, and A is dense in X^* , so there exists $f \in A$ such that $\|g - f\| < \epsilon$. Then

$$\begin{aligned} \limsup_{n \rightarrow \infty} |g(x_n - x)| &\leq \limsup_{n \rightarrow \infty} |f(x_n - x)| + \limsup_{n \rightarrow \infty} |(g - f)(x_n - x)| \\ &\leq \|g - f\| \limsup_{n \rightarrow \infty} (\|x_n\| + \|x\|) \\ &\leq 2M\epsilon \end{aligned}$$

where $M := \sup_{n \geq 1} \|x_n\| + \|x\| < \infty$. We see that since $\epsilon > 0$ is arbitrary, so $\limsup_{n \rightarrow \infty} |g(x_n - x)| = 0$ hence $x_n \xrightarrow{w} x$. ■

Note. We see that to show [weakly convergence](#), instead of checking for all $f \in X^*$, it's sufficient to check only $f \in A$ for some dense set $A \subseteq X^*$.

3.5.2 Weak Topology

[Weak convergence](#) is just a concept induced from [weak topology](#), so by studying it directly, we can consider other weak properties as we'll soon see.

Definition 3.5.3 (Weak topology). The *weak topology* on a [normed space](#) X is the weakest topology such that all maps $f \in X^*$ are continuous.

Note (Strong topology). To distinguish two natural topologies on X , the [norm](#) topology is sometimes called *strong topology* on X .

Intuitively, if f is continuous at x_0 , the preimage of an ϵ -ball $\{x \in X : |f(x) - f(x_0)| < \epsilon\}$ around $f(x_0)$ is open, i.e., the base of the [weak topology](#) are [cylinders](#) of the form

$$\{x \in X : |f_k(x - x_0)| < \epsilon, f_k \in X^*, \epsilon > 0, k = 1, 2, \dots, N\}$$

around $x_0 \in X$. In all, these cylinders form a local base of [weak topology](#) at point x_0 .

Remark (\mathbb{R}^∞ embedding). [Weak topology](#) is induced from the (infinite) products of reals.

Proof. Consider the embedding from X to an infinite product of \mathbb{R} such that

$$X \rightarrow \mathbb{R}^\infty, \quad x \mapsto (f_k(x))_{f_k \in X^*} = (f_1(x), f_2(x), \dots),$$

where we identify $x \in X$ by its value of uncountably many $f_k \in X^*$, i.e., $f_k(x)$. ⊗

Note (Relation between CW complex). The weak topology in [algebraic topology](#) is actually the strongest one, i.e., it's the [final topology](#). See [here](#) for a further discussion.

Although there are lots of difference between [weak](#) and the corresponding [strong](#) properties, some are equivalent.

Proposition 3.5.2 (Weak closedness). Let K be a **convex subset** of a **Banach space**. Then K is **weakly closed** if and only if K is **strongly closed**.

Proof. Clearly, **weak closure** implies **strong closure**.^a Conversely, if K is a **strongly closed convex set**, it is the intersection of all **hyperplane** containing K from **Corollary 2.4.1**, i.e., $K = \bigcap_{f,a} A_{f,a}$ where $A_{f,a} := \{x \in X : f(x) \leq a, f \in X^*\}$. Since it is **weakly closed** from $A_{f,a} = f^{-1}((-\infty, a])$,^b the intersection of them is also **weakly closed**, proving the result. ■

^aNotice that this case doesn't involve **convexity** since we only use the coarser relation between topologies.

^bSince f is continuous, f^{-1} maps closed set in \mathbb{R} $((-\infty, a])$ to **(weakly) closed set** in X with **weak topology**.

Lecture 17: Weak* Topology

There's another important case regarding $C(K)$.

27 Oct. 14:30

Theorem 3.5.1 (Weak convergence in $C(K)$). Let K be a **compact Hausdorff space** and $C(K)$ be the space of continuous functions $f: K \rightarrow \mathbb{R}$ with $\|f\| = \sup_{t \in K} |f(t)|$. Then a sequence $\{x_n\}_{n \geq 1}$ in $C(K)$ **converges weakly** to $x \in C(K)$ if and only if the sequence $\{x_n\}_{n \geq 1}$ is bounded in $C(K)$ and $\lim_{n \rightarrow \infty} x_n(t) = x(t)$ for all $t \in K$.

Proof. Suppose $\{x_n\}_{n \geq 1}$ **converges weakly** to x , then $\sup_{n \geq 1} \|x_n\| < \infty$. By **Proposition 3.5.2**, if $t \in K$, then $\delta_t \in C(K)^*$ where $\delta_t(x) = x(t)$. We see that

$$\lim_{n \rightarrow \infty} \delta_t(x_n) = \delta_t(x) \Rightarrow \lim_{n \rightarrow \infty} x_n(t) = x(t).$$

Conversely, assume $\sup_{n \geq 1} \|x_n\| < \infty$ and $\lim_{n \rightarrow \infty} x_n(t) = x(t)$ for all $t \in K$. We now need to show that $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ for all $f \in C(K)^*$. Recall that $C(K)^*$ is the space of bounded signed measures μ on K ,^a and $\|\mu\| = TV(\mu)$, we have

$$\lim_{n \rightarrow \infty} \int_K x_n d\mu = \int_K x d\mu,$$

from the **dominated convergence theorem**, proving the result. ■

^aThis is just a variant of **Riesz representation theorem**.

3.6 Weak* Topology

On X^* , there are two natural weaker topologies: one is the **weak topology** which makes all functionals in X^{**} continuous on X^* ; the other topology, called **weak* topology**, is only concerned with the continuity of functionals that come from $X \subseteq X^{**}$.

3.6.1 Weak* Convergence

We again start from convergence, and in this case, the **weak* convergence**.

Definition 3.6.1 (Weak* convergence). Let X be a **normed space**, a sequence $\{f_n\}_{n \geq 1}$ in X^* is **weak* converging** to f in X^* if $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for all $x \in X$.

Notation. If $\{f_n\}_{n \geq 1}$ is **weak* converging** to f , we write $f_n \xrightarrow{w^*} f$.

We see that if we consider the **weak convergence** for $f_k \in X^*$, we should test on all the double **dual** elements $x^{**} \in X^{**}$; but as said, **weak* convergence** only test on the subset $X \subseteq X^{**}$, so **Definition 3.6.1** becomes point-wise convergence.

Remark. The weak and weak* convergence coincides on X^* if X is reflexive, i.e., $X^{**} \equiv X$.

3.6.2 Weak* Topology

Similarly to Definition 3.5.3, we now defined the so-called weak* topology on X^* .

Definition 3.6.2 (Weak* topology). Let X be a normed space. The weak* topology on X^* is defined as the weakest topology in which point evaluation maps $f \mapsto f(x)$ from X^* to \mathbb{R} are continuous for all $x \in X$.

Note (Completeness). If $\{f_n\}_{n \geq 1}$ is weak* Cauchy, then uniform boundedness principle implies that there exists $f \in X^*$ such that $f_n \xrightarrow{w^*} f$.

Equivalently, the base of the weak* topology is given by the cylinders

$$\{f \in X^* : |(f - f_0)(x_k)| < \epsilon, x_k \in X, \epsilon > 0, k = 1, \dots, N\}$$

around $f_0 \in X^*$. So again, these cylinders form a local base of weak* topology at f_0 .

Remark (\mathbb{R}^∞ embedding). Weak* topology is again induced from the (infinite) products of reals.

Proof. Consider the embedding from X^* to an infinite product of \mathbb{R} such that

$$X^* \rightarrow \mathbb{R}^\infty, \quad f \mapsto (f(x_k))_{x_k \in X \subseteq X^*} = (f(x_1), f(x_2), \dots),$$

where we identify $f \in X^*$ by its value on uncountably many $x \in X \subseteq X^*$, i.e., $f(x_k)$. ⊗

As we mentioned before, since $X \subseteq X^{**}$, weak* topology is weaker than the weak topology on X^* . However, for reflexive spaces, these two topologies are of course equivalent.

The main result on weak* topology is Banach-Alaoglu theorem, which allows us to still have a weaker notion of (weak*) compactness of unit balls even though we know that it's not (strongly) compact from Riesz's theorem. The result depends on Tychonoff's theorem.

Theorem 3.6.1 (Tychonoff's theorem). Let $\{X_\gamma\}_{\gamma \in \Gamma}$ be a collection of any number^a of topological spaces X_γ . The Cartesian product $\prod_{\gamma \in \Gamma} X_\gamma$ can be equipped with the product topology whose base is formed by the sets of the form

$$\left\{ \prod_{\gamma \in \Gamma} A_\gamma : A_\gamma \text{ is open in } X_\gamma; \text{ all but finitely many of } A_\gamma \text{ equal } X_\gamma \right\}.$$

Then, if each X_γ is compact, then $\prod_{\gamma \in \Gamma} X_\gamma$ is compact in the product topology.

^aMay be countable or uncountable.

The proof is omitted here, but the upshot is that although the \mathbb{R}^∞ embedding may involve uncountably many \mathbb{R} , Tychonoff's theorem ensures that the compactness is still preserved.

Theorem 3.6.2 (Banach-Alaoglu theorem). Let X be a Banach space with dual X^* , then the unit ball $B_{X^*} = \{f \in X^* : \|f\| \leq 1\}$ in X^* is compact in the weak* topology.

Proof. Since $f \in B_{X^*} \Rightarrow |f(x)| \leq \|x\|$ for $x \in X$, so we can embed B_{X^*} into $\prod_{x \in X} [-\|x\|, \|x\|]$.^a Note that this is the product of compact spaces, so Tychonoff's theorem implies that this product is compact, and we only need to show B_{X^*} is weak* closed in $\prod_{x \in X} [-\|x\|, \|x\|]$. Observe that

$$B_{X^*} = \bigcap_{\substack{x, y \in X \\ a, b \in \mathbb{R}}} B_{x, y, a, b}, \text{ where } B_{x, y, a, b} = \{f \in K : f(ax + by) = af(x) + bf(y)\}$$

for $K := \prod_{x \in X} [-\|x\|, \|x\|]$,^b so it's sufficient to show $B_{x,y,a,b}$ is **weak*** closed in K . But since $B_{x,y,a,b}$ is the preimage of (a closed set) $\{0\}$ under mapping $f \mapsto f(ax + by) - af(x) - bf(y)$, which is continuous from the definition of **weak* topology**, hence we know all $B_{x,y,a,b}$ is (**weak***) closed as well, so is their intersection B_{X^*} . ■

^aRecall the \mathbb{R}^∞ embedding where we identify $f \in X^*$ by $(f(x_k))_{x_k \in X}$.

^bWhat we're claiming is that B_{X^*} consists of linear functions in K .

Let's see some applications of **Banach-Alaoglu's theorem**. Consider spaces $L^p(\mu)$ with $1 < p < \infty$, we know that these are **reflexive**, then the unit ball in $L^p(\mu)$ is **compact** in the **weak topology**. Namely, let $f_n \in L^p(\mu)$ for $n \geq 1$ be bounded, then there exists a subsequence f_{n_k} for $k \geq 1$ and $f \in L^p(\mu)$ such that

$$\lim_{k \rightarrow \infty} \int f_{n_k} g \, d\mu = \int f g \, d\mu$$

for all $g \in L^q(\mu)$ with $1/p + 1/q = 1$.

Another example is that let the **Banach space** be $C(K)$, the **dual** $C(K)^*$ is the space of finite signed measure μ on K with $TV(\mu) < \infty$. Let μ_n be a sequence of measures on K such that $\sup_{n \geq 1} TV(\mu_n) < \infty$, then $\{\mu_n\}_{n \geq 1}$ is bounded in $C(K)^*$. We see that there exists a subsequence μ_{n_k} for $k \geq 1$ such that $\mu_{n_k} \xrightarrow{w^*} \mu \in C(K)^*$, i.e.,

$$\lim_{n \rightarrow \infty} \int_K f \, d\mu_{n_k} = \int_K f \, d\mu$$

for all $f \in C(K)$.

Note. This is generally referred to as *weak convergence of measures*.

Chapter 4

Compact Operators and Spectral Theory

4.1 Compact Operators

Compact operators form an important class of bounded linear operators. On the one hand, they are almost finite rank operators,¹ so they share some properties of finite rank operators, which facilitates their study. On the other hand, the class of compact operators is wide enough to include integral and Hilbert-Schmidt operators, which are important in many applications.

Definition 4.1.1 (Compact operator). Let X, Y be Banach spaces, we say a bounded linear operator $T: X \rightarrow Y$ is compact if it maps bounded sets in X to precompact sets in Y .

Notation. The set of compact operators $T: X \rightarrow Y$ is denoted as $\mathcal{K}(X, Y)$.

In other words, the closure of $T(B_X)$ is compact for T being compact. Notice that since precompact sets are bounded, compact operators are always bounded, i.e., $\mathcal{K}(X, Y) \subseteq \mathcal{L}(X, Y)$. So indeed, we may remove the bounded condition in Definition 4.1.1.

Remark (Finite rank operator). An operator T is with finite rank if $\dim \operatorname{Im} T < \infty$. And we see that every finite rank operator $T \in \mathcal{L}(X, Y)$ is compact.

Proof. Since $T(B_X)$ is a bounded subset of a finite dimensional normed space $\operatorname{Im} T \subseteq Y$, so $T(B_X)$ is precompact by Heine-Borel theorem. *

4.1.1 Properties of Compact Operators

Let's study some basic properties of compact operators.

Proposition 4.1.1 (Properties of $\mathcal{K}(X, Y)$). Let X, Y be Banach spaces.

- (a) $\mathcal{K}(X, Y)$ is a closed linear subspace of $\mathcal{L}(X, Y)$.
- (b) If $T \in \mathcal{K}(X, Y)$ and S is bounded linear, then both TS and ST are compact.^a

^a S needs to have proper domain and range, i.e., TS is compact if $S: Z \rightarrow X$; ST is compact if $S: Y \rightarrow Z$.

Proof. Let's prove this one by one.

- (a) Linearity follows since the sum of two precompact sets is precompact. For closedness, let $T_n \in \mathcal{K}(X, Y)$, $T \in \mathcal{L}(X, Y)$ with $\lim_{n \rightarrow \infty} \|T_n - T\| = 0$, we need to show T is compact, which can be done by finding a finite ϵ -net of $T(B_X)$.

¹In the same way as compact sets are almost finite dimensional.

Given any $\epsilon > 0$, choose N such that $\|T_N - T\| < \epsilon/2$, i.e., $\|T_N x - Tx\| < \epsilon/2$ for every $x \in B_X$. This means $T_N(B_X)$ is an $\epsilon/2$ -net of $T(B_X)$. Now, since $T_N(B_X)$ is itself precompact, we can find another finite $\epsilon/2$ -net $\Omega_{\epsilon/2}$ of $T_N(B_X)$, hence $\Omega_{\epsilon/2}$ is a finite ϵ -net of $T(B_X)$.

- (b) Consider TS first. Since S maps unit balls to bounded sets, and T maps bounded sets to precompact sets, hence TS is compact. As for ST , since T maps unit balls to precompact sets, and because S is continuous, S maps precompact sets to precompact sets.

■

Remark (Operator ideal). If $\mathcal{K}(X, Y)$ satisfies the second property in Proposition 4.1.1, we call it an operator ideal.

Lecture 18: Hilbert-Schmidt Operators

1 Nov. 14:30

Corollary 4.1.1 (Isomorphisms are not compact). Let X be an infinite dimension normed space, then the identity map $I: X \rightarrow X$ is not compact. More generally, an invertible operator (i.e., isomorphism) $T \in \mathcal{L}(X, Y)$ is not compact.

Proof. I is not compact since Riesz theorem implies that B_X is not compact. And in general, if $T \in \mathcal{L}(X, Y)$ is compact and invertible, Proposition 4.1.1 implies that $T^{-1}T = I: X \rightarrow X$ is compact since T^{-1} is bounded and T is compact, contradiction. ■

Finally, an important observation is that since we know $\mathcal{K}(X, Y)$ is closed from Proposition 4.1.1, and any finite rank operators are compact, hence for any operator that can be approximated by finite rank operators is also compact, leading to the following.

Corollary 4.1.2 (Almost finite rank operators are compact). Suppose a linear operator $T: X \rightarrow Y$ can be approximated by finite rank operators $T_n \in \mathcal{L}(X, Y)$ as $\lim_{n \rightarrow \infty} \|T_n - T\| = 0$, then T is compact.

4.1.2 Hilbert-Schmidt Operators

This is a most frequently used class of compact operators in Hilbert spaces. As we will see, it covers the class of important operators, i.e., integral operators, motivating us to study this class.

Definition 4.1.2 (Hilbert-Schmidt operators). Let \mathcal{H} be a Hilbert space which is separable and $\{x_k\}_{k \geq 1}$ is an orthonormal basis in \mathcal{H} . A linear operator $T: \mathcal{H} \rightarrow \mathcal{H}$ is Hilbert-Schmidt if

$$\sum_{k=1}^{\infty} \|Tx_k\|^2 < \infty.$$

From Definition 4.1.2, the following naturally induced norm is defined.

Definition 4.1.3 (Hilbert-Schmidt norm). The Hilbert-Schmidt norm of a linear operator T is

$$\|T\|_{\text{HS}} = \left(\sum_{k=1}^{\infty} \|Tx_k\|^2 \right)^{1/2}.$$

To characterize Hilbert-Schmidt operators, we have the following.

Proposition 4.1.2. Let $T: \mathcal{H} \rightarrow \mathcal{H}$ be a linear operator on the separable Hilbert space \mathcal{H} with an orthonormal basis $\{x_k\}_{k \geq 1}$.

- (a) The **Hilbert-Schmidt norm** of T is independent of the choice of **orthonormal basis**.
- (b) If T is **Hilbert-Schmidt**, then T^* is **Hilbert-Schmidt** and $\|T\|_{\text{HS}} = \|T^*\|_{\text{HS}}$.
- (c) If T is **Hilbert-Schmidt**, then T is **bounded** on \mathcal{H} and $\|T\| \leq \|T\|_{\text{HS}}$.
- (d) If T is **Hilbert-Schmidt**, then T is **compact**.

Proof. We prove this in the following order.

- (c) Let $x \in \mathcal{H}$ such that $x = \sum_{k=1}^{\infty} a_k x_k$, $a_k \in \mathbb{C}$. From the **Parseval identity**, $\|x\|^2 = \sum_{k=1}^{\infty} |a_k|^2$ where $a_k = \langle x, x_k \rangle$, then

$$\|Tx\| = \left\| \sum_{k=1}^{\infty} a_k T x_k \right\| \leq \sum_{k=1}^{\infty} |a_k| \|T x_k\| \leq \left(\sum_{k=1}^{\infty} |a_k|^2 \right)^{1/2} \left(\sum_{k=1}^{\infty} \|T x_k\|^2 \right)^{1/2} = \|x\| \|T\|_{\text{HS}},$$

so T is **bounded** with $\|T\| \leq \|T\|_{\text{HS}}$.^a

- (b) Since

$$\sum_{k=1}^{\infty} \|T x_k\|^2 = \sum_{k,j} |\langle T x_k, x_j \rangle|^2 = \sum_{k,j} |\langle x_k, T^* x_j \rangle|^2 = \sum_j \|T^* x_j\|^2 = \|T^*\|_{\text{HS}}^2,$$

hence $\|T\|_{\text{HS}} = \|T^*\|_{\text{HS}}$.

- (a) Let $\{x'_k\}_{k \geq 1}$ be another **separate orthonormal basis**, then from the **Parseval identity** and $\|T\|_{\text{HS}} = \|T^*\|_{\text{HS}}$,

$$\|T\|_{\text{HS}} = \|T^*\|_{\text{HS}} = \sum_j \|T^* x_j\|^2 = \sum_{j,k} |\langle x'_k, T^* x_j \rangle|^2 = \sum_{j,k} |\langle T x'_k, x_j \rangle|^2 = \sum_k \|T x'_k\|^2,$$

i.e., the **Hilbert-Schmidt norm** is independent of the choice of basis.

- (d) From **Corollary 4.1.2**, we show that T is a limit of **finite rank operators**. Define for $N \geq 1$, T_N by $T_N x_k = T x_k$ for $k = 1, 2, \dots, N$, $T_N x_k = 0$ for $k > N$, hence T_N is **finite rank**. We then have

$$\|T - T_N\|_{\text{HS}}^2 = \sum_{k=N+1}^{\infty} \|T x_k\|^2 \Rightarrow \lim_{N \rightarrow \infty} \|T - T_N\|_{\text{HS}} = 0.$$

Since $\|\cdot\| \leq \|\cdot\|_{\text{HS}}$, $\lim_{N \rightarrow \infty} \|T - T_N\| = 0$ as desired.

■

^aThis implies that T^* is well-defined.

One important motivation of studying **compact operators** is because integral operators are **compact**, and furthermore, is **Hilbert-Schmidt**.

Proposition 4.1.3 (Hilbert-Schmidt integral operator). Let $k \in L^2([0, 1] \times [0, 1])$ such that

$$\int_0^1 \int_0^1 |k(t, s)|^2 dt ds < \infty,$$

and define an integral operator $T: L^2([0, 1]) \rightarrow L^2([0, 1])$ by

$$Tf(t) = \int_0^1 k(t, s) f(s) ds$$

for $0 < t < 1$ and $f \in L^2([0, 1]) = \mathcal{H}$. Then T is **Hilbert-Schmidt** on \mathcal{H} and $\|T\|_{\text{HS}} = \|k\|_{L^2([0, 1]^2)}$.

Proof. Let $K_t(s) := k(t, s)$ for $0 < s, t < 1$, then $Tf(t) = \langle K_t, f \rangle$. Note that $\|Tf(t)\| \leq \|K_t\|_2 \|f\|_2$ from [Cauchy-Schwarz](#), and also, notice that

$$\int_0^1 \|K_t\|_2^2 dt = \|k\|_2^2 \Rightarrow \|K_t\|_2 < \infty \text{ a.e. } t \in [0, 1].$$

Hence, $Tf(t)$ is defined for a.e. t . Let $\{x_k\}_{k \geq 1}$ be an [orthonormal basis](#) for $\mathcal{H} = L^2([0, 1])$,

$$\|T\|_{\text{HS}}^2 = \sum_{k=1}^{\infty} \|Tx_k\|_2^2 = \sum_{i=1}^{\infty} \int_0^1 |Tx_k(t)|^2 dt = \sum_{k=1}^{\infty} \int_0^1 |\langle K_t, x_t \rangle|^2 dt = \int_0^1 \sum_{k=1}^{\infty} |\langle K_t, x_t \rangle|^2 dt,$$

where the last equality follows from the [monotone convergence theorem](#). Further, from [Parseval](#),

$$\|T\|_{\text{HS}}^2 = \int_0^1 \sum_{k=1}^{\infty} |\langle K_t, x_t \rangle|^2 dt = \int_0^1 \|K_t\|_2^2 dt = \|k\|_2^2$$

by the definition of K_t and [Fubini's theorem](#) ■

4.1.3 Compactness of the Adjoint Operators

Recall the basic [duality](#) property for [bounded linear operators](#).

As previously seen. If $T \in \mathcal{L}(X, Y)$, then $T^* \in \mathcal{L}(Y^*, X^*)$ and $\|T^*\| = \|T\|$.

Indeed, a similar [duality](#) principle holds for [compact operators](#) as guaranteed by [Schauder's theorem](#).

Theorem 4.1.1 (Schauder's theorem). Let X, Y be [Banach spaces](#), then if $T \in \mathcal{K}(X, Y)$, $T^* \in \mathcal{K}(Y^*, X^*)$.

Proof. Without loss of generality, we prove that $T^*B_{Y^*}$ is [precompact](#) in X^* given $T \in \mathcal{K}(X, Y)$, i.e., consider $f_n \in B_{Y^*}$ for $n \geq 1$ such that $\{T^*f_n\}_{n \geq 1}$ has a convergent subsequence.

Claim. B_{Y^*} is [precompact](#) in Y^* , i.e., there is a convergent subsequence $\{f_{n_k}\}_{k \geq 1}$ of $\{f_n\}_{n \geq 1}$.

Proof. This can be done by embedding B_{Y^*} into $C(K)$ with $K := \overline{TB_X}$ by considering $f_n|_K \in C(K)$. Observe that $\{f_n|_K\}_{n \geq 1}$ is bounded and [equicontinuity](#): First, $\{f_n|_K\}_{n \geq 1}$ is bounded since $f_n \in B_{Y^*}$ implying $\|f_n\|_{Y^*} \leq 1$ for all $n \geq 1$, and so is $f_n|_K$; while for [equicontinuity](#), we have

$$|f_n|_K(y) - f_n|_K(y')| = |f(y_1 - y_2)| \leq \|y - y'\| \leq \|f_n\| \|y - y'\|$$

for all $y, y' \in Y$. In all, since K is [compact](#), with [Arzelà-Ascoli theorem](#), $\{f_n|_K\}_{n \geq 1}$ has a convergent subsequence $\{f_{n_k}|_K\}_{k \geq 1}$ in $C(K)$.^a ⊗

^aSame as saying B_{Y^*} is [precompact](#) as the original statement in [Arzelà-Ascoli theorem](#).

Hence, to show that $\{T^*f_n\}_{n \geq 1}$ has a convergent subsequence, consider $\{T^*f_{n_k}\}_{k \geq 1}$, we have^b

$$\|T^*f_{n_i} - T^*f_{n_j}\|_{X^*} = \sup_{x \in B_X} |f_{n_i}(Tx) - f_{n_j}(Tx)| = \sup_{y \in K} |f_{n_i}|_K(y) - f_{n_j}|_K(y)| \rightarrow 0$$

as $n_i, n_j \rightarrow \infty$ since f_{n_k} converges, i.e., $\{T^*f_{n_k}\}_{k \geq 1}$ is Cauchy in X^* hence converges. ■

^bThe last equality follows from the fact that TB_X is dense in K (i.e., $\overline{TB_X} = K$).

Lecture 19: Fredholm Alternative

4.2 Fredholm Theory

Fredholm theory studies operators of the form *identity plus compact*. They are conveniently put in the form $I - T$ for I being the identity operator on some Banach space X and $T \in \mathcal{K}(X, X)$.

Furthermore, Fredholm theory is motivated by two applications. One is for solving linear equations $\lambda x - Tx = b$, and in particular, integral equations.² Another is in spectral theory, where the spectrum of T consists of numbers λ for which the operator $\lambda I - T$ is not invertible.

4.2.1 Closed Image

Firstly, we see that the image of operators in the form $I - T$ has a closed image.

Lemma 4.2.1. Let X be a Banach space and $T \in \mathcal{K}(X, X)$, then the operator $I - T$ has closed image $\text{Im}(I - T)$.

Proof. Let $A = I - T$, since $\ker(A)$ is closed, we consider the injectivization $\tilde{A}: X / \ker(A) \rightarrow X$ be the induced operator $\tilde{A}([x]) = Ax$ for all $x \in X$. Since $\text{Im}(A) = \text{Im}(\tilde{A})$, it's sufficient to show $\text{Im}(\tilde{A})$ is closed, which is equivalent to show that \tilde{A} is bounded below from Proposition 3.1.2, i.e., $\exists c > 0$ such that

$$\|\tilde{A}[x]\| \geq c\|[x]\|$$

for all $x \in X$. Toward a contradiction, suppose \tilde{A} is not bounded below, then there exists a sequence $\{[x_k]\}_{k \geq 1}$ in $X / \ker(A)$ such that $\|[x_k]\| = 1$ for all $k \geq 1$ and $\|\tilde{A}[x_k]\| \rightarrow 0$ as $k \rightarrow \infty$, or $\tilde{A}([x_k]) \rightarrow 0$.

Now, choose $\{x_k \in [x_k]\}_{k \geq 1}$ such that $\|x_k\| \leq 2$, with $\inf_{y \in \ker(A)} \|x_k - y\| = 1$. From the fact that T is compact, $\{Tx_k\}_{k \geq 1}$ has a converging subsequence, so we can assume $Tx_k \rightarrow z$ as $k \rightarrow \infty$, with $x_k - Tx_k \rightarrow 0$, we have $x_k \rightarrow z$ as $k \rightarrow \infty$ and $Az = 0$, i.e., $z \in \ker(A)$. But from $\inf_{y \in \ker(A)} \|z - y\| = 1$, a contradiction since this distance should be 0 if $z \in \ker(A)$. ■

4.2.2 Fredholm Alternative

We now study a partial case of the so-called Fredholm alternative.

Theorem 4.2.1 (Fredholm alternative). Let X be a Banach space, $T: X \rightarrow X$ be compact. Then $I - T$ is injective if and only if $I - T$ is surjective.

Proof. Assume that $A := I - T$ is injective but not surjective, to have a contradiction, we just need to find a sequence $\{f_n\}_{n \geq 1}$ in X^* with $\|f_n\| = 1$ for all $n \geq 1$ such that the sequence $\{T^*f_n\}_{n \geq 1}$ has no converging subsequence, since it will contradict the fact that $T^*(B_{X^*})$ is precompact.

Claim. Let $Y_n := \text{Im}(A^n)$ for $n \geq 1$, then $Y_{n+1} \subsetneq Y_n$ for all n .

Proof. Since A is not surjective, $\text{Im}(A) \neq X$, i.e., $Y_1 \subsetneq Y_0 = X$. Consider $y \notin \text{Im}(A)$, and suppose $\text{Im}(A^{n+1}) = \text{Im}(A^n)$, then there exists x such that $A^{n+1}x = A^n y$, i.e., $A^n(Ax - y) = 0$. From the injectivity of A , $\ker(A^n) = 0$, so $Ax - y = 0$, implying $y \in \text{Im}(A)$. So Y_{n+1} is properly contained in Y_n for all $n \geq 1$. ⊗

Claim. Y_n are all closed.

Proof. Since $Y_n = \text{Im}(A^n) = \text{Im}((I - T)^n)$, where $(I - T)^n = I - ST = I - \tilde{T}$ for S being bounded and T being compact, hence $\tilde{T} = ST$ is compact from Proposition 4.1.1. The result follows from Lemma 4.2.1. ⊗

Now, since Y_n / Y_{n+1} is a Banach space, let $\tilde{f}_n: Y_n / Y_{n+1} \rightarrow \mathbb{R}$ be a bounded linear functional with $\|\tilde{f}_n\| = 1$, which can be found by the supporting functional theorem. Define $f_n: Y_n \rightarrow \mathbb{R}$ by $f_n(y) = \tilde{f}_n([y])$ for $y \in Y_n$ with $f_n(y) = 0$ when $y \in Y_{n+1}$, implying that $f_n \in Y_{n+1}^\perp$ from the Riesz

² T being an integral operator.

representation theorem. Finally, we extend f_n to $f_n: X \rightarrow \mathbb{R}$ with $\|f_n\| = 1$ by [Hahn Banach theorem](#) to avoid any domain issue.

We now show the sequence $\{T^*f_n\}_{n \geq 1}$ has no converging subsequence by shown that for $n > m \geq 1$, T^*f_n and T^*f_m are pairwise separated.

Claim. For $n > m \geq 1$, $(T^*f_n - T^*f_m)(x) = f_n(x)$.

Proof. We have

$$T^*f_n - T^*f_m = T^*(f_n - f_m) = (I - T^*)(f_m - f_n) + (f_n - f_m)$$

where $f_n \in Y_{n+1}^\perp$ and $f_m \in Y_{m+1}^\perp \subseteq Y_{n+1}^\perp$, so $f_n - f_m \in Y_{n+1}^\perp$. Now, observe that $(I - T)x = Ax \in Y_{n+1}$, hence $(I - T^*)(f_n - f_m) = (f_n - f_m)(I - T) = 0$. In all, we have that for $x \in Y_n$, $T^*(f_n - f_m)(x) = (f_n - f_m)(x) = f_n(x)$ from $m < n$ ⊗

This implies that

$$\|T^*f_n - T^*f_m\| = \sup_{\|x\|=1} \|(T^*f_n - T^*f_m)(x)\| = \sup_{\|x\|=1} \|f_n(x)\| = \|f_n\| = 1,$$

i.e., all terms in the sequence $\{T^*f_n\}$ are pairwise separated as desired.

Conversely, assume $I - T$ is surjective, we want to prove that $I - T$ is injective. Again, let $A := I - T$, we have $\text{Im}(A) = X$ and $\ker(A^*) = (\text{Im } A)^\perp$, implying $\ker(A^*) = \{0\}$, i.e., A^* is injective. Since $A^* = I - T^*$, where T^* is [compact](#), the previous result implies A^* is surjective. Note that $(\ker A)^\perp \supseteq (\text{Im } A^*) = X^*$, hence $\ker(A) = \{0\}$, so A is injective. ■

We see that the [spectrum](#) in infinite dimension is much more complicated compared to the finite dimension case.

Lecture 20: Spectrum Theory

[Fredholm alternative](#) does not hold if T is not [compact](#), see the following example.

8 Nov. 14:30

Example (Shift operator). Consider the shift operators on $\ell_2 = \{(a_1, a_2, \dots), a_j \in \mathbb{R}, \sum_{j=1}^\infty a_j^2 < \infty\}$. The right shift T_r is defined as

$$T_r(a_1, a_2, \dots) = (0, a_1, a_2, \dots),$$

which is injective but not surjective; while the left shift T_ℓ is defined as

$$T_\ell(a_1, a_2, \dots) = (a_2, a_3, \dots),$$

which is surjective but not injective.

The name [Fredholm alternative](#) is explained by the following application to solving linear equations of the form

$$\lambda x - Tx = b$$

where $T \in \mathcal{K}(X, X)$, $\lambda \in \mathbb{C}$, $b \in X$. One is interested in existence and uniqueness of solution. [Fredholm alternative](#) states that exactly one of the following statements holds for every $\lambda \neq 0$:

- either the homogeneous equation $\lambda x - Tx = 0$ has a nontrivial solution,
- or the inhomogeneous equation $\lambda x - Tx = b$ has a solution for every b , where this solution is automatically unique.

Also, this [alternative](#) is particularly useful for studying integral equations, since for the integral operator

$$(Tf)(t) = \int_0^1 k(t, x)f(s) \, ds,$$

the homogeneous Fredholm equation is

$$\lambda f(t) - \int_0^1 k(t, s) f(s) \, ds = 0,$$

while the inhomogeneous Fredholm equation (*of second kind*) is

$$\lambda f(t) - \int_0^1 k(t, s) f(s) \, ds = b(t).$$

4.3 Spectrum of Bounded Linear Operators

Studying **linear operators** through their spectral properties is a powerful approach in analysis and mathematical physics. Recall from linear algebra that the **spectrum** of a **linear operator** T acting on \mathbb{C}^n consists of the *eigenvalues* of T , which are the numbers $\lambda \in \mathbb{C}$ such that $Tx = \lambda x$ for some nonzero vector $x \in \mathbb{C}^n$; such x are called the *eigenvectors* of T .

There are at most n eigenvalues of T , or one can say exactly n counting multiplicities.³ Eigenvectors corresponding to different eigenvalues are linearly independent. However, the eigenvalues do not need to form a basis of \mathbb{C}^n . The dimension of the span of eigenvectors corresponding to a given eigenvalue (the eigenspace) may be strictly less than the multiplicity of that root. This happens, for example, for the Jordan block

$$T = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

An **orthonormal basis** of eigenvectors exists if and only if T is **normal**, i.e. $T^*T = TT^*$.

To start our formal discussion, we first formally define the notion of **spectrum points**.

Definition. Let X be a complex **Banach space**, and $T: X \rightarrow X$ be a **bounded linear operator**.

Definition 4.3.1 (Regular point). A number $\lambda \in \mathbb{C}$ is called a *regular point* of T if the operator $T - \lambda I$ is invertible, i.e., $(T - \lambda I)^{-1} \in \mathcal{L}(X, X)$.

Definition 4.3.2 (Spectrum point). A number $\lambda \in \mathbb{C}$ is called a *spectrum point* of T if it's not a **regular point**.

Notation. The set of **regular points** for T is denoted as $\rho(T)$, while the set of **spectrum points** is denoted as $\sigma(T)$.

Remark (Resolvent point). We sometimes call a **regular point** as a *resolvent point*.

From definitions, we know that $\sigma(T) = \mathbb{C} - \rho(T)$.

4.3.1 Classification of Spectrum

For operators T acting on a finite dimensional space, the **spectrum** consists of eigenvalues of T ; in infinite dimensions, this is not true, as there are various reason why $T - \lambda I$ may be non-invertible.

Let's first see some examples illustrating the fact that the **spectrum** is a richer concept in infinite-dimensional spaces than in finite-dimensional spaces.

Example (Uncountable number of eigenvalues). Consider the differential operator $T = d/dt$ acting on $C^1(\mathbb{C})$. Suppose λ is an eigenvalue of T with eigenvector $u \in C^1(\mathbb{C})$. This implies that the ordinary differential equation $u' = \lambda u$ holds. The solution has the form $u(t) = Ce^{\lambda t}$, so every $\lambda \in \mathbb{C}$ is an eigenvalue of T .

³Eigenvalues always exist by the **fundamental theorem of algebra**, as they are the roots of the characteristic polynomial $\det(T - \lambda I) = 0$.

Example (No eigenvalues). Consider a multiplication operator on $L^2([0, 1])$ acting as

$$(Tf)(t) = tf(t).$$

Suppose λ is an eigenvalue of T with eigenvector $f \in L^2([0, 1])$. This implies

$$tf(t) = \lambda f(t)$$

for all $t \in [0, 1]$. But this means $f = 0$, so T has no eigenvalues.

These reasons cause the following different types of [spectrum](#).

Definition (Classification of spectrum). Let T be a [bounded linear operator](#) on X .

Definition 4.3.3 (Point spectrum). The *point spectrum* $\sigma_p(T)$ contains λ such that $\ker(T - \lambda I) \neq \{0\}$, i.e., $T - \lambda I$ is not injective, i.e., $\lambda \in \sigma_p(T)$ if λ is an eigenvalue of T .

Definition 4.3.4 (Continuous spectrum). The *continuous spectrum* $\sigma_c(T)$ contains λ such that $\ker(T - \lambda I) = \{0\}$ ($\lambda \notin \sigma_p(T)$) and $\text{Im}(T - \lambda I)$ is dense in X , i.e., $\lambda \in \sigma_c(T)$ if λ is not an eigenvalue and $\text{Im}(T - \lambda I) \neq X$ but $\overline{\text{Im}(T - \lambda I)} = X$.^a

^aNote that by [inverse mapping theorem](#), if $\text{Im}(T - \lambda I) = X$, then $\lambda \in \rho(T)$.

Definition 4.3.5 (Residual spectrum). The *residual spectrum* $\sigma_r(T)$ is defined as $\sigma_r(T) := \sigma(T) - \sigma_p(T) - \sigma_c(T)$, i.e., $\lambda \in \sigma_r(T)$ if λ is not an eigenvalue and $\overline{\text{Im}(T - \lambda I)} \neq X$.

From the above, we have

$$\sigma(T) = \sigma_p(T) \sqcup \sigma_c(T) \sqcup \sigma_r(T).$$

Let's now compute and classify the [spectrum](#) of some basic [linear operators](#).

Example (Diagonal operator on ℓ_2). Let $\{\lambda_k\}_{k \geq 1}$ be a sequence in $\mathbb{C} \setminus \{0\}$ such that $\lim_{k \rightarrow \infty} \lambda_k = 0$. Define $T: \ell_2 \rightarrow \ell_2$ by

$$T(\{x_k\}_{k \geq 1}) = \{\lambda_k x_k\}_{k \geq 1}$$

where the sequence $\{\lambda_k\}_{k \geq 1}$ is bounded, hence T is a [bounded linear operator](#). Then, $(T - \lambda I)x = \{(\lambda_k - \lambda)x_k\}_{k \geq 1}$, so given $y = \{y_k\}_{k \geq 1}$,

$$(T - \lambda I)^{-1}y = \left\{ \frac{y_k}{\lambda_k - \lambda} \right\}_{k \geq 1},$$

which implies $(T - \lambda I)^{-1}$ is [bounded](#) on ℓ_2 if $\sup_{k \geq 1} |\lambda_k - \lambda|^{-1} < \infty$. Indeed, since $\lim_{k \rightarrow \infty} \lambda_k = 0$, $\sup_{k \geq 1} |\lambda_k - \lambda|^{-1} < \infty$ if $\lambda \notin \{\lambda_k\}_{k \geq 1} \cup \{0\}$. Further,

- if $\lambda = \lambda_k$ then $\ker(T - \lambda I) \neq \{0\}$;
- if $\lambda = 0$, then $\text{Im}(T - \lambda I)$ is dense in ℓ_2 , and notice that $\ker(T - \lambda I) = \{0\}$,^a so 0 is in the [continuous spectrum](#).

Hence, $\sigma(T) = \{\lambda_k\}_{k \geq 1} \cup \{0\}$ with $\sigma_p(T) = \{\lambda_k\}_{k \geq 1}$, $\sigma_c(T) = \{0\}$ with $\sigma_r(T) = \emptyset$.

^aNotice that $\lambda_k \in \mathbb{C} \setminus \{0\}$.

Let's revisit previous example on the multiplication operator on L^2 .

Example (Multiplication operator on L^2). Consider the multiplication on $L^2([0, 1])$ such that $Tf(t) =$

$tf(t)$ for $0 < t < 1$. Then

$$(T - \lambda I)f(t) = (t - \lambda)f(t) \Rightarrow (T - \lambda I)^{-1}g(t) = \frac{g(t)}{t - \lambda}$$

for $0 < t < 1$. Hence, $T - \lambda I$ is invertible if $\lambda \notin [0, 1]$, so $\sigma(T) = [0, 1]$. And if $\lambda \in [0, 1]$, then $\ker(T - \lambda I) = 0$, i.e., $\sigma_p(T) = \emptyset$. Lastly, since $\text{Im}(T - \lambda I)$ is dense in $L^2([0, 1])$ if $\lambda \in [0, 1]$, hence $\sigma_c(T) = [0, 1]$, so $\sigma_r(T) = \emptyset$.

4.3.2 Properties of Spectrum

In this section, let X denotes a Banach space and $T \in \mathcal{L}(X, X)$. We'll see that studying the spectrum of T is convenient via the so-called resolvent operator.

Definition 4.3.6 (Resolvent operator). To each regular point $\lambda \in \rho(T)$ for $T \in \mathcal{L}(X, X)$, the associated resolvent operator $R(\lambda) = (T - \lambda I)^{-1}$ is defined by $R: \rho(T) \rightarrow \mathcal{L}(X, X)$.^a

^aThis is the so-called resolvent function.

The resolvent operator can be computed in terms of series expansion involving T . This technique is based on the following simple lemma.

Lemma 4.3.1 (Von Neumann). Let $S \in \mathcal{L}(X, X)$ such that $\|S\| < 1$, then $I - S$ is invertible and $(I - S)^{-1}$ is given by the geometric series

$$(I - S)^{-1} = \sum_{k=0}^{\infty} S^k \text{ and } \|(I - S)^{-1}\| \leq \frac{1}{1 - \|S\|}.$$

Proof. From the inequality $\|S^k\| \leq \|S\|^k$ for all k , $\sum_{k=0}^{\infty} S^k$ converges absolutely. Hence,

$$(I - S) \sum_{k=0}^{\infty} S^k = \sum_{k=0}^{\infty} S^k (I - S) = I$$

by telescoping series. Finally,

$$\|(I - S)^{-1}\| \leq \sum_{k=0}^{\infty} \|S\|^k \leq \frac{1}{1 - \|S\|}.$$

■

Proposition 4.3.1. The resolvent set $\rho(T) \subseteq \mathbb{C}$ is open and containing $\{\lambda \in \mathbb{C}: |\lambda| > \|T\|\}$, with $\|R(\lambda)\| \leq 1/(|\lambda| - \|T\|)$.

Proof. Since

$$(T - \lambda I)^{-1} = -\lambda^{-1}(I - \lambda^{-1}T)^{-1} = -\lambda^{-1}(I - S),$$

by letting $S = T/\lambda$, we have $\|S\| < 1$ if $|\lambda| > \|T\|$, hence by Lemma 4.3.1,

$$\|(T - \lambda I)^{-1}\| \leq \frac{1}{|\lambda|} \frac{1}{1 - \|S\|} = \frac{1}{|\lambda|} \frac{1}{1 - |\lambda|^{-1}\|T\|} = \frac{1}{|\lambda| - \|T\|},$$

i.e., if $|\lambda| > \|T\|$, then $\lambda \in \rho(T)$. To show $\rho(T)$ is open, since

$$\frac{1}{x - \lambda} - \frac{1}{x - \mu} = \frac{\lambda - \mu}{(x - \lambda)(x - \mu)}$$

for all $\mu, \lambda \in \mathbb{C}$ and $x \in \mathbb{C}$, we can generalize this to^a

$$R(\lambda) - R(\mu) = (\lambda - \mu)R(\lambda)R(\mu)$$

since $R(\lambda), R(\mu)$ commutes. Hence,

$$R(\mu) = [I - (\mu - \lambda)R(\lambda)]^{-1} R(\lambda) = (I - S)^{-1} R(\lambda),$$

so $R(\mu)$ is **bounded** if $\|S\| < 1$, i.e., $|\mu - \lambda| \|R(\lambda)\| < 1$. We see that $\lambda \in \rho(T)$ implies that the disk $D(\lambda, r) \subseteq \rho(T)$ if $r \|R(\lambda)\| < 1$, i.e., $\rho(T)$ is open. ■

^aThis is known as the resolvent identity.

Proposition 4.3.1 implies the following.

Proposition 4.3.2. The **spectrum** $\sigma(T)$ is a closed bounded set with

$$\sigma(T) \subseteq \{\lambda \in \mathbb{C} \mid |\lambda| \leq \|T\|\}.$$

Proof. From **Proposition 4.3.1**, we have shown that $\sigma(T)$ is a closed set since $\rho(T)$ is open and $\mathbb{C} \setminus \rho(T) = \sigma(T)$, together with $\sigma(T)$ being bounded such that $\sigma(T) \subseteq \overline{D(0, \|T\|)}$. ■

4.3.3 Spectral Radius

The **spectrum** of any operator $T \in \mathcal{L}(X, X)$ is a bounded set by **Proposition 4.3.2**, and moreover, we have a quantitative bound $|\lambda| \leq \|T\|$ for all $\lambda \in \sigma(T)$. This bound is not always sharp, and we will try to come up with a sharper bound.

To do this, we first introduce the notion of **spectral radius**.

Definition 4.3.7 (Spectral radius). The *spectral radius* of an operator $T \in \mathcal{L}(X, X)$ is defined as

$$r(T) := \max_{\lambda \in \sigma(T)} |\lambda|.$$

From **Proposition 4.3.2**, $r(T) \leq \|T\|$, but the equality is not always achieved. However, we have the **Gelfand's formula**, which characterizes the asymptotic behavior of $r(T)$, essentially stating that

$$\|T^n\| \sim r(T)^n.$$

Theorem 4.3.1 (Gelfand's formula). Let $T \in \mathcal{L}(X, X)$ on a **Banach space** X , then

$$r(T) = \lim_{n \rightarrow \infty} \|T^n\|^{1/n} = \inf_{n \geq 1} \|T^n\|^{1/n}.$$

Proof. Let r be a large integer and $m \geq 1$ an integer with $r = am + b$ for $a \geq 0$, $0 \leq b < m$. Then

$$\|T^r\| = \|T^{am} T^b\| \leq \|T^m\|^a \|T^b\|,$$

hence

$$\|T^r\|^{1/r} \leq \|T^m\|^{a/r} \|T^b\|^{1/r} = \left(\|T^m\|^{1/m}\right)^{\frac{a}{a+b/m}} \|T^b\|^{1/r}.$$

Let $r \rightarrow \infty$, we have $\limsup_{r \rightarrow \infty} \|T^r\|^{1/r} \leq \|T^m\|^{1/m}$ for all $m \geq 1$ since $a/(a+b/m) \rightarrow 1$ and $\|T^b\|^{1/r} \rightarrow 1$, hence

$$\limsup_{r \rightarrow \infty} \|T^r\|^{1/r} = \liminf_{r \rightarrow \infty} \|T^r\|^{1/r} = \inf_{m \geq 1} \|T^m\|^{1/m}.$$

Lecture 21: Spectral Radius of a Bounded Linear Operator

10 Nov. 14:30

Proof of Theorem 4.3.1 (Continued). Now, the next goal is to show that $\sigma(T)$ is nonempty.

Claim. $\sigma(T)$ is nonempty

Proof. Toward a contradiction, assume that $\rho(T) = \mathbb{C}$, then $R(\lambda) = (T - \lambda I)^{-1}$, which is defined for all $\lambda \in \mathbb{C}$ such that

$$\|R(\lambda)\| \leq \frac{1}{|\lambda| - \|T\|}$$

if $|\lambda| > \|T\|$ by [Proposition 4.3.1](#), hence $\lim_{|\lambda| \rightarrow \infty} \|R(\lambda)\| = 0$. Let $f(\cdot)$ be a [bounded linear functional](#) on $\mathcal{L}(X, X)$, i.e., $f \in \mathcal{L}(X, X)^*$. Set $g(\lambda) = f(R(\lambda))$, for any $\lambda_0 \in \rho(T)$,

$$R(\mu) = \left(\sum_{n=0}^{\infty} (\mu - \lambda)^n R(\lambda)^n \right) R(\lambda),$$

implying that $g(\lambda)$ is a analytic function since it's locally given by a convergent power series. Also, since $\lim_{|\lambda| \rightarrow \infty} |g(\lambda)| = 0$, [Liouville's theorem](#) implies $g(\cdot) \equiv 0$, so $f(R(\lambda)) \equiv 0$ for all $f \in \mathcal{L}(X, X)^*$. From [Hahn-Banach theorem](#), $R(\lambda) = 0$, which is a contradiction since $R(\lambda)(T - \lambda I) = I$. \circledast

Claim. $r(T) \leq \lim_{n \rightarrow \infty} \|T^n\|^{1/n}$.

Proof. To see this, we use the fact that if $\lambda \in \mathbb{C}$, we have $\lambda^n \in \rho(T^n)$, i.e., $(T^n - \lambda^n I)^{-1}$ exists, then $\lambda \in \rho(T)$ because $(T^n - \lambda^n I) = (T - \lambda I) \sum_{j=0}^{n-1} \lambda^j T^{n-1-j}$, we have

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

Hence, $|\lambda^n| > \|T^n\|$, implying that $\lambda \in \rho(T)$, so $r(T) \leq \|T^n\|^{1/n}$ for all $n \geq 1$. \circledast

Finally, we show the following.

Claim. $r(T) \geq \limsup_{n \geq 1} \|T^n\|^{1/n}$.

Proof. We use the Taylor expansion

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

Let $f \in \mathcal{L}(X, X)^*$, set $g(\lambda) = f(R(\lambda))$, we have

$$g(\lambda) = - \sum_{n=0}^{\infty} \lambda^{-(n+1)} f(T^n),$$

where $g(\lambda)$ is analytic for $|\lambda| > r(T)$. Hence, the Laurent series for $g(\lambda)$ converges if $|\lambda| > r(T)$, i.e., $\sup_{n \geq 1} |\lambda^{-n} f(T^n)| < \infty$ if $|\lambda| > r(T)$, which is true for all $f \in \mathcal{L}(X, X)^*$. From the [uniform boundedness theorem](#), $\sup_{n \geq 1} |\lambda|^{-n} \|T^n\| < \infty$ if $|\lambda| > r(T)$, i.e., if $|\lambda| > r(T)$, then $|\lambda| \geq \limsup_{n \rightarrow \infty} \|T^n\|^{1/n}$, implying $r(T) \geq \limsup_{n \rightarrow \infty} \|T^n\|^{1/n}$. \circledast

In all, we conclude that $r(T) = \lim_{n \rightarrow \infty} \|T^n\|^{1/n}$, proving the result. \blacksquare

Note. [Gelfand's formula](#) is an improvement upon [Proposition 4.3.2](#).

Proof. Clearly, we see that

$$r(T) \leq \|T^n\|^{1/n} \leq \|T\|^{n \cdot 1/n} \leq \|T\|.$$

\circledast

4.4 Spectrum of Compact Operators

As [compact operators](#) are proxies of [finite rank operators](#), one is able to fully classify their [spectrum](#). First, for every $T \in \mathcal{K}(X, X)$, $0 \in \sigma(T)$ from T is not invertible because [isomorphism is not compact](#).

Theorem 4.4.1 (Point spectrum of compact operators). Let $T \in \mathcal{K}(X, X)$ on a [Banach space](#) X . Then for every $\epsilon > 0$, there exists a finite number of linearly independent eigenvectors corresponding to eigenvalues $\lambda \in \mathbb{C}$ with $|\lambda| > \epsilon$.

Proof. Toward a contradiction, we show that we can obtain a sequence $\{y_n\}_{n \geq 1}$ such that $\|y_n\| = 1$ for all $n \geq 1$, but the sequence $\{Ty_n\}_{n \geq 1}$ has no converging subsequence. This contradicts to the [compactness](#) of T .

From the assumption, there exists a sequence $\{x_k\}_{k \geq 1}$ in X such that $x_k \neq 0$ and are all linearly independent vectors with $Tx_k = \lambda_k x_k$ and $|\lambda_k| \geq \epsilon > 0$ for all $k \geq 1$. Let E_n be the span of $\{x_1, \dots, x_n\}$, so $E_1 \subseteq E_2 \subseteq \dots$. We can then choose $y_n \in E_n$ with $\|y_n\| = 1$ such that

$$\text{dist}(y_n, E_{n-1}) = \inf_{y \in E_{n-1}} \|y_n - y\| \geq 1/2.$$

Now, to show that $\{Ty_n\}_{n \geq 1}$ contains no Cauchy subsequences, we express y_n as a linear combination

$$y_n = \sum_{k=1}^n a_k^{(n)} x_k = a_n^{(n)} x_n + u_{n-1} \text{ for } u_{n-1} \in E_{n-1},$$

then

$$Ty_n = a_n^{(n)} \lambda_n x_n + v_{n-1} \text{ for } v_{n-1} = Tu_{n-1} \in E_{n-1}.$$

Let $n > m$, then

$$\begin{aligned} \|Ty_n - Ty_m\| &= \|\lambda_n a_n^{(n)} x_n + w_{n-1}\| && \text{where } w_{n-1} \in E_{n-1} \\ &= \|\lambda_n y_n + \tilde{y}\| && \text{where } \tilde{y} \in E_{n-1} \\ &= |\lambda_n| \|y_n + \tilde{y}'\| && \text{where } \tilde{y}' \in E_{n-1} \\ &\geq \frac{|\lambda_n|}{2} && \|y_n + \tilde{y}'\| \geq 1/2 \text{ when } \tilde{y}' \in E_{n-1} \\ &\geq \frac{\epsilon}{2}, \end{aligned}$$

so there are no converging subsequences for $\{Ty_n\}_{n \geq 1}$. We see that if $y_n \in B_X$ for $n \geq 1$, if T is [compact](#), TB_X is [precompact](#), which is a contradiction. ■

Remark. Consequently, from [Theorem 4.4.1](#), the [point spectrum](#) $\sigma_p(T)$ is at most countable, and it lies in a sequence that converges to 0. Also, each eigenvalue λ_k of T has finite multiplicity, i.e., $\dim \ker(T - \lambda_k I) < \infty$.

[Theorem 4.4.1](#) implies that there are countable many [point spectrums](#), and the only possible accumulation point is 0, and indeed, for a [compact operator](#), the only [spectrum](#) is either in $\sigma_p(T)$ or 0 as guaranteed by [Proposition 4.4.1](#).

Proposition 4.4.1 (Classification of spectrum of compact operators). Let $T \in \mathcal{K}(X, X)$, then $\sigma_p(T)$ is countable and $\sigma(T) = \sigma_p(T) \cup \{0\}$.

Proof. By noncompactness of unit ball from [Riesz's theorem](#), $0 \in \sigma(T)$ as noted at the beginning of the section. Recall that if $\lambda \neq 0$, $(T - \lambda I)$ is not surjective,^a and hence λ is an eigenvalue by [Fredholm alternative](#), so all others $0 \neq \lambda \in \sigma(T)$ is in $\sigma_p(T)$, i.e., $\sigma(T) = \sigma_p(T) \cup \{0\}$. ■

^aOtherwise, from [inverse mapping theorem](#), $T - \lambda I$ is invertible so $\lambda \in \rho(T)$, contradiction.

Recall that if $T \in \mathcal{L}(X, X)$, then $T^* \in \mathcal{L}(X^*, X^*)$, and we have $\|T^*\| = \|T\|$. Now, we want to find the relation between $\sigma(T^*)$ and $\sigma(T)$.

Theorem 4.4.2. Let $T \in \mathcal{L}(X, X)$ and $T^* \in \mathcal{L}(X^*, X^*)$, then we have the following.

- (a) $\sigma(T^*) = \sigma(T)$.
- (b) If $\lambda \in \sigma_r(T)$, then $\lambda \in \sigma_p(T^*)$.
- (c) If $\lambda \in \sigma_p(T)$, then $\lambda \in \sigma_p(T^*) \cup \sigma_r(T^*)$.

Proof. We prove this one by one.

- (a) We first show that $\rho(T) \subseteq \rho(T^*)$. Suppose $\lambda \in \rho(T)$, i.e., $(T - \lambda I)^{-1}$ exists, so $(T - \lambda I)^*$ is also invertible and

$$[(T - \lambda I)^*]^{-1} = (T^* - \lambda I)^{-1}.$$

Hence, $\lambda \in \rho(T^*)$, so we have $\rho(T) \subseteq \rho(T^*)$. To show that $\rho(T^*) \subseteq \rho(T)$, we need to show that if $S \in \mathcal{L}(X, X)$ and S^* is invertible, then S is invertible. We first show that S is injective, i.e., $\ker(S) = \{0\}$. Suppose not, then there exists $x \neq 0$ such that $Sx = 0$, implying $S^*f(x) = f(Sx) = 0$ for all $f \in X^*$. But since S^* is invertible, i.e., $\text{Im } S^* = X^*$, implying $g(x) = 0$ for all $g \in X^*$. Then from [Hahn-Banach theorem](#), $x = 0$, which is a contradiction. Next, we want to show S is surjective, i.e., $\text{Im}(S) = X$. Since S^* is invertible, there exists $\epsilon > 0$ such that

$$\|S^*f\|_{X^*} \geq \epsilon \|f\|_{X^*}$$

for all $f \in X^*$. We now use this to show that $\overline{SB_X}$ contains a ball of radius ϵ . Let $x_0 \in X$ lies outside $\overline{SB_X}$. Since $\overline{SB_X}$ is closed and [convex](#), the [separation theorem](#) states that there exists $f_0 \in X^*$ such that $f_0(x_0) > 1$ and $f_0(y) \leq 1$ for all $y \in \overline{SB_X}$. Note that $S^*f_0(x) = f_0(Sx) \leq 1$ for all $x \in B_X$, we have $\|S^*f_0\|_{X^*} \leq 1$. Hence, we conclude that

$$\epsilon < \epsilon f_0(x_0) \leq \epsilon \|f_0\|_{X^*} \|x_0\|_X \leq \|S^*f_0\|_{X^*} \|x_0\|_X \leq \|x_0\|_X,$$

so $\epsilon B_X \subseteq \overline{SB_X}$. Now, we use the argument from [open mapping theorem](#) to conclude that $\overline{SB_X} \subseteq S(2B_X)$. Hence, $S(2B_X) \subseteq \epsilon B_X$, i.e., S is surjective.

Lecture 22: Self-Adjoint Operators on Hilbert Spaces

15 Nov. 14:30

Proof of Theorem 4.4.2 (Continued). We now continue to prove [Theorem 4.4.2](#).

- (b) Let $S \in \mathcal{L}(X, X)$, then we already observed that $(\text{Im } S)^\perp = \ker(S^*)$. Assume that $\overline{\text{Im } S} \neq X$, by [separation theorem](#), there exists $f_0 \in X^*$, $f_0 \neq 0$ and $f_0 \in (\text{Im } S)^\perp = \ker(S^*)$.
- (c) We need to show that if $S \in \mathcal{L}(X, X)$ and $\ker(S) \neq \{0\}$, then $\text{Im } S$ is not dense in X . Assume $Sx = 0$ and $x \neq 0$, since $\ker S \neq \{0\}$, then

$$S^*f(x) = f(Sx) = 0$$

for all $f \in X^*$. Hence, if $g \in \text{Im } S^*$, then $g(x) = 0$. If $\text{Im } S^*$ is dense in X^* , conclude that $g(x) = 0$ for all g , implying that $x = 0$, contradiction. ■

Remark. The latter two claims in [Theorem 4.4.2](#) can be summarized as

$$\sigma_r(T) \subseteq \sigma_p(T^*) \subseteq \sigma_r(T) \cup \sigma_p(T).$$

Example (Shift operator). Let T be the left shift operator on ℓ_1 , i.e.,

$$T(a_1, a_2, \dots) = (a_2, a_3, \dots).$$

Notice that $\ell_1^* = \ell_\infty$,^a so T^* is the right shift operator on ℓ_∞ , i.e.,

$$T^*(a_1, a_2, \dots) = (0, a_1, a_2, \dots).$$

Since $\|T\| = \|T^*\| = 1$, hence $\lambda \in \rho(T)$ if $|\lambda| > 1$. Suppose $|\lambda| < 1$, then the vector $x_\lambda = (1, \lambda, \lambda^2, \dots)$ is in ℓ_1 such that $Tx_\lambda = \lambda x_\lambda$, so $\lambda \in \sigma_p(T)$. With $\sigma(T)$ is closed, we have

$$\sigma(T) = \{\lambda \in \mathbb{C} : |\lambda| \leq 1\}.$$

Next, we want to show that T^* has no **point spectrum**. Suppose $a = (a_1, a_2, \dots) \in \ell_\infty$ and $(T \pm \lambda I)a = 0$. Then

$$\lambda a_1 = 0, \quad \lambda a_2 - a_1 = 0, \quad \lambda a_3 - a_2 = 0, \quad \dots$$

We then see that if $\lambda \neq 0$, then $a \equiv 0$; if $\lambda = 0$, then $a \equiv 0$ as well, hence $\sigma_p(T^*)$ is empty. By [Theorem 4.4.2](#), $\lambda \in \sigma_p(T) \Rightarrow \lambda \notin \sigma_c(T^*)$. Now, consider

- $|\lambda| < 1$: we further have $\lambda \in \sigma_p(T)$ and $\lambda \notin \sigma_p(T^*)$, hence $\lambda \in \sigma_r(T^*)$.
- $|\lambda| = 1$: in this case, we want to show $\lambda \in \sigma(T) = \sigma(T^*)$. It's clear that $\lambda \notin \sigma_p(T)$. On the other hand, if $\lambda \in \sigma_r(T)$, then $\lambda \in \sigma_p(T^*)$ by [Theorem 4.4.2](#), with the fact that $\sigma_p(T^*) = \emptyset$, so $\sigma_r(T) = \emptyset$. So we conclude that if $|\lambda| = 1$, $\lambda \in \sigma_c(T)$.

Finally, we show that if $|\lambda| = 1$, then $\lambda \in \sigma_r(T^*)$. To do this, we shall find an open **ball** disjoint from $\text{Im}(T^* - \lambda I)$. Suppose $a = \{a_n\}_{n \geq 1}$ and $b = \{b_n\}_{n \geq 1}$ are in ℓ_∞ with $a = (\lambda I - T^*)b$, hence

$$a_1 = \lambda b_1, \quad a_2 = \lambda b_2 - b_1, \quad a_3 = \lambda b_3 - b_2, \quad \dots$$

This is equivalent to write

$$\begin{aligned} b_1 &= \frac{a_1}{\lambda} = \bar{\lambda} a_1, \\ b_2 &= \frac{a_1}{\lambda} + \frac{b_1}{\lambda} = \frac{a_2}{\lambda} + \frac{a_1}{\lambda^2} = \bar{\lambda}^2 (a_1 + \lambda a_2), \\ &\vdots \\ b_n &= \bar{\lambda}^{n+1} \sum_{m=1}^n \lambda^m a_m. \end{aligned}$$

Define $c = \{c_n\}_{n \geq 1}$ such that $c_n = \bar{\lambda}^n$. Suppose $d \in \ell_\infty$, $\|d - c\|_\infty < 1/2$. Then

$$\text{Re} \{\lambda^n d_n\} \geq \text{Re} \{\lambda^n c_n\} - \|d - c\|_\infty \geq \frac{1}{2}.$$

If $(\lambda I - T^*)e = d$ for some $e \in \ell_\infty$, then

$$e = \bar{\lambda}^{n+1} \sum_{m=1}^n \lambda^m d_m,$$

implying that $|e_n| \geq n/2$, i.e., $e \notin \ell_\infty$, contradiction. We then conclude that $\text{Im}(\lambda I - T^*)$ does not intersect **ball** centered at c with radius $1/2$, hence $\lambda \in \sigma_r(T^*)$.

^aWhile $\ell_\infty^* \neq \ell_1$.

Chapter 5

Self-Adjoint Operators on Hilbert Spaces

Throughout this chapter, \mathcal{H} will denote a [Hilbert space](#), and we will study [bounded self-adjoint operators](#) on \mathcal{H} , i.e., $T \in \mathcal{L}(\mathcal{H}, \mathcal{H}) =: \mathcal{L}(\mathcal{H})$.

5.1 Spectrum of Self-Adjoint Operators

Recall that in a [Hilbert space](#) \mathcal{H} , given $T \in \mathcal{L}(\mathcal{H})$, the [adjoint](#) $T^* \in \mathcal{L}(\mathcal{H})$ is defined by

$$\langle T^*x, y \rangle = \langle x, Ty \rangle$$

for $x, y \in \mathcal{H}$. An interesting situation is the following.

Definition 5.1.1 (Self-adjoint operator). Let $T \in \mathcal{L}(\mathcal{H})$ on a [Hilbert space](#) \mathcal{H} , then T is *self-adjoint* if for all $x, y \in \mathcal{H}$,

$$\langle Tx, y \rangle = \langle x, Ty \rangle.$$

There are lots of examples of [self-adjoint operators](#).

Example (Hermitian matrix). The linear operators on \mathbb{C}^n given by Hermitian matrices $A = (a_{ij})$ is [self-adjoint](#).

Proof. Since $a_{ij} = \overline{a_{ji}}$. *

Example (Integral operator). The integral operators T on $L^2([0, 1])$ with Hermitian symmetric kernels $k(s, t)$ given by

$$(Tf)(t) = \int_0^1 k(s, t)f(s) \, ds$$

is [self-adjoint](#).

Proof. Since $k(s, t) = \overline{k(t, s)}$. *

Also, the [orthogonal projection](#) P on \mathcal{H} is [self-adjoint](#).

Remark. Every $A \in \mathcal{L}(\mathcal{H})$ can be represented as $A = T + iS$ with T, S being [self-adjoint](#).

Proof. If $A = T + iS$, then $A^* = T - iS$. Solving these two equations gives

$$T = \frac{A + A^*}{2}, \quad S = \frac{A - A^*}{2i}.$$

*

5.1.1 Quadratic Form and Norm of Self-Adjoint Operators

An important object in the study of [self-adjoint operator](#) is the quadratic form $f: \mathcal{H} \rightarrow \mathbb{R}$ where

$$f(x) = \langle Tx, x \rangle$$

for $x \in \mathcal{H}$, where $f(\cdot)$ is real since $\langle Tx, x \rangle = \langle x, Tx \rangle = \overline{\langle Tx, x \rangle}$. Furthermore, $f(\cdot)$ determines T uniquely by the generalized [polarization identity](#)¹

$$\langle Tx, y \rangle = \frac{1}{4} [f(x+y) - f(x-y) + if(x+iy) - if(x-iy)].$$

Since $f(\cdot)$ determines T uniquely, we should be able to compute properties of T using f . The first property is the following.

Proposition 5.1.1. For every [self-adjoint operator](#) $T \in \mathcal{L}(\mathcal{H})$,

$$\|T\| = \sup_{\|x\|=1} |\langle Tx, x \rangle|.$$

Proof. Firstly, from [Cauchy-Schwarz](#), we have $|\langle Tx, x \rangle| \leq \|Tx\| \|x\|$, implying that

$$\sup_{\|x\|=1} |\langle Tx, x \rangle| \leq \sup_{\|x\|=1} \|Tx\| = \|T\|.$$

To get the equality, we use the [polarization identity](#), where

$$\operatorname{Re} \langle Tx, y \rangle = \frac{1}{4} [\langle T(x+y), x+y \rangle - \langle T(x-y), x-y \rangle].$$

Let $M := \sup_{\|x\|=1} |\langle Tx, x \rangle|$, we have

$$\operatorname{Re} \langle Tx, y \rangle \leq \frac{M}{4} [\|x+y\|^2 + \|x-y\|^2] = \frac{M}{4} [2\|x\|^2 + 2\|y\|^2],$$

where the last equality follows from the [parallelogram law](#). This implies that

$$\|T\| = \sup_{\|x\|=\|y\|=1} |\langle Tx, y \rangle| = \sup_{\|x\|=\|y\|=1} \operatorname{Re} \langle Tx, y \rangle \leq M.$$

■

5.1.2 Criterion of Spectrum Points

We now study the [spectrum](#) of [self adjoint operators](#) $T \in \mathcal{L}(\mathcal{H})$. The goal is to show that the whole [spectrum](#) of T is real, i.e., $\sigma(T) \subseteq \mathbb{R}$. Let's start with some basic facts.

Lemma 5.1.1. Let $T \in \mathcal{L}(\mathcal{H})$ be [self-adjoint](#), then $\sigma_p(T) \subseteq \mathbb{R}$ and $\sigma_r(T) = \emptyset$.

Proof. To show $\sigma_p(T) \subseteq \mathbb{R}$, suppose $Tx = \lambda x$, then

$$\begin{cases} \langle Tx, x \rangle = \langle \lambda x, x \rangle = \lambda \|x\|^2; \\ \langle x, Tx \rangle = \langle x, \lambda x \rangle = \bar{\lambda} \|x\|^2 \end{cases} \Rightarrow (\lambda - \bar{\lambda}) \|x\|^2 = 0.$$

Since $x \neq 0$, this implies that $\lambda - \bar{\lambda} = 0$, i.e., $\lambda \in \mathbb{R}$.

To show $\sigma_r(T) = \emptyset$, let $\lambda \in \sigma_r(T)$, then $\ker(T - \lambda I) = \{0\}$ and $\operatorname{Im}(T - \lambda I)$ is not dense in \mathcal{H} . Notice that since $\lambda \notin \sigma_p(T)$, then $\bar{\lambda} \notin \sigma_p(T)$ as we just proved. We then have

$$\{0\} = \ker(T - \bar{\lambda} I) = \ker(T - \lambda I)^*,$$

¹When $T = I$, we get back the usual [polarization identity](#).

with [Proposition 2.5.2](#),

$$\operatorname{Im}(T - \lambda I)^\perp = \ker(T - \lambda I)^* = \ker(T - \bar{\lambda}I) = \{0\},$$

hence $\operatorname{Im}(T - \lambda I)$ is dense in \mathcal{H} , a contradiction. So λ does not exist, proving that $\sigma_r(T) = \emptyset$. ■

Lecture 23: Spectral Theorems, Continuous Functional Calculus

17 Nov. 14:30

Lemma 5.1.2 (Invertibility criterion). Let T be a [bounded self-adjoint operator](#) on \mathcal{H} , then T is invertible if and only if T is bounded below, i.e., for all $x \in \mathcal{H}$, there exists $c > 0$ such that $\|Tx\| \geq c\|x\|$.

Proof. If T is invertible, then T is bounded below with $c = \|T^{-1}\|^{-1}$.

Conversely, if T is bounded below, then T is injective and $\operatorname{Im} T$ is closed by the [isomorphic embedding criterion](#). Since $\sigma_r(T)$ is empty, so $0 \neq \sigma_r(T)$, hence the injectivity of T implies that $\operatorname{Im} T$ is dense in \mathcal{H} , so $\operatorname{Im} T = \mathcal{H}$, i.e., T is surjective and injective. With [inverse mapping theorem](#), T is invertible with $\|T^{-1}\| \leq c^{-1}$. ■

We see that by applying the [invertibility criterion](#) for the operator $T - \lambda I$, we immediately obtain the following.

Corollary 5.1.1 (Criterion of spectrum points). Let $T \in \mathcal{L}(\mathcal{H})$ be a [self-adjoint operator](#), then $\lambda \in \sigma(T)$ if and only if the operator $T - \lambda I$ is not [bounded](#) below.

Finally, we introduce the following.

Definition. Let $T \in \mathcal{L}(\mathcal{H})$ on a [Hilbert space](#) \mathcal{H} .

Definition 5.1.2 (Approximate eigenvalue). A number $\lambda \in \sigma(T)$ for which $T - \lambda I$ is not bounded below is called an *approximate eigenvalue* of T .

Definition 5.1.3 (Approximate point spectrum). The set of all [approximate eigenvalues](#) of T is called the *approximate point spectrum* of T .

[Criterion of spectrum points](#) states that for [self-adjoint operators](#), the whole [spectrum](#) is the [approximate point spectrum](#).

The reason for the name *approximate* is the following. If λ is an eigenvalue, then $(T - \lambda I)x = 0$ for some x with $\|x\| = 1$. If λ is an [approximate eigenvalue](#), then $(T - \lambda I)x$ can be made *arbitrarily close* to 0 for some x with $\|x\| = 1$. So, eigenvalues of T form the [point spectrum](#) $\sigma_p(T)$ while the [approximate eigenvalues](#) of T form the [continuous spectrum](#) $\sigma_c(T)$.

Remark. $\lambda \in \sigma(T)$ is an [approximate point spectrum](#) if and only if there exists a sequence $\{x_n\}_{n \geq 1}$ in \mathcal{H} and $\|x_n\| = 1$, and

$$\lim_{n \rightarrow \infty} \|Tx_n - \lambda x_n\| = 0.$$

5.1.3 The Spectrum Interval

We now compute the *tightest* interval that contains the [spectrum](#) of a [self-adjoint operator](#) T . This interval can be computed from the quadratic form of T as follows.

Theorem 5.1.1 (Spectrum interval). Suppose $T \in \mathcal{L}(\mathcal{H})$ for T being a [self-adjoint operator](#). Then

- (a) $\sigma(T) \subseteq [m, M]$ for $m := \inf_{\|x\|=1} \langle Tx, x \rangle$, $M := \sup_{\|x\|=1} \langle Tx, x \rangle$.
- (b) The endpoints $m, M \in \sigma(T)$.

Proof. We prove this one by one.

(a) Let $\lambda \in \mathbb{C} - [m, M]$, and set d be

$$d := \text{dist}(\lambda, [m, M]) = \inf_{m \leq y \leq M} |\lambda - y| > 0.$$

Then we have

$$\|(T - \lambda I)x\| \geq |\langle (T - \lambda I)x, x \rangle| = |\langle Tx, x \rangle - \lambda| \geq d$$

since $\|x\| = 1$, which implies $T - \lambda I$ is bounded below, and hence $\lambda \in \rho(T)$.

(b) Without loss of generality, assume that $0 \leq m \leq M$ by considering $T - mI$ instead of T . Now, choose a sequence $\{x_n\}_{n \geq 1}$ in \mathcal{H} where $\|x_n\| = 1$ such that

$$\lim_{n \rightarrow \infty} \langle Tx_n, x_n \rangle = M.$$

By the [parallelogram law](#),

$$\|(T - MI)x_n\|^2 = \langle (T - MI)x_n, (T - MI)x_n \rangle = \|Tx_n\|^2 - 2M \langle Tx_n, x_n \rangle + M^2 \|x_n\|^2.$$

Since we already showed that $\|T\| = M$, i.e., $\|T\| = \sup_{\|x\|=1} \langle Tx, x \rangle$ from $\langle Tx, x \rangle \geq 0$, by letting $n \rightarrow \infty$, the right-hand side goes to ≤ 0 since $\|Tx_n\|^2 \leq M^2 \|x_n\|^2$ and $\langle Tx_n, x_n \rangle \rightarrow M$, we may conclude that

$$\lim_{n \rightarrow \infty} \|(T - MI)x_n\| = 0,$$

and hence $M \in \sigma(T)$. ■

As a consequence, $\sigma(T) = \|T\|$ for T being a [self-adjoint operator](#). This means that [Proposition 4.3.2](#) is tight, while [Gelfand's formula](#) is useless for [self-adjoint operators](#)! This observation leads to the following.

Corollary 5.1.2 (Spectral radius). Let $T \in \mathcal{L}(\mathcal{H})$ for T being a [self-adjoint operator](#). Then

$$r(T) = \max_{\lambda \in \sigma(T)} |\lambda| = \|T\|.$$

Proof. From the property of [spectrum interval](#), we know that $r(T) = \max(|m|, |M|) = \|T\|$. ■

5.2 Spectral Theorem for Compact Self-Adjoint Operators

[Compact self-adjoint operators](#) on a [Hilbert space](#) \mathcal{H} are proxies of Hermitian matrices. As we know from linear algebra, every Hermitian matrix has diagonal form in some [orthonormal basis](#) of \mathbb{C}^n , or equivalently, there exists an [orthonormal basis](#) of \mathbb{C}^n consisting of the eigenvectors. In this section, we generalize this fact to infinite dimensions, for all [compact self-adjoint operators](#) on \mathcal{H} .

5.2.1 Invariant Subspaces

Proposition 5.2.1 (Eigenvectors orthogonal). Let $T \in \mathcal{L}(\mathcal{H})$ be a [self-adjoint operator](#) on \mathcal{H} . Then its eigenvectors corresponding to distinct eigenvalues are [orthogonal](#).

Proof. If $Tx_1 = \lambda_1 x_1$ and $Tx_2 = \lambda_2 x_2$, then

$$\lambda_1 \langle x_1, x_2 \rangle = \langle Tx_1, x_2 \rangle = \langle x_1, Tx_2 \rangle = \lambda_2 \langle x_1, x_2 \rangle,$$

so if $\lambda_1 \neq \lambda_2$, then $\langle x_1, x_2 \rangle = 0$, proving the result. ■

Definition 5.2.1 (Invariant subspace). A subspace E of \mathcal{H} is called an *invariant subspace* of T if $T(E) \subseteq E$.

Example. Every eigenspace of T is *invariant*. More generally, the linear span of any subset of eigenvectors of T is an *invariant subspace*.

One of the most well-known open problems in functional analysis is the *invariant subspace problem*. It asks whether every operator $T \in \mathcal{L}(\mathcal{H})$ has a proper *invariant subspace*, i.e., different from $\{0\}$ and \mathcal{H} .

Proposition 5.2.2. Let $T \in \mathcal{L}(\mathcal{H})$ be a *self-adjoint operator* on \mathcal{H} . If $E \subseteq \mathcal{H}$ is an *invariant subspace* of T , then E^\perp is also an *invariant subspace* of T .

Proof. Let $x \in E^\perp$, and we need to check that $Tx \in E^\perp$ given E is *invariant*. Let's choose $y \in E$ arbitrarily, then we have

$$\langle Tx, y \rangle = \langle x, Ty \rangle = 0$$

since $x \in E^\perp$ and $y \in E$, hence $Ty \in E$, as required. ■

5.2.2 Spectral Theorem

The following result is known as the Hilbert-Schmidt theorem.

Theorem 5.2.1 (Spectral theorem for compact self-adjoint operator). Let T be a *compact self-adjoint operator* on \mathcal{H} and \mathcal{H} is *separable*. Then there exists an *orthonormal basis* of \mathcal{H} consisting of eigenvectors of T .

Proof. Let's first prove that T has at least one eigenvector. Firstly, since T is *compact*, by [Proposition 4.4.1](#),

$$\sigma(T) = \sigma_p(T) \cup \{0\}.$$

So if $\sigma(T) \neq \{0\}$, then $\sigma_p(T) \neq \emptyset$, i.e., T has an eigenvector. Otherwise, if $\sigma(T) = \{0\}$, then from [Theorem 5.1.1](#), $\|T\|$ or $-\|T\|$ is in $\sigma(T)$, so $r(T) = \|T\| = 0$, i.e., $T \equiv 0$. In this case, any *orthonormal basis* gives a basis of eigenvectors. With the fact that \mathcal{H} is *separable*, all such basis are at most countable, so from [Zorn's lemma](#), this family has a maximal element $\{\phi_k\}_{k \geq 1}$, so the result follows by showing that

$$E := \overline{\text{span}(\{\phi_k\}_{k \geq 1})} = \mathcal{H}.$$

Suppose $E \neq \mathcal{H}$. Since E is an *invariant subspace* of T , $E^\perp \neq \{0\}$ is also an *invariant subspace* of T by [Proposition 5.2.2](#). By using the first part of the proof for the restriction $T|_{E^\perp}$ which is a *compact self-adjoint operator* on E^\perp . It follows that $T|_{E^\perp}$, and thus T itself, has an eigenvector in E^\perp . But this contradicts the maximality of $\{\phi_k\}_{k \geq 1}$. ■

Finally, we introduce a new kind of operators called *normal operators*, where the above result generalizes to which.

Definition 5.2.2 (Normal operator). An operator T is *normal* if $TT^* = T^*T$.

Remark. The *spectral theorems for compact self-adjoint operators* extend to *normal operator*.

However, *spectrum* of *normal operators* do not have to be real,² we only have $\sigma(T) \subseteq \{\lambda \in \mathbb{C} : |\lambda| = 1\}$.

5.3 Continuous Functional Calculus

In this section, we develop the analogy between numbers and operators by introducing a *partial order* on the set of *self-adjoint operators* on $T \in \mathcal{L}(\mathcal{H})$, and we define an operator $f(T) \in \mathcal{L}(\mathcal{H})$ for every continuous function $f: \mathbb{C} \rightarrow \mathbb{C}$. This is the so-called *functional calculus* of operators.

²For example, the unitary operators $U^*U = UU^* = I$.

5.3.1 Positive Operators

Definition 5.3.1 (Positive operator). A **self-adjoint operator** $T \in \mathcal{L}(\mathcal{H})$ is called *positive* if $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$.

Remark (Positive semi-definite). In linear algebra, **positive operators** are called positive semi-definite.

We see that **positive operators** are generalizations of non-negative numbers, which correspond to operators on one-dimensional space \mathbb{C} .

Example. T^2 for every **self-adjoint** $T \in \mathcal{L}(\mathcal{H})$ is **positive**.

Proof. Since $\langle T^2x, x \rangle = \langle Tx, Tx \rangle \geq 0$. ⊗

Example. Hermitian matrices, or more generally, **compact self-adjoint operators** on \mathcal{H} with non-negative eigenvalues are **positive**.

Just like in the linear algebra, where we define $A \succeq B$ by $A - B \succeq 0$ (i.e., $A - B$ is positive semi-definite), we define the same thing for **self-adjoint operators**.

Definition 5.3.2 (Partial order). For **self-adjoint operators** $S, T \in \mathcal{L}(\mathcal{H})$, we say $S \leq T$ if $T - S \geq 0$.

Definition 5.3.2 defines a **partial order** on $\mathcal{L}(\mathcal{H})$, and we may restate the **spectrum interval theorem** with this new notion.

Theorem 5.3.1 (Spectrum interval). Let $T \in \mathcal{L}(\mathcal{H})$ be a **self-adjoint operator**, and let m, M be the smallest and the largest numbers such that $mI \leq T \leq MI$. Then $\sigma(T) \subseteq [m, M]$ and $m, M \in \sigma(T)$.

As an immediate corollary, T is **positive** if and only if its **spectrum** is positive.

Corollary 5.3.1. Let $T \in \mathcal{L}(\mathcal{H})$ be a **self-adjoint operator**, then $T \geq 0$ if and only if $\sigma(T) \subseteq [0, \infty)$.

5.3.2 Polynomials of Operators

We start to develop a functional calculus for **self-adjoint operators** $T \in \mathcal{L}(\mathcal{H})$ by defining polynomials of T , and then we extend the definition to continuous functions of T by approximation. Working with polynomials is straightforward, and the result of this subsection remain valid for every **bounded linear operator** T on a general **Banach space** X .

Definition 5.3.3 (Polynomial operator). Consider a polynomial $p(t) = a_0 + a_1t + \dots + a_nt^n$, then for an operator $T \in \mathcal{L}(\mathcal{H})$, we define

$$p(T) := a_0I + a_1T + \dots + a_nT^n.$$

We first note that if T is **self-adjoint**, then $p(T)$ is also **self-adjoint** if p is real since

$$\begin{aligned} \langle p(T)x, y \rangle &= \langle (a_0I + a_1T + \dots + a_nT^n)x, y \rangle \\ &= a_0 \langle x, y \rangle + a_1 \langle Tx, y \rangle + \dots + a_n \langle T^n x, y \rangle \\ &= \langle x, \overline{a_0}y \rangle + \langle x, \overline{a_1}Ty \rangle + \dots + \langle x, \overline{a_n}T^n y \rangle \\ &= \langle x, a_0y \rangle + \langle x, a_1Ty \rangle + \dots + \langle x, a_nT^n y \rangle \\ &= \langle x, (a_0I + a_1T + \dots + a_nT^n)y \rangle \\ &= \langle x, p(T)y \rangle. \end{aligned}$$

Moreover, we have the following properties for **polynomial operators**.

Proposition 5.3.1. Let p, q be 2 complex polynomials and an operator $T \in \mathcal{L}(\mathcal{H})$. Then the following holds.

- (a) $(ap + bq)(T) = ap(T) + bq(T)$ for $a, b \in \mathbb{C}$.
- (b) $(pq)(T) = p(T)q(T)$.
- (c) $p(T)^* = \overline{p}(T^*)$.^a

^a \overline{p} is the polynomial with coefficients that are complex conjugates of the coefficients of p .

The following example may serve us as a test case for many future results.

Example. Let T be a **self-adjoint** linear operator on an n -dimensional **Hilbert space**. In an **orthonormal basis** of eigenvectors, T can be identified with the $n \times n$ diagonal matrix

$$T = \text{diag}(\lambda_1, \dots, \lambda_n),$$

where λ_k are the eigenvalues of T . Then for every polynomial $p(t)$, we have

$$p(T) = \text{diag}(p(\lambda_1), \dots, p(\lambda_n)).$$

This can be generalized for all **compact self-adjoint operators** T on a general **Hilbert space** \mathcal{H} .

5.3.3 Spectral Mapping Theorem for Polynomial Operators

Let's study some important theorem for the **polynomial operators**.

Lemma 5.3.1 (Invertibility for polynomial operator). Suppose $T \in \mathcal{L}(\mathcal{H})$, then the operator $\rho(T)$ is invertible if and only if $p(t) \neq 0$ for all $t \in \sigma(T)$.

Proof. Let

$$p(t) = a_n(t - t_1)(t - t_2) \dots (t - t_n)$$

where t_1, \dots, t_n are zeros of $p(\cdot)$. Then

$$p(T) = a_n(T - t_1 I) \dots (T - t_n I).$$

Next, observe that if $S, R \in \mathcal{L}(\mathcal{H})$ and SR is invertible, then both S and R are invertible.^a Hence, if $p(T)$ is invertible, then $t_1, \dots, t_n \in \rho(T)$, and conversely, $p(t) \neq 0$ for $t \in \sigma(T)$. ■

^aSince if SR is invertible, then $S^{-1} = R(SR)^{-1}$.

The **spectrum** of a polynomial $p(T)$ can be easily computed from the **spectrum** of T .

Theorem 5.3.2 (Spectral mapping theorem for polynomial operator). Let $p(t)$ be a polynomial and $T \in \mathcal{L}(\mathcal{H})$. Then

$$\sigma(p(T)) = p(\sigma(T))$$

where $p(\sigma(T)) := \{p(t) : t \in \sigma(T)\}$.

Proof. Since $\lambda \in \sigma(p(T))$ if and only if $p(T) - \lambda I$ is not invertible, which from **Lemma 5.3.1**, it is equivalent to say $p(t) - \lambda = 0$ for some $t \in \sigma(T)$, i.e., $\lambda \in p(\sigma(T))$. ■

Using the **spectral mapping theorem for polynomial operator**, one can in particular easily compute the **norm** of **polynomial operator**.

Corollary 5.3.2 (Operator norm of polynomial operator). Suppose T is a **bounded self-adjoint operator** on \mathcal{H} and $p(t)$ is a polynomial with real coefficients. Then $p(T)$ is **self-adjoint operator** and

$$\|p(T)\| = \sup_{t \in \sigma(T)} |p(t)|.$$

Proof. Firstly, $p(T)$ is self-adjoint since $\bar{p} = p$ as noted in the beginning of the section, with the Corollary 5.1.2 and spectral mapping theorem for polynomial operator yielding

$$\|p(T)\| = r(p(T)) = \sup_{t \in \sigma(p(T))} |t| = \max_{t \in p(\sigma(T))} |t| = \max_{s \in \sigma(T)} |p(s)|.$$

■

Corollary 5.3.2 generalizes Corollary 5.1.2 which states that $r(T) = \|T\|$ for the spectral radius of self-adjoint operators T .

Lecture 24: The Universal Spectral Mapping Theorem

5.3.4 Continuous Functions of Operators

22 Nov. 14:30

We're now going to study the spectral mapping theorem, which is a generalization of finite dimensions self-adjoint T on \mathbb{R}^n , where there exists an orthonormal basis of eigenvectors. In this basis, the action of t is just multiplication by scalars. To generalize this to infinite dimensions, we need measure theory. In particular, we need Weierstrass approximation theorem.

Theorem 5.3.3 (Weierstrass approximation theorem). Let $K \subseteq \mathbb{R}$ be compact and $f: K \rightarrow \mathbb{R}$ be continuous, then for any $\epsilon > 0$, there exists a polynomial p^ϵ such that

$$\sup_{x \in K} |f(x) - p^\epsilon(x)| < \epsilon.$$

Let T be self-adjoint on \mathcal{H} and $f: \sigma(T) \rightarrow \mathbb{R}$ be a continuous function, then we can define a self-adjoint operator $f(T)$. Then we use Weierstrass approximation theorem, and we find polynomials $p^{\epsilon_n}(t)$ such that

$$p^{\epsilon_n}(t) \rightarrow f(t)$$

uniformly on $\sigma(T)$. This suggests us to define $f(T)$ as the limit of operator polynomials $p^{\epsilon_n}(T)$.

Definition 5.3.4 (Continuous function operator). The sequence $p^{\epsilon_n}(T)$ converges in $\mathcal{L}(\mathcal{H})$ to a limit that we call $f(T) \in \mathcal{L}(\mathcal{H})$, which is also self-adjoint.

Note (Well-defined). To show that Definition 5.3.4 is well-defined, i.e., we need to show that $p^{\epsilon_n}(T)$ indeed converges to $f(T)$, and $f(T)$ is self-adjoint and does not depend on the choice of the approximating polynomials p^{ϵ_n} .

Proof. To construct $f(T)$, choose a sequence $p^{\epsilon_n}(\cdot)$ of polynomial such that $\epsilon_n \rightarrow 0$ and

$$\sup_{x \in \sigma(T)} |f(x) - p^{\epsilon_n}(x)| < \epsilon \Rightarrow \sup_{x \in \sigma(T)} |p^{\epsilon_n}(x) - p^{\epsilon_m}(x)| < \epsilon_n + \epsilon_m.$$

From Corollary 5.3.2, $\|p(T)\| = \sup_{t \in \sigma(T)} |p(t)|$, so we have

$$\|p^{\epsilon_n}(T) - p^{\epsilon_m}(T)\| < \epsilon_n + \epsilon_m,$$

so the sequence $\{p^{\epsilon_n}(T)\}_{n \geq 1}$ is in $\mathcal{L}(\mathcal{H})$ is Cauchy, so there exists a limit $f(T)$ with

$$\lim_{n \rightarrow \infty} \|p^{\epsilon_n}(T) - f(T)\| = 0,$$

which shows that $f(T) \in \mathcal{L}(\mathcal{H})$ is unique. ⊗

By passing to the limit in the corresponding properties for polynomial operators as in Proposition 5.3.1, we see that these properties of $f(T)$ inherited from properties of $p(T)$ when $p(\cdot)$ is a polynomial, e.g.,

- (a) $(af + bg)(T) = af(T) + bg(T)$;
- (b) $(fg)(T) = f(T)g(T)$;

(c) $f(T)^* = \overline{f}(T)$ for $f: \sigma(T) \rightarrow \mathbb{C}$.

Note. The first two are also true for continuous complex functions $f, g: \sigma(T) \rightarrow \mathbb{C}$.

Proof. If $f: \sigma(T) \rightarrow \mathbb{C}$ is complex, we can then write $f = f_1 + if_2$ for $f_1, f_2: \sigma(T) \rightarrow \mathbb{R}$ such that

$$f(T) := f_1(T) + if_2(T).$$

⊛

5.3.5 Spectral Mapping Theorem

We will now generalize the [spectral mapping theorem for polynomial operators](#) to [continuous functions of an operator](#). It is based on the straightforward generalization of the [invertibility lemma for polynomial operators](#).

Lemma 5.3.2 (Invertibility). Let $T \in \mathcal{L}(\mathcal{H})$ be a [self-adjoint operator](#) and $f \in C(\sigma(T))$. Then the operator $f(T)$ is invertible if and only if $f(t) \neq 0$ for all $t \in \sigma(T)$.

Proof. ■

Now the [spectral mapping theorem](#) follows from [invertibility lemma](#) by the same argument as the corresponding result for [polynomial operators](#), i.e., [Theorem 5.3.2](#).

Theorem 5.3.4 (Spectral mapping theorem). Let $T \in \mathcal{L}(\mathcal{H})$ be a [self-adjoint operator](#) and $f \in C(\sigma(T))$. Then

$$\sigma(f(T)) = f(\sigma(T)).$$

This gives a simple way to create [positive operators](#).

Corollary 5.3.3. Let $T \in \mathcal{L}(\mathcal{H})$ be a [self-adjoint operator](#) and $f \in C(\sigma(T))$. If $f(t) \geq 0$ for all $t \in \sigma(T)$, then $f(T) \geq 0$.

Proof. ■

5.3.6 Square Root of Operators

Consider a [positive self-adjoint operator](#) $T \in \mathcal{L}(\mathcal{H})$, then $\sigma(T) \subseteq [0, \infty)$. The function $f(t) = \sqrt{t}$ is continuous on $[0, \infty)$, so we can define $f(T) = \sqrt{T}$. A simple observation leads to the following.

Proposition 5.3.2 (Square root of operator). Let $T \in \mathcal{L}(\mathcal{H})$ be [positive self-adjoint](#), i.e., $\langle Tx, x \rangle \geq 0$ for $x \in \mathcal{H}$. Then there exists a unique [positive self-adjoint operator](#) $\sqrt{T} \in \mathcal{L}(\mathcal{H})$ such that

$$(\sqrt{T})^2 = T.$$

Proof. Let $f: \sigma(T) \rightarrow \mathbb{R}$ where $\sigma(T) \subseteq \mathbb{R}^+$ since T is [positive](#). And since $f(t) = \sqrt{t}$ is continuous on $\sigma(T)$, so we can define $f(T)$ by $f(t)^2 = t$, i.e., $f(T)^2 = T$.^a ■

^aThe uniqueness is left as an exercise.

5.3.7 Modulus of Operators

Now, consider an arbitrary operator $T \in \mathcal{L}(\mathcal{H})$, which is not necessarily [self-adjoint](#). Then T^*T is a [positive self-adjoint operator](#), so it has a unique [positive](#) square root. This suggests the following definition.

Definition 5.3.5 (Modulus). Let $T \in \mathcal{L}(\mathcal{H})$, then the *modulus* of T is defined as $|T| := \sqrt{T^*T}$.

This generalizes the concept of modulus of complex numbers, i.e., $|z| = \sqrt{\bar{z}z}$ for $z \in \mathbb{C}$.

Lemma 5.3.3. For every operator $T \in \mathcal{L}(\mathcal{H})$ and vector $x \in \mathcal{H}$, one has

$$\| |T|x \| = \|Tx\|.$$

Proof. Since

$$\| |T|x \|^2 = \langle |T|x, |T|x \rangle = \langle |T|^2 x, x \rangle = \langle T^*Tx, x \rangle = \langle Tx, Tx \rangle = \|Tx\|^2,$$

which leads to the polar decomposition theorem. ■

5.3.8 Polar Decomposition

Lemma 5.3.3 motivates us to consider a map

$$U: |T|x \mapsto Tx$$

for $x \in \mathcal{H}$.

Let $U \in \mathcal{L}(\mathcal{H})$ be an isometry, i.e., $\|Ux\| = \|x\|$ for all $x \in \mathcal{H}$. Then this implies U is injective but not necessarily surjective.

Lemma 5.3.4. The operator $U \in \mathcal{L}(\mathcal{H})$ is an isometry if and only if $U^*U = I$.

Proof. Suppose $U^*U = I$, then

$$\|Ux\|^2 = \langle Ux, Ux \rangle = \langle U^*Ux, x \rangle = \langle x, x \rangle = \|x\|^2,$$

so U is an isometry. Conversely, suppose $\|Ux\| = \|x\|$ for all $x \in \mathcal{H}$, then this implies $U^*U = I$ since with [polarization identity](#),

$$\begin{aligned} \langle U^*Ux, y \rangle &= \frac{1}{4} [\langle U^*U(x+y), x+y \rangle - \langle U^*U(x-y), x-y \rangle \\ &\quad + i \langle U^*U(x+iy), x+iy \rangle - i \langle U^*U(x-iy), x-iy \rangle], \end{aligned}$$

and by the assumption, $\langle U^*Uz, z \rangle = \langle z, z \rangle$ for all $z \in \mathcal{H}$, so $\langle U^*Ux, y \rangle = \langle x, y \rangle$ for all $x, y \in \mathcal{H}$, so $U^*U = I$. ■

Recall the [unitary operator](#), where equivalently, we have that $U \in \mathcal{L}(\mathcal{H})$ is [unitary](#) if U is an isometry and U is surjective. In other words, U is [unitary](#) if and only if $U^*U = I = UU^*$, i.e., $U^* = U^{-1}$.

Proposition 5.3.3. Let $U \in \mathcal{L}(\mathcal{H})$ be [unitary](#), then

$$\sigma(U) \subseteq \{\lambda \in \mathbb{C} : |\lambda| = 1\}.$$

Proof. Since $\|U\| = \|U^{-1}\| = 1$, we know that $\|U\| \leq 1$ implies $\rho(U)$ contains $\{\lambda \in \mathbb{C} : |\lambda| > 1\}$; and $\|U^{-1}\| \leq 1$ implies that $\rho(U^{-1})$ contains $\{\lambda \in \mathbb{C} : |\lambda| > 1\}$, so $\lambda \in \rho(U^{-1}) \Leftrightarrow \lambda^{-1} \in \rho(U)$. So,

$$(U^{-1} - \lambda I)^{-1} = \lambda^{-1}U(\lambda^{-1}I - U)^{-1}.$$

If $\lambda \neq 0$, then we can conclude that $\rho(U)$ contains $\{\lambda \in \mathbb{C} : |\lambda| \neq 1\}$. ■

Theorem 5.3.5 (Polar decomposition). For every $T \in \mathcal{L}(\mathcal{H})$, there exists a unique bijective linear isometry $U \in \mathcal{L}(\text{Im}(|T|), \text{Im } T)$ such that $T = U|T|$.

Proof. Define U on $\text{Im}(|T|)$ by $U(|T|x) = Tx$ for all $x \in \mathcal{H}$.^a Then, U is an isometry since

$|T| = \sqrt{T^*T}$, and U is injective and U maps onto $\text{Im}(T)$, so U is surjective. ■

^aNotice that this makes sense since $|Tx| = 0 \Leftrightarrow Tx = 0$.

Polar decomposition generalizes the polar decomposition in the complex plane. The latter states that every $z \in \mathbb{C}$ can be represented as

$$z = e^{i \arg(z)} |z|,$$

where $e^{i \arg(z)}$ is a unit scalar (generalized by U), and $|z|$ is the modulus of z (generalized by $|T|$).

Theorem 5.3.6 (Polar decomposition for invertible operator). If $T \in \mathcal{L}(\mathcal{H})$ is invertible, then operator U is **unitary**.

Proof. $\text{Im}(T) = \mathcal{H}$, hence $\text{Im}(|T|) = \mathcal{H}$. ■

5.4 Borel Functional Calculus

Theorem 5.4.1 (Borel functional calculus). Let $\sigma(T) \subseteq [m, M]$ and $\mathcal{B}([m, M])$ be the linear space of bounded Borel measurable functions $f: [m, M] \rightarrow \mathbb{C}$ such that

$$\|f\|_\infty = \sup_{t \in [m, M]} |f(t)| < \infty.$$

Then we can define a **self-adjoint operator** $f(T) \in \mathcal{L}(\mathcal{H})$ with the following properties.

- (a) If $f(\cdot)$ is real-valued, then $f(T)$ is **self-adjoint**.
- (b) If $f: [m, M] \rightarrow \mathbb{C}$, $f \in \mathcal{B}([m, M])$, then $\|f(T)\| \leq \|f\|_\infty$.
- (c) Suppose $f_n \in \mathcal{B}([m, M])$ for all $n \geq 1$ and $f \in \mathcal{B}([m, M])$ such that $\sup_{n \geq 1} \|f_n\|_\infty < \infty$ and $f_n \rightarrow f$ point-wise, i.e., $\lim_{n \rightarrow \infty} f_n(t) = f(t)$ for all $t \in [m, M]$. Then $f_n(T)$ **converges strongly** to $f(T)$, i.e.,

$$\lim_{n \rightarrow \infty} \|f_n(T)x - f(T)x\| = 0$$

for all $x \in \mathcal{H}$.

- (d) Suppose $T, S \in \mathcal{L}(\mathcal{H})$ are **self-adjoint** and commute, i.e., $TS = ST$, and assume further that $\sigma(T), \sigma(S) \subseteq [m, M]$ and $f, g \in \mathcal{B}([m, M])$. Then $f(T)$ and $g(S)$ commute, i.e., $f(T) \cdot g(S) = g(S) \cdot f(T)$.

Lecture 25: Proofs of Borel Functional Calculus Theorem

We start by proving **Theorem 5.4.1**.

Proof. Let's prove this one by one.

- (a) We construct $f(T)$ using **Riesz representation for $C(K)$** where $K = [m, M]$ is **compact** Hausdorff. For $x, y \in \mathcal{H}$, define functional $F_{x,y}: C(K) \rightarrow \mathbb{C}$ by

$$F_{x,y}(f) = \langle f(T)x, y \rangle$$

where $f \in C([m, M])$. Since $F_{x,y} \in C(K)^*$, we have

$$|F_{x,y}(f)| \leq \|f(T)\| \|x\| \|y\| = \|f\|_\infty \|x\| \|y\|,$$

so $\|F_{x,y}\| \leq \|x\| \|y\|$. Then **Riesz representation** implies that there exists a unique Borel measure $\mu_{x,y}$ on $[m, M]$ such that

$$\langle f(T)x, y \rangle = \int_m^M f(\lambda) d\mu_{x,y}(\lambda)$$

for $f \in C([m, M])$. Now, we extend $F_{x,y}$ to all Borel measurable functions $f: [m, M] \rightarrow \mathbb{R}$

with $\|f\|_\infty < \infty$ by

$$Bf(x, y) = \int_m^M f(\lambda) d\mu_{x,y}(\lambda).$$

Notice that $TV(\mu_{x,y}) \leq \|x\|\|y\|$ by [Riesz representation](#), and hence $|Bf(x, y)| \leq \|f\|_\infty \|x\|\|y\|$.

Note that the function $[x, y] \rightarrow Bf(x, y)$ is linear in x and [sesquilinear](#)^a in y . Hence, there exists $f(T) \in \mathcal{L}(\mathcal{H})$ such that

$$Bf(x, y) = \langle f(T)x, y \rangle$$

by [Riesz representation for Hilbert space](#), so $f(T)$ is defined uniquely by the sesquilinear property.

- (b) $\|f(T)\| \leq \|f\|_\infty$ follows from

$$|Bf(x, y)| \leq \|f\|_\infty \|x\|\|y\|$$

for $x, y \in \mathcal{H}$.

- (c) Consider $f_n, f: [m, M] \rightarrow \mathbb{R}$ such that $\sup_{n \geq 1} \|f_n\|_\infty < \infty$ with $f_n(t) \rightarrow f(t)$ for $t \in [m, M]$. Then we have

$$\langle f_n(T)x, y \rangle = \int_m^M f_n(\lambda) d\mu_{x,y}.$$

From the dominated convergence theorem, we have

$$\lim_{n \rightarrow \infty} \int_m^M f_n(\lambda) d\mu_{x,y} = \int_m^M f(\lambda) d\mu_{x,y},$$

hence

$$\lim_{n \rightarrow \infty} \langle f_n(T)x, y \rangle = \langle f(T)x, y \rangle$$

for all $x, y \in \mathcal{H}$, which is saying that $f_n(T)$ [converges weakly](#) to $f(T)$. To show [strong convergence](#), note that

$$\begin{aligned} \|(f_n(T) - f(T))x\|^2 &= \langle (f_n(T) - f(T))x, (f_n(T) - f(T))x \rangle \\ &= \langle (f_n(T) - f(T))^2 x, x \rangle \\ &= \int_m^M (f_n(t) - f(t))^2 d\mu_{x,x} \rightarrow 0 \end{aligned}$$

by dominated convergence theorem. Hence, we conclude that

$$\lim_{n \rightarrow \infty} \|(f_n(T) - f(T))x\| = 0$$

for all $x \in \mathcal{H}$.

- (d) The commutativity follows from two approximation arguments. Firstly, it holds if f, g are polynomials, so by [Weierstrass approximation theorem](#), this holds for continuous $f, g: [m, M] \rightarrow \mathbb{R}$. Then, if this holds for continuous functions, it also follows for Borel functions by approximating Borel functions with continuous functions.^b

■

^aRecall that if $f(T)$ is sesquilinear, then $Bf(x, y) = \overline{Bf(y, x)}$.

^bThis is the so-called [Lusin's theorem](#).

5.4.1 Spectral Measures

From [Borel functional calculus](#), we have

$$\langle Tx, y \rangle = \int_m^M \lambda d\mu_{x,y}(\lambda)$$

for $x, y \in \mathcal{H}$. We can abstract this result to construct integrals with respect to spectral projections.

Let $E \subseteq [m, M]$ be a Borel set, and $\mathbb{1}_E$ be the indicator function such that

$$\mathbb{1}_E(\lambda) = \begin{cases} 1, & \text{if } \lambda \in E; \\ 0, & \text{if } \lambda \notin E. \end{cases}$$

Set $P_E = \mathbb{1}_E(T)$, from $(\mathbb{1}_E)^2 = \mathbb{1}_E$, we have $P_E^2 = P_E$, so P_E is indeed a projection.

Proposition 5.4.1. The spectral projections $E \rightarrow P_E$ for $E \in \mathcal{B}([m, M])$ satisfies $P_{[m, M]} = I$. Furthermore, if $E = \bigcup_{k=1}^{\infty} E_k$ and E_k are disjoint for all $k \geq 1$, then

$$P_E = \sum_{k=1}^{\infty} P_{E_k}$$

where convergence is in the **strong** sense.

Proof. It follows **Theorem 5.4.1** where if $f_n \rightarrow f$, then $f_n(T) \rightarrow f(T)$ **strongly**. ■

Remark. The mapping $E \rightarrow P_E$ for $E \in \mathcal{B}([m, M])$ is an operator valued measures.

Theorem 5.4.2 (Spectral theorem). Let $T \in \mathcal{L}(\mathcal{H})$ be **self-adjoint** and $E \rightarrow P_E$ be the associated spectral measure. Then

$$T = \int_{-\infty}^{\infty} \lambda \, dP_{\lambda}.$$

Proof. This is just a way of interpretation, i.e.,

$$\langle Tx, y \rangle = \int_{-\infty}^{\infty} \lambda \, [dP_{\lambda}x, y]$$

for $x, y \in \mathcal{H}$ where $[dP_{\lambda}x, y] = d\mu_{x, y}$. ■

Chapter 6

Epilogue

Lastly, we prove some left-out theorems, start with the [Riesz representation for \$C\(K\)\$](#) .

6.1 Proof of Riesz Representation for $C(K)$

This section is the proof about [Riesz representation for \$C\(K\)\$](#) . Let's first restate the theorem.

As previously seen (Riesz representation for $C(K)$). Let $E = C(K)$ be the space of continuous functions on [compact Hausdorff space \$K\$](#) . Then we have the following.

- (a) For every Borel regular signed measure on K , the [functional](#) $F(f) = \int_K f d\mu$ is a [bounded linear functional](#) on K .
- (b) Every [bounded linear functional](#) on $C(K)$ can be expressed as $F(f) = \int_K f d\mu$ for some measure μ , and $\|F\| = |\mu|(K)$, i.e., $TV(K)$.

Let's first outline the proof. The main theorem we're going to use is the [Urysohn's lemma](#) for the construction of continuous functions. [Urysohn's lemma](#) allows us to construct means of [positive linear functional](#) on $C(X)$ where X is [locally compact Hausdorff](#). Then, let $\Lambda: C(X) \rightarrow \mathbb{R}$ where Λ is linear and $\Lambda(f) \geq 0$ if $f(x) \geq 0$ for all $x \in X$. Now, we define means of an open set $V \subseteq X$ where

$$\mu(V) = \sup \{ \Lambda(f) : 0 \leq f \leq \mathbb{1}_V \}.$$

Then, we define μ for any subset $E \subseteq X$,

$$\mu(E) = \inf \{ \mu(V) : E \subseteq V, V \text{ open} \},$$

where μ is the outer measure, i.e., subadditive.

Lecture 26: Urysohn's Lemma

Now, we build the foundation toward proving the [Urysohn's lemma](#).

1 Dec. 14:30

Definition 6.1.1 (Topology). Given a nonempty set X of points, a family \mathcal{T} of subsets (the open sets) is called a *topology* if it satisfies the following.

- (a) \mathcal{T} contains X and the empty set \emptyset .
- (b) If $O_1, O_2 \in \mathcal{T}$, $O_1 \cap O_2 \in \mathcal{T}$.
- (c) If $O_d \in \mathcal{T}$ for $d \in \mathcal{F}$, $\bigcup_{d \in \mathcal{F}} O_d \in \mathcal{T}$.

Definition 6.1.2 (Topological space). A *topological space* (X, \mathcal{T}) is a nonempty set X of points together with a [topology](#) \mathcal{T} .

Definition 6.1.3 (Hausdorff). A **topological space** (X, \mathcal{T}) is *Hausdorff* if given two distinct points $x_1, x_2 \in X$, there exists disjoint open sets $O_1, O_2 \in \mathcal{T}$ such that $x_1 \in O_1, x_2 \in O_2$.

Definition 6.1.4 (Locally compact). A **Hausdorff** space (X, \mathcal{T}) is *locally compact* if for every $x \in X$, there exists open $O \in \mathcal{T}$ such that $x \in O$ and the closure \overline{O} of O is **compact**.

Remark (Closure). Formally, given a **topological space** (X, \mathcal{T}) , the closure \overline{O} of an open set $O \in \mathcal{T}$ is defined as

$$\overline{O} := \bigcap_{F \supseteq O} F,$$

where F is closed (complement of some open set).

Theorem 6.1.1. Let (X, \mathcal{T}) be a **topological space** and $K \subseteq X$ being **compact**, i.e., every open cover of K has a finite subcover. Suppose F is closed and $F \subseteq K$, then F is also compact.

Proof. If $\{V_\alpha\}_{\alpha \in \mathcal{F}}$ is an open cover of F , then $\{V_\alpha\}_{\alpha \in \mathcal{F}}$ and $X \setminus F$ is an open cover of K . Since K is **compact**, we can find a finite subcover of K , which also covers F since $F \subseteq K$, hence F is **compact** from definition. ■

Theorem 6.1.2. Suppose X is **Hausdorff**, $K \subseteq X$ being **compact** and $p \notin K$. Then there exists disjoint open sets U, W such that $p \in U, K \subseteq W$.

Proof. By the **Hausdorff** property, for every $q \in K$, there exists open sets U_p, V_q such that $p \in U_p, q \in V_q$ with U_p, V_q disjoint. Then $\{V_q\}_{q \in K}$ is an open cover of K , so we have a finite subcover V_{q_1}, \dots, V_{q_n} of K . Now, take

$$W = \bigcup_{j=1}^n V_{q_j} \supseteq K, \quad U = \bigcap_{j=1}^n U_{p_j} \ni p,$$

we have that U, W being open and disjoint. ■

Corollary 6.1.1. **Compact** subsets of **Hausdorff space** are closed.

Corollary 6.1.2. If F is closed and K is **compact**, then $F \cap K$ is compact.

Theorem 6.1.3. If $\{K_\alpha\}_{\alpha \in \mathcal{F}}$ is a collection of **compact** subsets of a **Hausdorff space** such that $\bigcap_{\alpha \in \mathcal{F}} K_\alpha$ is empty, then some finite subcollection of $K_\alpha, \alpha \in \mathcal{F}$, has empty intersection.

Proof. Consider any K_{α_0} with $\alpha_0 \in \mathcal{F}$. Then

$$K_{\alpha_0} \subseteq \bigcup_{\substack{\alpha \in \mathcal{F} \\ \alpha \neq \alpha_0}} (X \setminus K_\alpha).$$

This is an open cover of K_{α_0} , so there exists a finite subcovers, i.e.,

$$K_{\alpha_0} \subseteq (X \setminus K_{\alpha_1}) \cup \dots \cup (X \setminus K_{\alpha_n}),$$

leading to the fact that $\bigcap_{j=0}^n K_{\alpha_j} = \emptyset$. ■

Theorem 6.1.4. Let X be a **locally compact Hausdorff** space, and U is open, K is **compact** and $K \subseteq U$. Then there exists an open set V such that \overline{V} is **compact** and $K \subseteq V \subseteq \overline{V} \subseteq U$.

Proof. By the **locally compactness** property, every point of K has an open neighborhood with **compact** closure. Since K is covered by a finite union of these open neighborhoods, hence $K \subseteq G \subseteq \overline{G}$ and \overline{G} is **compact**. We see that if $U = X$ we're done since we can take $V = G$. Otherwise, note that for each $p \in X \setminus U$, $p \notin K$, so there exists open W_p and $p \notin \overline{W_p}$. Now, consider the family $(X \setminus U) \cap \overline{G} \cap \overline{W_p}$, $p \in X \setminus U$, with \overline{G} **compact**, we see that this is a family of **compact sets** with empty intersection. This means there is a finite number of these p_1, \dots, p_n have empty intersection, so by taking $V = G \cap W_{p_1} \cap \dots \cap W_{p_n}$,

$$K \subseteq V \subseteq \overline{V} \subseteq U.$$

■

Definition. Let X be a **topological space** and $f: X \rightarrow \mathbb{R}$.

Definition 6.1.5 (Lower semi-continuous). If $\{x \in X: f(x) > \alpha\}$ is open for all $\alpha \in \mathbb{R}$, then f is *lower semi-continuous*.

Definition 6.1.6 (Upper semi-continuous). If $\{x \in X: f(x) < \alpha\}$ is open for all $\alpha \in \mathbb{R}$, then f is *upper semi-continuous*.

Remark. A real function $f: X \rightarrow \mathbb{R}$ is continuous if and only if it is both **upper** and **lower semi-continuous**.

Remark. The characteristic functions of open sets are **lower semi-continuous**, while the characteristic functions of closed sets are **upper semi-continuous**.

Remark. The supremum of a family of **lower semi-continuous** is **lower semi-continuous**; while the infimum of a family of **upper semi-continuous** is **upper semi-continuous**.

Remark. $f(\cdot)$ is **lower semi-continuous** if $\{x \in X: f(x) \leq \alpha\}$ is closed for all $\alpha \in \mathbb{R}$. For metric spaces X , this is equivalent to

$$f(x) \leq \liminf_{x_n \rightarrow x} f(x_n).$$

Definition 6.1.7 (Support). Let (X, \mathcal{T}) be a **topological space**. The *support* of a function $f: X \rightarrow \mathbb{R}$ is the closure of the set $\{x \in X: f(x) \neq 0\}$.

Notation. The collection of all continuous functions on X with **compact support** is denoted as $C_c(X)$.

Theorem 6.1.5. Let X, Y be **topological spaces** and $f: X \rightarrow Y$ be continuous. If $K \subseteq X$ is **compact**, then $f(K)$ is **compact** in Y .

Proof. The open cover $\{O_\alpha\}_{\alpha \in \mathcal{F}}$ of $f(K)$ induces an open cover $\{f^{-1}(O_\alpha)\}_{\alpha \in \mathcal{F}}$ of K . Since K is **compact**, we can find a finite subcover $f^{-1}(O_{\alpha_1}), \dots, f^{-1}(O_{\alpha_n})$ of K , i.e.,

$$f(K) \subseteq O_{\alpha_1} \cup \dots \cup O_{\alpha_n}.$$

■

Remark. The range of any $f \in C_c(X)$ is **compact**, i.e., $f(X)$ is **compact**.

Theorem 6.1.6 (Urysohn's lemma). Let X be a **locally compact Hausdorff space** and V open in X , $K \subseteq V$ **compact**. Then there exists $f \in C_c(X)$ such that

$$\chi_K \leq f \leq \chi_V,$$

i.e., $f(x) = 1$ for $x \in K$, $0 \leq f(y) \leq 1$ for $y \in X$ and $f(y) = 0$ for $y \notin V$.

Proof. Set $r_1 = 0$, $r_2 = 1$, and let r_3, r_4 be any enumeration of the rational number r with $0 < r < 1$. By **Theorem 6.1.5**, we can find open sets V_0, V_1 with

$$K \subseteq V_1 \subseteq \bar{V}_1 \subseteq V_0 \subseteq \bar{V}_0 \subseteq V$$

such that \bar{V}_0 is **compact**. Now, define a sequence V_r for rationals r , $0 < r < 1$. Suppose $n \geq 2$ and $V_{r_1}, V_{r_2}, \dots, V_{r_n}$ have already been chosen such that if $r_i < r_j$, we have $\bar{V}_{r_j} \subseteq V_{r_i}$ and \bar{V}_{r_j} **compact** and V_{r_i} open. Let r_{n+1} be the next in enumerations of the rationals. Choose $V_{r_{n+1}}$ open with $\bar{V}_{r_{n+1}}$ **compact**. One of the

Hence, $\bar{V}_{r_j} \subseteq V_{r_i}$, where \bar{V}_{r_j} is **compact** and V_{r_i} is open. Let $V_{r_{n+1}}$ be open, $\bar{V}_{r_{n+1}}$ be **compact**, we have

$$\bar{V}_{r_j} \subseteq V_{r_{n+1}} \subseteq \bar{V}_{r_{n+1}} \subseteq V_{r_i}.$$

Continuing have by induction a countable set V_r , $0 \leq r \leq 1$, r , V_r is open and \bar{V}_r is **compact**, we have $\bar{V}_r \subseteq V_s$ if $r > s$, r rational. For each rational r , define function f_r

$$f_r(x) = \begin{cases} r, & \text{if } x \in V_r; \\ 0, & \text{otherwise,} \end{cases}$$

so f_r is **lower semi-continuous**; also, we define g_s

$$g_s(x) = \begin{cases} 1, & \text{if } x \in \bar{V}_s; \\ s, & \text{otherwise,} \end{cases}$$

so g_s is **upper semi-continuous**. Define $f = \sup_r f_r$, we know that f is **lower semi-continuous**; also, define $g = \inf_s g_s$, g is **upper semi-continuous**. Note that $f(x) = 1$ for $x \in K$, $f(x) = 0$ for $x \notin \bar{V}_0$ with $0 \leq f \leq 1$; while $g(x) = 1$ for $x \in K$, $g(x) = 0$ for $x \notin \bar{V}_0$. Suppose $f \equiv g$, hence f is continuous.

Claim. $f = g$

Proof. Suppose $f_r(x) > g_s(x)$, then $r > s$, implying $x \in V_r$. Also, $\bar{V}_s \subseteq V_s$, so $x \in V_s$ implies $g_s(x) = 1$, contradiction, hence $f \leq g$.

On the other hand, suppose $f(x) < g(x)$, then there exists rational r, s such that $0 < r < s < 1$ such that $f(x) < r < s < g(x)$. Since $f(x) < r$, $x \notin V_r$; and since $g(x) > s$, $x \in \bar{V}_s$, hence $g_s(x) = s$. Also, we have $\bar{V}_s \subseteq V_s$ since $s > 1$, which is a contradiction. \otimes

■

Lecture 27: Riesz Representation Theorem

We will see that the **Urysohn's lemma** leads to the construction of partition of unity.

6 Dec. 14:30

Theorem 6.1.7. Let V_1, \dots, V_n be open subsets of a **locally compact Hausdorff space** X , and $K \subseteq \bigcup_{j=1}^n V_j$ be **compact**. Then there exists functions h_i for $i = 1, \dots, n$ such that $h_i \in C_c(X)$, and $0 \leq h_i \leq 1$ for all i , with $\text{supp}(h_i) \subseteq V_i$ for all i . Moreover, we have

$$h_1(x) + h_2(x) + \dots + h_n(x) = 1$$

for $x \in K$.

Proof. Let $x \in K$, then $x \in V_i$ for some $i = i(x)$. Then there exists open set W_x such that

$$x \in W_x \subseteq \overline{W}_x \subseteq V_{i(x)}$$

by the previous result. Then, $\{W_x\}_{x \in K}$ is an open cover of K , hence there exists a finite subcover

$$K \subseteq W_{x_1} \cup W_{x_2} \cup \dots \cup W_{x_N}.$$

Now, for $1 \leq i \leq n$. Let H_i be the union of the W_{x_j} such that $\overline{W}_{x_j} \subseteq V_i$, so we have $\overline{H}_i \subseteq V_i$ for all $i = 1, \dots, n$, so $K \subseteq \bigcup_{i=1}^n H_i$. From [Urysohn's lemma](#), there exists g_i such that

$$\chi_{\overline{H}_i} \leq g_i \leq \chi_{V_i}$$

for $i = 1, 2, \dots, n$. Define the partition of unity h_1, \dots, h_n by

$$\begin{aligned} h_1 &= g_1, \\ h_2 &= (1 - g_1)g_2, \\ &\vdots \\ h_n &= (1 - g_1)(1 - g_2) \dots (1 - g_{n-1})g_n, \end{aligned}$$

hence $0 \leq h_i \leq \chi_{V_i}$ for all $i = 1, \dots, n$. Moreover, we have

$$h_1 + h_2 + \dots + h_n = 1 - (1 - g_1)(1 - g_2) \dots (1 - g_n).$$

We have $K \subseteq H_1 \cup \dots \cup H_n$, and when $x \in K$, $g_i(x) = 1$ for some i , i.e., $h_1(x) + \dots + h_n(x) = 1$. ■

Theorem 6.1.8 (Riesz representation theorem). Let X be [locally compact Hausdorff](#) space, and Λ be a positive linear functional on $C_c(X)$, i.e., $\Lambda(\cdot)$ is a linear functional on $C_c(X)$ and $\Lambda(f) \geq 0$ if $f(x) \geq 0$ for $x \in X$. Then there exists a σ -algebra \mathcal{M} in X which contains all Borel sets of X , and there exists a unique positive measure μ on \mathcal{M} which represents Λ in the following sense.

- (a) $\Lambda(f) = \int_X f \, d\mu$ for $f \in C_c(X)$.
- (b) $\mu(K) < \infty$ for all [compact](#) K .
- (c) $\mu(E) = \inf \{\mu(V) : E \subseteq V, V \text{ open}\}$.
- (d) The relation $\mu(E) = \sup \{\mu(K) : K \subseteq E, K \text{ compact}\}$ holds for every open set E , and for every $E \in \mathcal{M}$ with $\mu(E) < \infty$.

Proof. Let's first prove the uniqueness. (c) and (d) imply that the measure μ is determined by its values on [compact](#) sets K , so it's sufficient to prove that if μ_1, μ_2 are two such measures, $\mu_1(K) = \mu_2(K)$ for all [compact](#) K . From (c), for any [compact](#) $K \subseteq V$ and $\epsilon > 0$, there exists an open V such that $K \subseteq V$ with $\mu(V) < \mu(K) + \epsilon$. From [Urysohn's lemma](#), there exists $f \in C_c(X)$ such that $\chi_K \leq f \leq \chi_V$. Note that

$$\mu_1(K) = \int_X \chi_K \, d\mu_1 \leq \int_X f \, d\mu_1 = \Lambda(f) = \int_X f \, d\mu_2 \leq \int_X \chi_V \, d\mu_2 = \mu_2(V),$$

so $\mu_1(K) \leq \mu_2(V) < \mu_2(K) + \epsilon$. Let $\epsilon \rightarrow 0$, $\mu_1(K) \leq \mu_2(K)$, and similarly, $\mu_1(K) \geq \mu_2(K)$, so $\mu_1(K) = \mu_2(K)$.

To construct μ and \mathcal{M} , for every open set V in X , we define

$$\mu(V) = \sup \{\Lambda(f) : f \in C_c(X), 0 \leq f < \chi_V\}.$$

Then, define $\mu(E)$ for all subsets $E \subseteq X$ such that $\mu(E) = \inf \{\mu(V) : E \subseteq V\}$.

Note. Note that these two construction are consistent, i.e., if E is open, $\mu(E)$ is given by the first one.

Proof. Since from the first supremum definition, open sets V_1, V_2 such that $V_1 \subseteq V_2$, it implies $\mu(V_1) \leq \mu(V_2)$. \circledast

To establish additivity for the measure μ , we need to restrict to some σ -algebra of subsets of X . This is analogous of first defining outer measure and then the actual measure. To define \mathcal{M} , we first define \mathcal{M} as the class of all subsets $E \subseteq X$ such that $\mu(E) < \infty$, and

$$\mu(E) = \sup \{ \mu(K) : K \subseteq E, K \text{ compact} \}.$$

Then \mathcal{M} is given by the class of all subsets $E \subseteq X$ such that $E \cap K \in \widetilde{\mathcal{M}}$ for every compact K . We now want to show μ is a measure on \mathcal{M} .

Note. μ is monotone, i.e., $\mu(A) \leq \mu(B)$ for $A \subseteq B$. Also, $\mu(E) = 0$ implies $E \in \mathcal{M}$.

We then use the monotonicity of $\Lambda(\cdot)$ to prove the remaining properties, i.e., $f \leq g \Rightarrow \Lambda(f) \leq \Lambda(g)$, i.e., $g - f \geq 0 \Rightarrow \Lambda(g - f) = \Lambda(g) - \Lambda(f) \geq 0$.

For subadditivity, let E_n be subsets of X for all $n \geq 1$, then we want

$$\mu \left(\bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{n=1}^{\infty} \mu(E_n).$$

Lecture 28

8 Dec. 14:30

Proof of Theorem 6.1.8 (Cont.) Firstly, we show that $\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2)$ if V_1, V_2 are open. To do this, let $g \in C_c(X)$ and $0 \leq g \leq \chi_{V_1 \cup V_2}$. Recall that $\mu(V) = \sup_{0 \leq g \leq \chi_{V_1 \cup V_2}} \Lambda(g)$. Let $K = \text{supp}(g)$ such that K is compact and $K \subseteq V_1 \cup V_2$. Choose a partition of unity h_1, h_2 such that $0 \leq h_1, h_2 \leq 1$ and $0 \leq h_1 \leq \chi_{V_1}$, $0 \leq h_2 \leq \chi_{V_2}$ with $h_1(x) + h_2(x) = 1$ for $x \in K$ from the Urysohn's lemma. Then $g := h_1g + h_2g$, we have

$$\Lambda(g) = \Lambda(h_1g) + \Lambda(h_2g),$$

with $0 \leq h_1g \leq \chi_{V_1}$ and $0 \leq h_2g \leq \chi_{V_2}$, we have $\Lambda(h_1g) \leq \mu(V_1)$ and $\Lambda(h_2g) \leq \mu(V_2)$, hence

$$\Lambda(g) \leq \mu(V_1) + \mu(V_2)$$

for all $g \in C_c(X)$ such that $0 \leq g \leq \chi_{V_1 \cup V_2}$. Taking the supremum over g , we then get

$$\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2)$$

if V_1, V_2 are open. For general E_n , $n \geq 1$, subsets of X , we can assume that $\mu(E_n) < \infty$ for all $n \geq 1$. Hence, there exists open V_n such that $E_n \subseteq V_n$ such that

$$\mu(E_n) \leq \mu(V_n) + \frac{\epsilon}{2^n}$$

for $n \geq 1$. Let $V = \bigcup_{i=1}^{\infty} V_i$, which is open, then $E = \bigcup_{n=1}^{\infty} E_n \subseteq V$, i.e., $\mu(E) \leq \mu(V)$. Let $f \in C_c(X)$ and $0 \leq f < \chi_V$, we have $\mu(V) = \inf \{ \Lambda(f) : 0 \leq f < \chi_V \}$. Then let $K = \text{supp}(f)$ such that K is compact, the sets V_j for $j \geq 1$ is an open covering of K , hence there exists a finite subcover $K \subseteq \bigcup_{j=1}^N V_j$ for some $N < \infty$. This implies that

$$\Lambda(f) \leq \mu \left(\bigcup_{j=1}^N V_j \right) \leq \sum_{j=1}^N \mu(V_j).$$

Since $\mu(V_j) \leq \mu(E_j) + \epsilon/2^j$, we have

$$\Lambda(f) \leq \sum_{j=1}^N \mu(E_j) + \frac{\epsilon}{2^j} \leq \sum_{j=1}^{\infty} \mu(E_j) + \epsilon.$$

Take supremum over all f with $0 \leq f < \chi_V$, we have $\mu(V) \leq \sum_{j=1}^{\infty} \mu(E_j) + \epsilon$, i.e.,

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu(E_i) + \epsilon.$$

By letting $\epsilon \rightarrow 0$, we have subadditivity.

Recall that $\widetilde{\mathcal{M}}$ contains all subsets E of X such that $\mu(E) < \infty$. And

$$\mu(E) = \sup \{ \mu(K) : K \subseteq E, K \text{ compact} \}.$$

In particular, $\widetilde{\mathcal{M}}$ contains all compact subsets K of X and all open subsets V of X such that $\mu(V) < \infty$. Let $K \subseteq X$ and compact, suppose $f \in C_c(X)$ with $\chi_K \leq f$, with $V = \{x \in X : f(x) > 1/2\}$, which is open and $K \subseteq V$. Furthermore, let $g \in C_c(X)$ such that $0 \leq g < \chi_V$, hence $g \leq 2f$, which implies

$$\mu(K) \leq \mu(V) = \sup \{ \Lambda(g) : 0 \leq g < \chi_V \} \leq \Lambda(2f) < \infty.$$

We then conclude that $\mu(K) < \infty$, i.e., $K \in \widetilde{\mathcal{M}}$. Next, we show that if V is open and $\mu(V) < \infty$, then $\mu(V) = \sup \{ \mu(K) : K \subseteq V, K \text{ compact} \}$. This is clear if $\mu(V) = 0$, so assume $\mu(V) > 0$. Let α satisfy $0 < \alpha < \mu(V)$, then there exists $f \in C_c(X)$ such that $0 \leq f < \chi_V$, we have $\alpha < \Lambda(f)$. Set $K = \text{supp}(f)$, so K is compact. Then if W is open and $K \subseteq W$ have $\alpha < \Lambda(f) \leq \mu(W)$, so $\alpha < \mu(W)$ for all W containing K . With $\mu(K) = \inf \{ \mu(W) : K \subseteq W \}$, $\alpha \leq \mu(K)$ for any $\alpha < \mu(V)$. This leads to

$$\mu(V) = \sup \{ \mu(K) : K \subseteq V, K \text{ compact} \}.$$

Now, for $E \subseteq X$, $E \in \widetilde{\mathcal{M}}$ if $\mu(E) < \infty$ and $\mu(E) = \sup \{ \mu(K) : K \subseteq E, K \text{ compact} \}$, we have shown that if K is compact, we have $K \in \widetilde{\mathcal{M}}$; while if V is open and $\mu(V) < \infty$, we have $V \in \widetilde{\mathcal{M}}$. We now show the additivity, i.e., let E_1, E_2, \dots be in $\widetilde{\mathcal{M}}$ and the E_j , $j \geq 1$ disjoint, we want

$$\mu(E) = \sum_{j=1}^{\infty} \mu(E_j).$$

If in addition $\mu(E) < \infty$, then $E \in \overline{\mathcal{M}}$. First we show that if K_1, K_2, \dots is compact and disjoint, then $\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2)$. Let $K = K_1$, $U = X - K_2$, then $K \subseteq U$. By Theorem 6.1.4, there exists open V_1 such that

$$K \subseteq V_1 \subseteq \overline{V_1} \subseteq U.$$

Set $V_2 = X - \overline{V_1}$, then we have $K_1 \subseteq V_1$, $K_2 \subseteq V_2$ such that V_1 and V_2 are disjoint. For $\epsilon > 0$, let W be open, $K_1 \cup K_2 \subseteq W$, we have

$$\mu(W) < \mu(K_1 + K_2) + \epsilon.$$

With $W \cap V_1$ and $W \cap V_2$ are disjoint, there exists functions $f_1, f_2 \in C_c(X)$ such that

$$0 \leq f_1 < \chi_{W \cap V_1}, \quad 0 \leq f_2 < \chi_{W \cap V_2}$$

with

$$\Lambda(f_1) > \mu(W \cap V_1) - \epsilon, \quad \Lambda(f_2) > \mu(W \cap V_2) - \epsilon.$$

Then

$$\begin{aligned}
\mu(K_1) + \mu(K_2) &\leq \mu(W \cap V_1) + \mu(W \cap V_2) \\
&< \Lambda(f_1) + \Lambda(f_2) + 2\epsilon \\
&= \Lambda(f_1 + f_2) + 2\epsilon \\
&\leq \mu(W) + 2\epsilon \\
&< \mu(K_1 \cup K_2) + 3\epsilon
\end{aligned}$$

since V_1, V_2 are disjoint, with $0 \leq f_1 + f_2 < \chi_W$. Let $\epsilon \rightarrow 0$, we have

$$\mu(K_1) + \mu(K_2) \leq \mu(K_1 \cup K_2),$$

and from subadditivity, $\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2)$. Now, we show $\mu(E) = \sum_{j=1}^{\infty} \mu(E_j)$ where $E = \bigcup_{j=1}^{\infty} E_j$ such that $E_j \in \widetilde{\mathcal{M}}$ are disjoint. We may assume that $\mu(E) < \infty$, i.e., $\mu(E_i) < \infty$ for all i since $E_i \subseteq E$. Since $E_i \in \widetilde{\mathcal{M}}$, there exists a [compact](#) $H_i \subseteq E_i$ such that $\mu(H_i) > \mu(E_i) - \epsilon/2^i$ for all $i \geq 1$. Then, form $\bigcup_{i=1}^n H_i \subseteq E$, we have

$$\mu(E) \geq \mu\left(\bigcup_{i=1}^n H_i\right) = \sum_{i=1}^n \mu(H_i)$$

from finite additivity for [compact sets](#). Let $n \rightarrow \infty$,

$$\mu(E) \geq \sum_{i=1}^{\infty} \mu(H_i) \geq \sum_{i=1}^{\infty} \mu(E_i) - \frac{\epsilon}{2^i} = \sum_{i=1}^{\infty} \mu(E_i) - \epsilon,$$

by letting $\epsilon \rightarrow 0$, we have $\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$.

Finally, we want to show that if $\mu(E) < \infty$, then

$$\mu(E) = \sup \{ \mu(K) : K \subseteq E, K \text{ [compact](#) } \},$$

which further implies $E \in \widetilde{\mathcal{M}}$. We have shown that

$$\mu(E) = \sum_{i=1}^{\infty} \mu(E_i),$$

so if $\mu(E) < \infty$, for any $\epsilon > 0$, there exists an $N = N(\epsilon)$ such that

$$\sum_{i=1}^N \mu(E_i) \geq \mu(E) - \epsilon.$$

Since $\mu(H_i) \geq \mu(E_i) - \epsilon/2^i$ for all $i \geq 1$ for $H_i \subseteq E_i$ being [compact](#), this implies

$$\mu\left(\bigcup_{i=1}^N H_i\right) \geq \mu(E) - \epsilon,$$

hence we're done. ■

Appendix

Appendix A

Review

A.1 Midterm Review

A.1.1 Normed Spaces

Recall the [normed spaces](#), and the properties of which. In particular, focus on [convexity](#) and note that $x \mapsto \|x\|$ is a [convex function](#).

Example (Normed spaces). The spaces ℓ_p for $1 \leq p \leq \infty$ of sequences and $L^p(\Omega, \mathcal{F}, \mu)$ of integrable functions. Also, the space of continuous functions on compact [Hausdorff space](#) with supremum norm $C(K)$. Notice that

$$C(K) \subseteq L^\infty(K, \mathcal{F}).$$

Remark (Legendre transform). Recall the Legendre transform of [convex functions](#). The most general form is that let X be a [Banach space](#) and X^* its [dual space](#) with a [convex function](#) $f: X \rightarrow \mathbb{R}$ and $f^*: X^* \rightarrow \mathbb{R}$. We have

$$f^*(y^*) = \sup_{x \in X} [y^*(x) - f(x)].$$

Notice that f^* is [convex](#) and lower semi-continuous where $f^*: X^* \rightarrow \mathbb{R} \cup \{+\infty\}$.

A.1.2 Quotient Spaces

Let X be a [normed space](#) and E be a subspace of X . Then $X/E = \{[x] = x + E : x \in X\}$ if E is closed, then X/E is also a [normed space](#) with the [norm](#) $\|[x]\| := \inf_{y \in E} \|x - y\|$.

Remark. E need to be closed since we need $\|[x]\| = 0 \Rightarrow [x] = 0$.

A.1.3 Banach Spaces

Any [normed space](#) E can be completed to a [Banach space](#) \hat{E} by [Theorem 1.4.2](#).

Example. ℓ_p and L^p are [Banach spaces](#). For $x \in \ell_p$, $x = \{x_n, n \geq 1\}$ with

$$\|x\|_p = \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{1/p}.$$

Notice that [Minkowski inequality](#) is the triangle inequality for ℓ_p and L^p , and we can prove this using [Hölder's inequality](#) where we have

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

for $1/p + 1/q = 1$.

Remark (Proof of completeness of the ℓ_p spaces). This is easy for ℓ_p , but for L^p , we need to use **dominated convergence theorem**.

A.1.4 Inner Product Spaces and Hilbert Spaces

Notice that the **Hilbert spaces** are the completion of **inner product spaces**. Recall the **parallelogram law**

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

and the **Cauchy-Schwarz inequality**

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

Orthogonality

Recall the **orthogonal projection** P_E onto a closed subspace $E \subseteq \mathcal{H}$ is $P_E x = x(y)$ where $x(y)$ is the minimizer of $\min_{y \in E} \|x - y\|$.

Remark. P_E is the projection, i.e., $P_E^2 = P_E$, and $I - P_E$ is projection onto the **orthogonal complement** E^\perp of E in \mathcal{H} such that $\mathcal{H} = E \oplus E^\perp$. We see that

$$\|x\|^2 = \|P_E x\|^2 + \|(I - P_E)x\|^2$$

for $x \in \mathcal{H}$.

Consider the **orthogonal** or **orthonormal** sets of vectors x_k , $k = 1, 2, \dots$ in \mathcal{H} with the corresponding **Fourier series** being

$$S_n(x) := \sum_{k=1}^n \langle x, x_k \rangle x_k$$

such that

$$\|S_n(x)\|^2 = \sum_{k=1}^n |\langle x, x_k \rangle|^2.$$

If the set $\{x_k\}_{k=1}^\infty$ is **orthonormal**, then $S_n = P_{E_n}$ where E_n is the span of $\{x_1, \dots, x_n\}$, and

$$\|S_n x\|^2 = \|P_{E_n} x\|^2 \leq \|x\|^2,$$

which is the **Bessel's inequality**.

Remark. $S_n x \rightarrow S_\infty x$ in \mathcal{H} where $S_\infty = P_{E_\infty}$ and E_∞ is the closure of spaces E_n , $n \geq 1$.

The **orthonormal system** $\{x_k\}_{k \geq 1}$ is complete if $E_\infty = \mathcal{H}$. In that case, $\|x\|^2 = \|P_{E_\infty} x\|^2 = \sum_{k=1}^\infty |\langle x, x_k \rangle|^2$.

Remark. Proving completeness can be difficult.

Example (Haar basis). The Haar basis for $L^2([0, 1])$ is the Fourier basis $e^{2\pi n i x}$, $n \in \mathbb{Z}$ for $L^2([0, 1])$.

Proof. Let x_k , $k \geq 1$ be any arbitrary sequence of vectors in \mathcal{H} . We can then construct an **orthonormal** sequence y_k , $k \geq 1$ by applying Gram-Schmidt procedure. \circledast

A.1.5 Bounded Linear Functionals

Consider **bounded linear functionals** on a **Banach space** E , $f \in E^*$, $\|f\| = \sup_{\|x\|=1} |f(x)|$ and E^* is a **Banach space**. Recall that $f(\cdot)$ is essentially defined by $H = \ker(f)$ where H is a closed subspace of E with $\text{codim}(H) = 1$, i.e., $\dim E/H = 1$ and we have

$$\tilde{f}: E/H \rightarrow \mathbb{R}$$

is defined via $\tilde{f}([x]) = f(x)$ for $x \in E$, and $\tilde{f}(a[x]) = ca$ for some constant c .

A.1.6 Representation Theorem

The important representation theorem for [bounded linear functionals](#) is the [Riesz representation theorem](#). The easiest case is $E = \mathcal{H}$ being a [Hilbert space](#) and $E^* \equiv \mathcal{H}$. This implies [Radon-Nikodym theorem](#), where if we have $\nu \ll \mu$, then

$$\nu(E) = \int_E f \, d\mu, \quad f = \frac{d\nu}{d\mu} \in L^1(\mu)$$

for ν, μ being finite measures. Furthermore, the [Radon-Nikodym theorem](#) implies the [Riesz representation theorem](#) for ℓ_p and L^p with $1 \leq p < \infty$.

Remark. We have $E^* = \ell_q$ or L^q with $1/p + 1/q = 1$ for $1 \leq p < \infty$, and remarkably, $\ell_1^* = \ell_\infty$ but $\ell_\infty^* \neq \ell_1$.

Remark. The [Riesz representation theorem](#) for $C(K)$ is space of bounded Borel measures where for $g \in C(K)^*$,

$$g(f) = \int_K f \, d\mu$$

for $f \in C(K)$.

A.1.7 Hahn-Banach Theorem

Let E be a [Banach space](#) and E_0 be a subspace such that $f_0: E_0 \rightarrow \mathbb{R}$ a [bounded linear functional](#) on E_0 such that $\|f_0\| < \infty$. Then there exists an extension f of f_0 to E with $\|f\| = \|f_0\|$.

Remark. f is not necessary unique. Nevertheless, it is unique for [Hilbert spaces](#), or ℓ_p , L^p with $1 < p < \infty$.

Reflexivity

Consider the embedding $E \rightarrow E^{**}$ such that $x \mapsto x^{**}$, then E is [reflexive](#) if the embedding is surjective. Also, E is [reflexive](#) implies that

$$\|f\| = \sup_{\|x\|=1} |f(x)| = f(x_f)$$

for some $x_f \in E$ with $\|x_f\| = 1$ for every $f \in E^*$.

Remark. This is one way of showing some spaces is not [reflexive](#).

Separation Theorem

Recall the [separation theorem](#) for [convex sets](#) from a point. Given a [convex set](#) K and a point $x_0 \notin K$, there is a [hyperplane](#) such that $f(x_0) > f(k)$ for all $k \in K$. The [Minkowski functional](#) for [convex set](#) essentially makes [convex sets](#) unit [ball](#) for some semi-norm.

Bibliography

- [Lax02] P.D. Lax. *Functional Analysis*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2002. ISBN: 9780471556046. URL: <https://books.google.com/books?id=18VqDwAAQBAJ>.
- [RS80] M. Reed and B. Simon. *I: Functional Analysis*. Methods of Modern Mathematical Physics. Elsevier Science, 1980. ISBN: 9780125850506. URL: <https://books.google.com/books?id=hInvAAAAMAAJ>.