Lectures in Functional Analysis

Roman Vershynin

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A.

E-mail address: romanv@umich.edu

Preface

These notes are for a one-semester graduate course in Functional Analysis, which is based on measure theory. The notes correspond to the course *Real Analysis II*, which the author taught at University of Michigan in the Fall 2010. The course consists of about 40 lectures 50 minutes each.

The student is assumed to be familiar with measure theory (both Lebesgue and abstract), have a good command of basic real analysis (epsilon-delta) and abstract linear algebra (linear spaces and transformations).

The course develops the theory of Banach and Hilbert spaces and bounded linear operators. Main principles of are covered in depth, which include Hahn-Banach theorem, open mapping theorem, closed graph theorem, principle of uniform boundedness, and Banach-Alaoglu theorem. Fourier series are developed for general orthogonal systems in Hilbert spaces. Compact operators and basics of Fredholm theory are covered.

Spectral theory for bounded operators is studied in the second half of the course. This includes the spectral theory for compact self-adjoint operators, functional calculus and basic spectral theory of general (non-compact) operators, although the latter needs to be expanded a bit.

Topics not covered include: Krein-Milman theorem (although this can be done with one extra lecture), unbounded linear operators, and Fourier transform. Most applications to ODE and PDE are not covered, however the integral operators serve as a main example of operators in this course.

The material has been compiled from several textbooks, including Eidelman, Milman and Tsolomitis "Functional Analysis", Kirillov and Gvishiani "Theorems and problemsin functional analysis", Reed and Simon "Methods of modern mathematical physics. I. Functional analysis", V. Kadets "A course in functional analysis" (Russian), and P. Knyazev, "Functional analysis". Minor borrowings are made from Yoshida "Functional analysis", Rudin "Functional analysis", and Conway "A course in functional analysis". For some topics not covered, one may try R. Zimmer "Essential results of functional analysis".

Acknowledgement. The author is grateful to his students in the Math 602 course *Real Analysis II*, Winter 2010, who suggested numerous corrections for these notes. Special thanks are to Matthew Masarik for his numerous thoughtful remarks, corrections, and suggestions, which improved the presentation of this material.

Contents

CHAPTER 1

Banach and Hilbert spaces

1.1. Linear spaces and linear operators

Lec.1: 09/08/10

1.1.1. From individual functions to function spaces. In this section, our discussion is general and not very precise. For now, our goal is to see a big picture.

In antiquity, properties of individual *numbers* were of great importance. Whether $\sqrt{2}$ is rational, for example, was a source of a great discussion. Later, with the development of calculus in the XVII century, the focus shifted from numbers to *functions*. A function wraps up individual numbers into some assignment rule, and it is the assignment that now matters rather than individual numbers. Later, in the second half of XIX century, the interest of mathematicians shifted further from studying individual functions (their differentiability, integrability) to *function spaces*. One may say that a function space wraps up functions of interest into one geometric object. The geometry of the function space reflects important properties of functions. This lead to the development of the field of *functional analysis*.

In functional analysis, we view functions as *points* or vectors in a function space. Since we can add functions on a common domain (say $[a, b]$) by defining $(f + g)(x) := f(x) + g(x)$ and multiply them by defining $(af)(x) = af(x)$, we see that a function space is a *linear vector space*.

Furthermore, we can envision a kind of *distance* on a function space, which would quantify similarity (or dissimilarity) of functions. The choice of a distance depends on application. One choice of distance between f and g is the *"sup-distance"*

$$\|f - g\|_\infty := \sup_{x\in[a,b]} |f(x) - g(x)|.$$

This is clearly a metric, so a function space becomes not only a linear vector space but also a *metric space*. Such spaces will be called *normed spaces* later. Another natural choice of a distance would be the

$$\|f - g\|_1 := \int_a^b |f(x) - g(x)|\, dx$$

Heuristically, a control of $\|f-g\|_\infty$ forces the values of f and g stay close everywhere on $[a, b]$, while a control of $\|f - g\|_1$ forces the values of f and g stay close "on average".

1.1.2. Examples of linear vector spaces. Classical example of linear vector spaces given in linear algebra courses are $\mathbb{R}^n$ and $\mathbb{C}^n$. Here are some examples of *function spaces* that are linear vector spaces (check!):

1. $F = \{\text{all functions } \mathbb{R} \to \mathbb{R}\}$. This space is too large, and is never studied.
2. $\{\text{all solutions of a linear homogeneous PDE}\}$
3. $L_1[a, b] = \{\text{all Lebesgue integrable functions on } [a, b]\}$

4. $L_\infty[a,b] = \{$all bounded almost everywhere functions on $[a,b]\}$
5. $C[a,b] = \{$all continuous functions on $[a,b]\}$
6. $C^1[a,b] = \{$all continuously differentiable functions on $[a,b]\}$
7. $C^\infty[a,b] = \{$all infinitely differentiable functions on $[a,b]\}$
8. $\mathcal{P}(x) = \{$all univariate polynomials$\}$
9. $\mathcal{P}_n(x) = \{$all univariate polynomials of degree at most $n\}$

Also there are many natural examples of *sequence spaces* that are linear vector spaces (check!):

1. $s = \{$all sequences of real numbers $(a_n)_{n=1}^\infty\}$. This space is too large, and is never studied.
2. $\ell_1 = \{$all absolutely summable sequences, i.e. satisfying $\sum_n |a_n| < \infty\}$
3. $\ell_\infty = \{$all bounded sequences of real numbers$\}$
4. $c = \{$all convergent sequences of real numbers$\}$
5. $c_0 = \{$all sequences of real numbers converging to zero$\}$
6. $c_{00} = \{$all sequences of real numbers with finite support$\}$

1.1.3. Subspaces. A (linear) subspace of a linear vector space is a subset that is closed under the operations of addition of vectors and multiplication by scalars:

Definition 1.1.1. A subset A_1 of a linear vector space E is called a (linear) *subspace* if $x, y \in E_1$, $a, b \in \mathbb{R}$ implies $ax + by \in E_1$.

Example 1.1.2. One can check that the following set-theoretic inclusions hold:[1]

$$\mathcal{P}_n(x) \subset \mathcal{P}(x) \subset C^\infty[a,b] \subset C^1[a,b] \subset C[a,b] \subset L_\infty[a,b] \subset L_1[a,b] \subset F,$$
$$c_{00} \subset \ell_1 \subset c_0 \subset c \subset \ell_\infty \subset s.$$

Some of these inclusions encode non-trivial theorems of undergraduate analysis.

All these inclusions are *subspace* inclusions, i.e. $\mathcal{P}_n(x)$ is a subspace of $\mathcal{P}(x)$, etc. (check this!)

Exercise 1.1.3. **Let E be a linear vector space. Show that $\{0\}$ and E are subspaces of E. Show that the intersection of an arbitrary collection of subspaces of E is again a subspace of E.**

Lec. 2: 09/10/10

1.1.4. Hamel basis. As we know, every *finite dimensional* linear vector space E has a basis $\{x_1, \dots, x_n\}$. A basis is a maximal linearly independent subset of vectors in E. The number n of basis elements is called the dimension of E; this number is independent of the choice of the basis. Every vector $x \in E$ can be uniquely expressed as a linear combination of the basis elements:

$$x = \sum_{k=1}^{n} a_k x_i, \quad \text{for some } a_k \in \mathbb{R}. \tag{1.1}$$

The notion of basis can be generalized to arbitrary, possibly infinite dimensional linear vector spaces E.

[1] For consistency of these inclusions, we restrict the functions in $\mathcal{P}_n(x)$, $\mathcal{P}(x)$ and F onto $[a,b]$.

DEFINITION 1.1.4 (Hamel basis). A subset $\mathcal{X}$ of a linear vector space E is called a *Hamel basis* of E if every vector $x \in E$ can be uniquely expressed as a finite linear combination of some elements of $\mathcal{X}$:

$$x = \sum_{k=1}^{n} a_k x_k \tag{1.2}$$

for some nonzero scalars a_k and vectors $x_k \in \mathcal{X}$.

EXERCISE 1.1.5. **Show that each of the following two statements gives an equivalent definition of Hamel basis:**

(1) A Hamel basis is a maximal linearly independent[2] subset $\mathcal{X} \subset E$.

(2) A Hamel basis is a linearly independent subset of E which *spans* E. The latter means that the *linear span* of $\mathcal{X}$, defined as

$$\mathrm{Span}(\mathcal{X}) := \Big\{ x = \sum_{k=1}^{n} a_k x_k : \ a_k \in \mathbb{R}, \ x_k \in \mathcal{X}, \ n \in \mathbb{N} \Big\},$$

coincides with E.

Since we have no topology on E, we have to consider finite sums in (1.2). This requirement is too strong to be put in practice, which makes Hamel bases essentially impractical (except in theory). We will come across the more practical notion of Schauder basis later.

PROPOSITION 1.1.6. *Every linear vector space E has a Hamel basis.*

For finite dimensional spaces E, this result is usually proved in undergraduate linear algebra using induction. One keeps adding linearly independent elements to a set until it spans the whole E. This argument can be pushed into infinite dimensions as well, where the usual induction is replaced by *transfinite induction*. The transfinite induction is best done with Zorn's lemma (review a Wikipedia article on Zorn's lemma if you are uncomfortable with all notions it uses):

LEMMA 1.1.7 (Zorn's lemma). *A partially ordered set in which every chain has an upper bound contains a maximal element.*

PROOF OF PROPOSITION 1.1.6. Consider a family $\mathcal{F}$ of all linearly independent subsets of E, which is partially ordered by inclusion. We claim that $\mathcal{F}$ has a maximal element; this would obviously complete the proof by Exercise 1.1.5. We will get a maximal element from Zorn's lemma. Let us check its assumption. Consider a chain (X_α) of elements in $\mathcal{F}$. The elements X_α are linearly independent subsets of E totally ordered by inclusion. Therefore, their union $\cup_\alpha X_\alpha$ is again a linearly independent subset of E (check!) Hence this union is an element of $\mathcal{F}$, and it is clearly an upper bound for the chain (X_α). The assumption of Zorn's lemma is therefore satisfied, and the proof is complete □

As in the finite dimensional case, the cardinality of Hamel basis of E is called the *dimension* of E; one can show that the dimension is independent of the choice of a Hamel basis.

EXAMPLE 1.1.8. Here we consider some of the examples of linear vector spaces given in Section 1.1.2.

---

[2]Linear independence means that every *finite* subset of $\mathcal{X}$ is linearly independent in the ordinary sense. In other words, if $\sum_{k=1}^{n} a_k x_k = 0$ for some $a_i \in \mathbb{R}$, $x_k \in \mathcal{X}$, $n \in \mathbb{N}$, then all $a_k = 0$.

1. $\dim(\mathbb{R}^n) = n$, $\dim(\mathbb{C}^n) = n$.
2. $\dim(\mathcal{P}_n(x)) = n + 1$, the monomials $\{1, x, x^2, \ldots, x^n\}$ form a basis.
3. $\dim(\mathcal{P}(x)) = \infty$, the monomials $\{1, x, x^2, \ldots\}$ form a Hamel basis.
4. $\dim(c_{00}) = \infty$, the coordinate vectors $(0, \ldots, 0, 1, 0, \ldots)$ form a Hamel basis.

Remark 1.1.9. Unfortunately, the notion of Hamel basis is too strong. Except in spaces $\mathcal{P}(x)$ and c_{00} (which are isomorphic - why?) no explicit constructions are known in any other infinite dimensional vector space. It would be great to have a construction of a Hamel basis in $C[0, 1]$, for example. However, Hamel bases usually have to be uncountable; see a later exercise.

1.1.5. Quotient spaces. The notion of quotient space allows one easily to collapse some directions in linear vector spaces. One reason for doing this is when one has unimportant directions and would likes to neglect them; see the construction of L_1 below.

Definition 1.1.10 (Quotient space). Let E_1 be a subspace of a linear vector space E. Consider an equivalence relation on E defined as

$$x \sim y \quad \text{if} \quad x - y \in E_1.$$

The *quotient space* E/E_1 is then defined as the set of equivalence classes (*cosets*) $[x]$ for all $x \in E$.

The quotient space is a linear space, with operations defined as

$$[x] + [y] := [x + y], \quad a[x] := [ax] \quad \text{for } x, y \in E, \ a \in \mathbb{R}.$$

The dimension of the quotient space is called the *codimension* of E_1, thus

$$\operatorname{codim}(E_1) := \dim(E/E_1).$$

Exercise 1.1.11. **Prove that the operations above are well defined, and that quotient space is indeed a linear space.**

Remark 1.1.12. 1. Observe that $[x]$ is an affine subspace:

$$[x] = x + E_1 := \{x + h : \ h \in E_1\}.$$

2. The definition of the equivalence relation $x \sim y$ is meant to ignore the directions in E_1, and thus to identify points x, y if they only differ by a vector from E_1.

3. From undergraduate linear algebra we know that if E is finite dimensional then all of its subspaces E_1 satisfy

$$\dim(E_1) + \operatorname{codim}(E_1) = \dim(E).$$

Example 1.1.13 (Space L_1). The notion of quotient space comes handy when we define the *space of integrable functions* $L_1 = L_1(\Omega, \Sigma, \mu)$ where (Ω, Σ, μ) is an arbitrary measure space. We first consider

$$E := \{\text{all integrable functions } f \text{ on } (\Omega, \Sigma, \mu)\}.$$

To identify functions that are equal μ-almost everywhere, we consider the subspace we would like to neglect:

$$E_1 := \{\text{all functions } f = 0 \ \mu\text{-almost everywhere}\}$$

Then we define

$$L_1 = L_1(\Omega, \Sigma, \mu) := E/E_1.$$

This way, the elements of L_1 are, strictly speaking, not functions but classes of equivalences.[3] But in practice, one thinks of an $f \in L_1$ as a function, keeping in mind that functions that coincide μ-almost everywhere are "the same".

EXAMPLE 1.1.14 (Space L_∞). A similar procedure is used to define the *space of essentially bounded functions* $L_\infty = L_\infty(\Omega, \Sigma, \mu)$. A real valued function f on Ω is called essentially bounded if there exists a bounded function g on Ω such that $f = g$ μ-almost everywhere. Similar to the previous example, we consider the linear vector space

$$E := \{\text{all essentially bounded functions } f \text{ on } (\Omega, \Sigma, \mu)\}$$

and the subspace we would like to neglect:

$$E_1 := \{\text{all functions } f = 0 \ \mu\text{-almost everywhere}\}$$

Then we define

$$L_\infty = L_\infty(\Omega, \Sigma, \mu) := E/E_1.$$

EXAMPLE 1.1.15. As we know, the space c_0 of sequences converging to zero is a subspace of the space c of all convergent sequences. Let us observe that

$$\operatorname{codim}(c_0) = 1.$$

Indeed, every sequence $x \in c$ can be uniquely represented as

$$x = a\mathbf{1} + z \quad \text{for some } a \in \mathbb{R},\ z \in c_0$$

where $\mathbf{1} = (1, 1, \ldots)$. (How do we choose the value of a?). Hence

$$[x] = a[\mathbf{1}] + [z] = a[\mathbf{1}].$$

It follows that every element $[x] \in c/c_0$ is a constant multiple of the element $[\mathbf{1}]$. Therefore, $\dim(c/c_0) = 1$ as claimed.

This example shows that the space c_0 makes up almost the whole space c, except for one dimension given by the constant sequences. This explains why the space c is rarely used in practice; one prefers to work with c_0 which is almost the same as c but has the advantage that we know the limits of all sequences there (zero).

1.1.6. Linear operators. This is a quick review of the classical linear algebra concept.

DEFINITION 1.1.16 (Linear operator). A map $T : E \to F$ between two linear vector spaces E and F is called a *linear operator* if it preserves the operations of addition of vectors and multiplication by scalars, i.e.

$$T(ax + by) = aT(x) + bT(y) \quad \text{for all } x, y \in E,\ a, b \in \mathbb{R}.$$

The *kernel* and *image* of T is defined respectively as[4]

$$\ker(T) = \{x \in E : \ Tx = 0\}; \quad \operatorname{Im}(T) = \{Tx : \ x \in E\}.$$

[3] Even more strictly speaking, the representative functions f in L_1 may take infinite values, too. However, every integrable function is finite a.e. So every such function is equivalent to a function that is finite everywhere.

[4] One usually writes Tx instead of $T(x)$

EXAMPLE 1.1.17 (Differential operator). The simplest example of a *differential operator* is given by taking the derivative of a function:

$$T(f) = f'.$$

Such operator is well be defined e.g. on the space of polynomials $T : \mathcal{P}(x) \to \mathcal{P}(x)$. But usually one prefers to have a differential operator on a larger space; for example $T : C^1[0,1] \to C[0,1]$ is also well defined.

EXAMPLE 1.1.18 (Embedding and quotient map). Given a subspace E_1 of a linear vector space E, there are two canonical linear operators associated with it:
1. *Embedding* $\iota : E_1 \to E$, which acts as an identity $\iota(x) = x$;
2. *Quotient map* $q : E \to E/E_1$, which acts as $q(x) = [x]$.

EXAMPLE 1.1.19 (Shifts on sequence spaces). On any sequence space such as c_{00}, c_0, c, ℓ_∞, ℓ_1, one can define the right and left shift operators respectively as

$$R(x) = (0, x_1, x_2, \ldots); \quad L(x) = (x_2, x_3, \ldots) \quad \text{for } x = (x_1, x_2, \ldots).$$

EXERCISE 1.1.20. **Compute the images and kernels of the embedding, quotient map, and the shift operators in the examples above.**

1.1.7. Additional Exercises.

EXERCISE 1.1.21. **Show that the intersection of an arbitrary collection of subspaces of a linear vector space E is again a subspace of E.**

EXERCISE 1.1.22. **Show that every linearly independent subset of a linear vector space E can be extended to a Hamel basis of E.**

EXERCISE 1.1.23. **[Complementary subspaces] Let E_1 be a subspace of a linear vector space E. Prove that there exists a subspace E_2 of E such that**

$$E_1 \cap E_2 = \{0\}, \quad \mathrm{Span}(E_1 \cup E_2) = E.$$

(Hint: extend a Hamel basis from E_1 onto E; use the extension to construct E_2). Such subspaces E_1, E_2 are called *complementary* to each other. Show that E_1, E_2 are complentary if and only if every vector $x \in E$ can be uniquely represented as the sum

$$x = x_1 + x_2 \quad \text{for some } x_1 \in E_1,\ x_2 \in E_2.$$

EXERCISE 1.1.24. **[Injectivization] This is a linear version of the fundamental theorem on homomorphisms for groups. Consider a linear operator $T : E \to F$ acting between linear spaces E and F. The operator T may not be injective; we would like to make it into an injective operator. To this end, we consider the map $\tilde{T} : X/\ker T \to Y$ which sends every coset $[x]$ into a vector Tx, i.e. $\tilde{T}[x] = Tx$.**

(i) Prove that $\tilde{T}$ is well defined, i.e. $[x_1] = [x_2]$ implies $Tx_1 = Tx_2$.
(ii) Check that $\tilde{T}$ is a linear and injective operator.
(iii) Check that T is surjective then $\tilde{T}$ is also surjective, and thus $\tilde{T}$ is a linear isomorphism between $X/\ker T$ and Y.

(iv) Show that $T = \tilde{T} \circ q$, where $q : X \to X/\ker T$ is the quotient map. In other words, every linear operator can be represented as a composition of a surjective and injective operator.

1.2. Normed spaces

Lec. 3: 09/13/10

1.2.1. Definition and examples. A norm is a general concept of *length* of vectors. Once we have a norm we can geometrize analysis in some sense, because we would have a metric on our linear vector spaces. For example, this would allow us to study functions through geometry of function spaces.

A norm is an assignment of a non-negative number $\|x\|$ to every vector x in a linear vector space E. In order to have a meaning of length, this assignment must satisfy some natural axioms:

Definition 1.2.1 (Normed space). Let E be a linear vector space. A *norm* on E is a function $\|\cdot\| : E \to \mathbb{R}$ which satisfies the following axioms:

(i) $\|x\| \geqslant 0$ for all $x \in E$; $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|ax\| = |a|\|x\|$ for all $x \in E$, $a \in \mathbb{R}$ (or $\mathbb{C}$);
(iii) $\|x + y\| \leqslant \|x\| + \|y\|$ for all $x, y \in E$.

The linear vector space E equipped with the norm $\|\cdot\|$ is called a *normed space*, and denoted $X = (E, \|\cdot\|)$.

Axiom (iii) is called *triangle inequality* for the following reason. Given an arbitraty triangle in E with vertices $x, y, z \in E$, its lengths satisfy the inequality

$$\|x - z\| \leqslant \|x - y\| + \|y - z\|, \tag{1.3}$$

which follows from norm axiom (iii). For the usual Euclidean length on the plane, this is the ordinary triangle inequality.

The normed space is naturally a *metric space*, with the metric defined by

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

The norm axioms, and in particular triangle inequality (1.3), show that this is indeed a metric (check!)

Exercise 1.2.2. **[Normed spaces ℓ_∞, c, c_0, ℓ_1, $C(K)$, L_1, L_∞] Many of linear vector spaces introduced in Section 1.1.2 and Example 1.1.12 are in fact normed spaces. Check the norm axioms for them:**

1. **The space of bounded sequences ℓ_∞ is a normed space, with the norm defined as**

$$\|x\|_\infty := \sup_i |x_i|. \tag{1.4}$$

2. **The spaces c and c_0 are normed spaces, with the same sup-norm as in (1.4).**
3. **The space of summable sequences ℓ_1 is a normed space, with the norm defined as**

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

4. The space $C(K)$ of continuous functions on a compact topological space K is a normed spaces with the norm[5]

$$\|f\|_\infty := \max_K |f(x)|.$$

5. The space $L_1 = L_1(\Omega, \Sigma, \mu)$ is a normed space, with the norm defined as[6]

$$\|f\|_1 := \int_\Omega |f(x)|\, d\mu.$$

Note that ℓ_1 is a partial case of the space $L_1(\Omega, \Sigma, \mu)$ where $\Omega = \mathbb{N}$ and μ is the counting measure on $\mathbb{N}$.

6. The space $L_\infty = L_\infty(\Omega, \Sigma, \mu)$ is a normed space, with the norm defined as the *essential supremum*:

$$\|f\|_\infty := \operatorname*{esssup}_{t\in\Omega} |f(t)| := \inf_{g=f \text{ a.e.}} \sup_{t\in\Omega} |g(t)|.$$

Here the infimum is taken over all $g \in L_\infty$ that are equal to f μ-almost everywhere. Note that ℓ_∞ is a partial case of the space $L_\infty(\Omega, \Sigma, \mu)$ where $\Omega = \mathbb{N}$ and μ is the counting measure on $\mathbb{N}$.

Exercise 1.2.3. **[Essential supremum] Show that the norm in $L_\infty(\Omega, \Sigma, \mu)$ can be equivalently computed as**

$$\|f\|_\infty = \inf_{\mu(A)=0} \sup_{t\in\Omega\setminus A} |f(t)|$$

where the infimum is over all subsets $A \subset \Omega$ of measure zero.

Exercise 1.2.4. **[Continuity of norms] Prove that the norm assignment $x \mapsto \|x\|$ is a continuous function on the normed space. Specifically, show that if $\|x_n - x\| \to 0$ then $\|x_n\| \to \|x\|$.**

1.2.2. Convexity of norms and balls. The geometry of a normed space can be very different from that of the usual Euclidean geometry. The balls do not need to be round anymore. For example, the ball of ℓ_∞ looks like a cube (why?) Nevertheless, one important property still holds: the balls are always *convex* sets, and the norm is a convex function. The convexity considerations are very helpful when one works in normed spaces.

Let us first recall some notions coming from geometry of metric spaces.

Definition 1.2.5 (Balls, spheres of normed spaces). Let X be a normed space. A (closed) *ball* centered at a point $x_0 \in X$ and with radius $r > 0$ is defined as

$$B_X(x_0, r) := \{x \in X : \|x - x_0\| \leq r\}.$$

The (closed) *unit ball* of X is defined as

$$B_X := B_X(0,1) = \{x \in X : \|x\| \leq 1\}.$$

The *unit sphere* of X is the boundary of the unit ball, that is

$$S_X := \{x \in X : \|x\| = 1\}.$$

[5] The maximum is attained because K is compact.

[6] Formally, as we know the elements of L_1 are cosets $[f]$ rather than functions. We should define the norm $\|[f]\|_1 := \int_\Omega |f(x)|\, d\mu$ where f is an arbitrary element in the coset $[f]$. Check that this way, the norm is well defined. The same concerns the definition of L_∞ below.

Definition 1.2.6 (Convex functions and sets). Let E be a linear vector space. A function $f : E \to \mathbb{R}$ is *convex* if for all $x, y \in E$, $\lambda \in [0,1]$ one has

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

A set $K \subseteq E$ is *convex* if for all $x, y \in K$, $\lambda \in [0,1]$ one has

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

A geometric meaning of convexity is the following. A function f is convex on E if its graph restricted to any interval $[x,y] \subset E$ lies below the interval joining the points $(x, f(x))$ and $(y, f(y))$; see the picture. A set K is convex if, together with any two points $x, y \in K$, it contains the interval $[x, y]$.

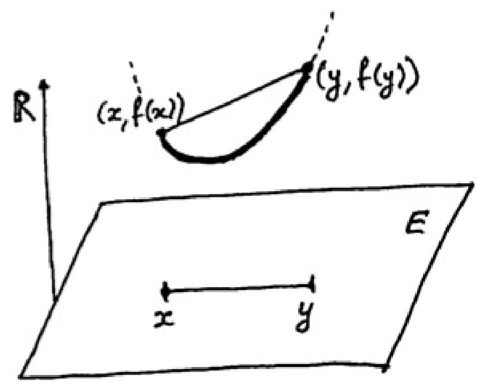

Figure 1.1. Convex function f on a linear vector space E

Proposition 1.2.7 (Norm axioms imply convexity). *Let X be a normed space. Then:*

1. *The function $x \to \|x\|$ is convex on X.*
2. *The unit ball B_X is a closed, origin-symmetric[7], and convex set in X.*

Proof. 1. Convexity of the norm follows from the norm axioms. Indeed, for every $x, y \in E$, $\lambda \in [0,1]$ we have

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

2. Closedness of B_X follows from continuity of the norm (see Exercise 1.2.4). Origin-symmetry follows from norm axiom (ii) with $\lambda = -1$. Finally, to prove convexity of B_X we choose arbitrary $x, y \in B_X$, $\lambda \in [0,1]$ and use inequality above to obtain

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

It follows that $\lambda x + (1-\lambda)y \in B_X$ as required. □

The converse to Proposition 1.2.7 also holds:

Proposition 1.2.8 (Convexity implies triangle inequality). *Let $x \mapsto \|x\|$ be a real-valued function defined on a linear vector space E. Assume that this function satisfies norm axioms (i) and (ii). Then:*

[7] Origin-symmetric means that $x \in B_X$ implies $-x \in B_X$

1. *If the function $x \to \|x\|$ is convex, then the triangle inequality is satisfied, and $\|\cdot\|$ is a norm on E.*
2. *If the sublevel set $\{x \in X : \|x\| \leqslant 1\}$ is convex, then $\|\cdot\|$ is a norm on E.*

PROOF. 1. Convexity ensures that for every $x, y \in E$, $\lambda \in [0,1]$ we have

$$\|\lambda x + (1-\lambda)y)\| \leqslant \lambda\|x\| + (1-\lambda)\|y\|.$$

Triangle inequality follows from this for $\lambda = 1/2$.

2. This statement is less trivial, and can not be obtained from the first one directly. Indeed, while it is true that the sublevel sets of a convex functions are convex sets, the converse statement may fail (construct an example!)

The assumption states that, for $u, v \in E$ we have:

$$\text{if } \|u\| \leqslant 1,\ \|v\| \leqslant 1,\ \lambda \in [0,1], \quad \text{then} \quad \|\lambda u + (1-\lambda)v\| \leqslant 1. \tag{1.5}$$

Let $x, y \in E$; we want to show that $\|x+y\| \leqslant \|x\| + \|y\|$. This is equivalent to

$$\left\| \frac{x}{\|x\|+\|y\|} + \frac{y}{\|x\|+\|y\|} \right\| \leqslant 1.$$

We obtain this inequality from (1.5) with

$$u = \frac{x}{\|x\|},\ v = \frac{y}{\|y\|},\ \lambda = \frac{\|x\|}{\|x\|+\|y\|}.$$

This completes the proof. □

Lec. 4: 09/15/2010

1.2.3. Spaces L_p. Minkowski inequality. We have already come across the spaces L_1 and ℓ_1. They are partial cases of a big family of spaces L_p and ℓ_p which we will study now.

Consider a measure space (Ω, Σ, μ) and an exponent $p \in [1, \infty)$. We define *the space of p-integrable functions* $L_p = L_p(\Omega, \Sigma, \mu)$ as the set of all measurable functions $f : \Omega \to \mathbb{R}$ such that

$$\int_\Omega |f(x)|^p \, d\mu < \infty.$$

PROPOSITION 1.2.9. *$L_p(\Omega, \Sigma, \mu)$ is a linear vector space for $p \in [1, \infty)$.*

PROOF. The only non-trivial point is that L_p is closed under addition, that is $f, g \in L_p$ implies $f + g \in L_p$. We will obtain this by the convexity of the function $z \mapsto |z|^p$ on $\mathbb{R}$ for $p \geqslant 1$. The convexity implies the pointwise inequality

$$\left| \frac{|f(t)+g(t)|}{2} \right|^p \leqslant \frac{|f(t)|^p + |g(t)|^p}{2}, \quad t \in \Omega.$$

Integrating both sides of this inequality, we obtain the required claim. □

We turn L_p into a normed space by defining

$$\|f\|_p := \Big(\int_\Omega |f(x)|^p \, d\mu \Big)^{1/p} \quad \text{for } f \in L_p(\Omega, \Sigma, \mu).$$

PROPOSITION 1.2.10. *$L_p(\Omega, \Sigma, \mu)$ is a normed spaces with the norm $\|f\|_p$ for $p \in [1, \infty)$.*

PROOF. Norm axioms (i) and (ii) are straightforward. Axiom (iii), triangle inequality, will follow from Proposition 1.2.8. To this end, it suffices to check that the sublevel set

$$B_p := \{f \in L_p : \|f\|_p \leqslant 1\}$$

is a convex set. To prove this, let us fix $f, g \in B_p$ and $\lambda \in [0,1]$. Since the function $z \mapsto |z|^p$ is convex on $\mathbb{R}$ for $p \geqslant 1$, we have a pointwise inequality

$$\big|\lambda f(t) + (1-\lambda)g(t)\big|^p \leqslant \lambda |f(t)|^p + (1-\lambda)|g(t)|^p.$$

Integrating both sides of this inequality implies

$$\int_\Omega \big|\lambda f(t) + (1-\lambda)g(t)\big|^p \, d\mu \leqslant \lambda \int_\Omega |f(t)|^p \, d\mu + (1-\lambda) \int_\Omega |g(t)|^p \, d\mu \leqslant \lambda + (1-\lambda) = 1.$$

We have showed that $\|\lambda f + (1-\lambda)g\|_p \leqslant 1$, hence $\lambda f + (1-\lambda)g \in B_p$. Therefore, the sublevel set B_p is convex. The proof is finished by Proposition 1.2.8. □

Writing out the triangle inequality $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$ in analytic form, we obtain the classical *Minkowski inequality*:

THEOREM 1.2.11 (Minkowski inequality in L_p). *Let $p \in [1, \infty)$. Then, for every two functions $f, g \in L_p(\Omega, \Sigma, \mu)$ one has*

$$\Big(\int_\Omega |f(t)+g(t)|^p \, d\mu\Big)^{1/p} \leqslant \Big(\int_\Omega |f(t)|^p \, d\mu\Big)^{1/p} + \Big(\int_\Omega |g(t)|^p \, d\mu\Big)^{1/p}.$$

□

1.2.4. Spaces ℓ_p and ℓ_p^n. An important partial case of the space $L_p(\Omega, \Sigma, \mu)$ is the space ℓ_p obtained by choosing $\Omega = \mathbb{N}$ and μ to be the counting measure on $\mathbb{N}$. Equivalently, for $p \in [1, \infty)$, *the space of p-summable sequences ℓ_p* is defined to consist of sequences $x = (x_i)_{i=1}^\infty$ for which

$$\sum_{i=1}^\infty |x_i|^p < \infty.$$

We turn ℓ_p into a normed space with the norm

$$\|x\|_p := \Big(\sum_{i=1}^\infty |x_i|^p\Big)^{1/p}.$$

Writing down Minkowski inequality for this specific measure space, we obtain:

THEOREM 1.2.12 (Minkowski inequality in ℓ_p). *Let $p \in [1, \infty)$. Then, for every two sequences $x, y \in \ell_p$ one has*

$$\Big(\sum_{i=1}^\infty |x_i + y_i|^p\Big)^{1/p} \leqslant \Big(\sum_{i=1}^\infty |x_i|^p\Big)^{1/p} + \Big(\sum_{i=1}^\infty |y_i|^p\Big)^{1/p}.$$

□

A remarkable family of *finite-dimensional* spaces $L_p(\Omega, \Sigma, \mu)$ is formed by considering Ω to be a finite set, say $\Omega = \{1, \ldots, n\}$ and μ to be the counting measure on Ω. The resulting space is called ℓ_p^n. The functions in ℓ_p^n can be obviously identified with vectors in $\mathbb{R}^n$. Thus $\ell_p^n = (\mathbb{R}^n, \|\cdot\|_p)$ with the norm

$$\|x\|_p := \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p}.$$

When $p = 2$, this space is the usual Euclidean space $\mathbb{R}^n$. However, for $p \neq 2$, the geometry of ℓ_p^n is quite different from Euclidean. Indeed, in two-dimensional spaces, the unit ball of ℓ_1^2 is a diamond with vertices $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$. The unit ball of ℓ_∞^2 is the square with vertices $(1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$.

EXERCISE 1.2.13. **[ℓ_∞ as the limit of ℓ_p] This exercise explains the index ∞ in the name of the spaces ℓ_∞, L_∞.**

1. Show that if $x \in \ell_{p_0}$ for some $p_0 \geqslant 1$ then

$$\|x\|_p \to \|x\|_\infty \quad \text{as } p_0 \leqslant p \to \infty.$$

2. Consider the space $L_\infty = L_\infty(\Omega, \Sigma, \mu)$ with finite total measure $\mu(\Omega)$. Show that if $f \in L_\infty$ then

$$\|f\|_p \to \|f\|_\infty \quad \text{as } p \to \infty.$$

1.2.5. Subspaces of normed spaces.

DEFINITION 1.2.14 (Subspace). Let X be a normed space. A subspace Y of X is a linear subspace equipped with the norm induced from that of X.

This concept should be familiar from topology, where a subspace is a subset of a topological space with the induced topology.

EXAMPLE 1.2.15. 1. The space of polynomials $\mathcal{P}(x)$ is a dense subspace of $C[0,1]$. This is the statement of Weierstrass approximation theorem.

2. The set of all continuous functions $C[0,1]$ forms a dense subspace of $L_1[0,1]$. (of course, both spaces are considered in the L_1 norm!) This follows from a theorem in measure theory that states that an integrable function can be approximated by a continuous function (why?)

EXERCISE 1.2.16. **1. Show that the set of convergent sequences c and the set of sequences converging to zero c_0 are closed subspaces of ℓ_∞.**

2. For all $p \in [1,\infty)$, show that the set of p-summable sequences ℓ_p is a closed subspace of ℓ_∞ but is a dense subspace of c_0.

1.2.6. Quotient spaces of normed spaces. In Section 1.1.5, we defined quotient spaces of linear vector spaces. If the ambient space is a normed space, then we can also induce the norm onto the quotient space as follows.

DEFINITION 1.2.17 (Quotient space of a normed space). Let X be a normed space and Y be a closed subspace of X. We define a norm on X/Y as follows. For every coset $[x] = x + Y$, we put

$$\|[x]\| := \inf_{y \in Y} \|x + y\|.$$

It is easy to understand the norm in the quotient space geometrically as the *distance* from the origin to the coset $[x]$. Indeed, let $\operatorname{dist}(b, A)$ denote the distance in X from a point b to a set A:

$$\operatorname{dist}(b, A) = \inf_{a \in A} \|b - a\|.$$

Then clearly

$$\|[x]\| = \operatorname{dist}(0, [x]). \tag{1.6}$$

Proposition 1.2.18. *The definition of* $\|[x]\|$ *above indeed produces a norm on* X/Y.

Proof. First we observe that the number $\|[x]\|$ is well defined, i.e. it does not depend on a choice of a representative x in the coset $[x]$. This clearly follows from the geometric definition (1.6).

Next, we have to check the three norm axioms.

(i) Assume that $\|[x]\| = 0$. Then, from the geometric definition (1.6) we see that 0 is a limit point of $[x]$. Since Y is closed, so is $[x] = x + Y$. Therefore $0 \in [x]$. Hence $[x] = [0]$, which verifies norm axiom (i).

(ii) Let $x \in X$ and $\lambda \in \mathbb{R}$. Then

$$\|[\lambda x]\| = \inf_{y \in Y} \|\lambda x + y\| = \inf_{y \in Y} \|\lambda x + \lambda y\| = \lambda \cdot \inf_{y \in Y} \|x + y\| = \lambda \|[x]\|.$$

This verifies norm axiom (ii).

(iii) Let us fix $x_1, x_2 \in X$; we want to show that $\|[x_1 + x_2]\| \leqslant \|[x_1]\| + \|[x_2]\|$. To this end, fix an arbitrary $\varepsilon > 0$. By the definition of the quotient norm there exist $y_1, y_2 \in Y$ so that

$$\|x_1 + y_1\| - \varepsilon \leqslant \|[x_1]\| \leqslant \|x_1 + y_1\|, \quad \|x_2 + y_2\| - \varepsilon \leqslant \|[x_2]\| \leqslant \|x_2 + y_2\|.$$

Using triangle inequality for the norm in X, we obtain

$$\|x_1 + x_2 + y_1 + y_2\| \leqslant \|x_1 + y_1\| + \|x_2 + y_2\| \leqslant \|[x_1]\| + \|[x_2]\| + 2\varepsilon.$$

We conclude that

$$\|[x_1 + x_2]\| = \inf_{y \in Y} \|x_1 + x_2 + y\| \leqslant \|x_1 + x_2 + y_1 + y_2\| \leqslant \|[x_1]\| + \|[x_2]\| + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof of norm axiom (iii). $\square$

Exercise 1.2.19. **Consider the subspace Y of $C(K)$ that consists of constant functions. Derive the following formula for the norm in the quotient space $C(K)/Y$:**

$$\|[f]\| = \frac{1}{2}\Big(\max_{t \in K} f(t) - \min_{t \in K} f(t)\Big) \quad \textbf{for } f \in C(K).$$

Exercise 1.2.20. **Derive the following formula for the norm in the quotient space ℓ_∞ / c_0:**

$$\|[a]\| = \limsup |a_i| \quad \textbf{for } a = (a_i)_{i=1}^\infty \in \ell_\infty.$$

1.2.7. Additional Exercises.

Exercise 1.2.21. **[Direct sum of two normed spaces] Let X and Y be two normed spaces. Consider their direct (Cartesian) product**

$$X \oplus_1 Y = \{(x, y) : \ x \in X, \ y \in Y\}.$$

Show that $X \oplus_1 Y$ is a normed space, with the norm defined as

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

EXERCISE 1.2.22. **[Minkowski functional] Consider a closed, convex, origin-symmetric set K in $\mathbb{R}^n$ with nonempty interior. Minkowski functional of K is the function defined on $\mathbb{R}^n$ by**

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

Show that $\|\cdot\|_K$ is a norm on $\mathbb{R}^n$, and the unit ball of this normed space is K.

EXERCISE 1.2.23. **[Seminorms] A seminorm on a linear vector space E is a function $\|\cdot\| : E \to \mathbb{R}$ which satisfies all norm axioms except the second part of axiom (i). That is, there may exist nonzero vectors x for which $\|x\| = 0$.**

Show that one can convert a seminorm into a norm by factoring out the zero directions. Mathematically, show that $\ker(p) := \{x \in E : \|x\| = 0\}$ is a linear subspace of E. Show that the quotient space $E/\ker(p)$ is a normed space, with the norm defined as

$$\|[x]\| := \|x\| \quad \textbf{for } x \in E.$$

Illustrate this procedure by constructing the normed space L_∞ from the semi-normed space of all essentially bounded functions, with the essential sup-norm.

EXERCISE 1.2.24. **[Convex hull] The *convex hull* of a subset A of a linear vector space X is defined to be the smallest convex set that contains A. The convex hull of A is denoted $\mathrm{conv}(A)$. (See the picture for an example).**

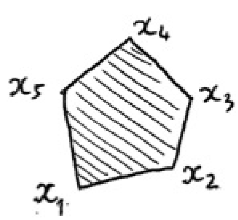

FIGURE 1.2. The shaded region is the convex hull $\mathrm{conv}(x_1, \ldots, x_5)$

A *convex combination* of vectors $x_1, \ldots, x_n$ in a linear vector space X is any vector of the form

$$x = \sum_{k=1}^{n} \lambda_k x_k$$

where $\lambda_k \geqslant 0$ are some numbers such that $\sum_{k=1}^{n} \lambda_k = 1$. Prove that $\mathrm{conv}(A)$ coincides with the set of all convex combinations of a finite number of vectors from A.

1.3. Banach spaces

Lec.5: 09/17/10

1.3.1. Definition. Completeness of $C(K)$. It turns out that the concept of normed space is deficient; many results in analysis can not be obtained just based on the norm axioms. An additional axiom is needed, which is *completeness*.

Recall that a metric space X is called complete if every Cauchy sequence in X converges to a point in X. For example, $\mathbb{R}$ is a complete metric space while $\mathbb{Q}$ is not.

Specializing to normed spaces X, recall that a sequence $(x_i)_{i=1}^{\infty}$ in X is *Cauchy* if

$$\|x_n - x_m\| \to 0 \quad \text{as } n, m \to \infty,$$

i.e. for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that

$$\|x_n - x_m\| < \varepsilon \quad \text{for all } n, m > N.$$

DEFINITION 1.3.1 (Banach space). A complete normed space X is called a Banach space.

EXERCISE 1.3.2. **[Subspaces, quotients of Banach spaces] Let X be a Banach space and Y be a (linear) subspace of X. Show that:**

1. Y is a Banach space if and only if Y is closed.
2. If Y is closed[8], then X/Y is a Banach space.

Many classical spaces are Banach spaces.

THEOREM 1.3.3. *For a compact topological space K, $C(K)$ is a Banach space.*

PROOF. Most proofs of completeness work out by reducing the problem to the completeness of $\mathbb{R}$. To this end, consider a Cauchy sequence (f_n) in $C(K)$, that is

$$\|f_n - f_m\|_\infty \to 0, \quad n, m \to \infty. \tag{1.7}$$

Therefore, for every $t \in K$, we have $|f_n(t) - f_m(t)| \to 0$. In other words, $(f_n(t))$ is a Cauchy sequence in $\mathbb{R}$ for every t. By completeness of $\mathbb{R}$, this sequence has a limit which we call $f(t)$. We have constructed a function $f(t)$ such that $f_n \to f$ pointwise.

We now claim that $f_n \to f$ uniformly, i.e. $\|f_n - f\|_\infty \to 0$. This would complete the proof, since the limit of a uniformly convergent sequence of continuous functions on the compact space K is a continuous function (as we know from the undergraduate analysis). Let us prove our claim then. By (1.7), for every ε there exists $N = N(\varepsilon)$ such that

$$|f_n(t) - f_m(t)| < \varepsilon \quad \text{for all } n, m > N, \ t \in K.$$

Letting $m \to \infty$ in this inequality (while keeping everything else fixed), we conclude that

$$|f_n(t) - f(t)| < \varepsilon \quad \text{for all } n, m > N, \ t \in K.$$

This means that $\|f_n - f\|_\infty \to \infty$, which is what we wanted. ☐

EXERCISE 1.3.4. **[Banach spaces ℓ_∞, c_0, L_∞]**

1. Show that ℓ_∞ and L_∞ are Banach spaces. (Hint: modify the proof of Theorem 1.3.3.)
2. Show that c_0 is a Banach space (Hint: use Exercise 1.2.16.)

[8] Recall that Y has to be closed in order for X/Y to be well defined.

3. Show that the set of functions

$$\{f \in C[0,1] : f(0) = f(1) = 0\}$$

is a Banach space with respect to the sup-norm. (Hint: check that this is a closed subspace of $C[0,1]$.)

4. Show that the set of polynomials $\mathcal{P}(x)$ restricted to $[0,1]$ is not a Banach space with respect to the sup-norm. (Hint: recall that $\mathcal{P}(x)$ is dense in $C[0,1]$ by Weierstrass approximation theorem.)

5. Show that c_{00} is not a Banach space with respect to any $\|\cdot\|_p$ norm, $1 \leqslant p \leqslant \infty$. (Hint: show that c_{00} is dense in ℓ_p for $p \in [1,\infty)$, and is not closed in ℓ_∞.)

Lec.6: 09/20/10

1.3.2. Series in Banach spaces. Completeness of L_p. We are going to give a useful criterion of completeness of normed spaces in terms of convergence of *series* rather than sequences. We shall use this criterion to prove the completeness of the L_p spaces.

Definition 1.3.5. Series Let (x_k) be a sequence of vectors in a normed space X. If the partial sums

$$s_n := \sum_{k=1}^{n} x_k$$

converge to some vector $x \in X$ as $n \to \infty$, then we say that the series $\sum_k x_k$ *converges* in X, and we write

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

A series $\sum_k x_k$ is called *absolutely convergent* if

$$\sum_{k=1}^{\infty} \|x_k\| < \infty.$$

Recall that in the scalar case, where $X = \mathbb{R}$ or $\mathbb{C}$, absolute convergence of series implies convergence (but not vice versa). As the following theorem shows, this happens precisely because of the completeness of $\mathbb{R}$ and $\mathbb{C}$.

Theorem 1.3.6 (Completeness criterion). *A normed space X is a Banach space if and only if every absolutely convergent series in X converges in X.*

Proof. *1. Necessity.* Let X be a Banach space, and consider an absolutely convergent series, i.e.

$$\sum_{k=1}^{\infty} \|x_k\| < \infty. \tag{1.8}$$

We want to prove that the series $\sum_k x_k$ converges. By completeness of X, it suffices to show that the partial sums of this series are Cauchy, i.e. that $\|s_n - s_m\| \to 0$ as $n > m \to \infty$. To this end, we use triangle inequality and our assumption (1.8) to obtain

$$\|s_n - s_m\| = \Big\| \sum_{k=n+1}^{m} x_k \Big\| \leqslant \sum_{k=n+1}^{m} \|x_k\| \to 0.$$

This completes the proof of necessity.

2. Sufficiency. Assume that X in incomplete; we want to construct a divergent series which is absolutely convergent. By incompleteness, there exists a Cauchy sequence (v_n) in X which diverges.

Every subsequence of (v_n) diverges (check!). Therefore, there exists a subsequence (w_n) of (v_n) which diverges but which is "rapidly Cauchy", i.e.

$$\|w_2 - w_1\| \leqslant \frac{1}{2}, \quad \|w_3 - w_3\| \leqslant \frac{1}{2^2}, \quad \|w_4 - w_4\| \leqslant \frac{1}{2^3}, \quad \dots \tag{1.9}$$

(Construct it!) It follows that the sequence (x_k) defined as

$$x_1 := w_2 - w_1, \quad x_2 := w_3 - w_2, \quad x_3 := w_4 - w_3, \quad \dots$$

forms the terms of an absolutely convergent series:

$$\sum_{k=1}^{\infty} \|x_k\| \leqslant \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots = 1.$$

Nevertheless, the partial sums

$$\sum_{k=1}^{n} x_k = w_{n+1} - w_1$$

diverge. So we have constructed an absolutely convergent series in X which diverges. This completes the proof. □

EXERCISE 1.3.7. **Validate the two missing steps in the proof of Theorem 1.3.6. Let X_n be a normed space.**

1. **Let (v_n) be a Cauchy sequence X which diverges. Prove that every subsequence of (v_n) diverges.**
2. **Let (v_n) be a Cauchy sequence in X. Construct a "rapidly Cauchy" subsequence (w_n) of (v_n), i.e. one that satisfies (1.9).**

THEOREM 1.3.8. *For every $p \in [1, \infty)$, the space $L_p = L_p(\Omega, \Sigma, \mu)$ is a Banach space.*

PROOF. Let functions (f_n) in L_p form the terms of an absolutely convergence series, i.e.

$$\sum_{k=1}^{\infty} \|f_k\|_p =: M < \infty.$$

By the completeness criterion, Theorem 1.3.6, it suffices to show that the series $\sum_k f_k$ converges in L_p.

Case 1: all $f_k \geqslant 0$ pointwise. The partial sums $\sum_{k=1}^{n} f_k$ form a pointwise non-decreasing sequence of functions. Denote the pointwise limit by $\sum_{k=1}^{\infty} f_k$; it may be infinite at some points.

The triangle inequality (which is Minkowski's inequality) implies that the partial sums are *bounded*:

$$\Big\| \sum_{k=1}^{n} f_k \Big\|_p \leqslant \sum_{k=1}^{n} \|f_k\|_p \leqslant M,$$

which in other words is

$$\int_\Omega \Big(\sum_{k=1}^{n} f_k \Big)^p \, d\mu \leqslant M^p.$$

We apply Lebesgue monotone convergence theorem for the sequence of functions $\big(\sum_{k=1}^n f_k\big)^p$ and get

$$\int_\Omega \Big(\sum_{k=1}^n f_k\Big)^p \, d\mu \to \int_\Omega \Big(\sum_{k=1}^\infty f_k\Big)^p \, d\mu.$$

The right hand side exists and is finite since the left hand side is bounded by M^p for all n. Thus $\sum_{k=1}^\infty f_k \in L_p$.

It remains to check that the series $\sum_{k=1}^\infty f_k \in L_p$ converges not only pointwise but also in L_p. By the pointwise convergence and positivity, the tails of this series

$$r_n := \sum_{k=n+1}^\infty f_k$$

form a pointwise non-increasing sequence of positive functions in L_p that pointwise converge to 0. Applying Lebesgue monotone convergence theorem for the sequence r_n^p, we conclude that $\|r_n\|_p \to 0$. We conclude that the series $\sum_{k=1}^\infty f_k$ converges in L_p as required.

Case 1: arbitraty f_k. As is usual in the theory of Lebesgue integration, the claim for arbitrary (not necessarily non-negative) functions f_k will follow by decomposing them into positive and negative parts:

$$f_k^+ := \max(f_k, 0), \quad f_k^- := \max(-f_k, 0) \quad \text{pointwise.}$$

Then $f_k = f_k^+ - f_k^-$. Since $0 \leqslant f_k^+ \leqslant |f_k|$, we have

$$\sum_{k=1}^\infty \|f_k^+\|_p \leqslant \sum_{k=1}^\infty \|f_k\|_p < \infty.$$

So, by the first part of the argument, the series $\sum_{k=1}^\infty f_k^+$ converges in L_p. Similarly we show that $\sum_{k=1}^\infty f_k^-$ converges in L_p. Therefore, $\sum_{k=1}^\infty f_k = \sum_{k=1}^\infty f_k^+ - \sum_{k=1}^\infty f_k^-$ converges in L_p. This completes the proof. □

Lec.7: 09/22/10

1.3.3. Completion. From a course in point-set topology we know that every metric space can be turned into a *complete* metric space. This procedure is called *completion*. We will illustrate it in the context of normed spaces.

THEOREM 1.3.9 (Completion). *Let X be a normed space. There exists a Banach space $\hat{X}$, called the* completion of X, *with the following properties. One can find a linear map $\iota : X \to \hat{X}$ such that:*

(i) $\|\iota x\| = \|x\|$ for all $x \in X$ (i.e. ι is an isometric embedding of X into $\hat{X}$;
(ii) $\mathrm{Im}(\iota)$ is dense in $\hat{X}$.

The completion of X is unique up to an isometry.[9]

EXERCISE 1.3.10. **Prove Theorem 1.3.9. Mimic the construction of $\mathbb{R}$ from $\mathbb{Q}$, or the more general construction of completion of metric spaces. Namely, consider the space Y of all Cauchy sequences (x_k) in X, equipped with the semi-norm**

$$\|(x_k)\| := \lim_k \|x_k\|.$$

[9]Precisely, this means that for any other completion $\hat{X}'$, there exists an injective and bijective linear map $T : \hat{X} \to \hat{X}'$ such that $\|Tx\| = \|x\|$ for all $x \in X$.

Then turn this space into a normed space $\hat{X}$ by taking quotient over the kernel of the semi-norm (see Exercise 1.2.23).

The concept of completion suggests an alternative construction of the space $L_p[a,b]$, $p \in [1,\infty)$. From measure theory we know that the set of continuous functions is a dense subset of $L_p[a,b]$ (why?). Furthermore, $L_p[a,b]$ is a complete normed space. Therefore (by the uniqueness of completion), $L_p[a,b]$ *is the completion of the space* $C[a,b]$ *in the* $\|\cdot\|_p$ *norm.*

This is an alternative definition of $L_p[a,b]$. This gives also an alternative construction of Lebesgue integral. Indeed, on for continuous functions, Riemann and Lebesgue integrals coincide. Therefore, the space $C[a,b]$ with $\|\cdot\|_1$ can be constructed using only Riemann integral. But its completion gives rise to Lebesgue integral.

1.3.4. Additional Exercises.

EXERCISE 1.3.11. **[Space of continuous periodic functions] Show that the set of functions**

$$\{f \in C[0,1] : \ f(0) = f(1)\}$$

is a Banach space with respect to the sup-norm. (Hint: identify this space with $C(\mathbb{T})$ where $\mathbb{T}$ is a one-dimensional torus.)

EXERCISE 1.3.12. **[Space of continuously differentiable functions] Show that the space $C^k[0,1]$ of k-times differentiable functions is not a Banach space with respect to the sup-norm.**

Show that $C^k[0,1]$ is a Banach space with respect to the norm

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

EXERCISE 1.3.13. **[Completeness of a direct sum] Let X and Y be Banach spaces. Show that the direct sum $X \oplus_1 Y$ defined in Exercise 1.2.21 is a Banach space.**

1.4. Inner product spaces

1.4.1. Definition. Cauchy-Schwarz inequality. Hilbert spaces form an important and simplest class of Banach spaces. Speaking imprecisely, Hilbert spaces are those Banach spaces where the concept of *orthogonality* of vectors is defined. Hilbert spaces will arise as complete inner product spaces.

DEFINITION 1.4.1 (Inner product space). Let E be a linear space over $\mathbb{C}$. An *inner product* on E is a function $\langle\cdot,\cdot\rangle : E \times E \to \mathbb{C}$ which satisfies the following three axioms:

(i) $\langle x,x\rangle \geqslant 0$ for all $x \in E$; $\langle x,x\rangle = 0$ if and only if $x = 0$;
(ii) $\langle ax+by,z\rangle = a\langle x,z\rangle + b\langle y,z\rangle$ for all $x,y,z \in E$ and $a,b \in \mathbb{C}$;
(iii) $\langle x,y\rangle = \overline{\langle y,x\rangle}$ for all $x,y \in E$.

The space E with an inner product is called an *inner product space.*

Inner products over $\mathbb{R}$ are defined similarly, except there is no conjugation in axiom (iii).

REMARK 1.4.2. The inner product is (congjugate) linear in the second argument:

$$\langle x, ay + bz\rangle = \overline{a}\langle x, y\rangle + \bar{b}\langle x, z\rangle.$$

This follows from axioms (ii) and (iii) of the inner product.

DEFINITION 1.4.3 (Orthogonality). If $\langle x, y\rangle = 0$ we say that vectors x and y are *orthogonal* and write $x \perp y$.

EXAMPLE 1.4.4. The canonical example of a finite-dimensional Hilbert space is the Euclidean space $\mathbb{C}^n$ equipped with the inner product

$$\langle x, y\rangle = \sum_{k=1}^{n} x_k \overline{y_k}.$$

THEOREM 1.4.5 (Cauchy-Schwarz inequality). *Let X be an inner product space. Then every two vectors $x, y \in X$ satisfy*

$$|\langle x, y\rangle| \leqslant \langle x, x\rangle^{1/2}\langle y, y\rangle^{1/2}.$$

PROOF. *Case 1: $\langle x, y\rangle \in \mathbb{R}$.* For all $t \in \mathbb{R}$, by the axioms of the inner product we have:

$$0 \leqslant \langle x + ty, x + ty\rangle = t^2\langle y, y\rangle + 2t\langle x, y\rangle + \langle x, x\rangle.$$

A quadratic polynomial that is everywhere non-negatve must have a non-positive discriminant, i.e.

$$\langle x, y\rangle^2 - \langle x, x\rangle\langle y, y\rangle \leqslant 0.$$

This is precisely Cauchy-Schwarz inequality.

Case 2: $\langle x, y\rangle \in \mathbb{C}$ arbitrary. We will multiply y by a unit scalar so that $\langle x, y\rangle$ becomes a real number, and use the first part. Indeed, polar decomposition

$$\langle x, y\rangle = |\langle x, y\rangle| e^{i \operatorname{Arg}\langle x, y\rangle}$$

implies that

$$|\langle x, y\rangle| = \langle x, y'\rangle \quad \text{where } y' = e^{i \operatorname{Arg}\langle x, y\rangle} y.$$

Now using the first part of the proof we conclude that

$$|\langle x, y\rangle| = \langle x, y'\rangle \leqslant \langle x, x\rangle\langle y', y'\rangle \leqslant \langle x, x\rangle\langle y, y\rangle$$

as required. □

COROLLARY 1.4.6 (Norm in an inner product space). *Let X be an inner product space. Then X is a normed space, with the norm defined as*

$$\|x\| := \langle x, x\rangle^{1/2}.$$

PROOF. Of the three norm axioms, only the triangle inequality is non-trivial. Let us check it. For $x, y \in X$ we have by Cauchy-Schwarz inequality that

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

This completes the proof. □

REMARK 1.4.7 (Pythagorean theorem). The calculation above clearly implies Pythagorean theorem: if $x \perp y$ then

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

REMARK 1.4.8 (Angle between vectors). The concept of inner product makes it possible to define the *angle between two vectors* x, y in an inner product space X. 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 $\theta(x, y)$ is the angle between x and y. Therefore, in a general inner product space X, it makes sense to define the angle between x, y by

$$\cos\theta(x, y) = \frac{\langle x, y\rangle}{\|x\|\|y\|}.$$

Cauchy-Schwarz inequality guarantees that the right hand side lies in $[-1, 1]$, so the angle exists. Nevertheless, the concept of angle is rarely used; one prefers to work with inner product directly.

Lec.8: 09/24/10

1.4.2. The space L_2. Hölder's inequality. The basic example of an inner product space is $L_2 = L_2(\Omega, \Sigma, \mu)$:

PROPOSITION 1.4.9 (Canonical inner product on L_2). *For $f, g \in L_2$, the quantity*

$$\langle f, g\rangle := \int f\bar{g}\, d\mu$$

is finite, and it defines an inner product on L_2. This inner product obviously agrees with the L_2 norm, i.e. $\|f\|_2 = \langle f, f\rangle^{1/2}$.

PROOF. The only non-trivial fact to prove is that $\langle f, g\rangle$ is finite, i.e. that $f\bar{g}$ is integrable. Since f, $\bar{g}$ and $f + \bar{g}$ belong to L_2, we have that f^2, $\bar{g}^2$ and $(f + \bar{g})^2 = f^2 + f\bar{g} + \bar{g}^2$ are integrable. Hence $f\bar{g}$ is integrable, as required. □

We can recast Cauchy-Schwarz inequality in this specific space L_2 as follows.

COROLLARY 1.4.10 (Cauchy-Schwarz inequality in L_2). *For every $f, g \in L_2$ one has*

$$\Big|\int f\bar{g}\, d\mu\Big| \leqslant \Big(\int |f|^2\, d\mu\Big)^{1/2}\Big(\int |g|^2\, d\mu\Big)^{1/2}.$$

The left hand side of Cauchy-Schwarz inequality can be replaced by the larger quantity $\int |fg|\, d\mu$. (This can be seen by applying Cauchy-Schwarz inequality for $|f|$, $|g|$.) Thus Cauchy-Schwarz inequality can be written as

$$\|fg\|_1 \leqslant \|f\|_2\|g\|_2.$$

Cauchy-Schwarz inequality is a partial case of the more general Hölder's inequality:

COROLLARY 1.4.11 (Hölder's inequality for functions). *Let $p, q \in (1, \infty)$ be adjoint, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Then for every $f \in L_p$ and $g \in L_q$ one has*

$$\Big|\int f\bar{g}\, d\mu\Big| \leqslant \Big(\int |f|^p\, d\mu\Big)^{1/p}\Big(\int |g|^q\, d\mu\Big)^{1/q}.$$

Before proving this result, note that similarly to Cauchy-Schwarz inequality, the left hand side in Hölder's inequality can be replaced by the larger quantity $\int |fg|\, d\mu$. Thus Hölder's inequality can be written as

$$\|fg\|_1 \leqslant \|f\|_p\|g\|_q. \tag{1.10}$$

PROOF OF COROLLARY 1.4.11. We are going to prove inequality (1.10). By rescaling, we can assume that $\|f\|_p = \|g\|_q = 1$ (In order to see this, divide both sides by $\|f\|_p$ and $\|g\|_q$). We shall use Young's inequality

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q} \quad \text{for } a, b \geqslant 0,$$

which follows by taking logarithm of both sides and using the concavity of the function $\log(x)$. Therefore, we have the pointwise inequality

$$|f(t)g(t)| \leqslant \frac{|f(t)|^p}{p} + \frac{|g(t)|^q}{q} \quad \text{for all } t \in \Omega.$$

Integrating yields

$$\int |fg| \leqslant \frac{1}{p} + \frac{1}{q} = 1$$

as required. □

Using Hölder's inequality we can clarify the scale of spaces $L_p = L_p(\Omega, \Sigma, \mu)$ for various p. Assume that μ is a *finite measure* (this is important!) Then L_∞ is the smallest space, L_1 is the largest, and all other L_p, $p \in [1, \infty)$ lie in between:

COROLLARY 1.4.12 (Scale of L_p spaces). *Let* (Ω, Σ, μ) *is a probability space, and* $1 \leqslant r \leqslant s \leqslant \infty$. *Then*

$$\|f\|_r \leqslant \|f\|_s \quad \textit{for all } f \in L_s = L_s(\Omega, \Sigma, \mu).$$

In particular, we have the inclusion

$$L_s \subseteq L_r.$$

PROOF. For $s = \infty$ this inequality is clearly true, so let $s < \infty$. We apply Hölder's inequality for $|f|^r$ and the constant function 1, and with $p = s/r$:

$$\begin{aligned}\|f\|_r^r = \int |f|^r \, d\mu = \int |f|^r \cdot 1 \, d\mu &\leqslant \Big(\int |f|^{r \cdot s/r} \Big)^{r/s} \Big(\int 1^q \, d\mu \Big)^{1/q} \\ &= \Big(\int |f|^s \Big)^{r/s} \mu(\Omega) = \|f\|_s^r.\end{aligned}$$

This completes the proof. □

1.4.3. The space ℓ_2. Recall that the space of square-summable sequences ℓ_2 is a particular case of $L_2 = L_2(\Omega, \Sigma, \mu)$ for the counting measure μ on $\Omega = \mathbb{N}$. Therefore, ℓ_2 is also an inner product space. The inner product formula reads as

$$\langle x, y \rangle = \sum_{i=1}^{\infty} x_i \overline{y_i} \quad \text{for } x = (x_i),\ y = (y_i) \in \ell_2.$$

Let us write out Cauchy-Schwarz inequality in this case:

COROLLARY 1.4.13 (Cauchy-Schwarz inequality in ℓ_2). *For every two sequences* $x = (x_i)$ *and* $y = (y_i)$ *in* ℓ_2 *one has*

$$\Big| \sum_{i=1}^{\infty} x_i \overline{y_i} \Big| \leqslant \Big(\sum_{i=1}^{\infty} |x_i|^2 \Big)^{1/2} \Big(\sum_{i=1}^{\infty} |y_i|^2 \Big)^{1/2}.$$

As before, the left hand side can be replaced by the larger quantity $\sum_{i=1}^{\infty} |x_i y_i|$. Similarly, Hölder's inequality in this case takes the following form:

COROLLARY 1.4.14 (Hölder's inequality for sequences). *Let $p, q \in (1, \infty)$ be adjoint, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Then for every $x = (x_i) \in \ell_p$ and $y = (y_i) \in \ell_q$ one has*

$$\sum_{i=1}^{\infty} |x_i y_i| \leqslant \Big(\sum_{i=1}^{\infty} |x_i^p| \Big)^{1/p} \Big(\sum_{i=1}^{\infty} |y_i^q| \Big)^{1/q}.$$

However, Corollary 1.4.12 on the scale of L_p (on finite measure spaces) does not hold for ℓ_p, because the counting measure on $\mathbb{N}$ is not finite. In fact, the scale is completely reversed in this case: ℓ_1 is the smallest space, ℓ_∞ is the largest, and the other ℓ_p, $p \in [1, \infty)$ lie in between:

COROLLARY 1.4.15 (Scale of ℓ_p spaces). *Let $1 \leqslant r \leqslant s \leqslant \infty$. Then*

$$\|x\|_s \leqslant \|x\|_r \quad \text{for all } x \in \ell_r.$$

In particular, we have the inclusion

$$\ell_r \subseteq \ell_s.$$

EXERCISE 1.4.16. **Prove Corollary 1.4.15. You don't have to use Hölder's inequality.**

1.4.4. The space of matrices. For arbitrary fixed dimensions $m, n \in \mathbb{N}$, we consider the space of matrices

$$M_{m,n} := \{\text{all } m \times n \text{ matrices with complex entries}\}.$$

One can turn $M_{m,n}$ into an inner product space by defining the trace inner product as

$$\langle A, B \rangle := \operatorname{tr}(AB^*) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \overline{b_{ij}}. \tag{1.11}$$

This is clearly an inner product. One way to see this is to identify $M_{m,n}$ with $\mathbb{C}^{mn}$ by concatenating the rows of a matrix $A \in M_{m,n}$ into a long vector in $\mathbb{C}^{mn}$. Then the canonical inner product in $\mathbb{C}^{mn}$ is the same as the right hand side of (1.11).

The norm defined by the inner product on $M_{m,n}$ is called *Hilbert-Schmidt* or *Frobenius* norm of matrices:

$$\|A\|_{\mathrm{HS}} = \langle A, A \rangle^{1/2} = \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \Big)^{1/2} \tag{1.12}$$

Note some similarity between the forms of the inner product in L_2, which is $\langle f, g \rangle = \int f \bar{g} \, d\mu$ and in $M_{m,n}$, which is $\langle A, B \rangle := \operatorname{tr}(AB^*)$ – the integral is replaced by the trace, functions by matrices, complex conjugation by transposition, and product of functions by product of matrices.

1.4.5. A space of random variables. The space L_2 arises in probability theory in a natural way as a *space of random variables* with finite variance. Indeed, consider a probability space $(\Omega, \Sigma, \mathbb{P})$. Recall that a random variable X is a measurable real-valued function defined on Ω. The expectation of X is, by definition, the integral of X:

$$\mathbb{E}X = \int X(\omega) \, d\mathbb{P}(\omega).$$

Therefore, the space $L_2(\Omega, \Sigma, \mathbb{P})$ consists of all random variables X with *finite second moment*:

$$\|X\|_2 = (\mathbb{E}X^2)^{1/2} < \infty.$$

EXERCISE 1.4.17. **Show that $X \in L_2(\Omega, \Sigma, \mathbb{P})$ if and only if X has *finite variance*:**

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 < \infty.$$

The concepts of covariance and correlation coefficient have some geometric meaning, too. Consider two random variables X and Y, and for simplicity assume that they have mean zero, i.e. $\mathbb{E}X = \mathbb{E}Y = 0$. Then the covariance of X and Y is nothing else than the inner product in L_2:

$$\mathrm{cov}(X, Y) := \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY = \langle X, Y\rangle.$$

Similarly, the correlation coefficient between X and Y is

$$\mathrm{corr}(X, Y) := \frac{\mathrm{cov}(X, Y)}{\mathrm{Var}(X)^{1/2}\,\mathrm{Var}(Y)^{1/2}} = \frac{\mathbb{E}XY}{(\mathbb{E}X^2)^{1/2}(\mathbb{E}Y^2)^{1/2}} = \frac{\langle X, Y\rangle}{\|X\|_2\|Y\|_2}.$$

Hence the correlation coefficient is nothing else as the cosine of the *angle* between random variables X and Y considered as vectors in L_2 (see Remark 1.4.8). This demonstrates the geometric meaning of correlation – the more random variables X and Y are correlated, the less the angle between them, and vice versa. Lec.9: 09/27

1.4.6. Parallelogram law. The parallelogram law in planar geometry states that for every parallelogram, the sums of squares of the diagonals equals the sum of squares of the sides. This statement remains to be true in all inner product spaces:

PROPOSITION 1.4.18 (Parallelogram law). *Let X be an inner product space. Then for every $x, y \in X$ one has*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{1.13}$$

PROOF. The result follows once we recall that $\|x + y\|^2 = \langle x + y, x + y\rangle = \|x\|^2 + 2\,\mathrm{Re}\langle x, y\rangle + \|y\|^2$ and similarly $\|x - y\|^2 = \|x\|^2 - 2\,\mathrm{Re}\langle x, y\rangle + \|y\|^2$. □

The parallelogram law characterizes inner product spaces. First recall that in inner product spaces, the inner product determines the norm ($\|x\| = \langle x, x\rangle^{1/2}$). Vice versa, the inner product is uniquely determined by the norm, and it can be reconstructed through the polarization identity:

PROPOSITION 1.4.19 (Polarization identity). *Let X be an inner product space. Then for every $x, y \in X$ one has*

$$\langle x, y\rangle = \frac{1}{4}\Big(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 + i\|x - y\|^2\Big). \tag{1.14}$$

EXERCISE 1.4.20. **Prove the polarization identity.**

The parallelogram law characterizes inner product spaces:

THEOREM 1.4.21 (Characterization of inner product spaces). *Let X be a normed space over $\mathbb{C}$ such that every two vectors $x, y \in X$ satisfy the parallelogram law* (1.13)*. Then X is an inner product space. Precisely, polarization formula* (1.14) *defines the inner product on X that agrees with the norm, i.e. $\|x\| = \langle x, x\rangle^{1/2}$.*

The proof of this result is deferred to the exercises for this section.

It follows from Theorem 1.4.21 that being an inner product space is a "local" property, since checking the parallelogram law involves just two (arbitrary) vectors. In particular, if all two-dimensional linear subspaces of a normed space X are inner-product spaces (with respect to some inner product, possibly different for each subspace), then X is an inner product space (and there the inner product on all subspaces is actually the same, induced from X!)

1.4.7. Additional Exercises.

EXERCISE 1.4.22. **[Direct sum of inner product spaces] Let X, Y be inner product spaces. Show that their direct sum**

$$X \oplus_2 Y := \{(x, y) : \ x \in X, \ y \in Y\}$$

is also an inner product space, with the inner product defined as

$$\langle (x_1, y_1), (x_2, y_2) \rangle = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle.$$

Derive a formula for the norm in $X \oplus_2 Y$. Show that if X, Y are Hilbert spaces then so is $X \oplus_2 Y$.

EXERCISE 1.4.23. **Show that the inner product is a continuous function on the direct sum $X \oplus_2 X$ of a Hilbert space X with itself.**

EXERCISE 1.4.24. **[Characterization of inner product spaces] Prove Theorem 1.4.21. For simplicity, do this for spaces over $\mathbb{R}$. (Formulate the polarization identity in this case). Axioms (i) and (iii) of inner product are simple to check. To prove axiom (iii) (linearity), you may follow this sequence of steps:**

- **(i) Show that $\langle x_1 + x_2, y \rangle + \langle x_1 - x_2, y \rangle = 2\langle x_1, y \rangle$ for all $x_1, x_2, y \in X$.**
- **(ii) Deduce the additivity property: $\langle u, y \rangle + \langle v, y \rangle = \langle u + v, y \rangle$ for all $u, v, y \in X$.**
- **(iii) From additivity property, deduce the multiplicativity property $\langle tx, y \rangle = t\langle x, y \rangle$ for all $x, y \in X$, $t \in \mathbb{R}$. Indeed, for $t \in \mathbb{N}$ this follows by induction; deduce the same for all $t \in \mathbb{Q}$ and extend to all $t \in \mathbb{R}$ by continuity.**

EXERCISE 1.4.25. **Show that $C(K)$, c_0, $L_p[0, 1]$, ℓ_p for $p \in [1, \infty]$, $p \neq 2$, are not inner product spaces. (More accurately, it is not possible to define an inner product on those spaces which would agree with their norms). Use Theorem 1.4.21.**

1.5. Hilbert spaces

1.5.1. Definition. Orthogonal complements. As in the case of normed spaces, we will really need an extra axiom – completeness – in order to study geometry of inner product spaces.

DEFINITION 1.5.1. A complete inner product space is called a *Hilbert space*.

EXAMPLE 1.5.2. As we know, $L_2(\Omega, \Sigma, \mu)$ is a Hilbert space. In particular, ℓ_2 is a Hilbert space. Also, $\mathbb{C}^n$ and $\mathbb{R}^n$ are Hilbert spaces; therefore the space of matrices $M_{m,n}$ is also a Hilbert space.

The geometry of Hilbert spaces is dominated by the concept of orthogonality. We will exploit it in detail now.

DEFINITION 1.5.3 (Orthogonal complement). Let A be a subset of an inner product space X. The *orthogonal complement of A* is defines as

$$A^\perp := \{x \in X : \langle x, a\rangle = 0 \text{ for all } a \in A\}.$$

PROPOSITION 1.5.4. *Let A be a subset of an inner product space X. Then $A^\perp$ is a closed linear subspace of X. Moreover,*

$$A^\perp \cap A \subseteq \{0\}.$$

PROOF. It is easy to check that $A^\perp$ is a linear subspace of X. To show that $A^\perp$ is a closed set, express it as

$$A^\perp = \bigcap_{a \in A} \{a\}^\perp.$$

Therefore[10] it suffices to check that $\{a\}^\perp$ is a closed set for every $a \in A$. So we fix $a \in A$ and consider a sequence $x_n \in \{a\}^\perp$ such that $x_n \to x$ for some $x \in X$. We would like to show that $x \in \{a\}^\perp$. To this end, notice that continuity of the inner product (Exercise 1.4.23) implies that

$$0 = \langle x_n, a\rangle \to \langle x, a\rangle.$$

Hence $\langle x, a\rangle = 0$ as required.

Finally, to show that $A^\perp \cap A \subseteq \{0\}$, consider $x \in A^\perp \cap A$; it follows that $\langle x, x\rangle = 0$ which implies $x = 0$. The proof is complete. □

1.5.2. Orthogonality principle. Orthogonal decompositions.

THEOREM 1.5.5 (Orthogonality principle). *Let Y be a closed linear subspace of a Hilbert space X, and let $x \in X$. Then the following holds.*

(i) There exists a unique closest point $y \in Y$ to x, i.e. such that

$$\|x - y\| = \min_{y' \in Y} \|x - y'\|.$$

(ii) The point y is the unique vector in Y such that

$$x - y \in Y^\perp.$$

The point y is called the orthogonal projection *of x onto the subspace Y.*

PROOF. (i). *Existence.* Denote the distance by

$$d := \inf_{y' \in Y} \|x - y'\|.$$

Let us choose a sequence (y_n) in Y which satisfies

$$\|x - y_n\| \to d.$$

Since Y is closed, it is a Banach space. Therefore it suffices to show that (y_n) is a Cauchy sequence. (By continuity of the norm it would follow for $y = \lim_n y_n$ that $\|x - y\| = d$.)

[10] Recall that the intersection of closed sets is closed.

To bound $\|y_n - y_m\|$, we use parallelogram law. We apply it for the parallelogram with vertices x, y_n, y_m (and whose fourth vertex is determined by these three, see the picture.)

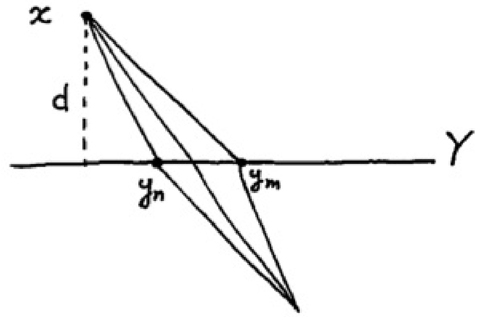

FIGURE 1.3. Proof of Theorem 1.5.5

Parallelogram law then yields

$$\|y_n - y_m\|^2 + 4\|x - \frac{1}{2}(y_n + y_m)\|^2 = 2\|x - y_n\|^2 + 2\|x - y_m\|^2.$$

By definition of d, we have $\|x - \frac{1}{2}(y_n + y_m)\| \geqslant d$, and by construction we have $\|x - y_n\| \to d$, $\|x - y_m\| \to d$. With this, we conclude that

$$0 \leqslant \liminf \|y_n - y_m\|^2 \leqslant 2d^2 + 2d^2 - 4d^2 = 0.$$

Therefore (y_n) is a Cauchy sequence as required.

Uniqueness. Suppose there existed two different closest points y_1, y_2' for x. Then the alternating sequence $(y_1, y_2, y_1, y_2, \ldots)$ would not be Cauchy, contradicting the argument above. Part (i) of Theorem is proved. Lec.10: 09/29

(ii). *Orthogonality.* Assume that $x - y \notin Y^\perp$, so

$$\langle x - y, y' \rangle \neq 0 \quad \text{for some } y' \in Y.$$

By multiplying y' by an appropriate complex scalar, we can assume that $\langle x - y, y' \rangle$ is a real number. We will show that by moving y in the direction of y', one can improve the distance from x, which will contradict the definition of y. Namely, for every $t \in \mathbb{R}$ the definition of y implies that

$$\|x - y\|^2 \leqslant \|x - y + ty'\|^2 = \|x - y\|^2 + 2t\langle x - y, y' \rangle + t^2\|y'\|^2.$$

This implies that the quadratic polynomial in t satisfies

$$\|y'\|^2 t^2 + 2\langle x - y, y' \rangle t \geqslant 0 \quad \text{for all } t \in \mathbb{R}.$$

This can only happens if $\langle x - y, y' \rangle = 0$, which contradicts our assumption.

Uniqueness. Suppose there are two vectors $y', y'' \in Y$ which satisfy $x - y' \in Y^\perp$, $x - y'' \in Y^\perp$. Since $Y^\perp$ is a linear subspace, subtracting yields that $y' - y'' \in Y^\perp$. But Y is also a linear subspace, so $y - y' \in Y$. Since $Y \cap Y^\perp = \{0\}$, it follows that $y - y' = 0$, hence $y' = y''$. Theorem is proved. $\square$

In the proof of part (i) of Theorem 1.5.5, we used *convexity* rather than linearity of Y. (Indeed, we needed that together with two points $y_n, y_m \in Y$ their midpoint $\frac{1}{2}(y_n + y_m)$ is contained in Y). Therefore, our argument implies the following more general result:

THEOREM 1.5.6 (Hilbert's projection theorem). *Given a closed convex set Y in a Hilbert space X and a point $x \in X$, there exists a unique closest point $y \in Y$.*

The map that takes x into the closest point y is called a *projection onto convex set* Y and is abbreviated POCS. This map appears in several applied fields.

The orthogonality principle immediately implies that a Hilbert space X can be decomposed into the orthogonal sum of a subspace Y and its complement $Y^\perp$:

COROLLARY 1.5.7 (Orthogonal decomposition). *Let X be a Hilbert space and Y be a closed subspace. Then every vector $x \in X$ can be uniquely represented as*

$$x = y + z, \quad y \in Y, \quad z \in Y^\perp.$$

This orthogonal decomposition is usually abbreviated as

$$X = Y \oplus Y^\perp.$$

DEFINITION 1.5.8 (Orthogonal projection). In the setting of Corollary 1.5.7, the map

$$P_Y : X \to X, \quad P_Y x = y$$

is called the *orthogonal projection* in X onto Y.

EXERCISE 1.5.9. **Show that the orthogonal projection P_Y is a linear map. Check that $\mathrm{Im}(P_Y) = Y$ and $\ker(P_Y) = Y^\perp$. Also check that the identity map I_X on X can be decomposed as**

$$I_X = P_Y + P_{Y^\perp}.$$

1.5.3. Additional Exercises.

EXERCISE 1.5.10. **Let A be a subset of a Hilbert space. Show that**

$$A^\perp = \bar{A}^\perp$$

where $\bar{A}$ denotes the closure of A.

EXERCISE 1.5.11. **[Projection onto constants] Let Y denote the subspace of constant functions in $L_2 = L_2(\Omega, \Sigma, \mu)$. Compute $P_Y f$ for an arbitrary function $f \in L_2$.**

1.6. Fourier series

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

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k) e^{ikt}$$

with Fourier coefficients

$$\widehat{f}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-ikt}\, dt \tag{1.15}$$

are called the Fourier coefficients of f. In order to make Fourier analysis rigorous, one has to understand what functions f can be written as Fourier series, and in what sense the Fourier series converges.

In order to do so, it is of great advantage to depart from this specific situation and carry out Fourier analysis in an abstract Hilbert space. We will regard the function $f(t)$ as a vector in the function space $L_2[-\pi, \pi]$. The exponential functions e^{-ikt} will form a set of orthogonal vectors in this space. Fourier series will then become an orthogonal decomposition of a vector f with respect to an orthogonal system of coordinates.

1.6.1. Orthogonal systems.

Definition 1.6.1 (Orthogonal system). A sequence (x_k) in a Hilbert space X is called an *orthogonal system* if

$$\langle x_k, x_l \rangle = 0 \quad \text{for all } k \neq l.$$

If additionally $\|x_k\| = 1$ for all k, the sequence (x_k) is called an *orthonormal system.*

Equivalently, (x_k) is an orthonormal system if

$$\langle x_k, x_l \rangle = \delta_{kl}$$

where δ_{kl} equals 1 if $k = l$ and 0 otherwise (it is called Kronecker's delta function).

Example 1.6.2 (Canonical basis of ℓ_2). In the space ℓ_2, consider the vectors

$$x_k = (0, \ldots, 0, 1, 0, \ldots)$$

whose all coordinates are zero except the k-th equals 1. The sequence $(x_k)_{k=1}^{\infty}$ is clearly an orthonormal system in ℓ_2.

Example 1.6.3 (Fourier basis in L_2). In the space[11] $L_2[-\pi, \pi]$, consider the exponentials

$$e_k(t) = \frac{1}{\sqrt{2\pi}} e^{ikt}, \quad t \in [-\pi, \pi]. \tag{1.16}$$

Then $(e_k)_{k=-\infty}^{\infty}$ is an orthonormal system in $L_2[-\pi, \pi]$ (check!).

Example 1.6.4 (Trigonometric system in L_2). Closely related to the Fourier basis is the trigonometric system. Note that we can write the exponentials from the previous example as

$$f_k(t) = \frac{1}{\sqrt{2\pi}} \big(\cos(kt) + i \sin(kt) \big).$$

Considering the real and imaginary parts separately, we see that the system

$$\Big\{ \frac{1}{\sqrt{2\pi}},\ \frac{1}{\sqrt{\pi}} \cos(t),\ \frac{1}{\sqrt{\pi}} \sin(t),\ \frac{1}{\sqrt{\pi}} \cos(2t),\ \frac{1}{\sqrt{\pi}} \sin(2t),\ \ldots \Big\}$$

is an orthonormal system in $L_2[-\pi, \pi]$ (check!)

[11] The space $L_2[-\pi, \pi]$ can be identified with $L_2(\mathbb{T})$ where $\mathbb{T}$ is the unit torus in $\mathbb{C}$. We can think of elements of this space as 2π-periodic functions.

EXERCISE 1.6.5. **Prove that an orthogonal system is a linearly independent set.**

1.6.2. Orthogonal series. The main interest in orthogonal systems is that they allow us to form orthogonal expansions of every vector $x \in X$. Such expansions are infinite series. So our first task will be to clarify when orthogonal series converge.

THEOREM 1.6.6 (Convergence of orthogonal series). *Let (x_k) be an orthogonal system in a Hilbert space X. Then the following are equivalent:*

(i) $\sum_k x_k$ *converges in X;*
(ii) $\sum_k \|x_k\|^2 < \infty$;
(iii) $\sum_k x_k$ *converges unconditionally in X, i.e. for every reordering of terms.*

In case of convergence, we have

$$\Big\| \sum_k x_k \Big\|^2 = \sum_k \|x_k\|^2. \tag{1.17}$$

The proof of this result is based on its finite version, which may be called the Pythagorean theorem in higher dimensions:

LEMMA 1.6.7 (Pythagorean theorem). *Let (x_k) be an orthogonal system in a Hilbert space X. Then for every $n \in \mathbb{N}$ one has*

$$\Big\| \sum_{k=1}^n x_k \Big\|^2 = \sum_{k=1}^n \|x_k\|^2. \tag{1.18}$$

PROOF. Using orthogonality, we see that the left hand side of (1.18)

$$\Big\langle \sum_{k=1}^n x_k, \sum_{k=1}^n x_k \Big\rangle = \sum_{k,l=1}^n \langle x_k, x_l \rangle = \sum_{k=1}^n \langle x_k, x_k \rangle$$

which equals the right hand side of (1.18). □

PROOF OF THEOREM 1.6.6. *(i) ⇔ (ii).* By the Cauchy criterion, the series $\sum_k x_k$ converges if and only if its partial sums form a Cauchy sequence in X, i.e.

$$\Big\| \sum_{k=n}^m x_k \Big\|^2 \to 0 \quad \text{as } n, m \to \infty. \tag{1.19}$$

Note that by Pythagorean theorem (Lemma 1.6.7), the quantity in (1.19) equals $\sum_{k=n}^m \|x_k\|^2$. So using Cauchy criterion again we see that (1.19) is equivalent to the convergence of the series $\sum_k \|x_k\|^2$, as required.

(ii) ⇒ (iii). The scalar series $\sum_k \|x_k\|^2$ converges absolutely, therefore also unconditionally (as we know from an analysis course). Hence, by the equivalence of (i) and (ii) proved above, the series $\sum_k x_k$ converges unconditionally.

(iii) ⇒ (i) is trivial.

The last part of the theorem, identity (1.17), follows by taking limit in Pythagorean identity (1.18). The theorem is proved. □

EXERCISE 1.6.8. **For what coefficients a_k does the series $\sum_{k=-\infty}^{\infty} a_k e^{ikt}$ converge in $L_2[-\pi, \pi]$?**

Lec.11: 10/01

1.6.3. Fourier series.

DEFINITION 1.6.9 (Fourier series). Consider an orthonormal system $(x_k)_{k=1}^{\infty}$ in a Hilbert space X and a vector $x \in X$. The *Fourier series* of x with respect to (x_k) is the formal series

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

The coefficients $\langle x, x_k \rangle$ are called the *Fourier coefficients* of x.

In order to understand the convergence of Fourier series, we will first focus on the finite case, and study the partial sums of Fourier series

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

LEMMA 1.6.10. *$S_n(x)$ is the orthogonal projection of x onto*[12] $\mathrm{Span}(x_1, \ldots, x_n)$.

PROOF. By the definition of the orthogonal projection (see the orthogonality principle, Theorem 1.5.5), the lemma states that

$$x - S_n(x) \in (\mathrm{Span}(x_1, \ldots, x_n))^{\perp}.$$

It suffices to check that $x - S_n(x) \perp x_k$ for all $k = 1, \ldots, n$. We have

$$\langle x - S_n(x), x_k \rangle = \langle x, x_k \rangle - \langle S_n(x), x_k \rangle.$$

By definition of $S_n(x)$ and orthonormality of (x_k) we see that $\langle S_n(x), x_k \rangle = \langle x, x_k \rangle$. Therefore we conclude that $\langle x - S_n(x), x_k \rangle = 0$ as required. □

Let us estimate the size of $S_n(x)$. Since $x - S_n(x) \perp S_n$, by Pythagorean theorem we have $\|S_n(x)\|^2 + \|x - S_n(x)\|^2 = \|x\|^2$. Hence

$$\|S_n(x)\|^2 \leqslant \|x\|^2.$$

On the other hand, by Pythagorean theorem and orthonormality,

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

Combining these two inequalities and letting $n \to \infty$, we arrive at the following result.

THEOREM 1.6.11 (Bessel's inequality). *Let (x_k) be an orthonormal system in a Hilbert space X. Then for every $x \in X$ one has*

$$\sum_k |\langle x, x_k \rangle|^2 \leqslant \|x\|^2.$$

This result along with the convergence criterion for orthogonal series, Theorem 1.6.6, shows that Fourier series always converge.

COROLLARY 1.6.12. *Let (x_k) be an orthonormal system in a Hilbert space X. Then the Fourier series $\sum_k \langle x, x_k \rangle x_k$ of every vector $x \in X$ converges in X.*

---

[12] Recall that the linear span of vectors $x_1, \ldots, x_n$ is defined as $\mathrm{Span}(x_1, \ldots, x_n) = \{x = \sum_{k=1}^{n} a_k x_k : a_k \in \mathbb{C}\}$.

In general, Fourier series of x needs not converge to x (why?) Still we can compute the point where it converges. Indeed, letting $n \to \infty$ in Lemma 1.6.10 and interpreting the conclusion via the orthogonality principle (Theorem 1.5.5), we conclude:

THEOREM 1.6.13 (Optimality of Fourier series). *Consider an orthonormal system $(x_k)_{k=1}^{\infty}$ in a Hilbert space X and a vector $x \in X$.*

(i) The Fourier series $\sum_k \langle x, x_k \rangle x_k$ is the orthogonal projection of x onto $\overline{\mathrm{Span}(x_k)}$ (the closure of the linear span).

(ii) 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.

EXERCISE 1.6.14. **Prove Theorem 1.6.13.**

1.6.4. Orthonormal bases. Fourier expansions. It is now easy to identify an extra condition so that the Fourier series of every vector x converges to x.

DEFINITION 1.6.15 (Complete systems). A system of vectors (x_k) in a Banach space X is called *complete* if

$$\overline{\mathrm{Span}(x_k)} = X.$$

A complete orthonormal system in a Hilbert space X is called an *orthonormal basis* of X.

THEOREM 1.6.16 (Fourier expansions). *Let (x_k) be an orthonormal basis of a Hilbert space X. Then every vector $x \in X$ can be expanded in its Fourier series:*

$$x = \sum_k \langle x, x_k \rangle x_k. \tag{1.20}$$

Consequently, Parseval's identity *holds:*

$$\|x\|^2 = \sum_k |\langle x, x_k \rangle|^2.$$

PROOF. The first part follows from the Optimality Theorem 1.6.13, since by completeness the orthogonal projection onto $\overline{\mathrm{Span}(x_k)} = X$ is the identity map in X. Parseval's identity follows from Fourier expansion (1.20), Pythagorean identity (1.17) for orthogonal series, and the normalization condition $\|x_k\| = 1$. □

EXERCISE 1.6.17. **Prove that Parseval's identity holds for an orthonormal system (x_k) if and only if (x_k) is complete. Therefore the equality cases of Bessel's inequality hold exactly when the system is complete.**

Now we describe some classical examples of complete sets and orthonormal bases.

EXAMPLE 1.6.18 (Monomials). Weierstrass approximation theorem states that the system of monomials $(t^k)_{k=0}^{\infty}$ is a complete system in $C[0,1]$. We claim that this is also a complete system in $L_2[0,1]$.

Indeed, $C[0,1]$ is dense in $L_2[0,1]$. This means that for every $f \in L_2[0,1]$ and $\varepsilon > 0$, there exists $g \in C[0,1]$ such that $\|f - g\|_2 \leqslant \varepsilon/2$. By Weierstrass approximation theorem, there exists $h \in \mathrm{Span}(t^k)_{k=0}^{\infty}$ such that $\|g - f\|_\infty \leqslant \varepsilon/2$. Hence $\|g - f\|_2 \leqslant \|g - f\|_\infty \leqslant \varepsilon/2$, so by triangle inequality we conclude that $\|f - h\|_2 \leqslant \varepsilon/2 + \varepsilon/2 = \varepsilon$. We have proved that $\mathrm{Span}(t^k)_{k=0}^{\infty}$ is dense in $L_2[0,1]$ as required.

EXAMPLE 1.6.19 (Exponentials). By a general version of Weierstrass approximation theorem (called Stone-Weierstrass theorem), the exponential monomials $(e^{itk})_{k\in\mathbb{Z}}$ is a complete system in $C[-\pi,\pi]$. Repeating the argument in Example 1.6.18, we can check that this is also a complete system in $L_2[-\pi,\pi]$.

Therefore, the system of exponentials

$$x_k(t) = \frac{1}{\sqrt{2\pi}} e^{ikt}, \quad k \in \mathbb{Z}, \tag{1.21}$$

forms an orthonormal basis of $L_2[-\pi,\pi]$. Reformulating Theorem 1.6.16 in this case, we obtain a basic result in classical Fourier analysis:

THEOREM 1.6.20 (Classical Fourier series). *Every function $f \in L_2[-\pi,\pi]$ can be represented by its Fourier series*

$$f(t) = \sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikt}, \quad \text{where} \quad \widehat{f}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t) e^{-ikt}\, dt.$$

The coefficients $\widehat{f}(k)$ are all finite; the Fourier series converges in $L_2[-\pi,\pi]$.

EXAMPLE 1.6.21 (Trigonometric system). In a similar way we can show that the trigonometric system considered in Example 1.6.4 is an orthonormal basis in $L_2[-\pi,\pi]$ (do this!) Therefore a version of Theorem 1.6.20 holds for the trigonometric system, and it reads as follows:

$$f(t) = \frac{a_0}{2} \sum_{k=1}^{\infty} \big[a_k \cos(kt) + b_k \sin(kt)\big]$$

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\, dt, \quad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)\, dt.$$

This again holds for every function $f \in L_2[-\pi,\pi]$; the coefficients a_k, b_k are all finite, and the Fourier series converges in $L_2[-\pi,\pi]$.

Lec.12: 10/04

1.6.5. Gram-Schmidt orthogonalization. There is a general way of creating an orthonormal basis (h_k) in a Hilbert space X out of some other, possibly non-orthogonal system (x_k). One orthogonalize the system (x_k) one element at a time. This procedure is called *Gram-Schmidt orthogonalization.*

So let us consider a linearly independent system of vectors $(x_k)_{k=1}^{\infty}$ in X. We define the system $(h_k)_{k=1}^{\infty}$ inductively as follows:

$$h_1 = \frac{x_1}{\|x_1\|}, \quad h_{n+1} = \frac{P_n x_{n+1}}{\|P_n x_{n+1}\|}, \quad n = 1, 2, \dots$$

where P_n denotes the orthogonal projection in X onto $\mathrm{Span}(h_1,\dots,h_n)^{\perp}$. Geometrically, one "rotates" the new vector x_{n+1} so it becomes orthogonal to the vectors h_k constructed earlier, normalizes it, and calls h_{n+1}; see the picture.

One can effectively compute the vectors $P_n x_{n+1}$ used in this process. Indeed, by Lemma 1.6.10, the orthogonal projection of a vector x onto $\mathrm{Span}(h_1,\dots,h_n)$ is the partial sum of Fourier series:

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

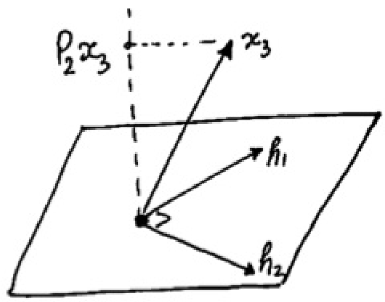

FIGURE 1.4. Gram-Schmidt orthogonalization

Now, $P_n(x)$ is the othogonal projection onto the orthogonal complement, so

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

So

$$P_n x_{n+1} = x_{n+1} - \sum_{k=1}^{n} \langle x_{n+1}, h_k \rangle h_k.$$

PROPOSITION 1.6.22. *Let (x_k) be a linearly independent system in a Hilbert space X. Then the system (h_k) obtained by Gram-Schmidt orthogonalization of (x_k) is an orthonormal system in X, and*

$$\operatorname{Span}(h_k)_{k=1}^n = \operatorname{Span}(x_k)_{k=1}^n \quad \textit{for all } n \in \mathbb{N}.$$

PROOF. The system (h_k) is orthonormal by construction. Also, we obviously have the inclusion of the subspaces $\operatorname{Span}(h_k)_{k=1}^n \subseteq \operatorname{Span}(x_k)_{k=1}^n$, and the dimensions of these subspaces both equal n by construction. Therefore, these subspaces are equal. □

EXERCISE 1.6.23. **[Legendre polynomials] Prove that Gram-Schmidt orthogonalization of the monomials $(t^k)_{k=0}^\infty$ in the space $L_2[-1,1]$ gives the system of *Legendre orthogonal polynomials***

$$P_k(t) = \frac{1}{2^k k!} \frac{d^k}{dt^k} (t^2 - 1)^k,$$

up to normalization constants. More precisely, $P_k(t)$ form an orthogonal basis in $L_2[-1,1]$ and $\|P_k\|_2^2 = 2/(2k+1)$.

In a similar way, by orthogonalization of monomials in different spaces, one produces other classical systems of orthogonal polynomials, such as Hermite (see Exercise 1.6.38 and Chebychev polynomials.

1.6.6. Existence of orthogonal bases. As a consequence of Proposition 1.6.22, we will obtain that every Hilbert space that is not "too large" has an orthonormal bases. Such are all *separable* Hilbert spaces.

Recall that a metric space is separable if it contains a countable dense subset.

EXERCISE 1.6.24. **Prove that the following Banach spaces X are separable by describing a countable dense subset in X: $C[0,1]$, $L_2[0,1]$, c_0, ℓ_p for $p \in [1,\infty)$.**

EXERCISE 1.6.25. **Prove that the spaces ℓ_∞ and L_∞ are not separable by constructing an uncountable separated subset A, i.e. such that**

$$\inf\{\|x-y\| : x, y \in A, \ x \neq y\} > 0.$$

For Banach spaces, separability is formally equivalent to a slightly stronger property:

LEMMA 1.6.26 (Separable spaces). *A Banach space X is separable if and only if it contains a system of vectors $(x_k)_{k=1}^\infty$ whose linear span is dense in X, i.e.*

$$\overline{\operatorname{Span}(x_k)} = X.$$

PROOF. *Necessity.* If X is separable, it contains a system of vectors $(z_k)_{k=1}^n$ whose linear span is dense in X. We construct (x_k) inductively as a subset of (z_k). Namely, we include z_1, and if $z_{n+1} \notin (z_1, \ldots, z_n)$ we include z_n for $n = 1, 2, \ldots$. By construction, (x_k) is linearly independent and $\operatorname{Span}(x_1, \ldots, x_n) = \operatorname{Span}(z_1, \ldots, z_n)$ for all $n = 1, 2, \ldots$ Letting $n \to \infty$ we conclude that $\overline{\operatorname{Span}(x_1, \ldots, x_n)} = \overline{\operatorname{Span}(z_1, \ldots, z_n)} = X$ as required.

Sufficiency. If $\operatorname{Span}(x_k)$ is dense in X, so is the set of all finite linear combinations $\sum_{k=1}^n a_k x_k$ with $a_k \in \mathbb{Q}$, which is a countable set. The lemma is proved. □

THEOREM 1.6.27. *Every separable Hilbert space has an orthonormal basis.*

PROOF. Let X be a separable Hilbert space. By Lemma 1.6.26, there is a system of vectors (x_k) in X such that $\overline{\operatorname{Span}(x_k)} = X$. Applying the Gram-Schmidt orthogonalization, we obtain an orthonormal system (h_k) in X. Using Proposition 1.6.22 and letting $n \to \infty$, we conclude that

$$\overline{\operatorname{Span}(h_k)} = \overline{\operatorname{Span}(x_k)} = X.$$

This proves the completeness of (h_k). Hence (h_k) is an orthonormal basis in X. □

EXERCISE 1.6.28. **Prove the converse for Theorem 1.6.27. Namely, if a Hilbert space X has an orthonormal basis then X is separable.**

REMARK 1.6.29. We developed the theory for countable orthogonal systems and bases. One can generalize it for systems of arbitrary cardinality, but we will not do this here.

1.6.7. Isometry of all separate Hilbert spaces. We are ready to show that all Hilbert spaces of the same cardinality have "the same geometry":

THEOREM 1.6.30. *All infinite-dimensional separable Hilbert spaces are isometric to each other. Precisely, for every such spaces X and Y, one can find a linear bijective map $T : X \to Y$ which preserves the inner product, i.e.*

$$\langle Tx, Ty \rangle = \langle x, y \rangle \quad \text{for all } x, y \in X. \tag{1.22}$$

PROOF. Let (x_k) and (y_k) be orthonormal bases of spaces X and Y respectively. Let T be the map that takes x_k to y_k. More precisely, define T by

$$T\Big(\sum_k a_k x_k\Big) = \sum_k a_k y_k.$$

Note that every $x \in X$ has the form $x = \sum_k a_k x_k$ for some (Fourier) coefficients a_k, so the definition makes sense. Also, by Parseval's identity,

$$\Big\| \sum_k a_k x_k \Big\|^2 = \sum_k |a_k|^2 = \Big\| \sum_k a_k y_k \Big\|^2. \tag{1.23}$$

Therefore, T is well defined on X, its inverse is also well defined as

$$T^{-1}\Big(\sum_k a_k y_k \Big) = \sum_k a_k x_k,$$

so T is bijective and clearly linear. Additionally, (1.23) shows that

$$\|Tx\| = \|x\| \quad \text{for all } x \in X.$$

So T preserves the norm. Since by polarization formula, the inner product is uniquely determined by the norm, T must also preserve the inner product, i.e. (1.22) holds. This completes the proof. □

Remark 1.6.31. It follows from Theorem 1.6.30 that $\|Tx\| = \|x\|$ for all $x \in X$. Thus by linearity,

$$\|Tx - Ty\| = \|x - y\| \quad \text{for all } x, y \in X.$$

Thus T preserves all pairwise distances; hence the name "isometry".

Remark 1.6.32. Since ℓ_2 and $L_2[0,1]$ are separable Hilbert spaces, it follows that every Hilbert space is isometric to ℓ_2 and $L_2[0,1]$.

1.6.8. Additional Exercises.

Exercise 1.6.33. **[Frames] Consider the space $L_2[a,b]$ where $-\pi < a < b < \pi$. Show that the system of exponentials (1.21) satisfies Parseval's identity for every function $x \in L_2[a,b]$, although it is not an orthogonal system. General systems that satisfy Parseval's identity are called *frames*. They are important in signal processing.**

Exercise 1.6.34. **[Haar system] Consider the function**

$$h(t) = \begin{cases} 1, & t \in [0, 1/2) \\ -1, & t \in [1/2, 1) \end{cases}$$

and define the functions

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

Together with the constant function 1, this system of functions is called the *Haar system*; the individual functions is called the *Haar wavelets* and the function $h(t)$ is called the Haar *mother wavelet*. Show that,, the Haar system is an orthonormal basis in $L_2[0,1]$. (Hint: first show that Haar system is complete in $C[0,1]$.)

EXERCISE 1.6.35. **[Rademacher system] Consider the functions**

$$r_k(t) = (-1)^{[2^k x]}, \quad k \in \mathbb{N} \cup \{0\},$$

where $[\cdot]$ denotes the integer part of a number. This system of functions is called the *Rademacher system*. Show that the Rademacher system is an orthonormal system in $L_2[0,1]$, but is not complete (therefore not an orthonormal basis).

EXERCISE 1.6.36. **[Walsh system] Consider the functions $w_A(t)$, $A \subset \mathbb{N} \cup \{0\}$ (indexed by subsets A rather than numbers!) defined by**

$$w_A(t) = \prod_{k \in A} r_k(t)$$

where $r_k(t)$ are the Rademacher functions. This system of functions is called the *Walsh system*. Show that the Walsh system is an orthonormal basis in $L_2[0,1]$.

EXERCISE 1.6.37. **[Characterization of complete systems] Consider a system of vectors (x_k) (not necessarily orthogonal) in a Hilbert space X. Prove that (x_k) is complete if and only if the only vector orthogonal to all of x_k is zero.**

EXERCISE 1.6.38. **[Hermite polynomials] Suppose $w(t)$ is a continuous weight function $\mathbb{R} \to \mathbb{R}_+$. Consider the Hilbert space $L_2(\mathbb{R}, w(t)\,dt)$, i.e. the measure on $\mathbb{R}$ is given by $w(t)\,dt$. The Gram-Schmidt orthogonalization of monomials $(t^k)_{k=0}^{\infty}$ produces a system orthogonal polynomials $P_k(t)$ with repsect to the weight $w(t)$, i.e.**

$$\int_{\mathbb{R}} P_k(t) P_l(t) w(t)\,dt = \delta_{kl}.$$

Now consider the weight $w(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, i.e. the standard normal density. Prove that the orthogonal polynomials with respect to this weight is the system of *Hermite polynomials*

$$P_k(t) = (-1)^k e^{t^2/2} \frac{d^k}{dt^k} e^{-t^2/2},$$

up to normalization constants. More precisely, $P_k(t)$ form an orthogonal basis in $L_2(\mathbb{R}, w(t)\,dt)$ and $\|P_k\|_2^2 = k!$.

EXERCISE 1.6.39. **[Space of almost periodic functions] A function $f : \mathbb{R} \to \mathbb{C}$ is called *almost periodic* if it is a superposition of a finite number of frequencies, i.e. f has the form**

$$f(t) = \sum_{k=1}^{n} a_k e^{i w_k t}, \quad \text{where } a_k \in \mathbb{C}, \quad w_k \in \mathbb{R}, \quad n \in \mathbb{N}.$$

Note that the frequencies w_k are allowed to take arbitrary real values. Denote the space of almost periodic functions by X_0, and equip it with the inner product

$$\langle f, g \rangle = \lim_{T \to \infty} \Big(\frac{1}{2T} \int_T^T |f(t)|^2\,dt \Big)^{1/2}.$$

Show that the inner product is well defined on X_0, and

$$\langle f, g\rangle = \sum_{k=1}^{n} a_k \overline{b_k} \quad \textbf{where } f(t) = \sum_{k=1}^{n} a_k e^{iw_k t},\ g(t) = \sum_{k=1}^{n} b_k e^{iw_k t}$$

(and where the frequencies w_k are all different). The completion X of the inner product space X_0 is called the *space of almost periodic functions.*

Prove that X is a non-separable Hilbert space by showing that the system of functions

$$\{e^{itw}, \quad w \in \mathbb{R}\}$$

is an orthonormal system in X.

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. Lec.13: 10/06

2.1. Bounded linear functionals

The most basic but rich class of linear operators are linear functionals, those that map a space into $\mathbb{R}$ or $\mathbb{C}$.

2.1.1. Definition and examples. At this moment, the topology does not matter, so we define linear functionals on general linear vector spaces.

Definition 2.1.1 (Linear functionals). Let E be a linear space over $\mathbb{C}$ (the real case is similar). A *linear functional* on E is a linear operator $f : E \to \mathbb{C}$. Equivalently, a function $f : E \to \mathbb{C}$ is a linear functional if

$$f(ax + by) = af(x) + bf(y) \quad \text{for all } x, y \in E, \; a, b \in \mathbb{C}.$$

Example 2.1.2 (Integration). The integral of an integrable function is a basic example of a linear functional. Specifically, the map

$$F(g) = \int g(t)\, d\mu$$

is clearly a linear functional on $L_1(\Omega, \Sigma, \mu)$.

Similarly, for a fixed weight function $w \in L^1[0, 1]$, the map

$$F(g) = \int_0^1 g(t)w(t)\, dt$$

defines a linear functional on $L_\infty[0, 1]$. (Check!)

Example 2.1.3 (Point evaluation functional). For a fixed $t_0 \in [0, 1]$, the map

$$F(g) = g(t_0) \tag{2.1}$$

is clearly a linear functional on $C[0, 1]$. It is called the *point evaluation functional* at t_0.

Physicists view the point evaluation functional as a special case of integration with weight:

$$g(t_0) = \int_0^1 g(t)\delta(t - t_0)\, dt \tag{2.2}$$

The weight here is given by *Dirac delta function* $\delta(t)$, which is zero for all values t except $\delta(t_0) = \infty$, and such that $\int_0^1 \delta(t)\, dt = 1$. Dirac delta function does not exist as a *function* $[0,1] \to \mathbb{R}$, and should be understood as a *linear functional*. The "integral" (2.2) of a function against the Dirac delta function should be understood as the point evaluation functional (2.1).

EXERCISE 2.1.4. **[Functionals on $\mathbb{C}^n$] Show that every linear functional f on $\mathbb{C}^n$ has the form**

$$f(x) = \sum_{k=1}^{n} x_k y_k = \langle x, y \rangle, \quad x = (x_1, \dots, x_n),$$

for some $y = (y_1, \dots, y_n) \in \mathbb{C}^n$.

EXAMPLE 2.1.5. More generally, we will soon show that every liner functional on a Hilbert space X has the form

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

for some $y \in X$. For now, we note that f defined this way is indeed a linear functional.

2.1.2. Continuity and boundedness. Dual space.

DEFINITION 2.1.6 (Continuity, boundedness). Let f be a linear functional on a normed space X.

(i) Recall that f is *continuous* if

$$x_n \to x \text{ in } X \quad \text{implies} \quad f(x_n) \to f(x).$$

(ii) f is called *bounded* if there exists a number C such that

$$|f(x)| \leqslant C\|x\| \quad \text{for all } x \in X.$$

PROPOSITION 2.1.7. *Continuity and boundedness of linear functionals are equivalent.*

PROOF. Assume that f is bounded, and let $x_n \to x$. Then

$$|f(x_n) - f(x)| = |f(x_n - x)| \leqslant C\|x_n - x\| \to 0.$$

Thus f is continuous.

Vice versa, assume that f is not bounded. Then we can find a sequence (x_n) of nonzero vectors in X such that

$$|f(x_n)| \geqslant n\|x_n\|, \quad n = 1, 2, \dots$$

Dividing both sides by $n\|x_n\|$ we obtain

$$\Big| f\Big(\frac{x_n}{n\|x_n\|}\Big)\Big| \geqslant 1, \quad n = 1, 2, \dots$$

On the other hand, $\frac{x_n}{n\|x_n\|} \to 0$ as the norm of these vectors equals $1/n$. This implies that f is not continuous. ☐

EXERCISE 2.1.8. **Prove that if f is continuous at a single point $x_0 \in X$ then f is continuous (everywhere on X).**

DEFINITION/PROPOSITION 2.1.9 (Dual space). Let X be a normed space. The space of all linear functionals f on X is a linear vector space. It is called the *dual space* and is denoted X^*. The dual space is a normed space, with the norm defined as

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

EXERCISE 2.1.10. **Prove the facts stated in this definition/proposition – that X^* is a normed space, the identity between the two norm definitions, and that $\|f\|$ indeed defines the norm on X^*.**

REMARK 2.1.11. The definition implies the following useful inequality:

$$|f(x)| \leqslant \|f\| \, \|x\| \quad \text{for all } x \in X, \; f \in X^*.$$

Also, $\|f\|$ is the smallest number in this inequality that makes it valid for all $x \in X$.

EXERCISE 2.1.12. **Compute the norms of the integration and the point evaluation functionals considered in Examples 2.1.2 and 2.1.3.**

2.1.3. Hyperplanes as level sets of linear functionals.

General functions, and in particular linear functionals f, on a linear vector space E may be visualized by describing their level sets

$$\{x \in X : \; f(x) = c\}$$

for various values $c \in \mathbb{C}$. The level set corresponding to $c = 0$ is the kernel of f.

It turns out that $\ker f$ is a *hyperplane*, i.e. a subspace of E of codimension 1. All other level sets of f are obviously the translates of $\ker f$, see the picture. Moreover, there is a canonical correspondence between the linear functionals and the hyperplanes in X. This is clarified in the following proposition.

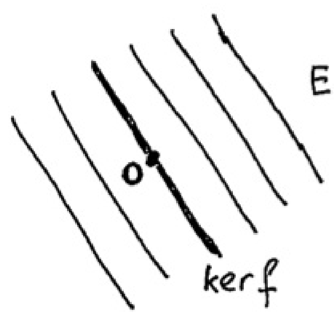

FIGURE 2.1. Level sets of a linear functional f on a linear vector space E

PROPOSITION 2.1.13 (Linear functionals and hyperplanes). *Let E be a linear vectors space.*

(i) For every linear functional f on E, $\ker f$ *is a hyperplane in E, i.e.* $\operatorname{codim}(\ker f) = 1$.

(ii) If $f, g \neq 0$ are linear functionals on E such that $\ker f = \ker g$, *then $f = ag$ for some scalar $a \neq 0$.*

(iii) For every hyperplane $H \subset E$ there exists a linear functional $f \neq 0$ such that $\ker f = H$.

Proof. (i) Follows from a linear version of the fundamental theorem on homomorphisms, Exercise 1.1.24. Indeed, the injectivization $\tilde{f} : E/\ker f \to \mathbb{C}$ of f establishes a linear bijection (isomorphism) between $E/\ker f$ and the range $\mathbb{C}$ of f. Thus $\dim(E/\ker f) = \dim(\mathbb{C}) = 1$, so $\ker f$ is a hyperplane in E.

(ii) Since $\ker f = \ker g =: H$, the injectivizations $\tilde{f}, \tilde{g} : E/H \to \mathbb{C}$ are linear functionals on the *one-dimensional* space E/H. A moment's thought yields that such linear functionals must be equal up to some constant factor a, i.e. $\tilde{f} = a\tilde{g}$. Then $\tilde{f}[x] = a\tilde{g}[x]$ for all $x \in E$. On the other hand, by construction of injectivization, $f(x) = \tilde{f}[x]$ and $g(x) = \tilde{g}[x]$. Therefore $f(x) = ag(x)$ as required.

(iii) Since $\dim(E/H) = 1$, we have

$$E/H = \{a[x_0] : \ a \in \mathbb{C}\}$$

for some $x_0 \in E$. Let $x \in E$ be arbitrary; then $[x] = a[x_0]$ for some $a = a(x) \in \mathbb{C}$, which implies $x = ax_0 + h$ for some $h \in H$. Let us define f on E by $f(x) = a$. Then f is a linear functional (check!), and clearly $\ker f = H$. □

Proposition 2.1.14. *Let f be a* bounded *linear functional, i.e. $f \in X^*$ then* $\ker f$ *is closed.*

Proof. $\ker f$ is the pre-image of the closed set $\{0\}$ under the continuous map f, so it must be closed. □

Remark 2.1.15. Using injectivization of f, one can show that the converse also holds. So, a linear functional f is bounded if and only if $\ker f$ is closed. It follows that the kernel of a linear functional is either closed or dense in X. (Why?)

Lec.14: 10/08

2.2. Representation theorems for linear functionals

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 and c_0) act by summation with weights.

2.2.1. Dual of a Hilbert space: Riesz representation theorem. We start by characterizing bounded linear functionals on a Hilbert space X. The following theorem says that every functional f acts as an inner product with some vector in X.

Theorem 2.2.1 (Riesz representation theorem). *Let X be a Hilbert space.*

(i) For every $y \in X$, the function

$$f(x) = \langle x, y \rangle, \quad x \in X \tag{2.3}$$

is a bounded linear functional on X, and its norm is $\|f\| = \|y\|$.

(ii) Conversely, for every bounded linear functional $f \in X^*$ there exists a unique vector $y \in X$ such that (2.3) *holds. Moreover, $\|f\| = \|y\|$.*

PROOF. (i) By Cauchy-Schwarz inequality,

$$|f(x)| = |\langle x, y\rangle| \leqslant \|x\| \, \|y\|, \quad x \in X.$$

Hence f is bounded, and $\|f\| \leqslant \|y\|$. Conversely, for $x = y$ we have

$$f(x) = \langle x, x\rangle = \|x\|^2.$$

Hence $\|f\| \geqslant \|y\|$.

(ii) Let $f \in X^*$. By Proposition 2.1.13, $\ker f$ is a hyperplane in X. Since f is bounded, $\ker f$ is closed (see Proposition 2.1.14). Therefore, X can be represented by the orthogonal decomposition

$$X = \ker f \oplus \mathrm{Span}(y_0) \quad \text{for some } y_0 \in (\ker f)^{\perp}.$$

(Use Theorem 1.5.7). Consider the map

$$g(x) := \langle x, y_0\rangle, \quad x \in X.$$

We have $g \in X^*$ by part (i). Moreover,

$$\ker g = \{y_0\}^{\perp} = \ker f.$$

Therefore, by Proposition 2.1.13, the functionals f and g are equal up to some constant factor a, that is $f = ag$. It follows that

$$f(x) = a\langle x, y_0\rangle = \langle x, a y_0\rangle,$$

and the conclusion follows with $y := a y_0$.

The uniqueness of y is simple. (Why?) □

REMARK 2.2.2. In a concise form, the statement of Riesz representation theorem can be expressed as

$$X^* = X.$$

Although X^* and X are formally different spaces, they can be canonically identified as in Riesz representation theorem.

Let us rewrite Riesz representation theorem for the Hilbert space L_2.

COROLLARY 2.2.3 ($L_2^* = L_2$). *Consider the space $L_2 = L_2(\Omega, \Sigma, \mu)$.*

(i) For every weight function $g \in L_2$, integration with weight

$$G(f) := \int f g \, d\mu, \quad f \in L_2$$

is a bounded linear functional on L_2, and its norm is $\|G\| = \|g\|_2 = (\int |g|^2 \, d\mu)^{1/2}$.

(ii) Conversely, every bounded linear functional $G \in L_2^*$ can be represented as integration with weight for some unique weight function $g \in L_2$. Moreover, $\|G\| = \|g\|_2$.

2.2.2. Application: proof of Radon-Nikodym theorem. Riesz representation theorem can be used to give a "soft" proof of Radon-Nikodym theorem in measure theory. This argument is due to von Neumann (1940).

Consider two measures μ, ν on the same σ-algebra. Recall that ν is called *absolutely continuous* with respect to μ, abbreviated $\nu \ll \mu$, if

$$\mu(A) = 0 \quad \text{implies} \quad \nu(A) = 0$$

for measurable sets A.

THEOREM 2.2.4 (Radon-Nikodym theorem). *Consider two finite measures*[1] μ, ν *such that* $\nu \ll \mu$. *Then* ν *can be expressed as*

$$\nu(A) = \int_A g \, d\mu \quad \text{for all measurable sets } A,$$

where $g \geqslant 0$ *is a measurable function. Moreover,* g *is uniquely determined* μ*-a.e.; it is called the Radon-Nikodym derivative and is denoted* $g =: d\nu/d\mu$.

PROOF. We shall prove the existence part only.

The linear functional

$$F(f) := \int f \, d\mu$$

is a bounded linear functional on the space $L_2(\mu)$, and therefore also on the space $L_2(\mu+\nu)$. By Riesz representation theorem, there exists $h \in L_2(\mu+\nu)$ such that

$$\int f \, d\mu = \int fh \, d(\mu+\nu) = \int fh \, d\mu + \int fh \, d\nu \quad \text{for all } f \in L_2(\mu+\nu). \tag{2.4}$$

Rearranging the terms, we obtain

$$\int fh \, d\nu = \int f(1-h) \, d\mu \quad \text{for all } f \in L_2(\mu+\nu). \tag{2.5}$$

We claim that

$$0 < h \leqslant 1 \quad \mu\text{-a.e.} \tag{2.6}$$

Indeed, consider the set $A = \{h \leqslant 0\}$ and the indicator function $f = \mathbf{1}_A$. In this case (2.4) becomes $\mu(A) = \int_A h \, d(\mu+\nu) \leqslant 0$, hence $\mu(A) = 0$. Similarly, consider the set $B = \{h > 1\}$ and the indicator function $f = \mathbf{1}_B$. If $\mu(B) > 0$ then (2.4) becomes $\mu(B) = \int_B h \, d(\mu+\nu) > (\mu+\nu)(B) \geqslant \mu(B)$, a contradiction. This proves (2.6).

Since $\nu \ll \mu$, we moreover have

$$0 < h \leqslant 1 \quad (\mu+\nu)\text{-a.e.} \tag{2.7}$$

By Monotone Convergence Theorem, one can show that (2.5) holds for arbitrary $(\mu+\nu)$-measurable functions f such that $f \geqslant 0$ $(\mu+\nu)$-a.e. (Indeed, consider the truncation $f_n(t) := \min(f(t), n)$ and let $n \to \infty$.) The convention is that if one side of (2.5) is infinite then the other is infinite, too.

Now, given a measurable set A, we choose f so that $fh = \mathbf{1}_A$. In other words, we consider

$$f := \frac{\mathbf{1}_A}{h}$$

and apply the identity in (2.5). We obtain

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

The proof is complete with $g := (1-h)/h$. □

Lec.15: 10/11

2.2.3. The dual of L_p. A version of the representation theorem for L_2, Corollary 2.2.3, holds in fact for all L_p spaces. In short, it states that $L_p^* = L_{p'}$ where p and p' are conjugate exponents as in Hölder's inequality, i.e.

$$\frac{1}{p} + \frac{1}{p'} = 1, \quad 1 < p, p' < \infty,$$

and $p' = \infty$ if $p = 1$. The rigorous statement is the following:

THEOREM 2.2.5 ($L_p^* = L_{p'}$). *Consider the space $L_p = L_p(\Omega, \Sigma, \mu)$ with finite of σ-finite measure μ, and where $1 \leqslant p < \infty$. Let p' be the conjugate exponent of p.*

(i) For every weight function $g \in L_{p'}$, integration with weight

$$G(f) := \int fg \, d\mu, \quad f \in L_p$$

is a bounded linear functional on L_p, and its norm is $\|G\| = \|g\|_{p'}$.

(ii) Conversely, every bounded linear functional $G \in L_p^*$ can be represented as integration with weight for some unique weight function $g \in L_{p'}$. Moreover, $\|G\| = \|g\|_{p'}$.

We shall prove Theorem 2.2.5 for the particular case of spaces ℓ_p. The general case is somewhat similar, and it relies on an application of Radon-Nykodim theorem. For the spaces ℓ_p, Theorem 2.2.5 reads as follows:

COROLLARY 2.2.6 ($\ell_p = \ell_{p'}$). *Let $1 \leqslant p < \infty$ and let p' be the conjugate exponent of p.*

(i) For every $y \in \ell_{p'}$, summation with weight

$$G(x) := \sum_{k=1}^{\infty} x_k y_k, \quad x \in \ell_p$$

is a bounded linear functional on ℓ_p, and its norm is $\|G\| = \|y\|_{p'}$.

(ii) Conversely, every bounded linear functional $G \in \ell_p^*$ can be represented as summation with weight for some unique weight $y \in \ell_{p'}$. Moreover, $\|G\| = \|y\|_{p'}$.

PROOF. We will only prove the case where $1 < p, p' < \infty$; the case $p = 1$, $p' = \infty$ is an exercise.

(i) By Hölder's inequality, we have

$$|G(x)| = \Big| \sum_k x_k y_k \Big| \leqslant \|x\|_p \|y\|_{p'}.$$

It follows that F is a bounded linear functional on ℓ_p, and $\|G\| \leqslant \|y\|_{p'}$. To prove the converse inequality, note that Hölder's inequality is sharp. Namely, for every $y \in \ell_{p'}$ there exists $x \in \ell_p$ such that

$$|G(x)| = \Big| \sum_k x_k y_k \Big| = \|x\|_p \|y\|_{p'}.$$

Indeed, one can check that this holds for $x = (x_k)$ defined as

$$x_k = e^{-i \operatorname{Arg}(y_k)} |y_k|^{p'-1}.$$

For this x, it follows that $\|G\| \geqslant \|y\|_{p'}$, so part (i) of the theorem is proved.

[1]The result can be extended (by decomposition) to σ-finite measures.

(ii) Consider $G \in \ell_p^*$. Let e_k denote as usual the coordinate vectors in ℓ_p, that is $e_k = (0, \dots, 0, 1, 0, \dots)$ with k-th coordinate equal 1. By the linearity and continuity of G we have

$$G(x) = G\Big(\sum_{k=1}^{\infty} x_k e_k \Big) = \sum_{k=1}^{\infty} x_k G(e_k), \quad x = (x_k) \in \ell_p.$$

We claim that the conclusion of (ii) follows for

$$y = (y_k) := (G(e_k)).$$

Obviously, $G(x) = \sum_k x_k y_k$ as required. To prove that $y \in \ell_{p'}$, consider

$$y^{(n)} := \sum_{k=1}^{n} y_k e_k \quad \text{and} \quad x^{(n)} := \sum_{k=1}^{n} e^{-i \operatorname{Arg}(y_k)} |y_k|^{p'-1} e_k, \quad n = 1, 2, \dots$$

Then using the equality case of Hölder's inequality, we see that

$$\|G\| \|x^{(n)}\|_p \geqslant |G(x^{(n)})| = \Big| \sum_{k=1}^{n} x_k y_k \Big| = \|x^{(n)}\|_p \, \|y^{(n)}\|_{p'}.$$

Canceling $\|x^{(n)}\|_p$ on both sides, we conclude that

$$\|y^{(n)}\|_{p'} = \Big(\sum_{k=1}^{n} |y_k|^{p'} \Big)^{1/p'} \leqslant \|G\|.$$

Letting $n \to \infty$ we conclude that $y \in \ell_{p'}$ and that $\|y\|_{p'} \leqslant \|G\|$. By part (i), we actually have $\|y\|_{p'} = \|G\|$. This completes the proof. □

Exercise 2.2.7. [$c_0^* = \ell_1$] **Prove that $c_0^* = \ell_1$. The meaning of this is the same as in Corollary 2.2.6, i.e. the functionals on c_0 are given by summation with weight from ℓ_1.**

2.2.4. The dual of $C(K)$. Finally, we state without proof the following characterization of bounded linear functionals on $C(K)$.

Theorem 2.2.8 ($C(K)^*$). *Consider the space $C(K)$ where K is a compact topological space.*

(i) For every Borel regular signed measure[2] μ on K, integration

$$G(f) := \int f \, d\mu, \quad f \in C(K)$$

is a bounded linear functional on $C(K)$, and its norm is the total variation $\|G\| = |\mu|(K)$.

(ii) Conversely, every bounded linear functional $G \in C(K)^*$ can be represented as integration with respect to a unique Borel regular signed measure μ on K. Moreover, $\|G\| = |\mu|(K)$.

---

[2]A signed measure μ is an extension of the notion of measure by allowing it to take on negative values. By Hahn decomposition theorem, a signed measure μ can be uniquely represented as $\mu = \mu_+ - \mu_-$ where μ_+ and μ_- are measures. The measure $|\mu| = \mu_+ + \mu_-$ is called the *variation* of μ, and the value $|\mu|(K)$ is called the *total variation* of μ.

2.2.5. Additional Exercises.

EXERCISE 2.2.9. **[Span of point evaluation functionals] Compute the norm of the following linear functional on $C[0,1]$:**

$$F(f) = \sum_{k=1}^{n} a_k f(t_k)$$

where a_k are fixed scalars, and $t_1, \ldots, t_n$ are fixed distinct points in $[0,1]$.

Deduce that the linear span of point evaluation functionals $\delta_{t_1}, \ldots, \delta_{t_n}$ in $C[0,1]^*$ is isometric to ℓ_1^n.

2.3. Hahn-Banach theorem

Lec.16: 10/13

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.

Let X be a normed space, and let X_0 be a subspace of X. Consider a bounded linear functional f_0 defined on X_0, i.e. $f_0 \in X_0^*$. An *extension* of f_0 to the whole space X is a bounded linear functional $f \in X^*$ whose restriction on X_0 coincides with f_0, i.e.

$$f|_{X_0} = f_0|_{X_0}, \quad \text{meaning that } f_0(x_0) = f(x) \text{ for all } x_0 \in X_0.$$

Constructing extensions is a nontrivial problem because of the continuity (=boundedness) requirement for f.

EXERCISE 2.3.1. **Show that if one does not require continuity of f, one can construct f using a Hamel basis.**

2.3.1. Extension by continuity.

Before we state Hahn-Banach theorem, let us address the simpler problem of extending a continuous linear functional from a *dense* subspace to the whole space.

PROPOSITION 2.3.2 (Extension by continuity). *Let X_0 be a dense subspace of a normed space X. Then every functional $f_0 \in X_0^*$ admits a unique extension $f \in X^*$. Moreover, $\|f\| = \|f_0\|$.*

PROOF. Let $x \in X$ be arbitrary. By density, we can find a sequence $(x_n) \subset X_0$ such that

$$x_n \to x.$$

Then $(f_0(x_n))$ is a Cauchy sequence, since

$$|f_0(x_n) - f_0(x_m)| \leqslant \|f_0\| \, \|x_n - x_m\| \to 0 \quad \text{as } n, m \to \infty.$$

By completeness of $\mathbb{R}$ or $\mathbb{C}$, the sequence converges. So we define

$$f(x) := \lim_n f_0(x_n)$$

and claim that f satisfies the conclusion of the proposition.

First of all, f is well defined, i.e. it depends only on x and not on the choice of an approximating sequence (x_n). Indeed, if $x_n' \to x$ then

$$|f_0(x_n) - f_0(x_n')| \leqslant \|f_0\| \, \|x_n - x_n'\| \to 0$$

since the limits of x_n and x'_n coincide.

Next, f is a linear functional on X, since for $x_n \to x$ and $y_n \to y$ we have

$$f(ax+by) = \lim_n f_0(ax_n + by_n) = a \lim_n f_0(x_n) + b \lim_n f_0(y_n) = af(x) + bf(y).$$

Finally, f is a bounded linear functional. Indeed, for $x_n \to x$ we have

$$|f(x)| = \lim_n |f_0(x_n)| \leqslant \|f_0\| \lim_n \|x_n\| = \|f_0\| \, \|x\|.$$

This shows that $f \in X^*$ and $\|f\| \leqslant \|f_0\|$. Note that the reverse inequality $\|f\| \geqslant \|f_0\|$ trivially holds (for any extension f). This completes the proof. $\square$

Example 2.3.3 (Lebesgue integral as an extension of Riemann integral). Let X_0 be the space of continuous functions $C[0,1]$ equipped with the norm

$$\|f\| = \int_0^1 |f(t)| \, dt \quad \text{(Riemann integral)}.$$

The Riemann integration

$$F_0(f) = \int_0^1 f(t) \, dt$$

is clearly a bounded linear functional on X_0. Let X denote the completion of $C[0,1]$. One can define Lebesgue integral as the extension $F \in X^*$ of F_0, and one can define $L_1[0,1] := X$.

Exercise 2.3.4. **Check that this indeed gives an equivalent definition of Lebesgue integral and of the space $L_1[0,1]$.**

2.3.2. Hahn-Banach theorem. Now we address the more difficult problem of extending linear funcitonals from *arbitrary* subspaces. The result is the same as for dense subspaces, except the extensions need not be unique.

Theorem 2.3.5 (Hahn-Banach theorem). *Let X_0 be a subspace of a normed space X. Then every functional $f_0 \in X_0^*$ admits an extension $f \in X^*$ such that $\|f\| = \|f_0\|$.*

Proof. We will prove this result for spaces over $\mathbb{R}$ only.

Step 1. Extension by one dimension. Let us first assume that X_0 is a hyperplane in X, i.e. $\operatorname{codim} X_0 = 1$. Fix a vector $z \in X \backslash X_0$. Every vector $x \in X$ is then uniquely represented as

$$x = az + x_0, \quad a \in \mathbb{R}, \, x_0 \in X_0.$$

Then an extension f has the form

$$f(x) = af(z) + f(x_0) = af(z) + f_0(x_0).$$

So f is determined by just one number

$$f(z) =: C.$$

Without loss of generality we can assume that $\|f_0\| = 1$ (by rescaling). We are looking for an extension f such that $\|f\| = 1$, which means that

$$|f(x)| \leqslant \|x\|, \quad x \in X.$$

This requirement is equivalent to

$$f(x) \leqslant \|x\|, \quad x \in X.$$

(Why?) Expressing x and f as above, we write this equivalently as

$$aC + f_0(x_0) \leqslant \|az + x_0\|, \quad a \in \mathbb{R},\ x_0 \in X_0. \tag{2.8}$$

The desired extension f will be constructed if we are able to show that this inequality has solutions in C.

For $a = 0$, inequality (2.8) trivially holds by the assumption that $\|f_0\| = 1$. For $a > 0$, inequality (2.8) can be written as

$$C \leqslant \left\| z + \frac{x_0}{a} \right\| - f_0\Big(\frac{x_0}{a}\Big), \quad a \in \mathbb{R}_+,\ x_0 \in X_0.$$

For $a = -b < 0$, inequality (2.8) can be written as

$$C \geqslant f_0\Big(\frac{x_1}{b}\Big) - \left\| \frac{x_1}{b} - z \right\|, \quad b \in \mathbb{R}_+,\ x_1 \in X_1.$$

Existence of a number $C \in \mathbb{R}$ that satisfies both these inequalities is equivalent to the inequality

$$\sup_{b \in \mathbb{R}_+,\, x_1 \in X_1} \Big(f_0\Big(\frac{x_1}{b}\Big) - \left\| \frac{x_1}{b} - z \right\| \Big) \leqslant \inf_{a \in \mathbb{R}_+,\, x_0 \in X_0} \Big(\left\| z + \frac{x_0}{a} \right\| - f_0\Big(\frac{x_0}{a}\Big) \Big).$$

Collecting the similar terms we see that this in turn is equivalent to the inequality

$$f_0\Big(\frac{x_0}{a} + \frac{x_1}{b}\Big) \leqslant \left\| z + \frac{x_0}{a} \right\| + \left\| \frac{x_1}{b} - z \right\|, \quad a, b \in \mathbb{R}_+;\ x_0, x_1 \in X_0.$$

But this inequality is true; it follows from the fact that $\|f_0\| = 1$, i.e.

$$f_0\Big(\frac{x_0}{a} + \frac{x_1}{b}\Big) \leqslant \left\| \frac{x_0}{a} + \frac{x_1}{b} \right\|,$$

and from the triangle inequality.

Step 2. Transfinite induction. This standard argument relies on Zorn's Lemma 1.1.7 in a similar way to the proof of Proposition 1.1.6 on the existence of Hamel bases. We consider the set Γ of all extensions of f_0, precisely

$$\Gamma = \big\{ (Y, g) :\ X_0 \subseteq Y \subseteq X \text{ a subspace},\ g \in Y^* \text{ an extension of } f_0,\ \|g\| = \|f_0\| \big\}.$$

The proof will be complete if we show that Γ contains an element (Y, g) with $Y = X$. We introduce a partial order on Γ by

$$(Y_1, g_1) \leqslant (Y_2, g_2) \quad \text{if} \quad Y_1 \subseteq Y_2,\ g_2 \text{ is an extension of } g_1.$$

Every chain $\big((Y_\alpha, g_\alpha)\big)_\alpha$ has an upper bound (Y, g) in Γ, namely

$$Y := \bigcup_\alpha Y_\alpha, \quad g(x) := g_\alpha(x) \text{ if } x \in Y_\alpha.$$

(Why is g a well defined bounded linear functional on Y?) Therefore, by Zorn's Lemma 1.1.7, there exists a maximal element (Y, g) in Γ.

We claim that $Y = X$. Indeed, otherwise by Step 1 we could extend g to a subspace larger than Y, which would contradict the maximality of (Y, g) in Γ. The proof is complete. □

EXERCISE 2.3.6. **Deduce Hahn-Banach theorem for spaces X over $\mathbb{C}$.**

Lec.17: 10/15

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

PROPOSITION 2.3.7 (Supporting functional). *Let X be a normed space. For every $x \in X$ there exists $f \in X^*$ such that*

$$\|f\| = 1, \quad f(x) = \|x\|.$$

The functional f is called the supporting functional *of x.*

PROOF. Consider the one-dimensional subspace $X_0 = \mathrm{Span}(x)$, and define a functional $f_0 \in X_0^*$ by

$$f_0(tx) = t\|x\|, \quad t \in \mathbb{R} \text{ or } \mathbb{C}.$$

Then $\|f_0\| = 1$. An extension $f \in X^*$ of f_0 guaranteed by Hahn-Banach theorem clearly satisfies the conclusion. □

EXERCISE 2.3.8. **Consider a normed space $X = (\mathbb{R}^2, \|\cdot\|)$ and a unit vector $x_0 \in X$. Let f be a supporting functional of x_0. Interpret geometrically the level set $\{x : f(x) = 1\}$ as a tangent hyperplane for the unit ball B_X at point x_0.**

Construct an example of a normed space for which the supporting functional of x is not unique.

Recall that the norm of a functional $f \in X^*$ is defined as

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

Generally it is not true that every functional f attains its norm on some vector x, i.e. that the supremum above can be replaced by the maximum.

EXERCISE 2.3.9. **Construct a bounded linear functional on $C[0,1]$ which does not attain its norm.**

However, every vector x does attain its norm on some functional $f \in X^*$, namely the supporting functional. This immediately follows from Proposition 2.3.7:

COROLLARY 2.3.10. *For every vector x in a normed space X, one has*

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

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

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

COROLLARY 2.3.11 (X^* separates the points of X). *For every two vectors $x_1 \neq x_2$ in a normed space X, there exists a functional $f \in X^*$ such that $f(x_1) \neq f(x_2)$.*

PROOF. The supporting functional $f \in X^*$ of the vector $x = x_1 - x_2$ must satisfy $f(x_1 - x_2) = \|x_1 - x_2\| \neq 0$, as required. □

2.3.4. Second dual space. Let X be a normed space as usual. The functionals f^* are designed to act on vectors $x \in X$ via $f : x \mapsto f(x)$. Vice versa, we can say that *vectors $x \in X$ act on functionals $f \in X^*$* via

$$x : f \mapsto f(x), \quad f \in X^*. \tag{2.9}$$

Thus a vector $x \in X$ can itself be considered as a function from X^* to $\mathbb{R}$.

This function (2.9) is clearly linear, so we may consider x as a linear functional on X^*. Also, the inequality

$$|f(x)| \leqslant \|x\| \, \|f\|$$

shows that this functional is bounded, so

$$x \in X^{**}$$

and the norm of x as a functional is $\|x\|_{X^{**}} \leqslant \|x\|$. Considering the supporting functional $f \in X^*$ of x we see that actually

$$\|x\|_{X^{**}} = \|x\|.$$

We demonstrated that there exists a *canonical embedding* of X into into X^{**}. We summarize this as follows.

Theorem 2.3.12 (Second dual space). *Let X be a normed space. Then X can be considered as a linear subspace of X^{**}. For this, a vector $x \in X$ is considered as a bounded linear functional on X^* via the action*

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

Example 2.3.13. As we know from Section 2.2.3, $c_0^* = \ell_1$ while $\ell_1^* = \ell_\infty$, so

$$c_0^{**} = \ell_\infty.$$

The space c_0 of sequences converging to zero is indeed canonically embedded into the larger space ℓ_∞ of all bounded sequences (and with the same sup-norm).

Definition 2.3.14 (Reflexive spaces). A normed space X is called *reflexive* if $X^{**} = X$ (under the canonical embedding).

Example 2.3.15. As we know from Section 2.2.3, $(L_p)^* = L_{p'}$ where $1 \leqslant p < \infty$ and p' is the conjugate index of p. Therefore, L_p is a reflexive space for $1 < p < \infty$.

One can show that L_1 and L_∞ are not reflexive spaces.

Proposition 2.3.16. *Let X be a reflexive space. Then every functional $f \in X^*$ attains its norm on X.*

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

The converse of Proposition 2.3.16 is also true. If every functional $f \in X^*$ on a Banach space X attains its norm, then X is reflexive. This is James' theorem (1971).

2.3.5. Hahn-Banach theorem for sublinear functions. A quick inspection of the proof of Hahn-Banach theorem in Section 2.3.2 reveals that we have not used all the norm axioms there. We used just these two – positive homogeneity and triangle inequality, precisely

(i) $\|ax\| = a\|x\|$ for all $x \in X$, $a \geqslant 0$;
(ii) $\|x + y\| \leqslant \|x\| + \|y\|$ for all $x, y \in X$.

Definition 2.3.17 (Sublinear function). Let X be a linear space. Functions $\|\cdot\| : X \to [0, \infty)$ that satisfy (i) and (ii) above are called *sublinear functions*.

It is an exercise to check that our proof of Hahn-Banach theorem is valid not only for norms but for sublinear functions:

Theorem 2.3.18 (Hahn-Banach theorem for sublinear functions). *Let X_0 be a subspace of a linear vector space X. Let $\|\cdot\|$ be a sublinear function on X. Consider a linear functional f on X such that*

$$f_0(x) \leqslant \|x\| \quad \text{for all } x \in X_0.$$

Then f_0 admits an extension f to the whole space X such that

$$f(x) \leqslant \|x\| \quad \text{for all } x \in X.$$

Exercise 2.3.19. **Prove Theorem 2.3.18 by modifying the proof of Hahn-Banach Theorem.**

Lec.18: 10/20

Considering sublinear functionals instead of norms offers us more flexibility in geometric applications. Sublinear functionals arise as Minkowski functionals of convex sets.

Definition 2.3.20. A subset K of a linear vector space X is called *absorbing* if

$$X = \bigcup_{t \geqslant 0} tK$$

where $tK = \{tx : x \in K\}$.

Exercise 2.3.21. **Let K be a subset of a normed space X such that $0 \in \mathring{K}$. Then K is an absorbing set.**

Proposition 2.3.22 (Minkowski functional). *Let K be a absorbing convex subset of a linear vector space X such that $0 \in K$. Then Minkowski functional*

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

is a sublinear functional on X.

Conversely, let $\|\cdot\|$ be a sublinear functional on a linear vector space X. Then the sub-level set

$$K = \{x \in X : \|x\| \leqslant 1\}$$

is an absorbing convex set, and $0 \in K$.

Exercise 2.3.23. **Prove Proposition 2.3.22.**

Proposition 2.3.22 should be compared to a similar result for norms that we proved earlier in Exercise 1.2.22.

2.3.6. Separation of convex sets. Hahn-Banach theorem has some remarkable geometric implications, which are grouped together under the name of *separation theorems*. Under some mild topological requirements, these results guarantee that two convex sets A, B can always be separated by a hyperplane. As we know from Section 2.1.3, the hyperplanes correspond to the level sets of linear functionals f. Therefore, we expect that a separation theorem for A, B would give us a linear functional f and a number C such that

$$f(a) \leqslant C \leqslant f(b), \quad a \in A,\, b \in B.$$

In this case, the sets A and B get separated by the hyperplane $\{x : f(x) = C\}$.

Let us start from the simpler case when one of the two sets is a point.

THEOREM 2.3.24 (Separating a point from a convex set). *Let K be an open convex subset of a normed space X, and consider a point $x_0 \notin K$. Then there exists a functional $f \in X^*$, $f \neq 0$, such that*

$$f(x) \leqslant f(x_0), \quad x \in K.$$

PROOF. Translating K if necessary, we can assume without loss of generality that $0 \in K$. (Why?) By Exercise 2.3.21, K is an absorbing set. Therefore, by Proposition 2.3.22, Minkowski functional $\|\cdot\|_K$ of K is a sublinear functional on X.

Since $0 \in K$ and K is open, it contains a centered ball $B_X(0, r)$ for some radius $r > 0$ (see the figure). The set inclusion $B_X(0, r) \subseteq K$ implies the inequality for the Minkowski functionals:

$$\frac{1}{r}\|x\| \geqslant \|x\|_K, \quad x \in K.$$

(Why?)

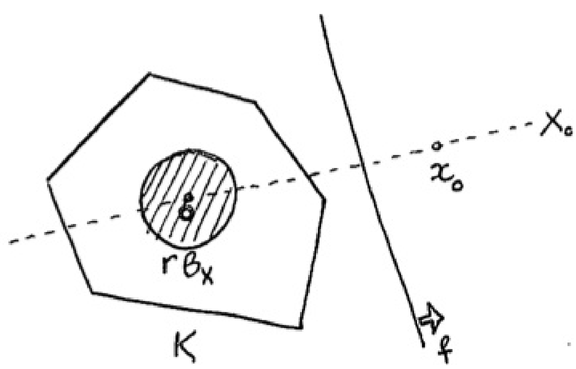

FIGURE 2.2. Separation of a point x_0 from the set K by a functional f

Consider the one-dimensional subspace

$$X_0 = \mathrm{Span}(x_0)$$

and define a linear functional f_0 on X_0^* by

$$f_0(tx) = t\|x\|, \quad t \in \mathbb{R}.$$

Then f_0 is dominated by $\|\cdot\|_K$ on X_0, since for $t \geqslant 0$ we have

$$f_0(tx_0) = \|tx_0\|_K;$$
$$f_0(-tx_0) = -tf(x_0) = -t\|x_0\|_K \leqslant 0 \leqslant \|tx_0\|_K.$$

By Hahn-Banach theorem for sublinear functionals (Theorem 2.3.18), f_0 admits an extension f onto the whole space X such that the domination is preserved, i.e.

$$f(x) \leqslant \|x\|_K, \quad x \in X.$$

To finish the proof, we need to check that f is bounded and that it separates x_0 from K as required. The boundedness follows from the inequality

$$f(x) \leqslant \|x\|_K \leqslant \frac{1}{r}\|x\|, \quad x \in X,$$

so $f \in X^*$. To check the separation, consider $x \in K$. Since $x_0 \notin K$, we have

$$f(x) \leqslant \|x\|_K \leqslant 1 \leqslant \|x_0\| = f_0(x_0) = f(x_0).$$

This completes the proof. □

Theorem 2.3.25 (Separation of open convex sets). *Let A, B be disjoint open convex subsets of a normed space X.*

(i) Assume that A is open. Then there exists a functional $f \in X^*$ and a number $C \in \mathbb{R}$ such that

$$f(a) \leqslant C \leqslant f(b), \quad a \in A,\ b \in B.$$

(ii) If both A and B are open, then the stronger inequality holds:

$$f(a) < C < f(b), \quad a \in A,\ b \in B.$$

Proof. (i) Consider the *Minkowski difference* set

$$K = A - B := \{a - b :\ a \in A,\ b \in B\}.$$

The set K is open and convex. (Check!) Since A and B are disjoint, $0 \in K$.

Using Theorem 2.3.24, we obtain a functional $f \in X^*$, $f \neq 0$ such that

$$f(a - b) \leqslant f(0) = 0, \quad a \in A,\ b \in B.$$

Hence $f(a) \leqslant f(b)$ for all $a \in A$, $b \in B$, so letting $C := \sup_{a \in A} f(a)$ we obtain

$$f(a) \leqslant C \leqslant f(b), \quad a \in A,\ b \in B.$$

Since A is open, by considering a small neighborhood of a in A (check!) we obtain the strict inequality

$$f(a) < C \leqslant f(b), \quad a \in A,\ b \in B.$$

(ii) This part is similar, and follows by considering small neighborhoods of a in A and of b in B. □

Exercise 2.3.26. **Fill in the details in the proof of Theorem 2.3.25.**

Corollary 2.3.27 (Separation of closed convex sets). *Let A, B be disjoint closed convex subsets of a normed space X. Assume B is compact. Then there exists a functional $f \in X^*$ and a number $C \in \mathbb{R}$ such that*

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

Proof. Let
$$r = \operatorname{dist}(A, B) := \inf_{a \in A,\, b \in B} \|a - b\|.$$
By the assumptions, $r > 0$. (Why?) Therefore, the open $r/3$-neighborhoods $A_{r/3}$ of A and $B_{r/3}$ of B are disjoint, open and convex sets. Applying Theorem 2.3.25, we obtain a functional $f \in X^*$ that separates the neighborhoods:
$$\sup_{a \in A_{r/3}} f(a) \leqslant \inf_{b \in B_{r/3}} f(b).$$
From this the conclusion easily follows. (How?) □

Exercise 2.3.28. **Fill in the details in the proof of Corollary 2.3.27.**

Remark 2.3.29. Suppose that in Theorem 2.3.24, the set K is either open (i.e. as stated) or closed. Then the strict separation holds:
$$f(x) < f(x_0), \quad x \in K.$$
Indeed, for open sets this follows from Theorem 2.3.25, while for closed sets this follows from Corollary 2.3.27.

2.3.7. Convex sets are intersections of half-spaces.

Corollary 2.3.30. *Every closed convex subset K of a normed space X is the intersection of all (closed) half-spaces that contain K.*

Recall that the half-space is what lies on one side of a hyperplane; therefore half-spaces have the form
$$\{x \in X : \ f(x) \leqslant a\}$$
for some $f \in X^*$, $a \in \mathbb{R}$. See the picture illustrating Corollary 2.3.30.

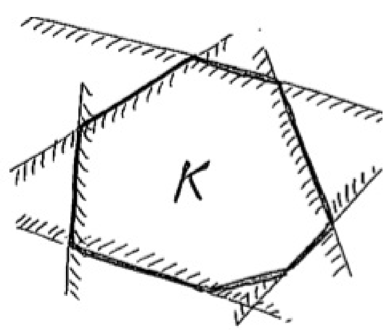

Figure 2.3. Convex set K is the intersection of half-spaces

Proof. K is trivially contained in the intersection of the half-spaces that contain K. To prove the reverse inclusion, choose a point $x_0 \notin K$ and use Separation Theorem 2.3.27 for $A = K$ and $B = \{x_0\}$. We thus obtain a functional $f \in X^*$ such that
$$a := \sup_{x \in K} f(x) < f(x_0).$$
It follows that the half-space $\{x \in X : \ f(x) \leqslant a\}$ contains K but not x_0. This completes the proof. □

2.3.8. Additional exercises.

EXERCISE 2.3.31. **[Closed convex sets that can not be strictly separated] Show that the compactness assumption in Corollary 2.3.27 is essential. Construct two closed convex sets on the plane $\mathbb{R}^2$ that can not be strictly separated.**

EXERCISE 2.3.32. **[Convex sets that can not be separated] Show that the openness assumption in Theorem 2.3.25 is essential.**

To this end, consider the linear space $\mathcal{P}$ of all polynomials in one variable and with real coefficients. Let the subset A consist of polynomials with negative leading coefficient, and let the subset B consists of polynomials with all non-negative coefficients. Show that A and B are disjoint convex subsets of $\mathcal{P}$, and that there does not exist a nonzero linear functional f on $\mathcal{P}$ such that

$$f(a) \leqslant f(b) \quad \text{for all } a \in A, b \in B.$$

(Hint: assume that for some $C \in \mathbb{R}$ one has $f(a) \leqslant C \leqslant f(b)$, $a \in A$, $b \in B$; deduce from $0 \in B$ that $C \leqslant 0$ and by considering monomials that $C \geqslant 0$.

EXERCISE 2.3.33. **[Functionals that annihilate a subspace] Let X_0 be a closed subspace of a normed space X. Prove that there exists a functional $f \in X^*$ such that**

$$f(x) = 0 \quad \text{for all } x \in X_0.$$

You may deduce this from Hahn-Banach theorem directly or from a separation theorem.

2.4. Bounded linear operators

Lec.19: 10/22

In this section, we shall study continuous linear operators $T : X \to Y$ between normed spaces X, Y. Linear functionals can be seen as a particular case of linear operators with Y being the scalar field, $\mathbb{R}$ or $\mathbb{C}$. So several results for linear operators will be generalizations of those we have already seen for linear functionals; there will be important differences though (e.g. a natural extension Hahn-Banach theorem fails for linear operators).

2.4.1. Operator norm. Continuity and boundedness. The notions of boundedness, continuity and norm of linear operators are similar to those for linear functionals given in Section 2.1.2:

DEFINITION 2.4.1 (Operator norm). A linear operator $T : X \to Y$ acting between normed spaces X and Y is called *bounded* if there exists a number C such that

$$\|Tx\| \leqslant C\|x\| \quad \text{for all } x \in X.$$

The *norm* of T is the smallest constant C in this inequality, so it is defined as

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\|.$$

REMARK 2.4.2. We always have the inequality

$$\|Tx\| \leqslant \|T\| \, \|x\|, \quad x \in X.$$

EXERCISE 2.4.3. **[Composition of operators] If $T : X \to Y$ and $S : Y \to Z$ are bounded linear operators between normed spaces, then $ST : X \to Z$ is a bounded linear operator, and**

$$\|ST\| \leqslant \|T\| \, \|S\|.$$

A version of Proposition 2.1.7 for linear functionals holds for linear operators, and with a similar proof:

PROPOSITION 2.4.4. *Continuity and boundedness of linear operators are equivalent.*

EXERCISE 2.4.5. **Prove that a linear operator $T : X \to Y$ is bounded if and only if it maps sequences that converge to zero to bounded sequences.**

2.4.2. Space of operators. Let X and Y be normed spaces. The space of all bounded linear operators $T : X \to Y$ equipped with the operator norm is denoted $L(X, Y)$.

As an example, the dual space is a space of operators that map to scalars, i.e.

$$X^* = L(X, \mathbb{R}).$$

PROPOSITION 2.4.6. *$L(X, Y)$ is a normed space. Moreover, if Y is a Banach space then $L(X, Y)$ is also a Banach space. In particular, the dual space X^* is always a Banach space, even if X is incomplete.*

PROOF. The norm axioms are straightforward. (Check!) To prove completeness, let $T_n \in L(X, Y)$ be a Cauchy sequence, i.e.

$$\|T_n - T_m\| \to 0, \quad n, m \to \infty.$$

Applying this to an arbitrary $x \in X$ and noting that

$$\|T_n x - T_n x\| \leqslant \|T_n - T_m\| \, \|x\| \to 0, \quad n, m \to \infty$$

we see that $(T_n x)$ is a Cauchy sequence in Y. By the completeness of Y it converges. Define the map T as

$$Tx := \lim_n T_n x.$$

We claim that T is the limit of T_n in $L(X, Y)$, which would complete the proof.

It is easy to check that $T : X \to Y$ is a linear operator. (Check!)

To show that T is bounded, we choose an arbitrary $x \in X$ and use continuity of the norm:

$$\|Tx\| = \lim_n \|T_n x\| \leqslant \sup_n \|T_n\| \|x\|.$$

Since a Cauchy sequence is always bounded (why?), $\sup_n \|T_n\| < \infty$. It follows that T is a bounded linear operator, i.e. $T \in L(X, Y)$.

It remains to show that $T_n \to T$ in $L(X, Y)$, i.e. in the operator norm. Since T is Cauchy, for every $\varepsilon > 0$ there exists a $N \in \mathbb{N}$ such that

$$\|T_n - T_m\| < \varepsilon \quad \text{for } n, m > N.$$

Applying this to an arbitrary $x \in X$ we obtain

$$\|T_n x - T_m x\| < \varepsilon \|x\| \quad \text{for } n, m > N.$$

Letting $m \to \infty$, we get

$$\|T_n x - Tx\| < \varepsilon \|x\| \quad \text{for } n > N.$$

Since x is arbitrary it follows that

$$\|T_n - T\| < \varepsilon \quad \text{for } n > N.$$

This means that $T_n \to T$ in $L(X, Y)$ as required. □

2.4.3. Operators on the finite-dimensional Euclidean space. Let $X = Y = \ell_2^n$; recall that ℓ_2^n is the Euclidean space $(\mathbb{R}^n, \|\cdot\|_2)$. As we know from linear algebra, T can be identified with its $n \times n$ matrix (t_{ij}), where

$$t_{ij} = \langle Te_j, e_i \rangle, \quad i, j = 1, \dots, n$$

and where (e_i) denotes the canonical basis of ℓ_2^n. This way, the i-th coordinate of the vector Tx can be computed as

$$(Tx)_i = \sum_{j=1}^{n} t_{ij} x_j. \tag{2.10}$$

Proposition 2.4.7. *Every linear operator $T : \ell_2^n \to \ell_2^n$ is bounded. Specifically,*

$$\|T\| \leqslant \|T\|_{\mathrm{HS}}$$

where $\|T\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm *of T defined as*

$$\|T\|_{\mathrm{HS}} = \Big(\sum_{i,j=1}^{n} |t_{ij}|^2 \Big)^{1/2}.$$

Recall that we already encountered Hilbert-Schmidt (or Frobenius) norm of matrices in (1.12).

Proof. Using (2.10) and applying Cauchy-Schwarz inequality, we obtain

$$\|Tx\|_2^2 = \sum_{i=1}^{n} |(Tx)_i|^2 = \sum_{i=1}^{n} \Big| \sum_{j=1}^{n} t_{ij} x_j \Big|^2 \leqslant \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} |t_{ij}|^2 \Big) \Big(\sum_{j=1}^{n} |x_j|^2 \Big) = \|T\|_{\mathrm{HS}}^2 \, \|x\|_2^2.$$

The claim follows. □

2.4.4. Hilbert-Schmidt integral operators. A similar construction in function spaces L_2 leads to the notion of Hilbert-Schmidt integral operators. To this end, consider a function $k(t, s) \in L_2([0, 1]^2)$ which we call the kernel. Define a linear operator $T : L_2[0, 1] \to L_2[0, 1]$ as

$$(Tf)(t) = \int_0^1 k(t, s) f(s)\, ds. \tag{2.11}$$

We can view this definition as a continuous version of (2.10), where kernel $k(t, s)$ can be considered as a continuous version of matrix. The operator T defined this way is called *Hilbert-Schmidt integral operator* with kernel $k(t, s)$.

Proposition 2.4.8. *A Hilbert-Schmidt integral operator $T : L_2[0, 1] \to L_2[0, 1]$ with kernel $k(t, s) \in L_2([0, 1]^2)$ is bounded. Specifically,*

$$\|T\| \leqslant \|k\|_2.$$

PROOF. Our argument is a continuous version of the proof of Proposition 2.4.7. Applying Cauchy-Schwarz and Fubini inequalities, we obtain

$$\begin{aligned}\|Tf\|_2^2 &= \int_0^1 dt \Big| \int_0^1 k(t,s)f(s)\,ds \Big|^2 \leqslant \int_0^1 dt \Big(\int_0^1 |k(t,s)|^2\,ds \Big) \Big(\int_0^1 |f(s)|^2\,ds \Big) \\ &= \|k\|_2^2\,\|f\|_2^2.\end{aligned}$$

The claim follows. $\square$

EXERCISE 2.4.9. **Why is $(Tf)(t)$ defined for almost all t?**

Lec.20: 10/25

REMARK 2.4.10 (Fredholm integral equations). Hilbert-Schmidt integral operators arise from the study of *Fredholm integral equations*. The so-called homogeneous Fredholm equation of the first kind is

$$g(t) = \int_0^1 k(t,s)f(s)\,ds.$$

Given a kernel $k(t,s)$ and the left hand side $g(t)$, the problem is to find the function $f(s)$.

Fredholm equations can be written as

$$Tf = g$$

where T is the corresponding Hilbert-Schmidt operator. Therefore, Fredholm equations are linear equations. They can be thought of as continuous versions of matrix linear equations $Ax = b$, where A is an $m \times n$ matrix, $b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$.

REMARK 2.4.11. The particular measure space $[0,1]$ does not play any role in the discussion above, and can be replaced with an arbitrary measure space.

2.4.5. Volterra operator and differential operators. Volterra operator $T : L_2[0,1] \to L_2[0,1]$ is defined as

$$(Tf)(t) = \int_0^t f(s)\,ds. \tag{2.12}$$

By Lebesgue differentiation theorem, Tf is the *antiderivative* of f.

Volterra operator is clearly a particular case of Hilbert-Schmidt integral operator with kernel $k(t,s) = \mathbf{1}_{\{s \leqslant t\}}$, and is therefore a bounded linear operator.

EXERCISE 2.4.12. **Compute the norm of Volterra operator.**

One would naturally think about the inverse of Volterra operator, which should be the *differential operator*

$$Df = f'.$$

Clearly D can only defined on an appropriate dense subspace of L_2, such as C^1, the space of continuously differentiable functions.

However, D is an unbounded linear operator on $L_2[-\pi,\pi]$.[3] To see this, consider the Fourier basis $(e_k)_{k\in\mathbb{Z}}$ of $L_2[-\pi,\pi]$, which we defined in Example 1.6.3 as

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

[3]The interval $[-\pi,\pi]$ is chosen only for convenience; a similar result holds for an arbitrary interval.

Differentiating yields

$$De_k = (ik)e_k, \quad k \in \mathbb{Z}.$$

Hence

$$\|De_k\|_2 = k\|e_k\|_2 = k, \quad k \in \mathbb{Z}.$$

Letting $k \to \infty$ implies that D is an unbounded operator.

This simple example suggests that various differential operators are unbounded even on good function spaces. They may be studied through their inverses, which are bounded integral operators.

2.4.6. Orthogonal projections and partial Fourier series. Let X_0 be a closed subspace of a Hilbert space X. Consider the orthogonal projection P in X onto X_0. As we know from Section 1.5.2, $P : X \to X$ is a linear map. Since Px and $x - Px$ are orthogonal vectors, we have

$$\|x\|^2 = \|Px\|^2 + \|x - Px\|^2, \quad x \in X,$$

so

$$\|Px\| \leqslant \|x\|, \quad x \in X.$$

We have shown:

PROPOSITION 2.4.13. *The orthogonal projection P in a Hilbert space X onto a closed subspace X_0 is a bounded linear operator.*

EXERCISE 2.4.14. **Show that $\|P\|$=1.**

As an example, consider the n-th partial sum of the Fourier series of a function $f \in L_2[-\pi, \pi]$:

$$S_n f = \sum_{k=-n}^{n} \langle f, e_k \rangle e_k$$

where e_k denotes the Fourier basis of exponentials (1.16). As we know from Lemma 1.6.10, S_n is an orthogonal projection onto $\mathrm{Span}(e_k)_{k=-n}^{n}$. This subspace is finite-dimensional and thus closed. (Why?) So by Proposition 2.4.13, S_n is a bounded linear operator in $L_2[-\pi, \pi]$.

A convenient and classical way to represent S_n is via convolution with Dirichlet kernel. Indeed,

$$(S_n f)(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(t-s) f(s)\, ds = \frac{1}{2\pi}(D_n * f)(t) \tag{2.13}$$

where

$$D_n(\theta) = \sum_{k=-n}^{n} e^{ik\theta} = \frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta} \tag{2.14}$$

is called the Dirichlet kernel.

EXERCISE 2.4.15. **Prove this identity.**

We see that S_n acts as a convolution with Dirichlet kernel D_n. In particular, S_n is a Hilbert-Schmidt operator with Dirichlet kernel $k(t, s) = \frac{1}{2\pi} D_n(t - s)$.

2.4.7. Isomorphisms, isometries. An important class of bounded linear operators is formed by isomorphisms. Recall that a transformation $T : X \to Y$ between topological spaces X and Y is called an *isomorphism* if T is invertible and both T, T^{-1} are continuous. Spaces X and Y are called *isomorphic* if there exists an isomorphism between them. So a linear map $T : X \to Y$ between normed spaces X and Y is an isomorphism if and only if T is invertible and $T \in L(X, Y)$, $T^{-1} \in L(Y, X)$.

EXERCISE 2.4.16. **Show that a surjective linear map $T : X \to Y$ between normed spaces is an isomorphism if and only if there exist $C, c > 0$ such that**

$$c\|x\| \leqslant \|Tx\| \leqslant C\|x\| \quad \text{for all } x \in X.$$

In other words, an isomorphism preserves all distances up to a multiplicative factor C/c. An *isometry* is an isomorphism T such that $\|T\| = \|T^{-1}\| = 1$, i.e.

$$\|Tx\| = \|x\| \quad \text{for all } x \in X.$$

Isometries preserve all distances exactly.

EXAMPLE 2.4.17. Isometries on a Hilbert space are commonly called *unitary operators*. As an example, in Section 1.6.7 we constructed an isometry T between any pair of infinite-dimensional separable Hilbert spaces X and Y.

EXAMPLE 2.4.18. The right shift $R(x_1, x_2, \ldots) = (0, x_1, x_2, \ldots)$ on any classical sequence space considered in Example 1.1.19 is not an isometry because R is not surjective, although $\|Rx\| = \|x\|$. The left shift $L(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$ is obviously not invertible but is bounded, and $\|L\| = 1$. (Why?)

EXERCISE 2.4.19. **[Isomorphisms preserve vital structures in normed spaces] Show that an isomorphism between normed spaces maps open sets to open sets, closed sets to closed sets, convegrent sequences to convergent sequences, complete spaces to complete spaces.**

Unfortunately, most infinite-dimensional Banach spaces are not isomorphic to each other. For example, A. Pelczynski proved that among the spaces c_0, $L_p[0, 1]$ and ℓ_q, $1 \leqslant p, q \leqslant \infty$, there are exactly two isomorphic ones, $L_2[0, 1]$ and ℓ_2, see [**8**].

However, as we know all separable Hilbert spaces are isomorphic (even isometric) to each other. We will also see in Section 3.1.5 that all finite-dimensional Banach spaces of same dimension are isomorphic (but not isometric to each other). The inverse mapping theorem of Section 3.1.2 below is a powerful tool to construct isomorphisms.

2.4.8. Extensions and projections. We will now address the extension problems for bounded linear operators. Similar to Section 2.3, we consider a normed space X and its subspace X_0 Consider an operator $T_0 \in L(X_0, Y)$ where Y is some normed space. An *extension* of T_0 to the whole space X is an operator $T \in L(X, Y)$ whose restriction on X_0 coincides with T_0, i.e.

$$T|_{X_0} = T_0|_{X_0}.$$

As we know from Section 2.3, every bounded linear functional can be extended from either dense or closed subspace to the whole space. For dense subspaces, we can extend by continuity, while for closed subspaces the extension is guaranteed by Hahn-Banach theorem.

For general linear operators, extension by continuity holds with the same proof as in Proposition 2.3.2:

Proposition 2.4.20 (Extension by continuity). *Let X_0 be a dense subspace of a normed space X, and Y be a Banach space. Then every operator $T_0 \in L(X_0, Y)$ admits a unique extension $T \in L(X, Y)$. Moreover, $\|T\| = \|T_0\|$.*

Unfortunately, extension from a closed subspace is not always possible, and Hahn-Banach theorem does not generalize to bounded linear operators. There is a simple geometric description of the situations when such extensions are possible. To state it we need a general notion of projections in normed space (not necessarily orthogonal).

Definition 2.4.21 (Projection). Let X_0 be a closed subspace of a normed space X. An operator $P \in L(X, X)$ is called a *projection* in X onto X_0 if

(i) $P(X) \subseteq X_0$;
(ii) $Px = x$ for all $x \in X_0$, i.e. $P|_{X_0} = I_{X_0}$.

Example 2.4.22. Any orthogonal projection in a Hilbert space is clearly a projection in this sense. However, even in a Hilbert space there is a plenty of non-orthogonal projections. (Construct one in a two-dimensional space.)

The following observation characterizes the subspaces from which extensions of linear operators are possible.

Proposition 2.4.23 (Extensions of operators and projections). *Let X_0 be a closed subspace of a normed space X. Then the following are equivalent:*

(i) There exists a projection in X onto X_0. In this case we say that X_0 is a complemented subspace *of X;*
(ii) For every normed space Y, every operator $T_0 \in L(X_0, Y)$ admits an extension $T \in L(X, Y)$.

Proof. Assume P is a projection in X onto X_0. Then for every operator $T_0 \in L(X_0, Y)$, the operator $T := T_0 P \in L(X_0, Y)$ is an extension.

Vice versa, consider the identity map $I : X_0 \to X_0$. Its extension $P : X \to X_0$ is clearly a projection in X onto X_0. □

Since every closed subspace in a Hilbert space is complemented, the extension problem in Hilbert spaces always has a positive solution.

Unfortunately, in general normed spaces there may be uncomplemented subspaces. Here is a synopsis of some of the known results, without proof:

Theorem 2.4.24 (Complemented subspaces). *In what follows, X_0 stands for a closed subspace of a Banach space X.*

(i) If either $\dim X_0 < \infty$ or $\operatorname{codim} X < \infty$ then X_0 is complemented in X.
(ii) c_0 is not complemented in ℓ_∞ [**10**].
(iii) Every complemented subspace X_0 of X is isomorphic to X if X is one of the spaces ℓ_p, $p \in [1, \infty]$ or c_0. These results are due to Pelczynski and Lindenstrauss [**9, 6**].
(iv) Every Banach space that is not isomorphic to a Hilbert space has an uncomplemented subspace. This is a theorem of Lindenstrauss and Tzafriri [**7**].

(v) *There exists a Banach space X without non-trivial complemented subspaces X_0 (i.e. such that* $\dim X_0 = \operatorname{codim} X_0 = \infty$*). This was proved in 1993 by Gowers and Maurey* [**5**].
(vi) *If X_0 is isomorphic to ℓ_∞ then it is complemented in X.*

Lec. 21: 10/27

2.4.9. Adjoint operators. The concept of adjoint operator 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 the $n \times n$ matrix $A^* = (\overline{a_{ji}})$. The transpose thus satisfies the identity

$$\langle A^* x, y\rangle = \langle x, Ay\rangle, \quad x, y \in \mathbb{C}^n. \tag{2.15}$$

Now we would like to extend this to a general definition of the adjoint T^* for a linear operator $T : X \to Y$ acting between normed spaces X and Y.

Definition 2.4.25 (Adjoint operator). Let $T \in L(X, Y)$. The adjoint operator $T^* \in L(Y^*, X^*)$ is defined by

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

In order to see a similarity with (2.15), we adopt the following alternative notation for the action of functionals, one that resembles the inner product:

$$\langle f, x\rangle := f(x), \quad f \in X^*, \quad x \in X. \tag{2.16}$$

Notice that if X is a Hilbert space, this notation agrees with the inner product by Riesz representation theorem (up to complex conjugation). In general, $\langle f, x\rangle$ does not define an inner product since the arguments are taken from different spaces. Then the definition of the adjoint reads as

$$\langle T^* f, x\rangle = \langle f, Tx\rangle, \quad f \in Y^*, \, x \in X$$

and we see that this is a general form of (2.15).

Remark 2.4.26 (Adjoint operators on Hilbert spaces). For operators T on a Hilbert space $X = Y = H$, the definition (2.4.25) of the adjoint operator T^* takes place with $\langle \cdot, \cdot\rangle$ denoting the inner product on H. This makes a small difference – the inner product is *conjugate* linear in the second argument, while the action of functionals (2.16) is just linear. So, for operators on a Hilbert space one has $(aT)^* = \overline{a} T^*$ for $a \in \mathbb{C}$, while the general definition of adjoint for Banach spaces incurs $(aT)^* = aT^*$.

One point has not been justified in Definition 2.4.25, why T^* is a bounded linear operator. We shall prove this now:

Proposition 2.4.27. *For every $T \in L(X, Y)$, we have $T^* \in L(Y^*, X^*)$, and*

$$\|T^*\| = \|T\|.$$

Proof. Denoting as usual S_X the unit sphere of X, and using notation (2.16), we have

$$\begin{aligned}\|T^*\| &= \sup_{f \in S_{Y^*}} \|T^* f\|_{X^*} = \sup_{f \in S_{Y^*}} \sup_{x \in S_X} |\langle T^* f, x\rangle| = \sup_{x \in S_X} \sup_{f \in S_{Y^*}} |\langle f, Tx\rangle| \\ &= \sup_{x \in S_X} \|Tx\|_Y \quad \text{(choosing } f \text{ as a supporting functional of } Tx\text{)} \\ &= \|T\|\end{aligned}$$

as required. $\square$

EXERCISE 2.4.28. **Let T be the Hilbert-Schmidt integral operator with kernel $k(t,s)$. Show that T^* is also the Hilbert-Schmidt integral operator with kernel $\overline{k(s,t)}$.**

EXERCISE 2.4.29. **Let R and L denote the right and left shift operators on ℓ_2. Prove that $R^* = L$.**

EXERCISE 2.4.30. **(i) Let $T \in L(X,Y)$ and $S \in L(Y,Z)$. Show that $(ST)^* = T^*S^*$.**
(ii) Let $S, T \in L(X,Y)$ and $a, b \in \mathbb{C}$. Show that $(aS + bT)^* = \bar{a}S^* + \bar{b}T^*$.
(iii) Let $T \in L(X,Y)$ be such that $T^{-1} \in L(Y,X)$. Show that $(T^{-1})^* = (T^*)^{-1}$.

The kernel and image of bounded linear operators are in a duality relation. To state it, we consider a generalization of the notion of orthogonal complement, which we studied in Section 1.5.1 for Hilbert spaces.

DEFINITION 2.4.31 (Annihilator). An *annihilator* of a subset A of a normed space X is the set $A^\perp \subseteq X^*$ defined as

$$A^\perp = \{f \in X^* : \langle f, x\rangle = 0 \text{ for all } x \in A\}.$$

PROPOSITION 2.4.32 (Duality of kernel and image). *Let $T \in L(X,Y)$. Then*

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

PROOF. Let $f \in Y^*$. Then $f \in \ker T^*$ means that $T^*f = 0$, which is equivalent to $\langle T^*f, x\rangle = \langle f, Tx\rangle = 0$ for all $x \in X$, which means that $f \in (\operatorname{Im} T)^\perp$. □

COROLLARY 2.4.33. *Let H be a Hilbert space, and $T \in L(H,H)$. Then the orthogonal decomposition holds:*

$$H = \overline{\operatorname{Im} T} \oplus \ker T^*.$$

PROOF. By Proposition 2.4.32, we have $(\overline{\operatorname{Im} T})^\perp = (\operatorname{Im} T)^\perp = \ker T^*$. (Why does the first identity hold?) By Proposition 1.5.7, the proof is complete. □

EXERCISE 2.4.34. **[Duality of kernel and image II] For a subset $A \subset X^*$, define the "pre-anihilator" as**

$$A_\perp = \{x \in X : \langle f, x\rangle = 0 \textbf{ for all } f \in A\}.$$

Let $T \in L(X,Y)$. Prove that

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

Deduce that

$$(\ker T)^\perp \supseteq \operatorname{Im} T^*.$$

Give an example of a linear operator such that $(\ker T)^\perp \neq \operatorname{Im} T^*$.

2.4.10. Application: ergodic theory. Ergodic theorems allow one to compute space averages as time averages. Let us first state and prove a preliminary form of von Neumann's ergodic theorem; its interpretation will follow.

THEOREM 2.4.35 (von Neumann ergodic theorem). *Let U be a unitary operator on a Hilbert space H. Let P denote the orthogonal projection onto the invariant subspace $\{x \in X : Ux = x\}$. Then, for all $x \in H$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n x = Px.$$

PROOF. It suffices to prove the result for $x \in \ker P$ and for $x \in \operatorname{Im} P$, because then the result will follow for all $x \in H$ by the orthogonal decomposition $H = \ker P \oplus \operatorname{Im} P$. (Check!)

For $x \in \operatorname{Im} P$ the result is trivial because in this case $U^n x = Ux = x$ and $Px = x$. So let $x \in \ker P$. We will first find a convenient representation of $\ker P$. By definition, $\operatorname{Im} P = \ker(I - U) = \ker(I - U^*)$ because for unitary operators, $Ux = x$ if and only if $U^* x = x$. (Check!) Therefore, using the duality between kernels and images, Corollary 2.4.33, we have

$$\ker P = (\ker(I - U^*))^\perp = \overline{\operatorname{Im}(I - U)}.$$

Therefore, every $x \in \ker P$ can be approximated arbitrarily well by vectors of the form $(I - U)y$. For $x = (I - U)y$, we arrive at a telescoping sum

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n x = \frac{1}{N}(x - U^N x) \to 0 \quad \text{as } N \to \infty.$$

This carries over to vectors $x \in \overline{\operatorname{Im}(I - U)}$ by a simple approximation argument. The proof is complete. ☐

EXERCISE 2.4.36. **Write down the approximation argument in this proof.**

Now we explain the implications of Theorem 2.4.35 for time and space averages. Consider first the simple example of a discrete dynamical system studied by Weyl. Let $\mathbb{T}$ denote the unit circle. We put a particle on the circle, and consider its consecutive rotations by some fixed angle $\theta \in (0, 2\pi)$. Equivalently, we consider the sequence

$$n\theta \mod 2\pi, \quad n = 0, 1, 2, \ldots$$

If $\theta/2\pi$ is rational then this sequence has finitely many values. If $\theta/2\pi$ is irrational, then the number of values is infinite. Moreover, in the latter case simulations suggest that the values become uniformly distributed for large N on the circle, see the picture.

This observation was formalized by the Weyl's ergodic theorem, which can be stated as follows. For every number $\theta \in (0, 2\pi)$ such that $\theta/2\pi$ is irrational, and every measurable subset $A \subseteq [0, 2\pi]$, one has

$$\frac{|\{n \leqslant N : n\theta \mod 2\pi \in A\}|}{N} \to \frac{\mu(A)}{2\pi} \quad \text{as } N \to \infty. \tag{2.17}$$

where μ denotes the Lebesgue measure.

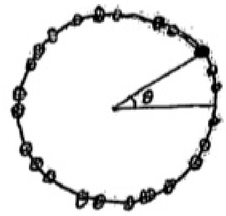

FIGURE 2.4. Rotations by angle θ are equidistributed on the circle

Theorem 2.4.35 implies a very general form of Weyl's ergodic theorem, for an arbitrary ergodic measure-preserving transformation of a probability space (instead of an irrational rotation of the circle).

DEFINITION 2.4.37 (Ergodic transformation). Let (Ω, Σ, μ) be a probability space. A transformation $T : \Omega \to \Omega$ is called *measure-preserving* if

$$\mu(T^{-1}(A)) = \mu(A)$$

for all measurable subsets $A \subseteq \Omega$.[4] A one-to-one, measure preserving transformation T is *ergodic* if the only functions $f \in L_2(\Omega, \Sigma, \mu)$ which satisfy $f(T\omega) = f(\omega)$ for almost all $\omega \in \Omega$ are the constant functions.

EXERCISE 2.4.38. **Show that T is ergodic if and only if for all measurable subset $A \subseteq \Omega$, $T^{-1}(A) = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.**

THEOREM 2.4.39. *Consider a measure-preserving, ergodic transformation T on a probability space (Ω, Σ, μ). For every $f \in L_2(\Omega, \Sigma, \mu)$ one has*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n\omega) = \int_\Omega f \, d\mu \tag{2.18}$$

where the convergence is in the L_2 norm.

PROOF. To see a connection with ergodic Theorem 2.4.35, define an operator $U : L_2(\Omega, \Sigma, \mu) \to L_2(\Omega, \Sigma, \mu)$ by

$$Uf = f \circ T, \quad \text{i.e.} \quad (Uf)(\omega) = f(T\omega), \quad \omega \in \Omega.$$

Since T is measure-preserving, U is a unitary operator. (Check!) Since T is ergodic, the invariant subspace of U is the space of constants. The orthogonal projection P in $L_2(\Omega, \Sigma, \mu)$ onto the subspace of constants is the integral in the right side of (2.18) (see Exercise 1.5.11). The claim then follows from Theorem 2.4.35. □

Weyl's ergodic theorem (2.17) is a partial case of Theorem 2.4.39 for Ω being the unit circle, T being the rotation of the circle by an irrational angle θ (why is T ergodic?) and for $f = \mathbf{1}_A$.

[4]Here $T^{-1}(A) = \{\omega \in \Omega : T\omega \in A\}$ is the preimage of A under T.

2.4.11. Additional Exercises.

EXERCISE 2.4.40. **[Integral operators in $C[0,1]$] Consider the integral operator T defined by (2.11) with kernel $k(t,s) \in C([0,1]^2)$. Then T maps continuous functions into continuous functions, i.e. $T : C[0,1] \to C[0,1]$. Compute the norm of T.**

EXERCISE 2.4.41. **[Schur's test] Assume that a kernel function $k(t,s)$ satisfies**

$$\sup_{t\in[0,1]} \int_0^1 |k(t,s)|\, ds =: M_1 < \infty,$$

$$\sup_{s\in[0,1]} \int_0^1 |k(t,s)|\, dt =: M_2 < \infty.$$

Show that the intergal operator (2.11) $T : L_2[0,1] \to L_2[0,1]$ with kernel $k(t,s)$ is bounded, and[5]

$$\|T\| \leqslant \sqrt{M_1 M_2}.$$

EXERCISE 2.4.42. **[Multiplication operator] Consider a multiplier function $k(t) \in C[0,1]$, and define a linear operator $T : C[0,1] \to C[0,1]$ by**

$$(Tf)(t) = k(t)\, f(t).$$

Show that T is a bounded linear operator and compute its norm.

EXERCISE 2.4.43. **[Annihilators] Let A, B be subsets of a normed space X. Prove the following:**

- **(i)** $A^\perp$ **is a closed linear subspace of** X^***.**
- **(ii) If** $A \subseteq B$ **then** $A^\perp \supseteq B^\perp$**.**
- **(iii)** $(A \cup B)^\perp = A^\perp \cap B^\perp$**. Give an example where** $(A \cap B)^\perp \neq A^\perp \cup B^\perp$**.**
- **(iv)** $A^\perp = (\operatorname{Span} A)^\perp$**.**
- **(v)** $(\bar{A})^\perp = A^\perp$**. (This property was used in the proof of Corollary 2.4.33.)**
- **(vi) Suppose** X_0 **is a closed linear subspace of** X**. Then** $X_0^\perp = \{0\}$ **is equivalent to** $X_0 = X$**.**

EXERCISE 2.4.44. **[Isomorphisms] Let X be a Banach space. Show that the isomorphisms on X form an open subset of $L(X,X)$, and that the inversion map $T \mapsto T^{-1}$ is continuous on this subset.**

EXERCISE 2.4.45. **[Sesquilinear forms] Let H be a Hilbert space. A *sesquilinear form* on H is a function $B : H \times H \to \mathbb{C}$ which is linear in the first argument and conjugate-linear in the second argument, i.e.**

$$B(a_1x_1 + a_2x_2, y) = a_1 B(x_1, y) + a_2 B(x_2, y),$$

$$B(x, b_1y_1 + b_2y_2) = \overline{b_1} B(x, y_1) + \overline{b_2} B(x, y_2).$$

An example of a sesquilinear form is $B(x,y) = \langle Tx, y\rangle$ where $T \in L(H,H)$.

[5]In fact, under the conditions of Schur's test, T is bounded as an operator $L_p[0,1] \to L_p[0,1]$ for all $p \in [1,\infty]$. Indeed, for $p = 1$ and $p = \infty$ this is an exercise. The result for intermediate p follows at once by Riesz-Thorin theorem (which we do not discuss here).

Consider a sesquilinear form $B(x, y)$ **which satisfies**

$$|B(x, y)| \leqslant M\|x\|\|y\|, \quad x, y \in H$$

for some number M**. Show that there exists an operator** $T \in L(H, H)$ **with** $\|T\| \leqslant M$ **and such that**

$$B(x, y) = \langle Tx, y\rangle \quad \textbf{for all } x, y \in H.$$

CHAPTER 3

Main principles of functional analysis

Lec. 22: 10/29

In this chapter we shall study the three theorems that, together with Hahn-Banach theorem, form the main principles of functional analysis. Those are the open mapping theorem, the uniform boundedness principle, and the closed graph theorem.

3.1. Open mapping theorem

3.1.1. Statement and proof. This result was proved by S. Banach.

Theorem 3.1.1 (Open mapping theorem). *Let X, Y be Banach spaces. Then every surjective linear operator $T \in L(X, Y)$ is an* open map*, i.e. T maps open sets in X to open sets in Y.*

The proof of the open mapping theorem relies on Baire category theorem, which states that every complete metric space M is a set of second category, i.e. M can not be represented as a countable union of nowhere dense sets. Recall that a subset $A \subseteq M$ is nowhere dense set if there is no neighborhood in X on which A is dense. Equivalently, A is nowhere dense if the interior of the closure of A is empty.

The open mapping theorem states that for every open set $U \subseteq X$, every $y \in TU$ is an interior point of TU. We claim that it suffices to show this for U being the unit ball B_X and for $y = 0$:

Claim 3.1.2. To prove the open mapping theorem, it suffices to find $\varepsilon > 0$ such that

$$TB_X \supseteq \varepsilon B_Y. \tag{3.1}$$

Proof of Claim. Let $U \subseteq X$ be open and choose $y \in TU$. We find $x \in U$ such that $y = Tx$. Since U is open, there exists $\delta > 0$ such that

$$U \supseteq x + \delta B_X.$$

Applying T to both sides and using (3.1), we conclude that

$$TU \supseteq T(x + \delta B_X) = y + \delta T B_X \supseteq y + \delta\varepsilon B_Y,$$

so y is an interior point of TU. $\square$

We will now prove the Claim. In view of application of Baire category theorem, we represent

$$X = \bigcup_{n\in\mathbb{N}} nB_X.$$

Therefore

$$Y = TX = \bigcup_{n\in\mathbb{N}} nTB_X.$$

By Baire category theorem, there exists $n \in N$ such that nTB_X is not a nowhere dense set. Thus TB_X is not a nowhere dense set, i.e. its closure has nonempty interior. So there exist $y \in Y$ and $\varepsilon > 0$ such that

$$\overline{TB_X} \supseteq y + \varepsilon B_Y.$$

By symmetry, $\overline{TB_X} \supseteq -y + \varepsilon B_Y$. Hence by convexity (check this!) we have

$$\overline{TB_X} \supseteq \varepsilon B_Y.$$

As we see, we have almost proved the Claim, except for the closure. Unfortunately, in general $\overline{K} \supseteq D$ does not imply $K \supseteq D$ even for convex and symmetric sets K, D in a Banach space. (Give a counterexample!) However, this is true for *perfectly* convex sets, defined as follows.

Definition 3.1.3 (Perfectly convex set). A set K in a Banach space Y is called *perfectly convex* if for every sequence $(x_k)_{k=1}^{\infty}$ and every numbers $\lambda_k \geqslant 0$ such that $\sum_{k=1}^{\infty} \lambda_k = 1$, one has $\sum_{k=1}^{\infty} \lambda_k x_k \in K$.

Convex sets satisfy this property only for *finite* sequences (x_k). (Why?) Therefore, every perfectly convex set is convex, but not vice versa. (Give an example.)

Lemma 3.1.4 (Perfectly convex sets). *Let K be a perfectly convex set in a Banach space Y. If $\overline{K} \supseteq \varepsilon B_Y$ for some $\varepsilon > 0$, then $K \supseteq \frac{\varepsilon}{2} B_Y$.*

Proof. Assume $B := \varepsilon B_Y \subseteq \overline{K}$; we would like to show that $\frac{1}{2}B \subseteq K$. The assumption clearly implies that

$$B \subseteq K + \frac{1}{2}B,$$

for the right side is the ε-neighborhood of K in Y. Iterating this inclusion gives

$$\begin{aligned} B &\subseteq K + \frac{1}{2}\Big(K + \frac{1}{2}B\Big) = K + \frac{1}{2}K + \frac{1}{4}B \\ &\subseteq K + \frac{1}{2}K + \frac{1}{4}\Big(K + \frac{1}{2}B\Big) = K + \frac{1}{2}K + \frac{1}{4}K + \frac{1}{8}B \subseteq \cdots \end{aligned}$$

Therefore[1]

$$B \subseteq K + \frac{1}{2}K + \frac{1}{4}K + \frac{1}{8}K + \cdots$$

By perfect convexity (check!), we have

$$\frac{1}{2}B \subseteq \frac{1}{2}K + \frac{1}{4}K + \frac{1}{8}K + \frac{1}{16}K + \cdots \subseteq K.$$

This proves the lemma. □

Exercise 3.1.5. **Verify the steps of the above proof where we used Minkowski sums and series of sets.**

Now we are ready to complete the proof of the open mapping theorem. By Lemma 3.1.4, it suffices to show that $K = TB_X$ is perfectly convex. This is easy to

---

[1] All sums and series of sets are in Minkowski sense. The sum of sets is defined as $\sum_k A_k = \{\sum_k a_k : \; a_k \in A_k\}$. The same for infinite sums (series), where we insist on the convergence of $\sum_k a_k$.

check. Indeed, consider any sequence $(Tx_k) \subseteq TB_X$ with $x_k \in B_X$, and numbers λ_k such that $\sum_k \lambda_k = 1$. Then

$$\sum_k \lambda_k T x_k = T\Big(\sum_k \lambda_k x_k\Big), \tag{3.2}$$

provided that the series $\sum_k \lambda_k x_k$ converges. It indeed converges absolutely:

$$\Big\| \sum_k \lambda_k x_k \Big\| \leqslant \sum_k \lambda_k \|x_k\| \leqslant \sum_k \lambda_k = 1.$$

By completeness of X, the series $\sum_k \lambda_k x_k$ converges to a vector in B_X. It follows that the right side of (3.2) belongs to $TB_X = K$, as required. This completes the proof of the open mapping theorem. □ Lec. 23: 11/1

3.1.2. Inverse mapping theorem. As an immediate consequence of the open mapping theorem, we obtain:

THEOREM 3.1.6 (Inverse mapping theorem). *Let X, Y be Banach spaces. Then every bijective linear operator $T \in L(X,Y)$ is an isomorphism, i.e. $T^{-1} \in L(Y,X)$.*

PROOF. The open mapping theorem states that the preimages of open sets under T^{-1} are open, hence T^{-1} is a continuous map. □

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

$$Tx = b \tag{3.3}$$

with $T \in L(X,Y)$ and $b \in Y$.

Assume a solution x *exists and is unique* for every right hand side b. Then, by inverse mapping theorem, the solution $x = x(b)$ is *continuous* with respect to b. In other words, the solution is stable under perturbations of the right hand side of (3.3).

In case T is not injective (but is surjective) in the inverse mapping theorem, one can still apply inverse mapping theorem to the injectivization of T:

COROLLARY 3.1.7 (Surjective operators are essentially quotient maps). *Let X, Y be Banach spaces. Then every surjective linear operator $T \in L(X,Y)$ is a composition of a quotient map and an isomorphism. Specifically,*

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

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

PROOF. Let $\tilde{T}$ be the injectivization of T constructed in Example 1.1.24. By construction, $T = \tilde{T}q$ and $\tilde{T}$ is injective. Since T is surjective, $\tilde{T}$ is also surjective. Therefore, by inverse mapping theorem $\tilde{T}$ is an isomorphism, completing the proof. □

3.1.3. Equivalent norms. Sometimes one wants to consider several norms on the same space, such as $\|\cdot\|_\infty$ and $\|\cdot\|_1$ on $C[0,1]$. In this case one naturally needs to compare them.

DEFINITION 3.1.8 (Equivalent norms). Two norms $\|\cdot\|$ and $|||\cdot|||$ on a linear vector space E are called *equivalent* if there exist $C, c > 0$ such that

$$c\|x\| \leqslant |||x||| \leqslant C\|x\| \quad \text{for all } x \in E.$$

In other words, the norms $\|\cdot\|$ and $|||\cdot|||$ are equivalent if the identity operator

$$I : (E, \|\cdot\|) \to (E, |||\cdot|||) \tag{3.4}$$

is an isomorphism.

EXERCISE 3.1.9. **Show that two norms on a linear vector space E are equivalent if and only if they generate the same topology on E.**

PROPOSITION 3.1.10 (Domination and equivalence of norms). *Consider two norms $\|\cdot\|$ and $|||\cdot|||$ on a linear vector space E such that E is complete with respect to both norms. Suppose that one norm* dominates *the other, i.e. one can find C such that*

$$|||x||| \leqslant C\|x\| \quad \text{for all } x \in E.$$

Then the two norms are equivalent.

PROOF. The claim follows from inverse mapping theorem applied to the identity map (3.4). □

Proposition 3.1.10 indicates that it is difficult to construct different good norms on the same space. Either the space will be incomplete or the norms need to be incomparable. This is a simple way to prove incompleteness of various spaces.

As an example, consider the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ on $C[0,1]$. On the one hand, $\|\cdot\|_1 \leqslant \|\cdot\|_\infty$. On the other hand, the norms are not equivalent – one can easily construct functions with $\|f\|_1 = 1$ and $\|f\|_\infty$ arbitrarily large. (Do this.) By Proposition 3.1.10, $C[0,1]$ must be incomplete with respect to one of these norms. Since it is complete with its natural norm $\|\cdot\|_\infty$, it follows that $C[0,1]$ is incomplete with respect to $\|\cdot\|_1$.

This argument is flexible and applies to a whole range of norms. It implies that there is essentially only one natural norm on $C[0,1]$, namely the sup-norm $\|\cdot\|_\infty$.

EXERCISE 3.1.11. **[Direct sums of normed spaces] Let X and Y be normed spaces and $p \in [1,\infty]$. Define the direct sum of $X \oplus_p Y$ as the Cartesian product $X \times Y$ equipped with the norm**

$$\|(x,y)\| := (\|x\|^p + \|y\|^p)^{1/p} \text{ if } p < \infty, \qquad \|(x,y)\| := \max(\|x\|, \|y\|) \text{ if } p = \infty.$$

Show that $X \oplus_p Y$ is a normed space, and all norms $\|(x,y)\|_p$, $p \in [1,\infty]$, are equivalent to each other.

For this reason, the index p is usually omitted from notation, and the space $X \oplus Y$ is called the *direct sum* of X and Y.

3.1.4. Isomorphic embeddings. As we know, the kernel of every operator $T \in L(X,Y)$ is always a closed subspace. The image of T may or may not be closed. The following result characterizes operators with closed images.

PROPOSITION 3.1.12 (Isomorphic embeddings). *Let $T \in L(X,Y)$ be an operator between Banach spaces X and Y. Then the following are equivalent:*

(i) T is an isomorphic embedding, *i.e. T acts as an isomorphism between spaces X and $\operatorname{Im} T \subseteq Y$;*
(ii) T is injective and it has closed image*;*
(iii) T is bounded below, *i.e. one can find $c > 0$ such that*

$$\|Tx\| \geqslant c\|x\| \quad \text{for all } x \in X.$$

PROOF. (i) ⇒ (ii). Recall that isomorphisms preserve completeness of spaces. Since X is complete, it follows that the subspace $\operatorname{Im} T \subseteq Y$ is complete. By Exercise 1.3.2, $\operatorname{Im} T$ is closed. Injectivity of T is an obvious consequence of the isomorphic embedding property.

(ii) ⇒ (iii). Considering T as an operator from X to $\operatorname{Im} T$ we see that T is injective and surjective. By inverse mapping theorem, T is an isomorphism.

(iii) ⇒ (i). We have

$$c\|x\| \leqslant \|Tx\| \leqslant C\|x\| \quad \text{for all } x \in X$$

where $C = \|T\|$. It follows that T is an isomorphic embedding. (How?) □

3.1.5. Finite dimensional normed spaces. Finite dimensional normed spaces are the simplest examples of Banach spaces. As we will see now, they are all complete, all isomorphic (but not isometric) to each other, all of their subspaces are closed and all linear operators on them are bounded.

THEOREM 3.1.13. *Every n-dimensional normed space X is isomorphic to the Euclidean space ℓ_2^n. Consequently, all n-dimensional normed spaces are isomorphic to each other.*

PROOF. We construct an isomorphism between X and ℓ_2^n which that identifies a basis of X with the canonical basis of ℓ_2^n. To this end, consider a basis $u_1, \dots, u_n$ of X, and define the operator $T : \ell_2^n \to X$ by

$$Tx = \sum_{k=1}^{n} x_k u_k \quad \text{for } x = (x_1, \dots, x_n) \in \ell_2^n.$$

Boundedness of T follows from triangle and Cauchy-Schwarz inequalities:

$$\|Tx\| \leqslant \sum_{k=1}^{n} |x_k| \, \|u_k\| \leqslant \Big(\sum_{k=1}^{n} |x_k|^2 \Big)^{1/2} \Big(\sum_{k=1}^{n} \|u_k\|^2 \Big)^{1/2} = M\|x\|,$$

where $M = \big(\sum_{k=1}^{n} \|u_k\|^2 \big)^{1/2}$.

Moreover, T is bijective because (u_k) is a basis of X. (Why?) By inverse mapping theorem, T is an isomorphism. □

EXERCISE 3.1.14. **Instead of using the inverse mapping theorem, give a proof based on the compactness of the unit sphere of ℓ_2^n.**

EXERCISE 3.1.15. **Any two n-dimensional normed spaces X and Y are isomorphic but not necessarily *isometric*. Show that X and Y are isometric if and only if $B_Y = TB_X$ for some invertible linear transformation $T : X \to Y$. Show that ℓ_2^2 is not isometric to ℓ_∞^2.**

REMARK 3.1.16 (Banach-Mazur distance). A quantitative notion of isomorphism is given by the *Banach-Mazur distance* $d(X, Y)$ between two isomorphic normed spaces X and Y. It is defined as

$$d(X, Y) = \inf\{\|T\| \, \|T^{-1}\| : \ T : X \to Y \text{ is an isomorphism}\}.$$

A theorem of F. John (see e.g. [**1**]) is a quantitative form of Theorem 3.1.13. It states that every n-dimensional normed space X satisfies

$$d(X, \ell_2^n) \leqslant \sqrt{n}.$$

It follows that for every two n-dimensional normed spaces X and Y, one has

$$d(X,Y) \leqslant n.$$

(Why?) E. Gluskin [**4**] proved in 1981 that this upper bound is asymptotically sharp, i.e. there exists an absolute constant $c > 0$ such that, for every $n \in \mathbb{N}$ one can construct n-dimensional normed spaces X_n and Y_n with

$$d(X_n, Y_n) \geqslant cn.$$

COROLLARY 3.1.17. *(i) Every finite dimensional normed space is a Banach space.*

(ii) Every finite dimensional subspace of a normed space is closed.

(iii) Every linear operator on a finite dimensional normed space is bounded.

(iv) Every two norms on a finite dimensional normed space are equivalent.

PROOF. (i) Since an n-dimensional normed space X is isomorphic to the complete space ℓ_2^n, X itself is complete.

(ii) Let Y be a finite-dimensional subspace of a normed space X. Then Y is complete by part (i). Therefore Y is closed. (Indeed, if a sequence $(y_n) \subset Y$ converges to $x \in X$ then (y_n) is Cauchy in Y, hence its limit x is in Y.)

(iii) By Proposition 2.4.7, every linear operator on ℓ_2^n is bounded, $\|T\| \leqslant \|T\|_{\mathrm{HS}}$. Since an n-dimensional normed space X is isomorphic ℓ_2^n, the same is true for linear operators on X.

(iv) Let $\|\cdot\|$ and $|||\cdot|||$ be two norms on X. By part (iii), the identity map $I : (X, \|\cdot\|) \to (X, |||\cdot|||)$ as well as its inverse are bounded. This completes the proof. □

EXERCISE 3.1.18. **[Operators from finite dimensional spaces] Let X and Y be normed spaces, and X be finite dimensional. Show that every linear operator $T : X \to Y$ is bounded. (Hint: identify X with ℓ_2^n by some isomorphism, and show boundedness of T similarly to the argument of Theorem 3.1.13.)**

Lec. 24: 11/3

3.2. Closed graph theorem

3.2.1. Statement and proof. Closed graph theorem is an alternative way to check whether a given linear operator is bounded. This result characterizes bounded operators in terms of their graphs.

DEFINITION 3.2.1 (Graph of an operator). Let $T : X \to Y$ be a linear operator between normed spaces X and Y. The *graph* of T is the following subset of the direct sum[2] $X \oplus_1 Y$:

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

Clearly, $\Gamma(T)$ is a linear subspace of the normed space $X \oplus_1 Y$. The main result of this section is that $\Gamma(T)$ is closed if and only if T is bounded.

Let us compare these two notions, boundedness (equivalently, continuity) of T and having closed graph. T is continuous if and only if

$$x_n \to x \in X \quad \text{implies} \quad Tx_n \to Tx. \tag{3.5}$$

[2]The direct sum of normed spaces is studied in Exercise 3.1.11. Instead of $X \oplus_1 Y$, one can choose to work with $X \oplus_p Y$ for any $p \in [1, \infty]$. As we know this defines the same topology, and will not affect the arguments in this section.

In contrast, $\Gamma(T)$ is closed if and only if

$$x_n \to x \in X \text{ and } Tx_n \to y \in Y \quad \text{imply} \quad y = Tx. \tag{3.6}$$

It is clear from these two lines that continuity always implies the closed graph property:

PROPOSITION 3.2.2. *Let $T : X \to Y$ be a linear operator between normed spaces X and Y. If T is bounded then $\Gamma(T)$ is closed.*

The opposite statement is nontrivial and requires completeness of both spaces X and Y:

THEOREM 3.2.3 (Closed graph theorem). *Let $T : X \to Y$ be a linear operator between Banach spaces X and Y. If $\Gamma(T)$ is closed then T is bounded.*

PROOF. The direct sum $X \oplus_1 Y$ is a Banach space (Exercise 1.3.13). The graph $\Gamma(T)$ is a closed linear subspace of $X \oplus_1 Y$, hence $\Gamma(T)$ is a Banach space itself.

Consider the linear operator

$$u : \Gamma(T) \to X, \quad u(x, Tx) := x.$$

Then u is a bounded, surjective and injective linear operator between two Banach spaces. (Check!) By the open mapping theorem, u^{-1} is bounded. This means that there exists a number M such that

$$\|x\| + \|Tx\| = \|(x, Tx)\| \leqslant M\|x\| \quad \text{for all } x \in X.$$

The inequality $\|Tx\| \leqslant M\|x\|$ implies that T is bounded. $\square$

3.2.2. Interpretation and an example. Recalling the interpretation of continuity (3.5) and closed graph property (3.6), we can make advantage of the closed graph property – the extra assumption that Tx_n converges in Y. So, to check the continuity of a linear operator T using the definition (3.5), *one can always assume for free that Tx_n converges.* Checking continuity no longer requires proving that the limit exists; it reduces to checking the *consistency* of the limits of x_n and Tx_n.

As an example, consider the simplest differential operator

$$T = \frac{d}{dt}, \quad T : C^1[0,1] \to C[0,1]$$

where $C^1[0,1]$ is considered as a subspace of $C[0,1]$, i.e. with respect to sup-norm.

LEMMA 3.2.4. *The differential operator T has closed graph.*

PROOF. Let $f_n \to f$ and $Tf_n = f_n' \to g$ in $C[0,1]$, i.e. uniformly. Then, by the theorem on differentiation under the limit,[3] $(\lim_n f_n)' = \lim_n f_n'$, i.e. $g = Tf$. This completes the proof. $\square$

Nevertheless, as we know the differential operator is unbounded. (Why?) This does not contradict the closed graph theorem, because $C^1[0,1]$ is not complete under the sup-norm. If we consider $C^1[0,1]$ under its natural norm $\|f\|_\infty + \|f'\|_\infty$ in which it is a Banach space (Exercise 1.3.12), then the differential operator will obviously be bounded.

---

[3]The theorem on differentiation under the limit states that $(\lim_n f_n)' = \lim_n f_n'$ provided that f_n' converges uniformly and $f_n(t_0)$ converges for some point t_0.

3.2.3. Symmetric operators on Hilbert spaces. A remarkable application of closed graph theorem is that the symmetry property of an operator always implies boundedness:

THEOREM 3.2.5 (Hellinger-Toeplitz). *Let $T : H \to H$ be a linear operator on a Hilbert space H. Suppose that*

$$\langle x, Tx\rangle = \langle Tx, y\rangle \quad \textit{for all } x, y \in H. \tag{3.7}$$

Then T is bounded.

PROOF. By the closed graph theorem, it suffices to check that the graph of T is closed. To this end we choose convergent sequences $x_n \to x$, $Tx_n \to y$ in H. We would like to show that $y = Tx$. It suffices to show that

$$\langle z, y\rangle = \langle z, Tx\rangle \quad \text{for all } z \in H.$$

(Why?) This follows by using continuity of the inner product and (3.7) twice:

$$\langle z, y\rangle = \lim_n \langle z, Tx_n\rangle = \lim_n \langle Tz, x_n\rangle = \langle Tz, x\rangle = \langle z, Tx\rangle.$$

The proof is complete. □

Hellinger-Toeplitz theorem identifies the source of considerable difficulties in mathematical physics. Many natural operators, such as differential, satisfy the symmetry condition (3.7) but are unbounded. Hellinger-Toeplitz theorem declares that such operators *can not be defined everywhere* on the Hilbert space. For example, one can never define a useful notion of differentiation that would make all functions in L_2 differentiable.

This explains that working with unbounded operators one has to always keep track of their domains. For example, a linear operator T on a Hilbert space H is called *symmetric* if the domain of T is dense in H, and (3.7) holds. An example of a symmetric operator is the differential operator on $L_2[0,1]$

$$T = i\frac{d}{dt}$$

with domain

$$\operatorname{Dom} T = \{f \in L_2[0,1] : \ f \in C^1[0,1],\ f(0) = f(1) = 1\}.$$

3.3. Principle of uniform boundedness

Lec.25: 11/5

3.3.1. Statement and proof. The principle of uniform boundedness is a result due to Banach and Steinhaus.

THEOREM 3.3.1 (Principle of uniform boundedness). *Let X and Y be Banach spaces. Consider a family of bounded linear operators $\mathcal{T} \subseteq L(X, Y)$. Assume that $\mathcal{T}$ is* pointwise bounded, *i.e.*

$$\sup_{T \in \mathcal{T}} \|Tx\| < \infty \quad \textit{for every } x \in X. \tag{3.8}$$

Then $\mathcal{T}$ is uniformly bounded, *i.e.*

$$\sup_{T \in \mathcal{T}} \|T\| < \infty.$$

Note that the reverse direction is trivially true – uniform boundedness implies pointwise boundedness. (Why?)

PROOF. For every $x \in X$, define

$$M(x) = \sup_{T \in \mathcal{T}} \|Tx\|.$$

Pointwise boundedness of $\mathcal{T}$ means that

$$X = \bigcup_{n \in \mathbb{N}} X_n, \quad \text{where} \quad X_n = \{x \in X : M(x) \leqslant n\}.$$

Baire category theorem implies that one of X_n is not a nowhere dense subset of X. Since X_n is closed (why?), X_n has nonempty interior. Summarizing, we have shown that there exist $n \in \mathbb{N}$, $x_0 \in X$ and $\varepsilon > 0$ such that

$$X_n \supseteq x_0 + \varepsilon B_X.$$

By symmetry of X_n, we also have $X_n \supseteq -x_0 + \varepsilon B_X$. Hence by convexity of X_n, we have

$$X_n \supseteq \varepsilon B_X.$$

(Check!) By definition of X_n this means that for $x \in X$,

$$\|x\| \leqslant \varepsilon \quad \text{implies} \quad \sup_{T \in \mathcal{T}} \|Tx\| \leqslant n.$$

It follows that for every $x \in X$,

$$\sup_{T \in \mathcal{T}} \|Tx\| \leqslant \frac{n}{\varepsilon} \|x\|.$$

This implies that

$$\sup_{T \in \mathcal{T}} \|T\| \leqslant \frac{n}{\varepsilon} < \infty$$

as required. □

EXERCISE 3.3.2. **Check that the sub-level sets X_n in the proof above are closed, convex and symmetric. (All these properties were used in the argument).**

REMARK 3.3.3 (Principle of condensation of singularities). Banach and Steinhaus called their Theorem 3.3.1 the *principle of condensation of singularities* for the following reason. Suppose a family $\mathcal{T} \subseteq L(X,Y)$ is not uniformly bounded. This means that the set of vectors $\{Tx : x \in B_X, T \in \mathcal{T}\}$ is unbounded. Theorem 3.3.1 states that $\mathcal{T}$ is not even pointwise 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.

REMARK 3.3.4 (Completeness). In the proof of Theorem 3.3.1, the completeness of only X was used. So the result still holds if X is a Banach space and Y is a normed space.

3.3.2. Weak and strong boundedness. Principle of uniform boundedness can be used to check whether a given set in a Banach space is bounded.

COROLLARY 3.3.5 (Weak and strong boundedness). *Let A be a subset of a Banach space X. Assume that A is* weakly bounded, *i.e.*

$$\sup_{x \in A} |f(x)| < \infty \quad \textit{for every } f \in X^*.$$

Then A is (strongly) bounded, i.e.

$$\sup_{x \in A} \|x\| < \infty.$$

Here again the reverse statement is trivially true – (strong) boundedness trivially implies weak boundedness.

PROOF. We embed X into X^{**} using the canonical embedding that we studied in Theorem 2.3.12. So we consider vectors $x \in A$ as bounded linear functionals on X^* acting as $x(f) := f(x)$, $f \in X^*$. Rewriting the weak boundedness assumption as $\sup_{x \in A} |x(f)| < \infty$ for $f \in X^*$, we may understand this assumption as pointwise boundedness of the family $A \subseteq X^{**} = L(X^*, \mathbb{R})$. The principle of uniform boundedness implies that $\sup_{x \in A} \|x\|_{X^{**}} = \sup_{x \in A} \|x\|_X < \infty$, as required. □

REMARK 3.3.6. Using Corollary 3.3.5, one can weaken the assumption (3.8) in the principle of uniform boundedness to the following one:

$$\sup_{T \in \mathcal{T}} |f(Tx)| < \infty \quad \text{for every } x \in X,\ f \in Y^*.$$

(Why?)

3.3.3. Application to convergence of Fourier series. A basic and classical question in Fourier analysis is – when does Fourier series of a function f on an interval converge to f?

Hilbert space technique provides a complete answer to this question in the space L_2. As we know from Theorem 1.6.20, the Fourier series of every function in $L_2[-\pi, \pi]$ converges to f in the L_2-norm.

In function spaces other than L_2, the answer to this problem is often nontrivial and even negative. Unfortunately, such is the situation in the space of continuous functions $C[-\pi, \pi]$. There exist continuous functions f whose Fourier series do not converge in $C[-\pi, \pi]$ (i.e. uniformly). This follows from a somewhat stronger result, which in turn is a consequence of the principle of uniform boundedness:

THEOREM 3.3.7 (Divergent Fourier series). *There exists a function $f \in C[-\pi, \pi]$ whose partial Fourier sums*

$$(S_n f)(t) = \sum_{k=-n}^{n} \hat{f}(k) e^{ikt}$$

forms an unbounded sequence of complex numbers at $t = 0$. In particular, the Fourier series of f is unbounded (hence divergent) at $t = 0$.

PROOF. Recall from (2.13) that the partial Fourier sums can be represented via convolution with Dirichlet kernel D_n:

$$(S_n f)(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(t - s) f(s)\, ds, \quad \text{where} \quad D_n(\theta) = \frac{\sin(n + \frac{1}{2})\theta}{\sin \frac{1}{2}\theta}.$$

We are interested in the behavior of $(S_n f)(0)$. These are obviously linear functionals on $C[-\pi, \pi]$, which we denote

$$\phi_n(f) := (S_n f)(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(s) f(s)\, ds.$$

(We used that D_n is an even function).

Since Dirichlet kernel D_n is continuous, ϕ_n are bounded linear functionals on $C[-\pi,\pi]$, and

$$\|\phi_n\| = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_n(s)|\, ds$$

(Why?)

On the other hand, evaluating these integrals by hand one can see that

$$\|\phi_n\| \to \infty \quad \text{as } n \to \infty. \tag{3.9}$$

(see the exercise below). Therefore, $(\phi_n)_{n\in\mathbb{Z}}$ is not a uniformly bounded family of linear functionals.

By principle of uniform boundedness, this family is not even pointwise bounded. This means that there exists a function $f \in C[-\pi,\pi]$ such that the set of numbers $\{\phi_n(f) : n \in \mathbb{N}\}$ is unbounded. This is exactly what we wanted to show. □

EXERCISE 3.3.8. **[L_1 norm of Dirichlet kernel] Show that**

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |D_n(s)|\, ds \geqslant c \log n$$

where $c > 0$ is an absolute constant. This validates step (3.9) in the argument above. Hint: show that the area under k-th bump of the graph of the Dirichlet kernel is $\gtrsim k$ (see the picture); as there are $\sim n$ bumps on $[-\pi,\pi]$ this will show that the total area of the bumps is $\sum_{k=1}^{n} 1/k \gtrsim \log n$.

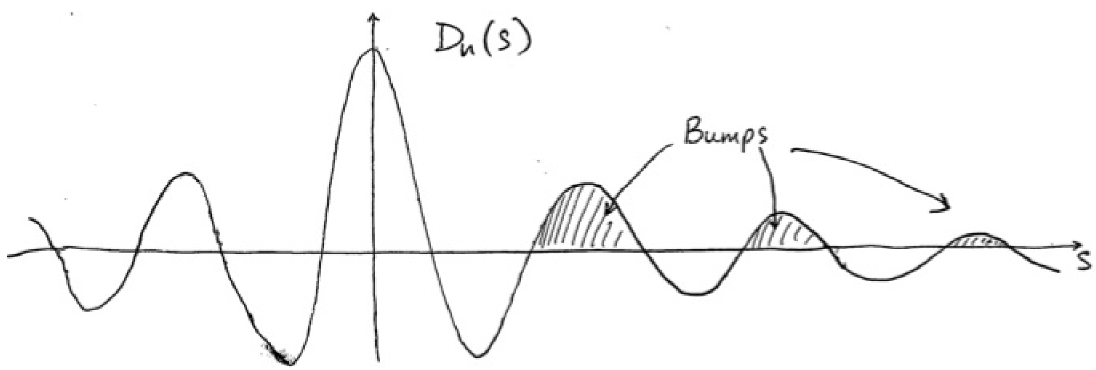

FIGURE 3.1. Dirichlet kernel

REMARK 3.3.9 (Convergence of Fourier series).
- (i) Analyzing the proof of Theorem 3.3.7 one can show that continuous functions whose Fourier series converge pointwise are rare. Precisely, the set of such functions is of *first Baire category* in $C[-\pi,\pi]$ (i.e. it is a countable union of nowhere dense sets).
- (ii) Nevertheless, for every continuous function f, and even for $f \in L_p[-\pi,\pi]$, $p > 1$, Fourier series converges to f almost everywhere. So the set of points of divergence is always small. This is a deep result of L. Carleson [**2**].
- (iii) For functions in L_1, Carleson result is generally false. Kolmogorov (apparently when he was 19-21!) constructed a function $f \in L_1[-\pi,\pi]$ whose Fourier series diverges everywhere.
- (iv) If f is *differentiable* at a point t, then Fourier series of f converges to f at t. This is called Dirichlet-Dini condition.

Lec. 26: 11/8

3.3.4. Schauder bases. The notion of Hamel basis, which we studied in Section 1.1.4, has a serious drawback. In all infinite-dimensional Banach spaces, Hamel bases are *uncountable*, see Exercise 3.3.22. This makes it difficult to use Hamel bases in practice. There exists an alternative notion of basis, which is more tailored to the needs of analysis:

DEFINITION 3.3.10 (Schauder basis). A sequence $(x_k)_{k=1}^{\infty}$ in a Banach space X is called a *Schauder basis* of X if every vector $x \in X$ can be uniquely expressed as a convergent series

$$x = \sum_{k=1}^{\infty} a_k x_k \tag{3.10}$$

for some scalars a_k.

EXERCISE 3.3.11. **Show that only separable spaces X may have Schauder bases.**

If (x_k) is a Schauder basis then the sequence (x_k) is clearly linear independent and complete (recall from Definition 1.6.15 that completeness means that $\mathrm{Span}(x_k)$ is dense in X.) However, the basis property is considerably stronger than linear independence and completeness. Completeness means that for arbitrary vector $x \in X$ and arbitrary precision $\varepsilon > 0$, one can find a linear combination of basis elements that approximates x well:

$$\Big\| x - \sum_{k=1}^{n} a_k x_k \Big\| \leqslant \varepsilon. \tag{3.11}$$

However, the coefficients $a_k = a_k(\varepsilon)$ may depend on ε. The limit $\lim_{\varepsilon \to 0} a_k(\varepsilon)$ generally does not exist, as will be clear from the next exercise. In contrast, the basis property guarantees that one can achieve approximation (3.11) with *the same* coefficients a_k independent of ε by only increasing the number of terms n in the linear combination.

EXERCISE 3.3.12. **[Completeness and basis property] In a Hilbert space H, find a complete and linearly independent sequence (x_k) which is not a Schauder basis. Hint: construct x_k so that it converges to some nonzero vector in H; show that such sequences are never Schauder bases.**

EXAMPLE 3.3.13 (Bases in sequence spaces). An orthogonal basis of a Hilbert space is a Schauder basis. (Check the uniqueness of represenation of x!)

In sequence spaces ℓ_p, $1 \leqslant p < \infty$ and c_0, the canonical (coordinate) system forms a Schauder basis. (Check!)

In ℓ_∞ there is no Schauder basis because this space is not separable.

EXAMPLE 3.3.14 (Bases in L_p). One can show (albeit somewhat nontrivially) that Fourier basis (1.16) and the trigonometric system each form Schauder bases in $L_p[-\pi, \pi]$, $1 < p < \infty$.

The Haar system defined in Exercise 1.6.34 also forms a Schauder basis in $L_p[-\pi, \pi]$, $1 \leqslant p < \infty$. (Check!)

In $L_\infty[-\pi, \pi]$ there is no Schauder basis because this space is not separable.

EXAMPLE 3.3.15 (Basis of the space of continuous functions). In $C[0,1]$, the natural candidates fail to be Schauder bases. The Fourier basis is not a Schauder basis – otherwise this would imply that the Fourier series of every continuous function would converge in $C[0,1]$ (why?), which would contradict Theorem 3.3.7.

The sequence of monomials $1, t, t^2, \dots$ is not a Schauder basis of $C[0,1]$ either. (Prove this!)

The most known Schauder basis of $C[0,1]$ is the so-called *Schauder system* of wavelets. Its mother wavelet $\phi(t)$ is obtained by integration of Haar mother wavelet, i.e.

$$\phi(t) = \int_0^t h(s)\, ds = \begin{cases} t, & t \in [0, 1/2) \\ 1-t, & t \in [1/2, 1) \end{cases}$$

Then we consider the translates and dilates of the mother wavelet:

$$\phi_{kl}(t) = \phi(2^k t - l), \quad k = 0, 1, 2, \dots, \quad l = 0, 1, 2, \dots, 2^k - 1.$$

Together with the constant function 1, the system of functions $\phi_{kl}(t)$ is called the *Schauder system*, see the picture. It forms a Schauder basis on $C[0,1]$. (Check!)

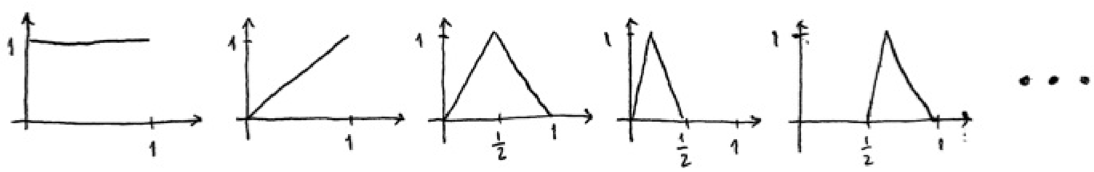

FIGURE 3.2. The first few functions of the Schauder system

EXERCISE 3.3.16. **[Verifying Schauder bases] Do all checks in the previous examples. For $C[0,1]$, note that the linear span of Schauder system consists of piecewise-linear functions on $[0,1]$ whose nodes are dyadic points.**

A major property of Schauder bases is the uniform bound on the partial sums:

THEOREM 3.3.17 (Partial sums of a Schauder basis). *Let (x_k) be a Schauder basis of a Banach space X. Then there exists a number M called the* basis constant *of (x_k) with the following property. The partial sums of the basis expansion* (3.10) *of every $x \in X$ satisfy*

$$\Big\| \sum_{k=1}^n a_k x_k \Big\| \leqslant M \|x\|, \quad n = 1, 2, \dots$$

PROOF. Consider the sequence space

$$E := \Big\{ a = (a_k)_{k=1}^\infty : \sum_{k=1}^\infty a_k x_k \text{ converges in } X \Big\} \tag{3.12}$$

with the norm

$$\|a\|_E := \sup_{n \in \mathbb{N}} \Big\| \sum_{k=1}^n a_k x_k \Big\|. \tag{3.13}$$

It is an exercise to check that E is a Banach space. We will show that X and E are isomorphic.[4]

To this end, we consider the synthesis operator $T : E \to X$ defined as

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

By construction, T is a bounded linear operator:

$$\|Ta\| \leqslant \|a\|_E.$$

Since (x_k) is a Schauder basis, T is surjective and injective. By the inverse mapping theorem, T is an isomorphism. Therefore one can find a number M such that

$$\|a\|_E \leqslant M\|Ta\|, \quad a \in E.$$

But this means that

$$\sup_{n\in\mathbb{N}} \Big\| \sum_{k=1}^{n} a_k x_k \Big\| \leqslant M\|x\|, \quad x \in X,$$

which we wanted to prove. □

Exercise 3.3.18. **[Space of coefficients] Let $(x_k)_{k=1}^{\infty}$ be a sequence of nonzero vectors in a Banach space X. Define the space of coefficients E by (3.12) and (3.13). Prove that E is a Banach space. (This property was used in the proof of Theorem 3.3.17.)**

Consider the partial sums of a basis expansion (3.10):

$$S_n(x) := \sum_{k=1}^{n} a_k x_k.$$

By Theorem 3.3.17, we see that S_n is a projection[5] in X onto $\mathrm{Span}(x_1, \ldots, x_n)$, and that S_n are uniformly bounded (by the basis constant):

$$\sup_{n\in\mathbb{N}} \|S_n\| = M < \infty.$$

Also, the coefficients $a_k = a_k(x)$ of the basis expansion (3.10) are obviously linear functionals on X. They are called *biorthogonal functionals* of the basis (x_k) and denoted x_k^*, i.e.

$$x_k^*(x) = a_k.$$

With this notation, the basis expansion of $x \in X$ looks as

$$x = \sum_{k=1}^{\infty} x_k^*(x) x_k$$

This resembles the Fourier series with respect to orthogonal bases in a Hilbert space, except now we discuss this in general Banach spaces.

Corollary 3.3.19 (Biorthogonal functionals). *The biorthogonal functionals (x_k^*) of a Schauder basis (x_k) are uniformly bounded:*

$$\sup_{k\in\mathbb{N}} \|x_k^*\| \, \|x_k\| < \infty.$$

[4]For an orthonormal basis (x_k) in a Hilbert space X, this statement means that X is isomorphic to $E = \ell_2$. We have proved this in Section 1.6.7.

[5]To recall the notion of projection in a Banach space, see Definition 2.4.21.

PROOF. We estimate the k-th term of basis expansion as

$$\|x_k^*(x)x_k\| = \|S_k(x) - S_{k-1}(x)\| \leqslant \|S_k(x)\| + \|S_{k-1}(x)\| \leqslant 2M\|x\|$$

where M is the basis constant. On the other hand, $\|x_k^*(x)x_k\| = |x_k^*(x)|\,\|x_k\|$. This clearly completes the proof. □

EXAMPLE 3.3.20. The biorthogonal functionals for an orthonormal basis in a Hilbert space coincide with the basis vectors, i.e. $x_k^* = x_k$. The biorthogonal functionals for the canonical (coordinate) basis of ℓ_p, $1 \leqslant p < \infty$ are the coordinate vectors in ℓ_q, where p and q are adjoint indices.

REMARK 3.3.21 (Basis problem). All classical separable Banach spaces are known to have Schauder bases. However, there exist separable Banach spaces that have no Schauder bases. The first constructions of such spaces was proposed by Enflo [**3**] as a negative solution to Banach's basis problem.

3.3.5. Additional Exercises.

EXERCISE 3.3.22. **[Hamel bases are uncountable] Show that a Hamel basis of an infinite-dimensional Banach space X is always uncountable. (This is the reason why Hamel bases are rarely used in practice). Hint: first prove that every finite-dimensional subspace of X is nowhere dense; then use Baire category theorem.**

3.4. Compact sets in Banach spaces

Lec. 27: 11/10

Compactness is a useful substitute of finite dimensionality. We assume that the reader is familiar with the notion of compactness from a basic course in topology, but we will give a brief review.

3.4.1. Review of compactness. By definition, a subset A of a topological space X is compact if every open cover of A contains a finite sub-cover. Precisely, if $A \subseteq \bigcup_\alpha U_\alpha$ for some collection of open sets U_α, then $A \subseteq \bigcup_{k=1}^n U_{\alpha_k}$ for some sub-collection. Some basic properties of compact sets are:

(i) compact subsets of a Hausdorff space are closed;
(ii) closed subsets of compact sets are compact;
(iii) images of compact sets under continuous maps are compact;
(iv) continuous functions on compact sets are uniformly continuous and they attain their maxima and minima.

A set A is called *precompact* if its closure $\overline{A}$ is compact.

In metric spaces X, a useful description of compact sets A can be given in terms of ε-nets. Recall that a subset $\mathcal{N}$ is an *ε-net* of A for some $\varepsilon > 0$ if for every $x \in A$ there exists $y \in \mathcal{N}$ such that $d(x, y) \leqslant \varepsilon$. Equivalently, $\mathcal{N}$ is an ε-net of A if A can be covered by balls of radius ε centered at points in $\mathcal{N}$.

THEOREM 3.4.1 (Compactness in metric spaces). *For a subset A of a metric space X, the following are equivalent:*

(i) A is precompact;
(ii) Every sequence (x_n) in A has a Cauchy subsequence (which furthermore converges in X if X is complete);

(iii) For every $\varepsilon > 0$, there exists a finite ε-net of A.

As a consequence, precompact sets in metric spaces are always bounded. The converse is true in all finite dimensional normed spaces:

THEOREM 3.4.2 (Heine-Borel). *A subset A of finite dimensional normed space X is precompact if and only if A is bounded.*

Actually, the classical Heine-Borel theorem is the statement for the specific space $X = \ell_2^n$. But as we know, all finite dimensional normed spaces X are isomorphic to ℓ_2^n (Theorem 3.1.13), so the general result is also true.

3.4.2. Compactness in infinite-dimensional normed spaces. In infinite-dimensional normed spaces, Heine-Borel theorem fails. For example, an orthonormal basis (x_k) of ℓ_2 is a bounded set but it is not precompact, because it does not have a convergent subsequence (as $\|x_k - k_j\| = \sqrt{2}$ for $k \neq j$).

Compact sets are almost finite dimensional. This heuristics, which is made rigorous in the following result, underlies many arguments in analysis:

LEMMA 3.4.3 (Approximation by finite dimensional subspaces). *A subset A of a normed space X is precompact if and only if A is bounded and, for every $\varepsilon > 0$, there exists a finite dimensional subspace Y of X which forms an ε-net of A.*

PROOF. *Necessity.* Let A be precompact. Choose a finite ε-net $\mathcal{N}_\varepsilon$ of A; then the subspace $Y := \mathrm{Span}(\mathcal{N}_\varepsilon)$ is finite-dimensional and forms an ε-net of A.

Sufficiency. Since A is bounded, $A \subseteq rB_X$ for some finite radius r. Since Y is an ε-net of A, it follows that $(r+\varepsilon)B_Y$ is also an ε-net of A. (Check!) Further, since Y is finite-dimensional, the set $(r+\varepsilon)B_Y$ is precompact by Heine-Borel theorem. So we have found a precompact ε-net of A. Therefore A itself is precompact. (Why?) □

By Heine-Borel theorem, the unit ball B_X of a finite-dimensional normed space X is compact. This never holds in infinite dimensions:

THEOREM 3.4.4 (F. Riesz). *The unit ball B_X of an infinite-dimensional normed space X is never compact.*

PROOF. Suppose B_X is compact. By approximation Lemma 3.4.3, we can find a finite dimensional subspace Y of X which forms an $\frac{1}{2}$-net of B_X, i.e.

$$\mathrm{dist}(x, Y) \leqslant \frac{1}{2} \quad \text{for all } x \in B_X. \tag{3.14}$$

Since X is infinite dimensional and Y is finite dimensional, the quotient space X/Y is nonzero. So we can find a coset $[x] \in X/Y$ with $\|[x]\| = 0.9$. Since by definition $\|[x]\| = \inf_{x \in [x]} \|x\|$, we can further choose a representative $x \in [x]$ such that $\|x\| \leqslant 1$. Summarizing, $\mathrm{dist}(x, Y) = \|[x]\| = 0.9$ and $x \in B_X$. This contradicts (3.14) and completes the proof. □

The next useful result states that pointwise convergence of operators implies uniform convergence on compacta. We say that a sequence of operators $T_n \in L(X, Y)$ between normed spaces X and Y *converges pointwise* to some $T \in L(X, Y)$ if

$$T_n x \to Tx \quad \text{for all } x \in X.$$

In contrast, we say that T_n *converges uniformly* to T on a subset $A \subseteq X$ if

$$\|T_n x - Tx\| \leqslant \varepsilon_n \to 0 \quad \text{for all } x \in A,$$

where $\varepsilon_n \geqslant 0$ is some sequence of numbers (that does not depend on x).

THEOREM 3.4.5 (Convergence on compacta). *Let X, Y be Banach spaces, and $T_n, T \in L(X,Y)$. Assume that $T_n \to T$ pointwise. Then $T_n \to T$ uniformly on every precompact subset $A \subset X$.*

PROOF. Since the sequence (T_n) is pointwise convergent, it is pointwise bounded. Therefore, by the principle of uniform boundedness, (T_n) is uniformly bounded, i.e. there exists a number M such that

$$\|T_n\| \leqslant M \quad \text{for all } n.$$

Let $\varepsilon > 0$ be arbitrary, and choose a finite ε-net $\mathcal{N}_\varepsilon$ of A. Since the set $\mathcal{N}_\varepsilon$ is finite, $T_n \to T$ uniformly on $\mathcal{N}_\varepsilon$. (Why?) So there exists a number N such that

$$\|T_n y - Ty\| \leqslant \varepsilon \quad \text{for all } n \geqslant N \text{ and } y \in \mathcal{N}_\varepsilon.$$

For every $x \in A$, we choose $y \in \mathcal{N}_\varepsilon$ with $\|x - y\| \leqslant \varepsilon$. Then

$$\begin{aligned}\|T_n x - Tx\| &= \|(T_n - T)x\| \leqslant \|(T_n - T)y\| + \|(T_n - T)(x - y)\| \\ &\leqslant \varepsilon + (\|T_n\| + \|T\|)\,\|x - y\| \leqslant \varepsilon + 2M\varepsilon.\end{aligned}$$

Summarizing, we have proved that for every $\varepsilon > 0$ there exists N such that $\|T_n x - Tx\| \leqslant (1 + 2M)\varepsilon$ for all $n \geqslant N$ and $x \in A$. This means that $T_n \to T$ uniformly 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).

THEOREM 3.4.6 (Compactness in spaces with basis). *Let X be a Banach space with a Schauder basis (x_k). A subset $A \subseteq X$ is precompact if and only if A is bounded and the basis expansions of vectors $x \in A$ converge uniformly, i.e.*

$$\Big\| \sum_{k>n} x_k^*(x) x_k \Big\| = \|x - S_n x\| \leqslant \varepsilon_n \to 0 \quad \textit{for all } x \in A,$$

where $\varepsilon_n \geqslant 0$ is some sequence of numbers (that does not depend on x).

PROOF. *Necessity.* Since (x_k) is a Schauder basis, this means that the projections $S_n \to I$ pointwise (where I is the identity operator in X). By Theorem 3.4.5, $S_n \to I$ uniformly on A, as required.

Sufficiency. Assume that $\|x - S_n x\| \leqslant \varepsilon$ for all $x \in A$. Since $S_n x \in \operatorname{Im}(S_n)$, this implies that $\operatorname{Im}(S_n)$ forms an ε-net of A. But $\operatorname{Im}(S_n)$ has finite dimension n. We conclude that A is precompact by Lemma 3.4.3. □

COROLLARY 3.4.7 (Compactness in ℓ_p). *A subset $A \subseteq \ell_p$, $p \in [1, \infty)$ is precompact if and only if A is bounded and has uniformly decaying tails, i.e.*

$$\sum_{k>n} |a_k|^p \leqslant \varepsilon_n \to 0 \quad \textit{for all } a = (a_k) \in A,$$

where $\varepsilon_n \geqslant 0$ is some sequence of numbers (that does not depend on a).

PROOF. The claim follows immediately by applying Theorem 3.4.6 for the canonical basis of ℓ_p. □

EXAMPLE 3.4.8. The *Hilbert cube* is the following subset of ℓ_2:

$$A = \{(a_k) \in \ell_2 : |a_k| \leqslant \frac{1}{k} \text{ for all } k \in \mathbb{N}\}.$$

By Corollary 3.4.7, the Hilbert cube is compact.

EXERCISE 3.4.9. **[Compactness in c_0] Prove that a subset $A \subseteq c_0$ is precompact if and only if there exists a vector $b \in c_0$ that majorizes all vectors $a \in A$, i.e.**

$$|a_k| \leqslant b_k \quad \textbf{for all } k \in \mathbb{N}.$$

(Hint: use Theorem 3.4.6).

Finally, we mention without proof compactness criteria in two other spaces, $C[a,b]$ and L_1.

THEOREM 3.4.10 (Arzela-Ascoli). *A subset $A \subseteq C[a,b]$ is precompact if and only if A is bounded and equicontinuous, i.e. for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$|s-t| \leqslant \delta \quad \textit{implies} \quad |f(s) - f(t)| \leqslant \varepsilon \quad \textit{for all } f \in A.$$

EXAMPLE 3.4.11. Arzela-Ascoli theorem implies that the set of differentiable functions f with $\|f'\|_\infty \leqslant 1$ is compact in $C[0,1]$.

THEOREM 3.4.12 (Compactness in L_1). *A subset $A \subseteq L_1[0,1]$ is precompact if and only if A is bounded and* uniformly bounded on average*, i.e. for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$|\tau| \leqslant \delta \quad \textit{implies} \quad \int_0^1 |f(t+\tau) - f(t)| \leqslant \varepsilon \quad \textit{for all } f \in A.$$

3.4.4. Additional Exercises.

EXERCISE 3.4.13. **[Stability of compactness] Prove that compactness in normed spaces is stable under linear operations:**

(i) If A, B are precompact sets in a normed space, then Minkowski sum $A + B$ is precompact;
(ii) If A is a precompact subset of X and $T \in L(X,Y)$ then $T(A)$ is a precompact set in Y.

EXERCISE 3.4.14. **[Convex hull of compact sets]**

(i) Show that the convex hull of a precompact set in a normed space is a precompact set.
(ii) Show that the closure of the convex hull of a compact set in a normed space is compact.
(iii) Construct an example showing that the convex hull of a compact set in a normed space does not need to be compact.

3.5. Weak topology

Lec. 28: 11/12

Every normed space X is a metric space, with the metric given by $d(x, y) = \|x - y\|$, $x, y \in X$. This topology on X is called *strong topology* . Thus a sequence $x_n \to x$ converges (strongly) in X if

$$\|x_n - x\| \to 0. \tag{3.15}$$

In addition to the strong topology, X carries a different topology called *weak topology.* We are going to study the weak topology in this section.

3.5.1. Weak convergence.

DEFINITION 3.5.1 (Weak convergence). A sequence (x_k) in a normed space X *converges weakly* to a vector $x \in X$ if

$$f(x_k) \to f(x) \quad \text{for every } f \in X^*.$$

The weak convergence is denoted $x_k \xrightarrow{\text{w}} x$.

Strong convergence (3.15) clearly implies weak convergence. The converse is generally not true:

EXAMPLE 3.5.2. An orthonormal system (x_k) in a Hilbert space X converges weakly to zero, and it clearly does not converge strongly. Indeed, by Riesz representation theorem the weak convergence to zero is equivalent to

$$\langle x_k, x \rangle \to 0 \quad \text{for every } x \in X.$$

This indeed follows from Bessel's inequality

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

Even though weak convergence is generally strictly weaker than strong convergence, there are several useful ties between weak and strong properties. Weak convergence clearly implies weak boundedness, which in turn implies strong boundedness by a consequence to the principle of uniform boundedness (Corollary 3.3.5):

PROPOSITION 3.5.3. *Weakly convergent sequences in Banach spaces are bounded.*

□

Moreover, we have a good control of the weak limit, given in the next two results.

PROPOSITION 3.5.4. *If $x_n \xrightarrow{w} x$ in a normed space then $\|x\| \leqslant \liminf_n \|x_n\|$.*

PROOF. Let $f \in X^*$ be a supporting functional of x, i.e. $\|f\| = 1$, $f(x) = \|x\|$. Then $f(x_n) \leqslant \|x_n\|$ for all n. Taking $\liminf$ of both sides, we conclude that

$$\liminf_n \|x_n\| \geqslant \liminf_n f(x_n) = f(x) = \|x\|$$

as required. □

LEMMA 3.5.5 (Mazur). *Let $x_k \xrightarrow{w} x$ in a normed space, then*[6] $x \in \overline{\operatorname{conv}(x_k)}$.

[6] Recall that $\operatorname{conv}(A)$ is the smallest convex set containing A, see Exercise 1.2.24.

PROOF. Suppose $x \notin K := \overline{\mathrm{conv}(x_k)}$. Using a separation theorem (Corollary refclosed convex separation), we can separate the closed convex set K from the point $\{x\}$. Namely, there exists a functional $f \in X^*$ such that

$$\sup_{y \in K} f(y) < f(x).$$

Since $x_k \in K$, this implies that

$$\sup_k f(x_k) < f(x),$$

which contradicts weak convergence. □

3.5.2. Criteria of weak convergence. Some known criteria of weak convergence in classical normed spaces rely on the following tool.

LEMMA 3.5.6 (Testing weak convergence on a dense set). *Let X be a normed space and $A \subseteq X^*$ be a dense set. Then $x_k \xrightarrow{w} x$ in X if and only if (x_k) is bounded and*

$$f(x_k) \to f(x) \quad \textit{for every } f \in A.$$

PROOF. *Nesessity* follows by Proposition 3.5.3.

Sufficiency will be proved by a standard approximation argument. Consider arbitrary $g \in X^*$ and $\varepsilon > 0$. By density, we can choose $f \in A$ such that $\|g - f\| \leqslant \varepsilon$. Then

$$\begin{aligned} \limsup_k |g(x_k - x)| &\leqslant \limsup_k |f(x_k - x)| + \limsup_n |(g - f)(x_k - x)| \\ &\leqslant 0 + \|g - f\| \limsup_k (\|x_k\| + \|x\|) \leqslant M\varepsilon \end{aligned}$$

where $M = \sup_k \|x_k\| + \|x\| < \infty$ by the boundedness assumption. Since $\varepsilon > 0$ is arbitrary, we conclude that $\limsup_k |g(x_k - x)| = 0$. Hence $g(x_k) \to g(x)$ as required. □

THEOREM 3.5.7 (Weak convergence in c_0 and ℓ_p). *Let $X = c_0$ or $X = \ell_p$, $p \in (1, \infty)$. Then $x_k \xrightarrow{w} x$ in X if and only if the sequence (x_k) is bounded and converges to x* pointwise, *i.e.*

$$x_k(i) \to x(i) \quad \textit{for every } i \in \mathbb{N}.$$

PROOF. *Necessity.* If $x_n \xrightarrow{\mathrm{w}} x$ then by applying coordinate functionals $e_i^* \in X^*$ (i.e. those acting as $e_i^*(x) = x(i)$) we see that $x_k(i) \to x(i)$ as required.

Sufficiency. We are given that (x_k) is bounded and that $f(x_k) \to f(x)$ for every coordinate functional $f = e_i^*$. By linearity, we get $f(x_k) \to f(x)$ for every $f \in \mathrm{Span}(e_i^*)_{i=1}^\infty$.

On the other hand, the representation theorems (Corollary 2.2.6 and Exercise 2.2.7) state that $X^* = \ell_1$ if $X = c_0$ and $X^* = \ell_{p'}$ if $X = \ell_p$. The functionals $e_i^* \in X^*$ get identified with the coordinate vectors $(0, \ldots, 0, 1, 0, \ldots)$, which shows that $\mathrm{Span}(e_i^*)_{i=1}^\infty$ is dense in X^*. (Why?)

The proof is finished by applying Lemma 3.5.6 to $A = \mathrm{Span}(e_i^*)_{i=1}^\infty$. □

EXERCISE 3.5.8. **Consider the sequence $x_k = (1, \ldots, 1, 0, 0, \ldots)$ (with k ones) in ℓ_∞. Use Mazur's lemma to show that x_k does not converge weakly. Deduce that the criterion of weak convergence in ℓ_p, $p \in (1, \infty)$, given in Theorem 3.5.7 fails for ℓ_∞. (There is no useful criterion of weak convergence in ℓ_∞.**

EXERCISE 3.5.9. **State and prove a similar criterion of weak convergence in spaces with Schauder basis.**

A similar criterion of weak convergence holds in spaces of continuous functions.

THEOREM 3.5.10 (Weak convergence in $C(K)$). *Let K be a compact topological space. Then $x_k \xrightarrow{w} x$ in $C(K)$ if and only if the sequence of functions $x_k(t)$ is uniformly bounded*[7] *and converges to $x(t)$* pointwise, *i.e.*

$$x_k(t) \to x(t) \quad \textit{for every } t \in K.$$

PROOF. *Necessity.* Boundedness of (x_k) follows from weak convergence as before. Pointwise convergence follows by applying point evaluation functionals $\delta_t \in C(K)^*$ (acting as $\delta_t(x) = x(t)$).

Sufficiency. We need to show that $f(x_k) \to f(x)$ for all $f \in C(K)^*$. By representation Theorem 2.2.8, this is equivalent to claiming that

$$\int_K x_n \, d\mu \to \int_K x \, d\mu \tag{3.16}$$

for every Borel regular signed measure μ. On the other hand, our assumptions are that the sequence of functions $x_n(t)$ is uniformly bounded and it converges to $x(t)$ pointwise. The Lebesgue dominated convergence theorem implies (3.16). □

A similar criterion of weak convergence holds in L_p spaces. However, it does not make sense to consider the values of functions $x \in L_p$ in individual points. Instead, we shall consider integrals of $x(t)$ over short intervals.

THEOREM 3.5.11 (Weak convergence in L_p). *Let $p \in (1, \infty)$. A sequence $x_k \xrightarrow{w} x$ in $L_p[0,1]$ if and only if the sequence (x_k) is bounded in L_p and*

$$\int_a^b x_k(t)\, dt \to \int_a^b x_k(t)\, dt \quad \textit{for every interval } [a,b] \subseteq [0,1].$$

PROOF. One notices that the set of characteristic functions $\mathbf{1}_{[a,b]}(t)$ for $[a,b] \subseteq [0,1]$ spans the set of step functions, which is dense in $(L_p)^* = L_{p'}$. (Why?) The argument is finished similarly to Theorem 3.5.7. □

REMARK 3.5.12. The same criterion holds for $L_p(\mathbb{R})$. (Why?)

EXAMPLE 3.5.13 (Sliding bumps). A good example of weakly convergent but strongly divergent sequence of functions is formed by a sliding bump in $L_p(\mathbb{R})$, $p \in (1, \infty)$. Consider a function $x \in L_p(\mathbb{R})$ with compact support. Then the sequence $x_k(t) = x(t-k)$ converges weakly to zero by Theorem 3.5.11.

REMARK 3.5.14 (Weak convergence in ℓ_1). The weak convergence in ℓ_1 is unusual. (Note that this space is not covered by Theorem 3.5.7.) The weak and strong convergence in ℓ_1 are equivalent. This result is due to I. Schur, and is called *Schur property* of ℓ_1.

EXERCISE 3.5.15. **[Weak convergence in finite dimensional spaces] Show that all finite-dimensional normed spaces X have Schur property, so the weak and strong convergence in X coincide.**

Lec. 29: 11/15

3.5.3. Weak topology. Now we broaden the picture and study the *weak topology* on X which defines weak convergence. This way, in addition to weak convergence, we could be able to study other weak properties, such as weak boundedness, weak compactness and so on.

DEFINITION 3.5.16 (Weak topology). The *weak topology* on a normed space X is defined as the weakest topology in which all maps $f \in X^*$ (i.e. $f : X \to \mathbb{R}$) are continuous.

Equivalently, the base of the weak topology is given by the *cylinders*, which are the sets of the form

$$\{x \in X : \ |f_k(x - x_0)| < \varepsilon, \, k = 1, \ldots, N\}$$

where $x_0 \in X$, $f_k \in X^*$, $\varepsilon > 0$, and $n \in \mathbb{N}$. So, these cylinders form a local base of weak topology at point x_0.

To distinguish the two natural topologies on X, the norm topology is sometimes called *strong*.

EXERCISE 3.5.17. **Check the equivalence of the two ways to define the weak topology in the definition above.**

REMARK 3.5.18. The cylinders are clearly open in the strong topology. Hence the weak topology is weaker than the strong topology.

REMARK 3.5.19. In infinite-dimensional spaces X, the cylinders are rather large as they contain subspaces of finite codimension $\{x \in X : \ f_k(x - x_0) = 0, \, k = 1, \ldots, N\}$. This shows that in infinite dimensions, *weakly open sets are unbounded.*

EXERCISE 3.5.20. **Prove that in an infinite dimensional normed space X, weak topology is strictly weaker than the strong topology. Why does not this contradict Schur property of $X = \ell_1$ mentioned in Remark 3.5.14?**

Nevertheless, some weak and strong properties are equivalent. For example, *weak boundedness and strong boundedness are equivalent.* This follows from the principle of uniform boundedness, see Corollary 3.3.5. (Check that what we called "weak boundedness" is indeed boundedness in the weak topology).

Also, weak closedness and strong closedness are equivalent for convex sets:

PROPOSITION 3.5.21 (Weak closedness). *Let K be a convex set in a normed space X. Then K is weakly closed if and only if K is (strongly) closed.*

PROOF. *Necessity* is trivial.

Sufficiency. Assume K is closed and convex. By Corollary intersection of half-spaces to Hahn-Banach theorem, K is the intersection of the closed half-spaces that contain K. Each closed half-space has the form

$$A_{f,a} = \{x \in X : \ f(x) \leqslant a\}$$

for some $f \in X^*$ and $a \in \mathbb{R}$. Hence $A_{f,a}$ is weakly closed.[8] The intersection K of the closed half-spaces is therefore automatically weakly closed. □

[7] That is, $\sup_k \|x_k\|_\infty < \infty$.

[8] Indeed, $A_{f,a}$ is the preimage of the closed subset $(-\infty, a]$ under the function $f : X \to \mathbb{R}$, which is continuous by the definition of weak topology.

Remark 3.5.22. Convexity assumption is critical in Proposition 3.5.21. Otherwise the result would claim that the weak and strong topologies are equivalent, which is false.

3.6. Weak* topology. Banach-Alauglu's theorem

On X^*, there are two natural weaker topologies. The weak topology that we already considered makes all functionals in X^{**} continuous functions on X^*. The other topology, called *weak* topology*, is only concerned with continuity of functionals that come from $X \subset X^{**}$.

3.6.1. Weak* convergence.

Definition 3.6.1 (Weak* convergence). Let X be a normed space. A sequence of functionals (f_k) in X^* *weak* converges* to a functional $f \in X^*$ if

$$f_k(x) \to f(x) \quad \text{for every } x \in X.$$

The weak convergence is denoted $f_k \xrightarrow{\text{w*}} f$.

In other words, the weak* convergence of functionals is *pointwise* convergence.

While weak convergence of functionals $f_k \in X^*$ is tested on all functionals from X^{**}, weak* convergence of f_k is tested on the subset $X \subseteq X^{**}$. Therefore, *weak convergence implies weak* convergence* in X^*. Of course, for reflexive spaces, weak and weak* convergence coincide.

Example 3.6.2 (Weak convergence of measures). In probability theory, one says that a sequence of regular Borel measures μ_n on $\mathbb{R}$ converges weakly to a Borel regular measure μ if

$$\int f \, d\mu_n \to \int f \, d\mu \quad \text{for every } f \in C(\mathbb{R}).$$

Assume that the measures μ_n and μ are compactly supported, say on an interval $[a,b]$. By the representation theorem for $(C[a,b])^*$, Theorem 2.2.8, this convergence is nothing different from

$$\mu_n \xrightarrow{\text{w*}} \mu \quad \text{in } (C[a,b])^*.$$

Summarizing, the weak convergence of measures in probability theory is actually the weak* convergence of measures acting as linear functionals on $C[a,b]$.

Example 3.6.3 (Dirac delta function). Recall that we understand Dirac delta function $\delta(t)$ as the point evaluation functional at zero, see Example 2.1.3. Equivalently, Dirac delta function may be identified with the probability measure on $\mathbb{R}$ with the only atom at the orgigin. Therefore *Dirac delta function is the weak limit of uniform measures on* $[-\frac{1}{n}, \frac{1}{n}]$ *as* $n \to \infty$.

This gives a natural way to approximate Dirac delta function $\delta(t)$ (which does not exists as a function on $\mathbb{R}$) by genuine functions $\delta_n(t)$, which are the probability distribution functions of the uniform measures on $[-\frac{1}{n}, \frac{1}{n}]$, see the picture.

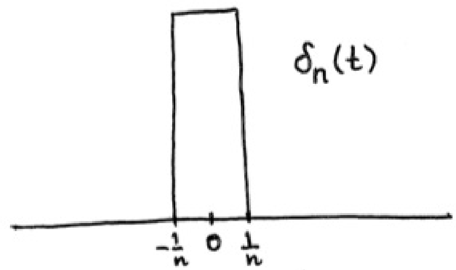

FIGURE 3.3. Approximation of Dirac delta function $\delta(t)$ by functions $\delta_n(t)$

3.6.2. Weak* topology. Banach-Alaoglu's theorem. Similarly to the Definition 3.5.16, we will consider the weak* topology on X^*.

DEFINITION 3.6.4 (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 points $x \in X$.

Equivalently, the base of the weak topology is given by the *cylinders*, which are the sets of the form

$$\{f \in X^* : \ |(f - f_0)(x_k)| < \varepsilon, \, k = 1, \ldots, N\}$$

where $f_0 \in X^*$, $x_k \in X$, $\varepsilon > 0$, and $n \in \mathbb{N}$. So, these cylinders form a local base of weak topology at f_0.

Looking at the first part of this definition and recalling that $x \in X \subseteq X^{**}$, we see that *weak* topology is weaker than the weak topology* on X^*. However, for reflexive spaces the two topologies are of course equivalent.

The main result on weak* topology is Alaoglu's theorem. It allows one to bring back to life compactness arguments in infinite-dimensional normed spaces X, even though the unit ball of such X is always not compact.

THEOREM 3.6.5 (Banach-Alaoglu). *For every normed space X, the closed unit ball B_{X^*} is weak* compact.*

The proof will be based on *Tychonoff's theorem* that states that the product of any collection of compact topological spaces is compact. Let us briefly recall this result.

Consider a collection of $(X_\gamma)_{\gamma \in \Gamma}$ of any number (countable or uncountable) 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

$$\Big\{ \prod_{\gamma \in \Gamma} A_\gamma : \ A_\gamma \text{ is open in } X_\gamma; \ \text{ all but finitely many of } A_\gamma \text{ equal } X_\gamma \Big\}.$$

Tychonoff's theorem states that if each X_γ is compact then $\prod_{\gamma \in \Gamma} X_\gamma$ is compact in the product topology.

PROOF OF BANACH-ALAOGLU'S THEOREM. We shall embed B_{X^*} into the product space of intervals

$$K := \prod_{x \in X} [-\|x\|, \|x\|] = \big\{ f : X \to \mathbb{R} : \ |f(x)| \leqslant \|x\| \text{ for all } x \in X \big\}.$$

equipped with the product topology.

This is the weakest topology in which the point evaluation maps $f \mapsto f(x)$ from K to $\mathbb{R}$ are continuous for all $x \in X$. (Check!) We identify a functional $f \in B_{X^*}$ with the element of the product space $(f(x))_{x \in X} \in K$. With this identification, the weak* topology on B_{X^*} coincides with the product topology on K. (Why?) Therefore, this identification is a homeomorphic embedding of B_{X^*} into K.

It remains to check that B_{X^*} is a weak* closed subset of K; the proof will then be finished by Tychonofff's theorem. This is simple. Indeed, note that B_{X^*} consists of the *linear* functions in K. So we can represent

$$B_{X^*} = \bigcap_{x,y \in X,\ a,b \in \mathbb{R}} B_{x,y,a,b}, \quad \text{where } B_{x,y,a,b} = \big\{ f \in K :\ f(ax+by) = af(x)+bf(y) \big\}.$$

Each set $B_{x,y,a,b}$ is the preimage of the weak* closed set $\{0\}$ under the map $f \mapsto f(ax+by)-af(x)-bf(y)$ which, as we know, is continuous in the product topology.[9] Therefore all sets $B_{x,y,a,b}$ are weak* closed, and so is their intersection B_{X^*}. This completes the proof. □

Lec. 30: 11/17

3.6.3. Universality of space $C(K)$. As an application of Banach-Alaoglu's theorem, we will show that the space of continuous functions $C(K)$ is *universal* in the sense that it contains every Banach space X as a subspace.[10]

THEOREM 3.6.6 (Universality of $C(K)$). *Every Banach space X can be isometrically embedded into $C(K)$ for some compact topological space K.*

PROOF. Let $K := B_{X^*}$ equipped with weak* topology. By Banach-Alaoglu's theorem, K is indeed compact. We define the embedding $X \to C(K)$ by associating every $x \in X$ the point evaluation function

$$x(f) := f(x), \quad f \in K. \tag{3.17}$$

Recall that the point evaluation function is indeed in $C(K)$ by the definition of weak* topology. The map defined by (3.17) is linear by construction. Finally, this map is an isometric embedding; indeed

$$\|x\|_{C(K)} = \max_{f \in K = B_{X^*}} |f(x)| = \|x\|_X,$$

where the last inequality uses a consequence of Hahn-Banach theorem, Corollary 2.3.10. □

EXERCISE 3.6.7. **[Universality of ℓ_∞] Show that ℓ_∞ is a universal space for all separable Banach spaces. In other words, show that every separable Banach space X isometrically embeds into ℓ_∞.**

Hint: Consider a dense subset $(x_k)_{k=1}^\infty$ of S_X, choose supporting functionals $f_k \in S_{X^*}$ of x_k, and define the embedding $X \to \ell_\infty$ by $x \mapsto (f_k(x))_{k=1}^\infty$.

[9] Recall that the point evaluation maps are continuous in the product topology.

[10] A little disclaimer is that the compact topological space K may depend on X; otherwise the result is false for spaces X of too large cardinality.

CHAPTER 4

Compact operators. Elements of 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 (in the same way as compact sets are "almost" finite dimensional). So compact operators do 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.

4.1.1. Definition. Integral operators. Throughout this section, X, Y will denote normed spaces.

DEFINITION 4.1.1 (Compact operators). A linear operator $T : X \to Y$ is called *compact* if it maps bounded sets in X to precompact sets in Y. The set of compact operators is denoted $K(X,Y)$.

EXERCISE 4.1.2. **Show that T is compact if and only if it maps B_X to a precompact set in Y.**

Since precompact sets are bounded, compact operators are always bounded, i.e. $K(X,Y) \subseteq L(X,Y)$.

EXAMPLE 4.1.3. Every *finite rank* operator[1] $T \in L(X,Y)$ is compact. Indeed, $T(B_X)$ 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.

The next example is one of the main motivation to study compact operators.

PROPOSITION 4.1.4 (Integral operators are compact). *Consider the integral operator $T : C[0,1] \to C[0,1]$ defined as*

$$(Tf)(t) = \int_0^1 k(t,s) f(s)\, ds$$

with kernel $k(t,s) \in C([0,1]^2)$. Then T is a compact operator.

PROOF. We need to show that $K := T(B_{C[0,1]})$ is a precompact subset of $C[0,1]$, see Exercise 4.1.2. By Arzela-Ascoli Theorem 3.4.10, this would follow from (uniform) boundedness and equicontinuoity of the set K.

The (uniform) boundedness of K follows from the boundedness of T. (Why?) To prove equicontinuity, we let $\varepsilon > 0$ and choose $\delta > 0$ such that

$$|t_1 - t_2| \leqslant \delta \quad \text{implies} \quad |k(t_1,s) - f(t_2,s)| \leqslant \varepsilon \quad \text{for all } s \in [0,1].$$

[1] T is a finite rank operator if $\dim \operatorname{Im} T < \infty$.

(We can do this by continuity of the kernel $k(t,s)$.) Now, for every $f \in B_{C[0,1]}$, we obtain by triangle inequality that

$$\big|(Tf)(t_1) - (Tf)(t_2)\big| \leqslant \int_0^1 |k(t_1,s) - f(t_2,s)|\,|f(s)|\,ds \leqslant \varepsilon$$

as $|f(s)| \leqslant 1$ for all s. This shows that the set K is equicontinuous, and therefore precompact. □

Exercise 4.1.5. **Show that Volterra operator (2.12) is compact on $C[0,1]$, even though its kernel is discontinuous. See Exercise 4.1.19 for a more general result.**

4.1.2. Basic properties of compact operators.

Proposition 4.1.6 (Properties of $K(X,Y)$). *(i) The set of compact operators $K(X,Y)$ is a closed linear subspace of $L(X,Y)$.*

(ii) $K(X,Y)$ is an operator ideal. *This means that if $T \in K(X,Y)$ then the compositions ST and TS are both compact for every bounded linear operator S.*

Proof. (i) *Linearity* follows from the observation that the Minkowski sum of two precompact sets is precompact (see exercise below).

Closedness. Consider a sequence $T_n \in K(X,Y)$ such that $T_n \to T$ in $L(X,Y)$; we want to prove that $T \in K(X,Y)$. Let $\varepsilon > 0$ and choose $n \in \mathbb{N}$ such that $\|T_n - T\| \leqslant \varepsilon$. This means that

$$\|T_n x - Tx\| \leqslant \varepsilon \quad \text{for every } x \in B_X.$$

This shows that $T_n(B_X)$ is a precompact ε-net of $T(B_X)$. Since ε is arbitrary, $T(B_X)$ is itself precompact. (Why?)

(ii) is straightforward and is left as an exercise. □

Exercise 4.1.7. **Prove that Minkowski sum $A + B$ of two precompact subsets A, B of a normed space is a precompact set.**

Corollary 4.1.8 (Isomorphisms are not compact). *Let X be an infinite dimensional normed space. Then the identity operator on X is not compact. More generally, any isomorphism $T : X \to Y$ is not compact.*

Proof. For the identity operator on X, the result follows from F. Riesz's Theorem 3.4.4 on non-compactness of B_X. As for the general statement, if an isomorphism $T : X \to Y$ were compact then the identity operator $T^{-1}T$ would also be compact by Proposition 4.1.6, which would be a contradiction. □

As we know, finite rank operators are compact (Example 4.1.3). More generally, since $K(X,Y)$ is closed, it follows that any operator that can be approximated by finite rank operators is also compact:

Corollary 4.1.9 (Almost finite rank operators are compact). *Suppose a linear operator $T : X \to Y$ can be approximated by finite rank operators $T_n \in L(X,Y)$, i.e.*

$$\|T_n - T\| \to 0 \quad \text{as } n \to \infty.$$

Then T is compact. □

EXERCISE 4.1.10. **Fix a sequence of real numbers $(\lambda_k)_{k=1}^\infty$, and define the linear operator $T : \ell_2 \to \ell_2$ by**

$$Tx = (\lambda_k x_k)_{k=1}^\infty.$$

For what multiplier sequences (λ_k) is the operator T (a) well defined? (b) bounded? (c) compact?

4.1.3. 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 Hilbert-Schmidt *integral* operators discussed in Section 2.4.4.

DEFINITION 4.1.11 (Hilbert-Schmidt operators). Let H be a separable Hilbert space, and let (x_k) be an orthonormal basis of H. A linear operator $T : H \to H$ is called a *Hilbert-Schmidt operator* if

$$\sum_{k=1}^\infty \|Tx_k\|^2 < \infty.$$

The quantity

$$\|T\|_{\mathrm{HS}} := \Big(\sum_{k=1}^\infty \|Tx_k\|^2\Big)^{1/2}$$

is called the *Hilbert-Schmidt norm* of T.

EXERCISE 4.1.12. **For operators in finite dimensional spaces $T : \mathbb{C}^m \to \mathbb{C}^n$, (equivalently, for $m \times n$ matrices), we already gave a definition of Hilbert-Schmidt norm in (1.12). Verify that the two definitions do agree.**

PROPOSITION 4.1.13. *The definition of Hilbert-Schmidt operator and of the Hilbert-Schmidt norm does not depend on the choice of an orthonormal basis of H.*

PROOF. Assume that $\sum_k \|Tx_k\|^2 < \infty$ for some orthonormal basis (x_k) of H. Using Parseval's identity twice, we obtain

$$\sum_k \|Tx_k\|^2 = \sum_{k,j} |\langle Tx_k, x_j\rangle|^2 = \sum_{k,j} |\langle x_k, T^* x_j\rangle|^2 = \sum_k \|T^* x_j\|^2. \tag{4.1}$$

Let (x'_k) be another orthonormal basis of H. Then a similar argument gives

$$\sum_j \|T^* x_j\|^2 = \sum_{j,k} |\langle x'_k, T^* x_j\rangle|^2 = \sum_{j,k} |\langle Tx'_k, x_j\rangle|^2 = \sum_k \|Tx'_k\|^2.$$

This completes the proof. □

REMARK 4.1.14. As a byproduct of the proof, we obtained in (4.1) that

$$\|T^*\|_{\mathrm{HS}} = \|T\|_{\mathrm{HS}}.$$

Lec. 31: 11/19

As we know from Proposition 2.4.7, linear operators on finite-dimensional spaces satisfy $\|T\| \leqslant \|T\|_{\mathrm{HS}}$. The same holds in general:

PROPOSITION 4.1.15. *Hilbert-Schmidt operators are bounded, and $\|T\| \leqslant \|T\|_{\mathrm{HS}}$.*

PROOF. Let T be a Hilbert-Schmidt operator on a Hilbert space H, and (x_k) be an orthonormal basis of H. It suffices to prove that the restriction of T on the dense subspace $\mathrm{Span}(x_k)$ of H is bounded and has norm at most $\|T\|_{\mathrm{HS}}$; the result would then follow by extension by continuity (Proposition 2.4.20).

So let $x \in \mathrm{Span}(x_k)$, which means that $x = \sum_k a_k x_k$ for some scalars a_k (finite sum). Then using triangle inequality and Cauchy-Schwarz inequality, we obtain

$$\|Tx\| = \Big\|\sum_k a_k T x_k\Big\| \leqslant \sum_k |a_k|\,\|Tx_k\| \leqslant \Big(\sum_k |a_k|^2\Big)^{1/2}\Big(\sum_k \|Tx_k\|^2\Big)^{1/2} = \|x\|\,\|T\|_{\mathrm{HS}}.$$

This completes the proof. □

An important example of Hilbert-Schmidt operators are integral operators on L_2, which we introduced in Section 2.4.4.

PROPOSITION 4.1.16 (Hilbert-Schmidt integral operators). *Consider the integral operator $T : L_2[0,1] \to L_2[0,1]$ defined as*

$$(Tf)(t) = \int_0^1 k(t,s) f(s)\,ds$$

with kernel $k(t,s) \in L_2([0,1]^2)$. Then T is a Hilbert-Schmidt operator, and

$$\|T\|_{\mathrm{HS}} = \|k\|_2.$$

PROOF. We will view the integral in the definition of T as the inner product of f with the kernel k. Specifically, consider the function $K_t(s) = k(t,s)$; then

$$(Tf)(t) = \langle K_t, f\rangle \quad \text{for every } t \in [0,1].$$

Let us fix some orthonormal basis (x_k) of $L_2[0,1]$. Then

$$\begin{aligned}
\|T\|_{\mathrm{HS}}^2 &= \sum_k \|Tx_k\|_2^2 = \sum_k \int_0^1 |(Tx_k)(t)|^2\,dt = \sum_k \int_0^1 |\langle K_t, x_k\rangle|^2\,dt \\
&= \int_0^1 \sum_k |\langle K_t, x_k\rangle|^2\,dt \quad \text{(by monotone convergence theorem)} \\
&= \int_0^1 \|K_t\|_2^2\,dt \quad \text{(by Parseval's identity)} \\
&= \|k\|_2^2 \quad \text{(by definition of } K_t \text{ and Fubini theorem).}
\end{aligned}$$

This completes the proof. □

4.1.4. Compactness of the adjoint operator. Recall the basic duality property for bounded linear operators: if $T \in L(X,Y)$ then $T^* \in L(Y^*,X^*)$ and $\|T^*\| = \|T\|$. A similar duality principle holds for compact operators:

THEOREM 4.1.17 (Schauder). *Let X and Y be Banach spaces. If $T \in K(X,Y)$ then $T^* \in K(Y^*,X^*)$.*

PROOF. Given $f \in Y^*$, we are seeking a bound on

$$\|T^* f\|_{X^*} = \sup_{x \in B_X} |(T^*f)(x)| = \sup_{x \in B_X} |f(Tx)| = \sup_{y \in K} |f(y)| \tag{4.2}$$

where $K := \overline{T(B_X)}$. (Taking the closure here is justified by continuity of f).

We shall interpret the identity (4.2) in topological terms. Indeed, we know that K is compact, and we need to prove that $G := T^*(B_{Y^*})$ is precompact in X^*. Let us embed the subset $G \subseteq X^*$ into $C(K)$ and use Arzela-Ascoli theorem. Namely, we define the embedding $U : G \to C(K)$ by

$$U(T^* f) := f|_K$$

(making some selection of f; it does not matter in which way). Then identity (4.2) implies that

$$\|T^* f\|_{X*} = \|f|_K\|_{C(K)} \quad \text{for every } f \in Y^*,$$

which shows that U is an isometric (thus homeomorphic) embedding.

Now, $U(G)$ is (uniformly) bounded in $C(K)$ as

$$\|T^* f\|_{X*} \leqslant \|T^*\| \, \|f\|_{Y*} \leqslant \|T\| \quad \text{for every } f \in B_{Y*}.$$

Moreover, $U(G)$ is equicontinuous. Indeed, for every $f \in B_{Y*}$ and for $y_1, y_2 \in K$ we have

$$\big|f|_K(y_1) - f|_K(y_2)\big| = |f(y_1 - y_2)| \leqslant \|f\|_{X*} \, \|y_1 - y_2\| \leqslant \|y_1 - y_2\|.$$

Arzela-Ascoli theorem completes the proof. □

Remark 4.1.18 (For future). Consider proving the reverse direction in Schauder's theorem. Also consider proving that compact operators map weak Cauchy sequences to strongly convergent.

4.1.5. Additional Exercises.

Exercise 4.1.19. **[Compactness of integral operators] Consider an integral operator T with kernel $k(t,s) : [0,1]^2 \to \mathbb{R}$ which satisfies the following:**

(i) for each $s \in [0,1]$, the function $k_s(t) = k(t,s)$ is integrable in t;
(ii) the map $s \mapsto k_s$ is a continuous map from $[0,1]$ to $L_1[0,1]$.

Show that the integral operator T is compact in $C[0,1]$.

4.2. Fredholm theory

Lec. 32: 11/22

Fredholm theory studies operators of the form "identity plus compact". They are conveniently put in the form $I - T$ where I is the identity operator on some Banach space X and $T \in K(X,X)$.

Fredholm theory is motivated by two applications. One is for solving linear equations $\lambda x - Tx = b$, and in particular integral equations (T being an integral operator). Another related application is in spectral theory, where the spectrum of T consists of numbers λ for which the operator $\lambda I - T$ is invertible. We will discuss both applications in detail later.

4.2.1. Closed image.

Theorem 4.2.1. *Let X be a Banach space and $T \in K(X,X)$. Then operator $I - T$ has closed image.*

PROOF. The argument relies on the characterization of injective operators with closed image, Proposition 3.1.12. So we consider the operator $A = I - T$ and its injectivization $\tilde{A} : X/\ker A \to X$. Since $\operatorname{Im} A = \operatorname{Im} \tilde{A}$, it suffices to show that $\tilde{A}$ is bounded below.

Suppose the contrary, that there exists cosets $[x_k]$ with $\|[x_k]\| = 1$ and such that

$$\tilde{A}[x_k] \to 0.$$

We can find representatives $x_k \in [x_k]$ with, say, $\|x_k\| \leqslant 2$ and such that

$$\operatorname{dist}(x_k, \ker A) = \|[x_k]\| = 1 \quad \text{but} \quad Ax_k = \tilde{A}[x_k] \to 0.$$

So we have $Ax_k = x_k - Tx_k \to 0$. By compactness of T, we can assume (passing to a subsequence if necessary) that $Tx_k \to z$ for some $z \in X$. It follows that $x_k \to z$. Therefore $Ax_k \to Az$; but we already know that $Ax_k \to 0$, thus $Az = 0$. So $z \in \ker A$. Furthermore, since $x_k \to z$ it follows that $\operatorname{dist}(x_k, \ker A) \to 0$. This contradiction completes the proof. □

4.2.2. Fredholm alternative. We now state and prove a partial case of the so-called Fredholm alternative. The more general theorem is given in Exercise 4.2.4.

THEOREM 4.2.2 (Fredholm alternative). *Let X be a Banach space and $T \in K(X, X)$. Then operator $A = I - T$ is injective if and only if A is surjective.*

PROOF. *Necessity.* Assume that A is injective but not surjective. Consider the subspaces of X

$$Y_n := \operatorname{Im}(A^n), \quad n = 0, 1, \ldots$$

Then

$$Y_0 \supset Y_1 \supset Y_2 \supset \cdots$$

is a chain of *proper* inclusions. Indeed, the first inclusion $X \supset \operatorname{Im}(A)$ is proper by assumption; the claim follows by induction. (Check this!)

Furthermore, Y_n are *closed* subspaces of X. Indeed, by Newton's binomial expansion we see that $A^n = (I - T)^n$ has the form $A = I - T_1$ for some compact operator T_1, so the claim follows from Theorem 4.2.1.

By Hahn-Banach theorem (see Exercise 2.3.33) we can find functionals

$$f_n \in Y_n^* \quad \text{such that} \quad \|f_n\| = 1, \quad f_n \in Y_{n+1}^\perp.$$

We can extend f_n, again by Hahn-Banach theorem, so that $f_n \in X^*$. We are going to show that the sequence $(T^* f_n)$ has no convergent subsequences. This will contradict the compactness of T^* and, by Schauder's Theorem 4.1.17, the compactness of T.

To this end, let us fix $n > m$ and compute the pairwise distances

$$d_{n,m} := \|T^* f_n - T^* f_m\| = \|T^*(f_n - f_m)\| = \|(I - T)^*(f_n - f_m) + f_m - f_n\|.$$

So

$$d_{n,m} \geqslant \sup_{x \in B_{Y_n}} |\langle T^* f_n - T^* f_m, x\rangle| = \sup_{x \in B_{Y_n}} |\langle f_n - f_m, (I - T)x\rangle + \langle f_m - f_n, x\rangle|.$$

Now, $(I - T)x = Ax \in Y_{n+1}$ while $f_n - f_m \in Y_{n+1}^\perp$ by construction, so $\langle f_n - f_m, (I - T)x\rangle = 0$. Further, $x \in Y_n$ while $f_m \in Y_n^\perp$, so $\langle f_m, x\rangle = 0$. Therefore

$$d_{n,m} \geqslant \sup_{x \in B_{Y_n}} |\langle f_n, x\rangle| = 1$$

by construction. It follows that the terms of the sequence $(T * f_n)$ are pairwise separated, so there can not be any convergent subsequence. This completes the proof of the necessity direction.

Sufficiency will follow from a duality argument. We use the relations

$$(\operatorname{Im} A)^{\perp} = \ker A^*, \quad (\ker A)^{\perp} \supseteq \operatorname{Im} A^* \tag{4.3}$$

which we proved in Proposition 2.4.32 and Exercise 2.4.34. So, assume that $A = I - T$ is surjective. Then $A^* = I - T^*$ is injective by (4.3). Since T^* is compact by Schauder's theorem, the first part of the proof gives that A^* is surjective. This implies that A is injective by (4.3). The proof is complete. □

REMARK 4.2.3 (Compactness is essential). Fredholm alternative does not hold for non-compact operators in general. For example, the right shift operator in ℓ_2 is injective but not surjective; the left shift operator in ℓ_2 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 $T \in K(X, X)$, $\lambda \in \mathbb{C}$, $b \in X$. One is interested in existence and uniqueness of solution. Theorem 4.2.2 states that exactly one 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; this solution is automatically unique.

This alternative is particularly useful for studying integral equations, since for the integral operator $(Tf)(t) = \int_0^1 k(t, s) f(s)\, ds$, 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

$$\lambda f(t) - \int_0^1 k(t, s) f(s)\, ds = b(t).$$

4.2.3. Additional exercises.

EXERCISE 4.2.4. **[General Fredholm alternative] Let X be a Banach space and $T \in K(X, X)$. Show that operator $A = I - T$ satisfies**

$$\dim \ker A = \dim \ker A^* = \operatorname{codim} \operatorname{Im} A = \operatorname{codim} \operatorname{Im} A^*.$$

4.3. Spectrum of a bounded linear operator

Lec. 33: 11/24

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. Eigenvalues always exist by the fundamental theorem of algebra, as they are the roots of the characteristic polynomial $\det(T - \lambda I) = 0$. There are at most n eigenvalues of T, or one can say exactly n counting multiplicities. Eigenvectors corresponding to different eigenvalues are linearly independent.[2]

4.3.1. Examples and definition of spectrum. In infinite-dimensional normed spaces, the spectrum is a richer concept than in finite-dimensional spaces. Let us illustrate the difference on two examples.

Example 4.3.1 (Uncoutnable number of eigenvalues). Consider the differential operator

$$T = \frac{d}{dt}$$

acting, for example, on $C^1(\mathbb{C})$. To compute the spectrum of T, we solve the ordinary differential equation $u' = \lambda u$. The solution has the form

$$u(t) = Ce^{\lambda t}.$$

Therefore, *every* $\lambda \in \mathbb{C}$ *is an eigenvalue of* T.

Example 4.3.2 (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 means that the following identity holds

$$tf(t) = \lambda f(t) \quad \text{for all } t \in [0,1].$$

It follows that $f = 0$. Therefore, T *has no eigenvalues.*

Definition 4.3.3 (Spectrum). Let X be a normed space and $T \in L(X,X)$. A number $\lambda \in \mathbb{C}$ is called a *regular point* if $T - \lambda I$ is invertible as a bounded linear operator,[3] i.e. $(T - \lambda I)^{-1} \in L(X,X)$. All other λ are called *spectrum points.* The set of all regular points is denoted $\rho(T)$ and is called the *resolvent set of* T. The set of all spectrum points is denoted $\sigma(T)$ and is called the *spectrum of* T.

---

[2] Recall however that 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. $T^*T = TT^*$.

[3] In the future, we will often say "invertible" instead of 'invertible as a bounded linear operator".

4.3.2. 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 reasons why $T - \lambda I$ may be non-invertible. These reasons are listed in the following definition:

DEFINITION 4.3.4 (Classification of spectrum). Let X be a normed space and $T \in L(X, X)$.

(i) The *point spectrum* $\sigma_p(T)$ is the set of all eigenvalues of T, i.e. the numbers $\lambda \in C$ satisfying
$$\ker(T - \lambda I) \neq 0.$$

(ii) The *continuous spectrum* $\sigma_c(T)$ is the set of all $\lambda \in \mathbb{C}$ such that
$$\ker(T - \lambda I) = 0 \quad \text{and} \quad \mathrm{Im}(T - \lambda I) \text{ is dense in } X.$$

(iii) The *residual spectrum* $\sigma_r(T)$ is the set of all $\lambda \in \mathbb{C}$ such that
$$\ker(T - \lambda I) = 0 \quad \text{and} \quad \mathrm{Im}(T - \lambda I) \text{ is not dense in } X.$$

So the spectrum of T can be expressed as a disjoint union
$$\sigma(T) = \sigma_p(T) \cup \sigma_c(T) \cup \sigma_r(T).$$

Let us now compute and classify the spectrum of come basic linear operators.

EXAMPLE 4.3.5 (Diagonal operator on ℓ_2). Let us fix some sequence $\lambda_k \to 0$ in $\mathbb{C}\backslash\{0\}$, and consider the operator T on ℓ_2 defined as
$$T((x_k)_{k=1}^{\infty}) = (\lambda_k x_k)_{k=1}^{\infty}.$$

As $(T - \lambda I)x = ((\lambda_k - \lambda)x_k)_{k=1}^{\infty}$, we have $(T - \lambda I)^{-1}y = \left(\frac{y_k}{\lambda_k - \lambda}\right)_{k=1}^{\infty}$. It follows that $(T - \lambda I)^{-1}$ is a bounded operator is and only if λ is not in the closure of $\{\lambda_k\}_{k=1}^{\infty}$, which is $\{\lambda_k\}_{k=1}^{\infty} \cup \{0\}$.

All λ_k are clearly the eigenvalues of T as $Te_k = \lambda_k e_k$ for the canonical basis (e_k) of ℓ_2. 0 is not an eigenvalue since T is injective (as all $\lambda_k \neq 0$). So 0 is either in continuous or residual spectrum. Now, $\mathrm{Im}\, T$ is dense in ℓ_2 (why?), so 0 is in the continuous spectrum. Our conclusion is:
$$\sigma_p(T) = \{\lambda_k\}_{k=1}^{\infty}, \quad \sigma_c(T) = \{0\}, \quad \sigma_r(T) = \varnothing.$$

EXAMPLE 4.3.6 (Multiplication operator on L_2). Let us consider the same multiplication operator as in Example 4.3.2, i.e.
$$(Tf)(t) = tf(t).$$

As $(T - \lambda I)f(t) = (t - \lambda)f(t)$, we have
$$(T - \lambda I)^{-1}y(t) = \frac{1}{t - \lambda}y(t). \tag{4.4}$$

If $\lambda \notin [0, 1]$ then the function $\frac{1}{t-\lambda}$ is bounded, thus $(T - \lambda I)^{-1}$ is a bounded operator. Therefore such λ are regular points. Conversely, if $\lambda \in [0, 1]$ then $\frac{1}{t-\lambda} \notin L_2[0, 1]$ because of the non-integrable singularity at 0. Hence $T - \lambda I$ is not invertible (at $y(t) \equiv 1$). Hence all such λ are regular points. Therefore, $\sigma(T) = [0, 1]$.

As we noticed in Example 4.3.2, T has no eigenvalues. It follows from (4.4) that $\mathrm{Im}(T - \lambda I)$ is dense in $L_2[0, 1]$. (Check!) Our conclusion is:
$$\sigma_p(T) = \varnothing, \quad \sigma_c(T) = [0, 1], \quad \sigma_r(T) = \varnothing.$$

REMARK 4.3.7. If Dirac delta function $\delta(t)$ was a genuine function in L_2, then its translates $\delta_\lambda(t) := \delta(t-\lambda)$ would be the eigenvectors of the multiplication operator on L_2:

$$T\delta_\lambda = \lambda\delta_\lambda$$

and δ_λ would be the eigenfunctions of T. The situation would be similar to the discrete multiplication operator from Example 4.3.5.

EXAMPLE 4.3.8 (Shift operator). Consider the right and left shift operators on ℓ_2, acting on a vector $x = (x_1, x_2, \dots)$ as

$$R(x) = (0, x_1, x_2, \dots), \quad L(x) = (x_2, x_3, \dots).$$

Since R is clearly injective but $\operatorname{Im} R$ is not dense in ℓ_2 (why?), 0 is in the residual spectrum of R. One can show that

$$\sigma_p(R) = \varnothing, \quad \sigma_c(R) = \{\lambda \in \mathbb{C} : |\lambda| = 1\}, \quad \sigma_r(R) = \{\lambda \in \mathbb{C} : |\lambda| < 1\};$$
$$\sigma_p(L) = \{\lambda \in \mathbb{C} : |\lambda| < 1\}, \quad \sigma_c(R) = \{\lambda \in \mathbb{C} : |\lambda| = 1\}, \quad \sigma_r(R) = \varnothing.$$

EXERCISE 4.3.9. **Prove the claims about the spectra of shift operators made in Example 4.3.8.**

4.4. Properties of spectrum. Spectrum of compact operators.

Lec. 34: 11/29

Throughout this section, X denotes a Banach space and $T \in L(X, X)$.

4.4.1. Resolvent operator. Spectrum is bounded. Studying the spectrum of T is convenient via the so-called resolvent operator:

DEFINITION 4.4.1 (Resolvent operator). To each regular point $\lambda \in \rho(T)$ we associate the operator

$$R(\lambda) = (T - \lambda I)^{-1}.$$

$R(\lambda)$ is called the *resolvent operator* of T. So the resolvent is a function $R : \rho(T) \to L(X, X)$.

The resolvent operator can be computed in terms of series expansion involving T. This technique is based on the following simple lemma:

LEMMA 4.4.2 (Von Neumann). *Consider an operator $S \in L(X, X)$ such that $\|S\| < 1$. Then $I - S$ is invertible, and it can be expressed as a convergent series in $L(X, X)$:*

$$(I - S)^{-1} = \sum_{k=0}^{\infty} S^k, \quad \|(I - S)^{-1}\| \leqslant \frac{1}{1 - \|S\|}.$$

PROOF. The series $\sum_{k=0}^{\infty} S^k$ converges absolutely because $\|S^k\| \leqslant \|S\|^k$ while $\|S\| < 1$. Furthermore,

$$(I - S)\sum_{k=0}^{\infty} S^k = \sum_{k=0}^{\infty} S^k (I - S) = I$$

as telescoping series. (Check!) Finally,

$$\|(I - S)^{-1}\| \leqslant \sum_{k=0}^{\infty} \|S\|^k \leqslant \frac{1}{1 - \|S\|}.$$

This completes the proof. □

PROPOSITION 4.4.3. *The spectrum $\sigma(T)$ is a bounded set. Specifically,*

$$\sigma(T) \subseteq \{\lambda \in \mathbb{C} : \ |\lambda| \leqslant \|T\|\}.$$

PROOF. The claim is that $|\lambda| > \|T\|$ implies $\lambda \in \rho(T)$, i.e. the operator $T - \lambda I$ is invertible. We express $T - \lambda I = \lambda(\frac{1}{\lambda}T - I)$. Since $\|\frac{1}{\lambda}T\| < 1$, von Neumann's Lemma 4.4.2 implies that the operator in question is indeed invertible. □

Furthermore, writing the series expansion of the inverse of $T - \lambda I = -\lambda(I - \frac{1}{\lambda}T)$ according to von Neumann's lemma, we immediately obtain:

PROPOSITION 4.4.4 (Bound on the resolvent). *If $|\lambda| > \|T\|$ then*

$$R(\lambda) = -\sum_{k=1}^{\infty} \lambda^{-k-1} T^k, \quad \|R(\lambda)\| \leqslant \frac{1}{|\lambda| - \|T\|}. \qquad \square$$

4.4.2. Resolvent identity. Spectrum is closed.

PROPOSITION 4.4.5 (Resolvent identity). *For all $\lambda, \mu \in \rho(T)$ we have*

$$R(\lambda) - R(\mu) = (\lambda - \mu) R(\lambda) R(\mu).$$

PROOF. In the scalar case, the following identity clearly holds:

$$\frac{1}{x - \lambda} - \frac{1}{x - \mu} = \frac{\lambda - \mu}{(x - \lambda)(x - \mu)}$$

for all $x, y \in \mathbb{C}$, $\lambda, \mu \neq x$. It is an exercise to generalize it to the operator-valued case, which yields the resolvent identity. (Do this!) □

Solving the resolvent identity for $R(\mu)$ we obtain for $\lambda, \mu \in \rho(T)$ that

$$R(\mu) = \left[I - (\mu - \lambda) R(\lambda)\right]^{-1} R(\lambda) \tag{4.5}$$

COROLLARY 4.4.6. *The regular set $\rho(T)$ is an open set. Equivalently, the spectrum $\sigma(T)$ is a closed set.* □

PROOF. Let $\lambda \in \rho(T)$. Von Neumann's lemma implies that the operator $I - (\mu - \lambda) R(\lambda)$ is invertible for all μ in a suitably small neighborhood of λ. So for such μ, the right hand side of (4.5) defines a bounded linear opearator. One can check that in this case identity (4.5) holds (do this!) and therefore $\mu \in \rho(T)$. □

4.4.3. Resolvent is an analytic function. Spectrum is nonempty. The proof of Corollary 4.4.6 gives us a bit more information about the resolvent than we have noticed. Let us go fo back to identity (4.5) and write the series expansion of the inverse of $(\lambda - \mu) R(\lambda) - I$ according to von Neumann's lemma. We immediately obtain:

COROLLARY 4.4.7 (Resolvent expansion). *The resolvent $R(\lambda)$ is an analytic operator-valued function on its domain $\rho(T)$. Specifically, $R(\mu)$ can be expressed as a convergent power series in a small neighborhood of any point $\lambda \in \rho(T)$:*

$$R(\mu) = \sum_{k=1}^{\infty} (\mu - \lambda)^{k-1} R(\lambda)^k. \qquad \square \tag{4.6}$$

REMARK 4.4.8. It follows that for every functional $f \in L(X, X)^*$, the function $f(R(\mu))$ is a usual (i.e. complex-valued) analytic function on $\rho(T)$.

THEOREM 4.4.9. *The spectrum $\sigma(T)$ is a nonempty set.*

PROOF. We shall deduce this result from Liouville's theorem in complex analysis.[4] To this end, assume that $\sigma(T) = \varnothing$, hence $\rho(T) = \mathbb{C}$ and the resolvent $R(\lambda)$ is an *entire* function (i.e. analytic on the whole complex plane).

Claim. $\mathbb{R}(\lambda)$ is also *bounded* function on $\mathbb{C}$ with $R(\lambda) \to 0$ as $\lambda \to \infty$.

Indeed, by Proposition 4.4.4, $\mathbb{R}(\lambda)$ is a bounded in the annulus $|\lambda| \geqslant 2\|T\|$ and vanishes at infinity. Since $R(\lambda)$ is a continuous function by Corollary 4.4.7, $R(\lambda)$ is also bounded in the disc $|\lambda| \leqslant 2\|T\|$.

Claim. By Liouville's theorem, $R(\lambda) = 0$ everywhere.

Indeed, we fix a functional $f \in L(X, X)^*$ and apply the usual Liouville's theorem for the bounded entire function $f(R(\lambda))$. It follows that $f(R(\lambda))$ is constant, and since it must vanish at infinity it is zero everywhere. The claim follows.

The last claim contradicts the fact that $R(\lambda)$ is an invertible operator. $\square$

Summarizing our findings, we can state that *the spectrum of every bounded linear operator is a nonempty compact subset of* $\mathbb{C}$.

Lec. 35: 12/1

4.4.4. Spectral radius. The spectrum of any operator $T \in L(X, X)$ is a bounded set by Proposition 4.4.3, and moreover we have a quantitative bound $|\lambda| \leqslant \|T\|$ for all $\lambda \in \sigma(T)$. This bound is not always sharp, and we will try to come up with a sharp bound.

DEFINITION 4.4.10. The *spectral radius* of an operator $T \in L(X, X)$ is defined as

$$r(T) = \max\big\{|\lambda| : \ \lambda \in \sigma(T)\big\}.$$

THEOREM 4.4.11 (Gelfand's formula). *For every operator $T \in L(X, X)$ acting on a Banach space X, one has*

$$r(T) = \lim_n \|T^n\|^{1/n} = \inf_n \|T^n\|^{1/n}.$$

EXERCISE 4.4.12. **Clearly $r(T) \leqslant \|T\|^{n \cdot 1/n} \leqslant \|T\|$, so Gelfand's formula is an improvement upon Proposition 4.4.3. Give an example where $r(T) < \|T\|$.**

REMARK 4.4.13. Gelfand's formula gives an asymptotics for the growth of operator powers $\|T^n\| \sim r(T)^n$.

PROOF. *Upper bound.* Let $\lambda \in \sigma(T)$. One can easily show that $\lambda^n \in \sigma(T^n)$, see Exercise 4.4.27.[5] Therefore, using Proposition 4.4.3, we have $|\lambda^n| \leqslant \|T^n\|$, so $|\lambda| \leqslant \|T^n\|^{1/n}$. We conclude that

$$r(T) \leqslant \inf_n \|T^n\|^{1/n}.$$

Lower bound will be obtained as an application of the theory of Laurent series in complex analysis.[6] Consider the function $f(R(\lambda))$ where $R(\lambda)$ is the resolvent of T and $f \in L(X, X)^*$ is an arbitrary functional. Recall two facts:

---

[4] Liouville's theorem states that an entire (i.e. analytic on $\mathbb{C}$) and bounded function $f : \mathbb{C} \to \mathbb{C}$ is constant everywhere.

[5] This is a partial case of the spectral mapping theorem we will study later.

[6] Specifically, we shall use the following theorem of compex analysis. Consider a Laurent series

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k.$$

(i) $f(R(\lambda))$ is an analytic function on the annulus $|\lambda| > r(T)$ by Corollary 4.4.4;
(ii) $f(R(\lambda))$ is represented by a convergent Laurent series

$$f(R(\lambda)) = -\sum_{k=1}^{\infty} \lambda^{-k-1} f(T^k) \tag{4.7}$$

in the smaller annulus $|\lambda| > \|T\|$.

By the theory of convergence of Laurent series (outlined in the footnote), the series (4.7) converges in the larger annulus $|\lambda| > r(T)$. So the terms of the series are bounded for such λ:

$$\sup_n \left|\lambda^{-n-1} f(T^n)\right| < \infty.$$

This means that the sequence $(\lambda^{-n-1}T^n)$ is weakly bounded in $L(X, X)$. By a consequence of the principle of uniform boundedness (Corollary 3.3.5), this sequence is (strongly) bounded, i.e.

$$\sup_n \left\|\lambda^{-n-1} T^n\right\| =: K < \infty.$$

Taking n-th root and rearranging the terms, we obtain $\|T^n\|^{1/n} \leqslant K^{1/n}\lambda^{1+1/n}$ for all n. It follows that $\limsup_n \|T^n\|^{1/n} \leqslant |\lambda|$. Since this happens for all λ such that $|\lambda| > r(T)$, we have proved that

$$\limsup_n \|T^n\|^{1/n} \leqslant r(T).$$

So, putting this together with the upper bound, we have proved that

$$r(T) \leqslant \inf_n \|T^n\|^{1/n} \leqslant \liminf_n \|T^n\|^{1/n} \leqslant \limsup_n \|T^n\|^{1/n} \leqslant r(T).$$

This completes the proof. □

4.4.5. Spectrum of compact operators. As compact operators are proxies of finite rank operators, one is able to fully classify their spectrum. First of all, for every $T \in K(X, X)$ one has

$$0 \in \sigma(T)$$

since T is not invertible by Corollary 4.1.8.

EXERCISE 4.4.14. **Construct three examples of compact operators for which 0 is in the point, continuous, and residual spectrum respectively.**

THEOREM 4.4.15 (Point spectrum of compact operators). *Let $T \in K(X, X)$ be a compact operator on a normed space X. For every $\varepsilon > 0$ there exists a finite number of linearly independent eigennvectors corresponding to eigenvalues λ_k with $|\lambda_k| > \varepsilon$.*

Consequently, the point spectrum $\sigma_p(T)$ is at most countable, and it lies in a sequence that converges to zero.

It also follows that each eigenvalue λ_k of T has finite multiplicity, i.e. $\dim\ker(T-\lambda_k I) < \infty$.

There exist unique radii $r, R \in \mathbb{R} \cap \{\infty\}$ such that the Laurent series converges in the annulus

$$A = \{z \in \mathbb{C} : \ r < |z - z_0| < R\}$$

and diverges outside the closure of A. Moreover, there exists at least one point on the inner boundary $\{z \in \mathbb{C} : \ |z - z_0| = r\}$ of A and at least one point on the outer boundary $\{z \in \mathbb{C} : |z - z_0| = R\}$ of A such that $f(z)$ can not be analytically continues to those points.

PROOF. Clearly, the second and third claims of the theorem follow from the first one (why?). So, assume the contrary, that there exist $\varepsilon > 0$ and an infinite sequence of linearly independent vectors $(x_k)_{k=1}^{\infty}$ such that

$$Tx_k = \lambda_k x_k, \quad \text{where } |\lambda_k| > \varepsilon.$$

Consider the subspaces $E_n = \mathrm{Span}(x_k)_{k=1}^n$; then $E_1 \subset E_2 \subset \cdots$ is a sequence of proper inclusions. Therefore we can choose vectors

$$y_n \in E_n, \ \|y_n\| = 1, \ \mathrm{dist}(y_n, E_{n-1}) \geqslant \frac{1}{2}.$$

(Why? Think about E_n/E_{n-1}.)

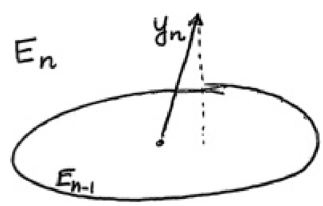

FIGURE 4.1. Construction of subspaces E_n and vectors y_n

We will show that the sequence $(Ty_n)_{n=1}^{\infty}$ contains no Cauchy subsequences, which will contradict compactness of T. To this end, we express y_n as a linear combination

$$y_n = \sum_{k=1}^{n} a_k^{(n)} x_k = a_n^{(n)} x_n + u_{n-1}, \quad \text{where } u_{n-1} \in E_{n-1}.$$

Then

$$Ty_n = \lambda_n a_n^{(n)} x_n + v_{n-1}, \quad \text{where } v_{n-1} = Tu_{n-1} \in E_{n-1}.$$

Now we are ready to estimate $\|Ty_n - Ty_m\|$ for $n > m$. Since $Ty_m \in E_m \subseteq E_{n-1}$, we obtain

$$\begin{aligned}\|Ty_n - Ty_m\| &= \|\lambda_n a_n^{(n)} x_n + w_{n-1}\| \quad \text{where } w_{n-1} \in E_{n-1}\\ &= \|\lambda_n y_n + z_{n-1}\| \quad \text{where } w_{n-1} \in E_{n-1}\\ &\geqslant |\lambda_n| \ \mathrm{dist}(y_n, E_{n-1}) \geqslant \frac{\varepsilon}{2}.\end{aligned}$$

It follows that $(Ty_n)_{n=1}^{\infty}$ contains no Cauchy subsequences as claimed. The proof is complete. □

PROPOSITION 4.4.16 (Classification of spectrum of compact operators). *Let $T \in K(X, X)$ be a compact operator on a Banach space X. Then*

$$\sigma(T) = \sigma_p(T) \cup \{0\}.$$

PROOF. As we already noticed in the beginning of this section, $0 \in \sigma(T)$. Let now $\lambda \in \sigma(T)$, $\lambda \neq 0$. Fredholm alternative (Theorem 4.2.2) states that either $T - \lambda I$ is not injective (in which case $\lambda \in \sigma_p(T)$) or $T - \lambda I$ is both injective and surjective. In the latter case $T - \lambda I$ is invertible by the inverse mapping theorem, which means that $\lambda \notin \sigma(T)$. The proof is complete. □

4.4.6. Spectrum of unitary operators. In this subsection, H denotes a Hilbert space.

DEFINITION 4.4.17. An operator $U \in L(H, H)$ is called a *unitary* operator if U is a bijective isometry on H. The latter means that U is bijective and

$$\|Ux\| = \|x\| \quad \text{for all } x \in H.$$

EXAMPLE 4.4.18. Examples of unitary operators include:

(i) operators on $\mathbb{C}^n$ and $\mathbb{R}^n$ given by $n \times n$ unitary complex matrices and orthogonal real matrices; in particular rotations, symmetries, and permutations of coordinates in $\mathbb{C}^n$ and $\mathbb{R}^n$;
(ii) right shift R on ℓ_2 (but not left – why?)
(iii) an isometry between any pair of separable Hilbert spaces established in Theorem 1.6.30.

REMARK 4.4.19. A unitary operator U preserves all pairwise distances, i.e. $\|Ux - Uy\| = \|x - y\|$. Moreover, by polarization identity 1.4.19, U also preserves the inner products:

$$\langle Ux, Uy\rangle = \langle x, y\rangle \quad \text{for all } x, y \in H.$$

PROPOSITION 4.4.20. *An operator $U \in L(H, H)$ is unitary if and only if*

$$U^*U = UU^* = I,$$

i.e. if and only if U is invertible and $U^{-1} = U^*$.

This identity is analogous to the one for unit complex numbers, $\bar{z}z = z\bar{z} = 1$.

PROOF. *Necessity.* If U is unitary then $\langle U^*Ux, y\rangle = \langle Ux, Uy\rangle = \langle x, y\rangle$. Since this holds for arbitrary $x, y \in H$, it follows that $U^*U = I$. Similarly one proves that $UU^* = I$.

Sufficiency. Since U is invertible it is bijective. The isometry property follows because $\|Ux\|^2 = \langle Ux, Ux\rangle = \langle U^*Ux, x\rangle = \langle x, x\rangle = \|x\|^2$. The proof is complete. □

PROPOSITION 4.4.21 (Spectrum of unitary operators). *The spectrum of a unitary operator $U \in L(H, H)$ lies on the unit circle:*

$$\sigma(U) \subseteq \{\lambda \in \mathbb{C} : |\lambda| = 1\}.$$

PROOF. The isometry property implies that $\|U\| = \|U^{-1}\| = 1$. Therefore the spectral radius is $r(U) \leqslant \|U\| \leqslant 1$ by Proposition spectrum bounded. On the other hand, if $|\lambda| < 1$ then the operator $U^{-1}(U - \lambda I) = I - \lambda U^{-1}$ is invertible by von Neumann's Lemma 4.4.3 since $\|\lambda U^{-1}\| = |\lambda| < 1$. It follows that $U - \lambda I$ is invertible. □

EXERCISE 4.4.22. **Show that eigenvectors of a unitary operator U that correspond to distinct eigenvalues are orthogonal.**

4.4.7. Additional exercises. In the following two exercises, one can work over $\mathbb{R}$. Similar results hold over $\mathbb{C}$. The only difference is that for Hilbert spaces, one has to take complex conjugation in appropriate places (which ones?), see Remark 2.4.26.

EXERCISE 4.4.23. **[Spectrum of adjoint I] Let $T \in L(X,X)$. Prove that $\sigma(T^*) = \sigma(T)$. Here the bar stands for complex conjugation rather than for closure.**

EXERCISE 4.4.24. **[Spectrum of adjoint II] Let $T \in L(X,X)$**

- **(i) Prove that if $\lambda \in \sigma_p(T)$ and $\lambda \notin \sigma_p(T^*)$ then $\lambda \in \sigma_r(T^*)$. (Hint: use the duality relations from Proposition 2.4.32 and Exercise 2.4.34 for the operator $T - \lambda I$.)**
- **(ii) Prove that**

$$\sigma_r(T) \subseteq \sigma_p(T^*) \subseteq \sigma_r(T) \cup \sigma_p(T).$$

Deduce that if X is reflexive, then $\sigma_r(T^*) \subset \sigma_p(T)$. Deduce that self-adjoint bounded linear operators in Hilbert space do not have residual spectrum.

EXERCISE 4.4.25. **[General multiplication operator on L_2] Consider a general multiplication operator T acting on $L_2[0,1]$ as**

$$(Tf)(t) = g(t)f(t)$$

where $g \in L_\infty[0,1]$ is some fixed multiplier function.

EXERCISE 4.4.26. **[Invertibility of a composition] Let $S, T \in L(X,X)$. Prove that the operator ST is invertible if and only if both S and T are invertible.**

EXERCISE 4.4.27. **[Partial case of spectral mapping theorem] Let $T \in L(X,X)$. Prove that $\lambda \in \sigma(T)$ implies $\lambda^n \in \sigma(T^n)$. (Hint: (i) factor $T^n - \lambda^n I = S(T - \lambda I)$ for some $S \in L(X,X)$. (ii) Show that for $U, V \in L(X,X)$, the operator UV is invertible if and only if both U and V are invertible.)**

EXERCISE 4.4.28. **[Spectrum of projections] Compute the spectrum of a projection $P \in L(X,X)$ on a Banach space X.**

CHAPTER 5

Self-adjoint operators on Hilbert space

Lec. 36: 12/03

Throughout this chapter, H will denote a Hilbert space, and we will study bounded self-adjoint operators T on H.

5.1. Spectrum of self-adjoint operators

5.1.1. Definition and examples. Let T be a bounded linear operator on a Hilbert space, i.e. $T \in L(H,H)$. Recall from Section 2.4.9 that the adjoint operator $T^* \in L(H,H)$ is defined by $\langle T^*x, y\rangle = \langle x, Ty\rangle$ for $x, y \in H$.

DEFINITION 5.1.1. An operator $T \in L(H,H)$ is called *self-adjoint* if $T^* = T$, i.e.

$$\langle Tx, y\rangle = \langle x, Ty\rangle, \quad x, y \in H.$$

EXAMPLE 5.1.2. Examples of self-adjoint operators include:

(i) linear operators on $\mathbb{C}^n$ given by Hermitian matrices (a_{ij}), i.e. such that $a_{ij} = \overline{a_{ji}}$;

(ii) integral operators $(Tf)(t) = \int_0^1 k(s,t) f(s)\, ds$ on $L_2[0,1]$ with Hermitian symmetric kernels, i.e. such that $k(s,t) = \overline{k(t,s)}$;

(iii) orthogonal projections P on H. (Why?)

Every bounded linear operator can be decomposed into two self-adjoint operators:

LEMMA 5.1.3. *Every operator $A \in L(H,H)$ can be uniquely represented as*

$$A = T + iS$$

where $T, S \in L(H,H)$ are self-adjoint operators.

PROOF. If $A = T + iS$ then $A^* = T - iS$. Solving these two equations we see that the lemma holds with $T = \frac{A+A^*}{2}$ and $S = \frac{A-A^*}{2i}$. $\square$

EXERCISE 5.1.4. **Prove that the set of self-adjoint operators forms a closed linear subspace in $L(H,H)$.**

5.1.2. The quadratic form and the norm of a self-adjoint operator. It is convenient to study self-adjoint operators $T \in L(H,H)$ through the quadratic form

$$f(x) = \langle Tx, x\rangle, \quad x \in H.$$

One sees immediately that this quadratic form is real-valued, i.e. $f(x) \in \mathbb{R}$ for all $x \in H$.[1]

[1] This follows from the identity $\langle Tx, x\rangle = \overline{\langle x, Tx\rangle} = \overline{\langle Tx, x\rangle}$.

Furthermore, the quadratic form $f(T)$ determines the operator T uniquely. This follows from the polarization identity that relates bilinear and quadratic forms:

$$\langle Tx, y\rangle = \frac{1}{4}\big[f(x+y) - f(x-y) + if(x+iy) - if(x-iy)\big] \tag{5.1}$$

This is a slight generalization of the polarization identity of Proposition 1.4.19, where $T = I$. (Prove it!)

The norm of T can be conveniently computed from the quadratic form $f(T)$:

Proposition 5.1.5 (Norm of a self-adjoint operator). *For every self-adjoint operator $T \in L(H,H)$, one has*

$$\|T\| = \sup_{x \in S_H} |\langle Tx, x\rangle|.$$

Proof. The lower bound follows by definition of the operator norm:

$$\|T\| = \sup_{x \in S_H} \|Tx\| = \sup_{x,y \in S_H} |\langle Tx, y\rangle| \geqslant \sup_{x \in S_H} |\langle Tx, x\rangle| =: M.$$

It remains to show that the inequality here is actually the identity. To this end, we note that

$$\sup_{x,y \in S_H} |\langle Tx, y\rangle| = \sup_{x,y \in S_H} \operatorname{Re}\langle Tx, y\rangle$$

and use the real part of polarization identity (5.1):

$$\begin{aligned}
\operatorname{Re}\langle Tx, y\rangle &= \frac{1}{4}\big[\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle\big] \\
&\leqslant \frac{M}{4}\big[\|x+y\|^2 + \|x-y\|^2\big] \quad \text{(by the definition of } M) \\
&\leqslant \frac{M}{4}\big[2\|x\|^2 + 2\|x\|^2\big] \quad \text{(by the parallelogram law)} \\
&\leqslant M \quad (\text{as } \|x\| = \|y\| = 1).
\end{aligned}$$

This completes the proof. □

5.1.3. Criterion of spectrum points. We would like to study the spectrum of self-adjoint operators $T \in L(H,H)$. An easy observation is that all eigenvalues of T must be real, that is

$$\sigma_p(T) \subseteq \mathbb{R}.$$

Indeed, if λ is an eigenvalue with an eigenvector x then $\langle Tx, x\rangle = \langle \lambda x, x\rangle = \lambda\langle x, x\rangle$ which must be the same as $\langle x, Tx\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\langle x, x\rangle$. This shows that $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.

We will soon prove that the whole spectrum of T is real, i.e. $\sigma(T) \subseteq \mathbb{R}$, and moreover we will compute the smallest interval containing $\sigma(T)$. Let us start with ruling out the residual spectrum:

Proposition 5.1.6 (No residual spectrum). *Let $T \in L(H,H)$ be a self-adjoint operator. Then $\sigma_r(T) = \varnothing$.*

Proof. Let $\lambda \in \sigma_r(T)$. This means that $\ker(T - \lambda I) = 0$ while $\operatorname{Im}(T - \lambda I)$ is not dense in H. Since λ is not an eigenvalue, $\overline{\lambda}$ is not an eigenvalue either (recall that all eigenvalues of T must be real). Using this and the duality relation (2.4.32), we obtain that $(\operatorname{Im}(T - \lambda I))^\perp = \ker(T - \lambda I)^* = \ker(T - \overline{\lambda} I) = 0$. It follows that $\operatorname{Im}(T - \lambda I)$ is dense in H, which is a contradiction. □

PROPOSITION 5.1.7 (Invertibility criterion). *Let $T \in L(H,H)$ be a self-adjoint operator. T is invertible*[2] *if and only if T is bounded below, i.e. there exists $c > 0$ such that*

$$\|Tx\| \geqslant c\|x\| \quad \textit{for all } x \in H.$$

PROOF. If T is invertible then T is bounded below with $c = 1/\|T^{-1}\|$. To prove the converse, assume that T is bounded below. Then by the criterion of isomorphic embedding (Proposition 3.1.12), T is injective and $\operatorname{Im} T$ is closed in H. On the other hand, since $0 \notin \sigma_r(T) = \varnothing$, injectivity of T implies that $\operatorname{Im} T$ is dense in H. It follows that $\operatorname{Im} T = H$. So T is injective and surjective, thus T is invertible by the inverse mapping theorem. □

Applying this result for the operator $T - \lambda I$, we immediately obtain

COROLLARY 5.1.8 (Criterion of spectrum points). *Let $T \in L(H,H)$ be a self-adjoint operator. Then $\lambda \in \sigma(T)$ if and only if the operator $T - \lambda I$ is not bounded below.*

REMARK 5.1.9 (Approximate point spectrum). A number $\lambda \in \sigma(T)$ for which $T - \lambda I$ is not bounded below is called an *approximate eigenvalue* of T, and the set of all approximate eigenvalues is called the *approximate point spectrum* of T. Corollary 5.1.8 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 $x \in S_H$. If λ is an approximate eigenvalue then $(T - \lambda I)x$ can be made *arbitrarily close* to zero for some $x \in S_H$. So, eigenvalues of T form the point spectrum $\sigma_p(T)$ while the approximate eigenvalues of T form the continuous spectrum $\sigma_c(T)$.

5.1.4. The spectrum interval. Now we 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:

THEOREM 5.1.10 (Spectrum interval). *Let $T \in L(H,H)$ be a self-adjoint operator. Then*

(i) The spectrum of T is real, and moreover $\sigma(T) \subseteq [m, M]$ where

$$m = \inf_{x \in S_H} \langle Tx, x\rangle, \quad M = \sup_{x \in S_H} \langle Tx, x\rangle.$$

(ii) The endpoints $m, M \in \sigma(T)$.

PROOF. *(i)* Let $\lambda \in \mathbb{C} \backslash [m, M]$; since the interval is closed we have

$$d := \operatorname{dist}(\lambda, [m, M]) > 0.$$

Given $x \in S_H$, we use Cauchy-Schwarz to obtain the lower bound

$$\|(T - \lambda I)x\| \geqslant \left|\langle (T - \lambda I)x, x\rangle\right| = \left|\langle Tx, x\rangle - \lambda\right| \geqslant d$$

where the last inequality follows because $\langle Tx, x\rangle \in [m, M]$ by definition. We have shown that $T - \lambda I$ is bounded below. By the criterion of spectrum points (Proposition 5.1.8), we conclude that $\lambda \notin \sigma(T)$.

(ii) Let us show that $M \in \sigma(T)$; the claim for m can be proved similarly. (Do this!) Without loss of generality we can assume that $0 \leqslant m \leqslant M$ (This follows by

[2]As usual, by being invertible we mean continuously invertible, i.e. $T^{-1} \in L(H, H)$.

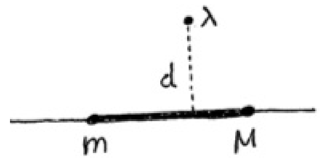

FIGURE 5.1. Distance from λ to the spectrum interval $[m, M]$ is d.

a translation argument, namely by considering $T - mI$ instead of T. Check this!) So $\|T\| = M$.

Let us choose a sequence of vectors $x_n \in S_H$ so that $\langle Tx_n, x_n\rangle \to M$. Then

$$\begin{aligned}\|(T - MI)x_n\|^2 &= \big\langle (T - MI)x_n, (T - MI)x_n \big\rangle \\ &= \|Tx_n\|^2 - 2M\langle Tx_n, x_n\rangle + M^2\|x_n\|^2.\end{aligned}$$

Now, $\|Tx_n\|^2 \leqslant \|T\|^2 = M^2$, $\langle Tx_n, x_n\rangle \to M$ and $\|x_n\|^2 = 1$. It follows that

$$\limsup \|(T - MI)x_n\|^2 \leqslant M^2 - 2M^2 + M^2 = 0$$

so $T - MI$ is not bounded below. Therefore $M \in \sigma(T)$. The proof is complete. □

As a consequence of this result, the spectral radius $r(T)$ of a self-adjoint operator equals $\|T\|$, so Proposition 4.4.3 is tight and Gelfand's formula is useless for self-adjoint operators:

COROLLARY 5.1.11 (Spectral radius). *Let $T \in L(H, H)$ be a self-adjoint operator. Then*

$$r(T) = \max_{\lambda \in \sigma(T)} |\lambda| = \|T\|.$$

PROOF. By the properties of the spectrum interval in Theorem 5.1.10,

$$r(T) = \max(|m|, |M|) = \|T\|$$

as claimed. □

5.1.5. Additional Exercises.

EXERCISE 5.1.12. **Consider a self-adjoint operator $P \in L(H, H)$ such that $P^2 = P$. Prove that P is an orthogonal projection.**

5.2. Spectral theorem for compact self-adjoint operators

Lec.37: 12/06

Compact self-adjoint operators on a Hilbert space H are proxies of Hermitian matrices. As we know from linear algebra, every Hermitian matrix has diagonal form in some orthonormal basis of $\mathbb{C}^n$. Equivalently, such for such a matrix 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 H.

5.2.1. Invariant subspaces.

PROPOSITION 5.2.1 (Eigenvectors orthogonal). *Let $T \in L(H, H)$ be a self-adjoint operator. 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

$$\lambda_1 \langle x_1, x_2 \rangle = \langle Tx_1, x_2 \rangle = \langle x_1, Tx_2 \rangle = \lambda_2 \langle x_1, x_2 \rangle.$$

(In the last identity we used that λ_2 is always real, so there is no conjugation). It follows that if $\lambda_1 \neq \lambda_2$ then $\langle x_1, x_2 \rangle = 0$ as claimed. □

DEFINITION 5.2.2 (Invariant subspace). A subspace E of H is called an *invariant subspace* of T if $T(E) \subseteq E$.

EXAMPLE 5.2.3. 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 $T \in L(H, H)$ has a proper invariant subspace (i.e. different from $\{0\}$ and H).

PROPOSITION 5.2.4. *Let $T \in L(H, H)$ be self-adjoint. If $E \subseteq H$ is an invariant subspace of T then $E^\perp$ is also an invariant subspace of T.*

PROOF. Let $x \in E^\perp$; we need check that $Tx \in E^\perp$. So let us choose $y \in E$ arbitrarily. Then $\langle Tx, y \rangle = \langle x, Ty \rangle = 0$ since $x \in E^\perp$ and $y \in E$ so $Ty \in E$. □

5.2.2. Spectral theorem. The following result is known as the Hilbert-Schmidt theorem.

THEOREM 5.2.5 (Spectral theorem for compact self-adjoint operators). *Let T be a compact self-adjoint linear operator on a separable Hilbert space H. Then there exists an orthonormal basis of H consisting of eigenvectors of H.*

PROOF. Let us first prove that T has at least one eigenvector.
By Proposition 4.4.16,

$$\sigma(T) = \sigma_p(T) \cup \{0\}.$$

If $\sigma(T) \neq 0$ then $\sigma_p(T) \neq \varnothing$, so T has an eigenvector. If $\sigma(T) = 0$ then by Corollary 5.1.11 we have $\|T\| = r(T) = 0$, so $T = 0$ and every vector in H is an eigenvector of T.

We will complete the proof by induction. Consider the family of all orthonormal sets in H consisting of eigenvectors of T. All such sets are at most countable since H is separable. By Zorn's lemma, this family has a maximal element $(\phi_k)_{k=1}^\infty$. (Check!) It remains to show that $E := \overline{\operatorname{Span}(\phi_k)} = H$.

Suppose $E \neq H$. Since E is an invariant subspace of T (check!), $E^\perp \neq \{0\}$ is also an invariant subspace of T by Proposition 5.2.4. So we can use the first part of the proof for the restriction $T|_{E^\perp}$ which is a compact self-adoint operator on $E^\perp$. It follows that $T|_{E^\perp}$ (and thus T itself) has an eigenvector in $E^\perp$. This contradicts the maximality of (ϕ_k). The proof is complete. □

5.2.3. Diagonalization. Spectral Theorem 5.2.5 allows us to always represent compact self-adoint operators $T \in L(H, H)$ in a *diagonal form*, similarly to the one for Hermitian matrices.

Let (ϕ_k) be an orthonormal basis of eigenvectors of T. Then $T\phi_k = \lambda\phi_k$ where λ_n are the eigenvalues. We can identify the space H with ℓ_2 by identifying (ϕ_k) with the canonical basis (e_k) of ℓ_2 (recall Section 1.6.7). With this identification, T becomes a multiplication operator acting on ℓ_2 as $Te_k = \lambda_k e_k$; equivalently

$$T((x_k)_{k=1}^{\infty}) = (\lambda_k x_k)_{k=1}^{\infty}.$$

We see that T now has a quite simple form, which we studied in Example 4.3.5.

In literature, one comes across various forms of spectral Theorem 5.2.5. We mention two of them. Let as before (ϕ_k) denote an orthonormal basis of eigenvectors of T with corresponding eigenvalues λ_k. Orthogonal basis expansion gives

$$x = \sum_k \langle x, \phi_k\rangle \phi_k, \quad x \in H.$$

Applying operator T for both sides and using that $T\phi_k = \lambda_k\phi_k$ we obtain that

$$Tx = \sum_n \lambda_n \langle x, \phi_n\rangle \phi_n, \quad x \in H.$$

This is sometimes written as

$$T = \sum_k \lambda_k \phi_k \otimes \phi_k \tag{5.2}$$

where $\phi_k \otimes \phi_k \in L(H, H)$ denotes the *elementary tensor*, which is the rank-one orthogonal projection in H onto the span of the vector ϕ_k.

Furthermore, by Theorem 4.4.15 all eigenvalues λ_k have finite multiplicity. So, grouping in (5.2) the tensors $\phi_k \otimes \phi_k$ that correspond to the same λ_k, we obtain the following decomposition of T:

$$T = \sum_k \lambda_k P_k \tag{5.3}$$

where P_k denotes the orthogonal projection in H onto the (finite-dimensional) eigenspace corresponding to the eigenvalue λ_k.

REMARK 5.2.6 (Normal operators). One can show that spectral Theorem 5.2.5 holds also for compact *normal* linear operators T on H. Recall that T is normal if $T^*T = TT^*$. The only difference between the spectral properties of self-adjoint and normal operators is that the eigenvalues of normal operators need not be real.

5.2.4. Separation of variables. We illustrate Spectral Theorem 5.2.5 with a purely analytic consequence. The following result shows how one can separate variables of a general function $k(t, s)$. It is also due to Hilbert and Schmidt.

THEOREM 5.2.7 (Separation of variables). *Consider a function $k(t, s) \in L_2([0, 1]^2)$ such that $k(t, s) = \overline{k(s, t)}$. There exists an orthonormal basis (ϕ_n) of $L_2[0, 1]$ and numbers $\lambda_k \to 0$ such thath*

$$k(t, s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t) \phi_n(s).$$

The convergence of this series is understood in $L_2([0, 1]^2)$.

PROOF. Consider the integral operator $(Tf)(t) = \int_0^1 k(t,s)f(s)\,ds$ on $L_2[0,1]$. Let (ϕ_n) be an orthonormal basis of its eigenvectors. Then the functions

$$\psi_{nm} = \phi_n(t)\phi_m(s), \quad n,m = 1,2,\dots$$

form an orthonormal basis of $L_2([0,1]^2)$. (Check!)

Let us write the basis expansion of our function in $L_2([0,1]^2)$:

$$k = \sum_{n,m} \langle k, \psi_{nm}\rangle \psi_{nm}.$$

Now we compute the coefficients

$$\begin{aligned}
\langle k, \psi_{nm}\rangle &= \int_0^1 \int_0^1 k(t,s)\phi_n(t)\phi_m(s)\,dt\,ds \\
&= \int_0^1 \Big(\int_0^1 k(t,s)\phi_m(s)\,ds \Big) \phi_n(t)\,dt \quad \text{(by Fubini theorem)} \\
&= \int_0^1 (T\phi_m)(t)\phi_n(t)\,dt = \langle T\phi_m, \phi_n\rangle = \lambda_m \langle \phi_m, \phi_n\rangle \\
&= \begin{cases} \lambda_m, & n = m \\ 0, & n \neq m. \end{cases}
\end{aligned}$$

Therefore

$$k = \sum_n \langle k, \psi_{nn}\rangle \psi_{nn} = \sum_n \lambda_n \phi_n(t)\phi_n(s)$$

as claimed. □

5.2.5. Additional exercises.

EXERCISE 5.2.8. **[Spectral theorem for general compact operators] Let T be a compact linear operator on a separable Hilbert space H. Show that there exist an orthonormal basis (ϕ_k) of H, an orthonormal system (ψ_k) in H, and a sequence of numbers $\lambda_k \geqslant 0$, $\lambda_k \to 0$ such that**

$$T = \sum_n \lambda_n \phi_k \otimes \psi_k.$$

The numbers λ_k are called *singular values* of T and the vectors ϕ_k and ψ_k are called left (resp. right) *singular vectors* of T.

(Hint: Choose (ψ_k) to be an orthonormal basis of eigenvectors of T^*T. Write the basis expansion of $x \in H$ and apply T to both sides.)

5.3. Positive operators. Continuous functional calculus

Lec. 38: 12/8

It is a powerful idea to work with linear operators by analogy with complex numbers. For example, both numbers and operators can be added and multiplied, so both $\mathbb{C}$ and $L(H,H)$ are algebras (over $\mathbb{C}$). Moreover, complex conjugate corresponds to operator adjoint, which makes $L(H,H)$ a *-algebra. Operator algebras is a rich topic which we omit in this course.

In this section, we develop the analogy between numbers and operators in the following way. We introduce a partial order on the set of self-adjoint operators on $T \in L(H,H)$, and we define an operator $f(T) \in L(H,H)$ for every continuous function $f : \mathbb{C} \to \mathbb{C}$. In other words, we develop what is called the *functional calculus* of operators.

5.3.1. Positive operators.

Definition 5.3.1. A self-adjoint operator $T \in L(H, H)$ is called *positive*[3] if

$$\langle Tx, x\rangle \geqslant 0 \quad \text{for all } x \in H.$$

Positive operators are generalizations of non-negative numbers (which correspond to operators on one-dimensional space $\mathbb{C}^1$).

Example 5.3.2. Examples of positive operators include:

(i) T^2 for every self-adjoint $T \in L(H, H)$, as $\langle T^2x, x\rangle = \langle Tx, Tx\rangle \geqslant 0$;
(ii) Hermitian matrices with non-negative eigenvalues;
(iii) More generally, compact self-adjoint operators on H with non-negative eigenvalues. (Why?)

Definition 5.3.3 (Partial order). For self-adjoint operators $S, T \in L(H, H)$, we shall say that $S \leqslant T$ if $T - S \geqslant 0$.

This defines a partial order on $L(H, H)$.

Let us restate Theorem 5.1.10 on the spectrum interval in these new terms:

Theorem 5.3.4 (Spectrum interval). *Let $T \in L(H, H)$ be a self-adjoint operator. Let m, M be the smallest and the largest numbers such that*

$$mI \leqslant T \leqslant MI.$$

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 5.3.5. *Let $T \in L(H, H)$ be a self-adjoint operator. Then $T \geqslant 0$ if and only if $\sigma(T) \subseteq [0, \infty)$.* □

5.3.2. Polynomials of an operator. We start to develop a functional calculus for self-adjoint operators $T \in L(H, H)$. We begin by defining polynomials of T, 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 5.3.6 (Polynomials of an operator). Consider a polynomial $p(t) = a_0 + a_1t + \cdots a_nt^n$. For an operator $T \in L(H, H)$, we define

$$p(T) = a_0I + a_1T + \cdots a_nT^n.$$

If T is self-adjoint operator then $p(T)$ is also self-adjoint. (Check!) Moreover, for two polynomials f and g, one has

$$(af + bg)(T) = a \cdot f(T) + b \cdot g(T), \quad (fg)(T) = f(T)g(T), \quad \overline{f}(T) = f(T^*).$$

This last property states in other words that for a fixed $T \in L(H, H)$, the map $p \mapsto p(T)$ is an $*$*-algebra homomorphism* from $P[t]$ into $L(H, H)$.

The following example may serve us as a test case for many future results.

[3]In linear algebra, positive operators are called positive semidefinite.

EXAMPLE 5.3.7. 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

$$T = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

where λ_k are the eigenvalues of T. Then for every polynomial $p(t)$ we have

$$p(T) = \operatorname{diag}(p(\lambda_1), \dots, p(\lambda_n)).$$

(Check!) This example can be generalized for all compact self-adjoint operators T on a general Hilbert space H. (Do this!)

5.3.3. Spectral mapping theorem for polynomials.

LEMMA 5.3.8 (Invertibility). *Let $p(t)$ be a polynomial and $T \in L(H, H)$. Then the operator $p(T)$ is invertible if and only if $p(t) \neq 0$ for all $t \in \sigma(T)$.*

PROOF. Let us factorize $p(t) = a_n(t - t_1) \cdots (t - t_n)$ where t_k are the roots of $p(t)$. Then $p(T) = a_n(T - t_1 I) \cdots (T - t_n I)$. By Exercise 4.4.26, $p(T)$ is invertible if and only if all factors $T - t_k I$ are invertible. This in turn is equivalent to the fact that all roots $t_k \notin \sigma(T)$. □

The spectrum of a polynomial $p(T)$ can be easily computed from the spectrum of T:

THEOREM 5.3.9 (Spectral mapping theorem). *Let $p(t)$ be a polynomial and $T \in L(H, H)$. Then*[4]

$$\sigma(p(T)) = p(\sigma(T)).$$

PROOF. For every complex number λ, we have $\lambda \in \sigma(p(T))$ if and only if the operator $p(T) - \lambda I = (p - \lambda)(T)$ is not invertible. By the invertibility Lemma 5.3.8, this is equivalent to the condition that $(p - \lambda)(t) = 0$ for some $t \in \sigma(T)$, which means that $p(t) = \lambda$ for some $t \in \sigma(T)$. The latter is equivalent to $\lambda \in p(\sigma(T))$. □

Using the spectral mapping theorem, one can in particular easily compute the norms of operator polynomials:

COROLLARY 5.3.10 (Operator norm of polynomials). *Let $p(t)$ be a polynomial and $T \in L(H, H)$ be a self-adjoint operator. Then*

$$\|p(T)\| = \max_{t \in \sigma(T)} |p(t)|.$$

This result generalizes the identity $r(T) = \|T\|$ for the spectral radius of self-adjoint operators T proved in Corollary 5.1.11.

PROOF. Let us apply Corollary 5.1.11 for the operator $p(T)$. Then spectral mapping theorem yields

$$\|p(T)\| = r(p(T)) = \max_{t \in \sigma(p(T))} |t| = \max_{t \in p(\sigma(T))} |t| = \max_{s \in \sigma(T)} |p(s)|$$

as claimed. □

---

[4] Here we use the notation $p(\sigma(T)) := \{p(t) : t \in \sigma(T)\}$.

5.3.4. Continuous functions of an operator. Let $T \in L(H,H)$ be a self-adjoint operator, and $f(t)$ be a continuous function on $\sigma(T)$. We would like to define $f(T) \in L(H,H)$. To this end, we use Weierstrass approximation theorem,[5] and we find polynomials $p_n(t)$ such that

$$p_n(t) \to f(t) \quad \text{uniformly on } \sigma(T). \tag{5.4}$$

This suggests us to define $f(T)$ as the limit of operator polynomials $p_n(T)$:

Definition/Proposition 5.3.11 (Continuous functions of an operator).

(i) The sequence $p_n(T)$ converges in $L(H,H)$ to a limit that we call $f(T)$.
(ii) The operator $f(T) \in L(H,H)$ is self-adjoint, and it does not depend on the choice of the approximating polynomials p_n.

Proof. (i) By completeness of $L(H,H)$ it suffices to check that $(p_n(T))$ is a Cauchy sequence in the operator norm. Computing the operator norm with the help of Corollary 5.3.10 and using (5.4), we obtain the desired conclusion that

$$\|p_n(T) - p_m(T)\| = \|(p_n - p_m)(T)\| = \max_{t\in\sigma(T)} |(p_n - p_m)(t)| \to 0 \quad \text{as } n, m \to \infty.$$

(ii) Since the operators $p_n(T)$ are self-adjoint, and the self-adjoint operators form a closed subset of $L(H,H)$ (Exercise 5.1.4), $f(T)$ is also self-adjoint. Furthermore, repeating the estimate in part (i), one sees that for any other approximating sequence of polynomials $q_n(t)$ one has $\|p_n(T) - q_n(T)\| \to 0$ as $n \to \infty$. It follows that the limit $f(T)$ must be the same whether one chooses $p_n(T)$ or $q_n(T)$ as an approximating sequence. $\square$

By passing to the limit in the corresponding properties for polynomials, one sees that for two polynomials f and g we have

$$(af + bg)(T) = a \cdot f(T) + b \cdot g(T), \quad (fg)(T) = f(T)g(T), \quad \overline{f}(T) = f(T^*).$$

(Check!) This property states in other words that for a fixed $T \in L(H,H)$, the map $f \mapsto f(T)$ is an *$*$-algebra homomorphism* from $C(\sigma(T))$ into $L(H,H)$.

Example 5.3.12. Consider an invertible self-adjoint operator $T \in L(H,H)$; then $\sigma(T) \in [m,M]$ with $m > 0$. Consider the function $f(t) = 1/t$, which is continuous on $[m,M]$. Then $f(T) = T^{-1}$ (Check!) In other words, we have the remarkable identity of reciprocal and inverse:

$$\frac{1}{T} = T^{-1}.$$

Lec. 39: 12/10

5.3.5. Spectral mapping theorem. We will now generalize the spectral mapping Theorem 5.3.9 from polynomials to continuous functions of an operator. It is based on the straightforward generalization of the invertibility Lemma 5.3.8:

Lemma 5.3.13 (Invertibility). *Let $T \in L(H,H)$ 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 $t \in \sigma(T)$.*

[5] Formally, since Weierstrass theorem is typically stated for functions on an interval, we first extend $f(t)$ to a continuous function on some interval $[m,M] \supseteq \sigma(T)$ (for example, the spectral interval). Such extension can be done e.g. by Tietze extension theorem.

PROOF. *Sufficiency.* If $f(t) \neq 0$ for all $t \in \sigma(T)$ then $1/f \in C(\sigma(T))$, so $1/f(T)$ is the inverse of T (see Example 5.3.12).

Necessity. Assume that $f(t_0) = 0$ for some $t_0 \in \sigma(T)$. Choose a sequence of polynomials $p_n(t)$ such that $p_n(t) \to f(t)$ uniformly on $s(T)$. Without loss of generality, we can assume that $p_n(t_0) = f(t_0) = 0$. (Justify this by a slight translation!) Then $p_n(T) \to f(T)$ in $L(H,H)$.

On the other hand, the invertibility Lemma 5.3.8 for polynomials, the operators $p_n(T)$ are not invertible. Since the non-invertible operators form a closed subset of $L(H,H)$ (see Exercise 2.4.44), it follows that $f(T)$ is not invertible. This completes the proof. □

Now the spectral mapping theorem follows from invertibility Lemma 5.3.13 by the same argument as the corresponding result for polynomials, Theorem 5.3.9:

THEOREM 5.3.14 (Spectral mapping theorem). *Let $T \in L(H,H)$ be a self-adjoint operator and $f \in C(\sigma(T))$. Then*

$$\sigma(f(T)) = f(\sigma(T)). \qquad \square$$

This gives a simple way to create positive operators:

COROLLARY 5.3.15. *Let $T \in L(H,H)$ be a self-adjoint operator and $f \in C(\sigma(T))$. If $f(t) \geqslant 0$ for all $t \in \sigma(T)$ then $f(T) \geqslant 0$.*

PROOF. By Corollary 5.3.5 it suffices to check that $\sigma(f(T)) \subseteq [0,\infty)$. The latter follows from the spectral mapping theorem, as $\sigma(f(T)) = f(\sigma(T)) \subseteq [0,\infty)$. □

EXERCISE 5.3.16. **Prove the converse statement in Corollary 5.3.15.**

It follows that the algebra homomorphism from $C(\sigma(T))$ into $L(H,H)$ preserves not only addition and multiplication, but also the order:

$$f(t) \leqslant g(t) \text{ for all } t \in \sigma(T) \quad \text{implies} \quad f(T) \leqslant g(T).$$

EXERCISE 5.3.17. **[Further properties of continuous functions of an operator] Let $T, S \in L(H,H)$ be self-adjoint operators and $f, g \in C(\sigma(T))$. Prove that:**

(i) $\overline{f}(T) = f(T)^*$**;**
(ii) $\|f(T)\| = \sup_{t\in\sigma(T)} |f(t)|$ **(this generalizes Corollary 5.3.10, and follows by the same argument);**
(iii) If operators S and T commute then operators $f(S)$ and $g(T)$ commute. (Check this for polynomials and pass to a limit.)

5.3.6. Square root of an operator. Consider a positive self-adjoint operator $T \in L(H,H)$. Then $\sigma(T) \subseteq [0,\infty)$. The function $f(t) = \sqrt{t}$ is continuous on $[0,\infty)$, so we can define $f(T) = \sqrt{T}$.

Since $(\sqrt{t})^2 = t$, the algebra homomorphism property implies that $(\sqrt{T})^2 = T$. Since $\sqrt{t} \geqslant 0$, Corollary 5.3.15 implies that $\sqrt{T}$ is a positive self-adjoint operator. Summarizing, we have proved the following (except uniqueness):

PROPOSITION 5.3.18 (Square root of an operator). *For every positive self-adjoint operator $T \in L(H,H)$, there exists a unique positive self-adjoint operator $\sqrt{T} \in L(H,H)$ such that*

$$(\sqrt{T})^2 = T.$$

EXERCISE 5.3.19. **Prove the uniqueness of $\sqrt{T}$.**

As an application of operator square root, we will now prove the following result, whose formulation has nothing to do with functional calculus

PROPOSITION 5.3.20. *Let $S, T \in L(H, H)$ be a pair of commuting self-adjoint operators. If $S \geqslant 0$ and $T \geqslant 0$ then $ST \geqslant 0$.*

For compact operators S, T this would follow because in this case one can find a common basis of eigenvectors in H. (Do this!) In general, I don't know any proof that won't involve functional calculus.

PROOF. We want to show that $\langle STx, x\rangle \geqslant 0$ for every $x \in H$. Since $\sqrt{S}$ and $\sqrt{T}$ are communing self-adjoint operators (see Exercise 5.3.17), and $\sqrt{ST} = \sqrt{S}\sqrt{T}$. (Why?) So we have

$$\langle STx, x\rangle = \langle \sqrt{S}\sqrt{S}\sqrt{T}\sqrt{T}x, x\rangle = \langle \sqrt{T}\sqrt{S}\cdot\sqrt{S}\sqrt{T}x, x\rangle = \langle \sqrt{S}\sqrt{T}x, \sqrt{S}\sqrt{T}x\rangle \geqslant 0$$

as required. □

5.3.7. Modulus of an operator. Now consider an arbitrary operator $T \in L(H, H)$. Then T^*T is a positive self-adjoint operator (check!) so it has a unique positive square root. Hence we define

$$|T| := \sqrt{T^*T}, \quad T \in L(H, H)$$

which we call the *modulus* of T. This generalizes the concept of modulus of complex numbers,

$$|z| = \sqrt{\bar{z}\, z}, \quad z \in \mathbb{C}.$$

EXAMPLE 5.3.21.
(i) For a diagonal operator $T = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ on $\mathbb{C}^n$, one has $|T| = \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_n|)$. (Check!)
(ii) For the multiplication operator $(Tf)(t) = g(t)f(t)$ on $L_2[0, 1]$, one has $(|T|f)(t) = |g(t)|f(t)$. (Check!)
(iii) For the right shift operator R on ℓ_2, we saw that $R^* = L$ (the left shift), so $|R| = I$.

LEMMA 5.3.22. *For every operator $T \in L(H, H)$ and vector $x \in H$, one has*

$$\||T|x\| = \|Tx\|.$$

PROOF. This is a matter of straightforward computation:

$$\||T|x\|^2 = \langle |T|x, |T|x\rangle = \langle |T|^2x, x\rangle = \langle T^*Tx, x\rangle = \langle Tx, Tx\rangle = \|Tx\|^2. \quad \square$$

Lec. 40: 12/13

5.3.8. Polar decomposition. Lemma 5.3.22 motivates us to consider a map

$$U : |T|x \mapsto Tx, \quad x \in H.$$

The following properties are obvious:

(i) U is well defined;
(ii) U is a linear operator.;
(iii) U is an isometry, i.e. $\|Uy\| = \|y\|$ for all $y \in \operatorname{Im}(|T|)$ (by Lemma 5.3.22);
(iv) $\operatorname{Im}(U) = \operatorname{Im}(T)$.

We have proved the following result (except uniqueness):

Theorem 5.3.23 (Polar decomposition). *For every operator $T \in L(H, H)$, there exists a unique bijective linear isometry $U \in L(\mathrm{Im}(|T|), \mathrm{Im}(T))$ such that*

$$T = U|T|.$$

The uniqueness of U follows easily: $Tx = U|T|x$ means that U takes $|T|x$ to Tx, thus U is uniquely determined on $\mathrm{Im}(|T|)$.

Theorem 5.3.23 generalizes the polar decomposition in the complex plane. The latter states that every $z \in \mathbb{C}$ can be represented as

$$z = e^{i \operatorname{Arg}(z)} |z|.$$

Here $e^{i \operatorname{Arg}(z)}$ is a unit scalar (generalized by U), and $|z|$ is the modulus of z (generalized by $|T|$).

Remark 5.3.24. In general, U can not be extended to a bijective linear isometry on the whole space H. Indeed, if T is the right shift on ℓ_2 then $|T| = I$, so the polar decomposition yields $U = T$. Although $U = T$ is defined on the whole ℓ_2, its image is not even dense in ℓ_2, so U is not bijective on ℓ_2.

However, for invertible operators T, it is possible to extend U to a bijective isometry on the whole space:

Theorem 5.3.25 (Polar decomposition for invertible operators). *For every operator $T \in L(H, H)$, there exists a unique unitary operator $U \in L(H, H)$ such that*

$$T = U|T|.$$

Proof. Since T is invertible, T^* is invertible, so T^*T is invertible, and finally $|T| = \sqrt{T^*T}$ is invertible. (Why? Use Example 4.4.26.) Therefore $\mathrm{Im}(T) = \mathrm{Im}(|T|) = H$, and the claim follows from Theorem 5.3.23. □

Remark 5.3.26. This second form of the polar decomposition holds also for all normal operators (those with $T^*T = TT^*$), compact + scalar operators, and generally for all operators for which $\dim \ker(T) = \dim \ker(T^*)$.

5.4. Borel functional calculus. Spectral theorem for self-adjoint operators

In this section, we extend functional calculus to bounded Borel functions of operators. This is done primarily to define the spectral projections, which are indicator functions of an operator. Once we have spectral projections, we formulate the spectral theorem for general (not necessarily compact) self-adjoint operators.

As usual, $T \in L(H, H)$ will denote a fixed self-adjoint operator on a Hilbert space H. Let us fix an interval $[m, M]$ which contains the spectrum $\sigma(T)$, e.g. the spectrum interval.

5.4.1. Borel functional calculus. We consider the linear space of bounded Borel complex-valued functions on $[m, M]$; denote this space $\mathcal{B}[m, M]$. We would like to define $f(T)$ for $f \in \mathcal{B}[m, M]$, so that this extends the definition of $f(T)$ for continuous functions $f \in C[m, M]$. The difficulty is that Borel functions are only pointwise (but not uniform) limits of continuous functions.

Theorem 5.4.1 (Borel functional calculus). *Let $T \in L(H, H)$ be a self-adjoint operator on a Hilbert space H. For each $f \in \mathcal{B}[m, M]$ one can define a self-adjoint operator $f(T) \in L(H, H)$ such that*

(i) $(af+bg)(T) = a \cdot f(T) + b \cdot g(T)$, $(fg)(T) = f(T)g(T)$, $\overline{f}(T) = f(T^*)$. *In other words, the map* $f \mapsto f(T)$ *is an* $*$-algebra homomorphism *from* $\mathcal{B}[m,M]$ *into* $L(H,H)$.
(ii) *If* $f_n, f \in \mathcal{B}[m,M]$ *satisfy* $\sup_n \|f_n\|_\infty < \infty$ *and* $f_n \to f$ *pointwise then* $f_n(T) \to f(T)$ *pointwise in* $L(H,H)$.
(iii) $\|f(T)\| \leqslant \|f\|_\infty$.
(iv) *If* T *and* S *commute then* $f(T)$ *and* $g(s)$ *commute for* $f, g \in \mathcal{B}[m,M]$.

Proof. *Construction of* $f(T)$. For a fixed pair of vectors $x, y \in H$, define a linear functional $F_{x,y}$ on $C[m,M]$ by

$$F_{x,y}(f) := \langle f(T)x, y \rangle, \quad f \in C[m,M].$$

This functional is bounded:

$$|F_{x,y}(f)| \leqslant \|f(T)\|\|x\|\|y\| \leqslant \|f\|_\infty \|x\|\|y\|,$$

where the last inequality follows by the spectral mapping Theorem 5.3.14. We have shown that

$$F_{x,y} \in (C[m,M])^* \quad \text{and} \quad \|F_{x,y}\| \leqslant \|x\|\|y\|.$$

By representation Theorem 2.2.8 for $C(K)^*$, there exists a unique Borel regular signed measure $\mu_{x,y}$ on $[m,M]$ with total variation $|\mu_{x,y}|([m,M]) \leqslant \|x\|\|y\|$, and such that

$$\langle f(T)x, y \rangle = \int_m^M f(\lambda)\, d\mu_{x,y}(\lambda), \quad f \in C[m,M].$$

We extend this to bounded Borel functions by defining

$$(Bf)(x,y) := \int_m^M f(\lambda)\, d\mu_{x,y}(\lambda), \quad f \in \mathcal{B}[m,M].$$

(Since $\mu_{x,y}$ is a Borel measure, the integral is defined.) One quickly checks using the definition of $\mu_{x,y}$ that $(Bf)(x,y)$ is linear in x and conjugate linear in y. So B is what is called a *sesquilinear form.* Moreover, this form is bounded:

$$|(Bf)(x,y)| \leqslant \|f\|_\infty \cdot |\mu_{x,y}|([m,M]) \leqslant \|f\|_\infty \|x\|\|y\|.$$

It follows (see Exercise 2.4.45) that B is given by a linear operator that we shall call $f(T) \in L(H,H)$ acting as

$$(Bf)(x,y) = \langle f(T)x, y \rangle, \quad f \in \mathcal{B}[m,M],\ x, y \in H,$$

and that $\|f(T)\| \leqslant \|f\|_\infty$. We thus defined $f(T)$ for all bounded Borel functions f, and we verified (iii).

(ii) with weak convergence. Let us try to prove (ii). By the dominated convergence theorem, for every $x, y \in H$ we have

$$\langle f_n(T)x, y \rangle = \int_m^M f_n(\lambda)\, d\mu_{x,y}(\lambda) \to \int_m^M f(\lambda)\, d\mu_{x,y}(\lambda) = \langle f(T)x, y \rangle.$$

This establishes a somewhat weaker conclusion than in (ii), namely that $f_n(T)x \to f(T)x$ weakly for all $x \in H$. We will prove the actual statement (ii) in a minute.

(i). All these identities hold for continuous functions f, and they are preserved under pointwise limits. As every Borel function is a pointwise limit of continuous functions, the conclusion follows by applying (ii) with weak convergence. (Check!) The same argument proves (iv), and it also establishes xthat $f(T)$ is a self-adjoint operator.

(ii) completed. Now we complete the proof of (ii) by showing that $f_n(T)x \to f_n(T)x$ in H. To this end, we bound

$$\begin{aligned}
\|(f_n - f)(T)x\|^2 &= \big\langle (f_n - f)(T)x, (f_n - f)(T)x \big\rangle \\
&= \big\langle ((f_n - f)(T))^2 x, x \big\rangle \\
&= \big\langle ((f_n - f)^2(T))x, x \big\rangle \quad \text{(by part (i), we have } (g^2)(T) = (g(T))^2) \\
&\to 0
\end{aligned}$$

where the last line follows by part (ii) with weak convergence, as $(f_n - f)^2 \to 0$ pointwise. This completes the proof. □

Exercise 5.4.2. **Show that if $f(t) \geqslant 0$ for all t then $f(T) \geqslant 0$.**

5.4.2. Spectral measures. Let us fix a self-adjoint operator $T \in L(H, H)$. In the proof of Borel functional calculus, we constructed Borel regular measures $\mu_{x,y}(T)$ for every pair of vectors $x, y \in H$. They are called *spectral measures* for T. We showed that spectral measures satisfy the identity

$$\langle f(T)x, y \rangle = \int_m^M f(\lambda)\, d\mu_{x,y}(\lambda) \tag{5.5}$$

valid for every bounded Borel function $f \in \mathcal{B}[m, M]$. In particular, the bilinear form of T can be reproduced using spectral measures as

$$\langle Tx, y \rangle = \int_m^M \lambda\, d\mu_{x,y}(\lambda). \tag{5.6}$$

Exercise 5.4.3. **Compute the spectral measures for the diagonal matrix $T = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ acting as an operator on $\mathbb{C}^n$.**

Exercise 5.4.4. **Let T be a multiplication operator in $L_2[0, 1]$ by a function $g \in L_\infty[0, 1]$. Show that for $f \in \mathcal{B}[0, 1]$, the operator $f(T)$ is the multiplication operator in $L_2[0, 1]$ by the function $f(g(t))$.**

5.4.3. Spectral projections. Let E be a Borel subset of $[m, M]$, and we consider the indicator function

$$\mathbf{1}_E(t) = \begin{cases} 1, & t \in E \\ 0, & t \notin E. \end{cases}$$

We shall consider the operator $P_E = \mathbf{1}_E(T) \in L(H, H)$.

Since $\mathbf{1}_E(t)^2 = \mathbf{1}_E(t)$ and $\mathbf{1}_E(t)$ is real valued, the homomorphism property (iii) in Theorem 5.4.1 yields that P_E is an orthogonal projection in H (see Exercise 5.1.12).

The projections P_E are called *spectral projections* associated with T.

Exercise 5.4.5. **Show that if $E_1 \subseteq E_2$ then $P_{E_1} \leqslant P_{E_2}$ and $\operatorname{Im}(E_1) \subseteq \operatorname{Im}(E_2)$.**

Proposition 5.4.6 (Projection-valued measure). *The spectral projections P_E associated with an operator T have the following properties:*

(i) $P_\varnothing = 0$, $P_{[m,M]} = I$;

(ii) For every decomposition $[m, M] = \bigcup_{k=1}^{\infty} E_k$ into disjoint sets Borel E_k, one has

$$I = \sum_{n=1}^{\infty} P_{E_n},$$

where the pointwise convergence of the series.

These properties are strikingly similar to the axioms of a Borel probability measure on $[m, M]$. However, the usual probability measures are functions $\lambda : E \mapsto \lambda(E)$ with values in $[0, 1]$, while P_E takes values in $L(H, H)$, more specifically in the operator interval between 0 and I. For this reason, the assignment $P_\lambda : E \mapsto P_E$ for Borel subsets $E \subseteq [m, M]$ is called a *projection-valued measure* for T.

The projection-valued measure can be reduced to the usual measure in a simple way. For $x \in H$, the assignment $E \mapsto \langle P_E x, x\rangle$ is clearly a usual Borel measure on $[m, M]$. This follows from Proposition 5.4.6. Moreover, (5.5) shows that this is nothing else than the spectral measure $\mu_{x,x}$:

$$\langle P_\lambda(E)x, x\rangle = \langle P_E x, x\rangle = \int_m^M \mathbf{1}_E(t)\, d\mu_{x,x}(\lambda) = \mu_{x,x}(E).$$

For this reason, the projection-valued measure P_λ itself, rather than $\mu_{x,x}$, is often called the *spectral measure* associated with the operator T.

5.4.4. Spectral theorem for self-adjoint operators.

THEOREM 5.4.7 (Spectral theorem). *Let $T \in L(H, H)$ be a self-adjoint operator on a Hilbert space H. Then*

$$T = \int_{-\infty}^{\infty} \lambda\, dP_\lambda$$

where P_λ is the projection-valued measure associated with T. The integral in fact can be taken over a finite interval $[m, M]$ containing the spectrum of T.

The integration with respect to the operator-valued measure P_λ is understood in the sense that

$$\langle Tx, x\rangle = \int_m^M \lambda \langle dP_\lambda x, x\rangle;$$

As we noted, $\langle dP_\lambda x, x\rangle$ is just the spectral measure $\mu_{x,x}$, so the last integral is the usual Lebesgue integral.

PROOF. With this remark, Theorem 5.4.7 is a reformulation of a partial case of (5.6):

$$\langle Tx, x\rangle = \int_m^M \lambda\, d\mu_{x,x}(\lambda). \qquad \square$$

Theorem 5.4.7 should be compared to the spectral Theorem 5.3 for *compact* self-adjoint operators T. According to this theorem, T can be decomposed into the sum

$$T = \sum_k \lambda_k P_k$$

where λ_k are the eigenvalues of T and P_k denotes the orthogonal projection in H onto the eigenspace corresponding to the eigenvalue λ_k, see (5.3).

As we know, for general operators T (not necessarily compact), the spectrum of T may no longer be countable. So the sum is replaced by the integral in Theorem 5.4.7.

Bibliography

[1] K. Ball, *An elementary introduction to modern convex geometry.* Flavors of geometry, 1–58, Math. Sci. Res. Inst. Publ., 31, Cambridge Univ. Press, Cambridge, 1997.

[2] L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Mathematica 116 (1966), 135–157.

[3] P. Enflo, *A counterexample to the approximation problem in Banach spaces.*, Acta Math. 130 (1973), 309–317.

[4] E. Gluskin, *The diameter of the Minkowski compactum is roughly equal to n*, Funktsional. Anal. i Prilozhen. 15 (1981), 72–73.

[5] W. T. Gowers, B. Maurey, *The unconditional basic sequence problem*, J. Amer. Math. Soc. 6 (1993), 851–874

[6] J. Lindenstrauss, Joram, *On complemented subspaces of m*, Israel J. Math. 5 (1967), 153–156.

[7] J. Lindenstrauss, L. Tzafriri, *On the complemented subspaces problem*, Israel J. Math. 9 (1971), 263–269.

[8] J. Lindenstrauss, L. Tzafriri, *Classical Banach spaces. I and II.* Springer-Verlag, , 1977, 1979.

[9] A. Pelczkynski, *Projections in certain Banach spaces*, Studia Math. 19 (1960), 209–228.

[10] R. S. Phillips, *On linear transformations*, Trans. Amer. Math. Soc. 48 (1940), 516–541.

Index