Introduction to Functional Analysis

Yen Do

Fall 2015
Preface

This is the accompanying expository notes for an introductory course in Functional Analysis that I was teaching at UVA. The goal of the course is to study the basic principles of linear analysis, including the spectral theory of compact and self-adjoint operators. This is not a monograph or a treatise and of course no originality is claimed. The prerequisite is some basic knowledge about real analysis and topology. Some preliminary understanding of functional analysis is beneficial but not required.
Contents

I Linear Spaces 9

1 Basic facts 11
1.1 Metric spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
1.1.1 Separability . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
1.2 Linear spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
1.2.1 The Hahn-Banach extension theorem . . . . . . . . . . . . . . . . . . . 15
1.2.2 The HB theorem with symmetry constraints . . . . . . . . . . . . . . . 17
1.3 Topological spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
1.3.1 Compactness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
1.3.2 Continuous functions on LCH spaces . . . . . . . . . . . . . . . . . . . 19
1.3.3 Proof of Urysohn’s lemma . . . . . . . . . . . . . . . . . . . . . . . . . 20
1.3.4 Proof of Tietze’s theorem . . . . . . . . . . . . . . . . . . . . . . . . . 21
1.3.5 Partition of unity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22

2 Overview of normed linear spaces 23
2.1 Linear functionals and dual spaces . . . . . . . . . . . . . . . . . . . . . . . . 25
2.1.1 Linear functionals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
2.1.2 Dual spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
2.1.3 Reflexivity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
2.2 (Non)compactness of the unit ball . . . . . . . . . . . . . . . . . . . . . . . . 28

3 Geometry and topology on NLS 31
3.1 Topologies on normed linear spaces . . . . . . . . . . . . . . . . . . . . . . . 31
3.1.1 The Banach–Alaoglu theorem . . . . . . . . . . . . . . . . . . . . . . . 31
3.1.2 The sequential Banach-Alaoglu theorem . . . . . . . . . . . . . . . . . 32
3.1.3 Weak compactness of the closed unit ball in reflexive spaces . . . . . . 33
3.2 Uniform convexity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.3 Isometries between normed linear spaces . . . . . . . . . . . . . . . . . . . . 35

4 Basis on Hilbert spaces and Banach spaces 39
4.1 Basic properties . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
4.2 Orthogonal basis for a Hilbert space . . . . . . . . . . . . . . . . . . . . . . . 40
4.3 Schauder basis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
##CONTENTS

5 Bounded maps on Banach spaces 45
  5.1 The Baire category theorem .................................. 45
  5.2 The principle of uniform boundedness ............................ 45
  5.3 The open mapping theorem .................................... 46
  5.4 The closed graph theorem ..................................... 47
  5.5 Uniform boundedness of canonical Schauder projections ....... 47
  5.6 $L^p$ spaces .................................................... 48
    5.6.1 Separability ............................................... 48
    5.6.2 Duality .................................................... 48
    5.6.3 Basic facts about bounded operators on $L^p$ ............... 50

6 Bounded continuous functions and dual spaces 53
  6.1 Riesz representation theorems ................................ 53
    6.1.1 The dual space of $L^\infty$ ................................ 53
    6.1.2 Measures on a locally compact Hausdorff spaces ......... 54
    6.1.3 The dual space of $C_0(X)$ and $C_c(X)$ ................... 54
  6.2 The Stone-Weierstrass approximation theorem .................. 56

7 Locally convex spaces 59
  7.1 Two equivalent definitions of LCS ................................ 59
    7.1.1 Basic properties ....................................... 60
    7.1.2 Linear maps ............................................. 61
  7.2 The Hyperplane separation theorem .............................. 61
  7.3 The Krein-Milman theorem .................................... 62
  7.4 Inductive limit and weak solutions ............................. 63

II Spectral analysis for linear operators 65

8 Elementary spectral theory 67
  8.1 Spectrum and resolvent set .................................. 67
  8.2 Functional calculus and spectral mapping ....................... 68
  8.3 Examples of operators and their spectra ....................... 69
  8.4 Adjoint operators and spectrum ................................ 70

9 Compact operators 73
  9.1 Compact operators ........................................... 73
  9.2 Compactness of integral operators ............................. 75
  9.3 Spectral properties of compact operators ....................... 76
    9.3.1 Riesz’s theorem ......................................... 76
    9.3.2 Spectral properties .................................... 77
10 Bounded self-adjoint operators
   10.1 Diagonalization form ................................................. 79
   10.2 Projection-valued measure and spectral projection ............... 81
   10.3 Spectral representation and decomposition ....................... 82

11 Spectral theory for unbounded self-adjoint operator ............. 85

12 Fredholm determinant
   12.1 Motivation ................................................................. 89
   12.2 Fredholm's approach for integral operators ....................... 90
      12.2.1 Convergence and Continuity of the determinant ............. 91
   12.3 Fredholm determinant for operators on Hilbert spaces .......... 95
Part I

Linear Spaces
Chapter 1
Preliminaries

Generally speaking, a space is a set with some structures, which could be
- algebraic (add, multiply)
- geometric (distance, angle, convexity)
- topological (open sets, continuity)

Our goal is to study spaces of functions and their structures using analytic tools. One could think of this study as one place where “analysis” meets “linear algebra” and “geometry/topology”.

Our focus will be on linear spaces with some notion of geometry/topology.

Some times further understanding of these structures could be obtained via looking at functions on these spaces, which could takes value in $\mathbb{R}$, $\mathbb{C}$, or even in other spaces with similar structure. We refer to these functions are operators/maps/functionals depending on the nature of the target spaces.

In this chapter we will recall several basic facts about metric spaces, linear spaces, and topological spaces.

1.1 Metric spaces

Here we have an additional geometric structure (distance) on the given set $M$, namely $d$ measures the distance between two elements of $M$, such that

- (positive) $d(x, y) \geq 0$ equality iff $x = y$, and
- (symmetric) $d(x, y) = d(y, x)$, and
- (triangle inequality) $d(x, y) + d(y, z) \geq d(x, z)$.

A sequence $(x_n)$ converges to $x$ in $(M, d)$ if $\lim_{n \to \infty} d(x_n, x) = 0$ A sequence $(x_n)$ is called a Cauchy sequence in $(M, d)$ if $\lim_{\min(m,n) \to \infty} d(x_n, x_m) = 0$.

It is clear that “convergence” $\Rightarrow$ “Cauchy”.

If the reverse holds then the metric space is called complete.
Definition 1. A metric space \((M, d)\) is complete if any Cauchy sequence is convergent. (Note that the sequence and the limit are required to be in \(M\)).

Examples: \(M = \{\) continuous functions on \([0, 1]\) \}, let
\[
\begin{align*}
    d_1(f, g) &= \sup_{x \in [0,1]} |f(x) - g(x)| \\
    d_2(f, g) &= \int_0^1 |f(x) - g(x)| \, dx
\end{align*}
\]
Then \((M, d_1)\) is complete but \((M, d_2)\) is not complete.

Question: if \((M, d)\) is incomplete can we add more stuffs to make it complete?

Definition 2. \(M_1 \subset M_2\) is dense in \(M_2\) wrt metric \(d\) if every \(m \in M_2\) could be approximated by a sequence in \(M_1\). In other words for some sequence \((a_n)\) in \(M_1\) we have
\[
\lim_{n \to \infty} d(a_n, m) = 0.
\]

Theorem 1. If \((M, d)\) is a metric space then there exists a complete metric space \((\tilde{M}, \tilde{d})\) and a map \(h : M \to \tilde{M}\) such that
\[
d(m_1, m_2) = \tilde{d}(h(m_1), h(m_2))
\]
and \(h(M)\) is dense in \(\tilde{M}\).

(such a \(h\) is called an isometry between \(M\) and \(h(M)\).)

Ideas of the proof: One starts by considering the set of all Cauchy sequences in \(M\). Inside the new space \(\tilde{M}\) these sequences should converge to a limit, although two different sequences may have the same limit. The idea is to identify these sequences together using some equivalence relation and build \(\tilde{M}\) from there.

More specifically, for two Cauchy sequences \(\mathbf{x} = (x_n)\) and \(\mathbf{y} = (y_n)\) in \(M\) let
\[
d_0(\mathbf{x}, \mathbf{y}) = \lim_{n \to \infty} d(x_n, y_n).
\]
It is clear that such limit exists since \(d(x_n, y_n)\) is a Cauchy sequence of real numbers and it is not hard to verify that this is indeed a metric.

If \(d_0(\mathbf{x}, \mathbf{y}) = 0\) we say that \(\mathbf{x}\) and \(\mathbf{y}\) are equivalent, and we then let \(\tilde{M}\) be the set of all equivalence classes, and \(\tilde{d}\) be the induced metric on \(\tilde{M}\).

Turns out \((\tilde{M}, \tilde{d})\) is complete (this requires some work using the Cantor diagonal trick), and we could embed \(M\) inside \(\tilde{M}\) using the following isometry: for \(x \in M\),
\[
h(x) = [(x, x, \ldots)]
\]
the equivalence class of \((x, x, x \ldots)\) in \(\tilde{M}\). □

It is not hard to see that if the given metric has some reasonable symmetries then they are preserved in the completion space. For instance, this applies if the space is linear and the metric is dilation invariant or translation invariant (for instance if the metric is induced from a norm). Thus we could easily see that the completion theorem also holds for normed linear spaces.
1.1. METRIC SPACES

1.1.1 Separability

We recall that a topological space is separable if there exists a countable dense set. We say that this set separates the space; for instance the rationals separate the real numbers.

A metric space is said to be separable if the topology induced from the metric is separable.

Note that there is also the concept of second countable, which says that one could find a countable collection open sets from which one could get all open sets in the topology simply by taking union.

This would imply separability (just take one point from each such open set), however for metric/metrizable spaces these two are equivalent.

Separability is important in constructive proofs, for instance we could typically avoid Zorn’s lemma/the axiom of choice if separability is available.

**Basic properties:**
1. All compact metric spaces are separable.
2. If the given topological space is an union of a countable number of separable subspaces, then it is also separable.

Combining Properties 1 and 2, it is clear that an Euclidean space $\mathbb{R}^I$ (consisting of functions from $I$ to $\mathbb{R}$) is separable if any only the dimension $|I|$ is countable.

3. A metric space is **not** separable if there is an uncountable collection of functions such that the distance between any two is at least 1. (Indeed, if one could find a countable dense set then at least an element of this dense set has to be close to two functions in the given collection, which by the triangle inequality would imply that the distance between the two has to be small, contradictory to the given hypothesis).

Using Property 3, we could show that $L^\infty[0,1]$ (or $L^\infty(\mathbb{R})$ etc.) is not separable. To see this, one could simply take the collection $C$ of characteristic functions of intervals $[a,b] \subset [0,1]$ where $a,b$ are irrationals – there are uncountably many such functions and any two elements of $C$ are of $L^\infty$ distance at least 1, contradiction.

Using a similar line of reasoning, one could see that the space of all signed Borel measures on $[0,1]$ (with norm being the total mass) is not separable. Here simply use the (uncountable) collection of delta measures at points of $[0,1]$.

4. We could also define separability for measure spaces by defining a metric on the underlying $\sigma$-algebra. Namely, given a $\sigma$-algebra $\mathcal{A}$ on $X$ and a corresponding measure $\mu$ we could define a metric on $\mathcal{A}$ by

$$p(A, B) = \mu(A \Delta B)$$

the measure of the symmetric difference

$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

and we say that $(X, \mathcal{A}, \mu)$ is separable if the metric space $(\mathcal{A}, p)$ is separable.

**Lemma 1.** If a $\sigma$-algebra is generated from a countable collection of sets, then it is separable.

**Proof.** Let $A$ be a countable generating set, let $A'$ consist of all finite intersections of elements of $A$, and $A''$ consist of all finite unions of element of $A'$, clearly $A''$ is still countable and it
is an algebra of set (i.e. the union and intersection of any two elements are still inside $A''$).
Clearly $A''$ also generates the given $\sigma$-algebra $\mathcal{A}$, thus to show separability it suffices to show
that given any $E \in \mathcal{A}$ and any $\epsilon > 0$ one could find $F \in A''$ such that $\mu(E \Delta F) < \epsilon$. This is
done by looking at the set of all such $E$ and show that it is actually a $\sigma$ algebra containing $A''$, thus by minimality it has to contain $\mathcal{A}$ and the desired claim follows. □

Exercise: $\mathbb{R}^n$ with the usual Lebesgue measure is separable.

**Theorem 2.** Let $(X, \mathcal{A}, \mu)$ be a separable measure space. Then for any $1 \leq p < \infty$ the space $L^p(X, \mathcal{A}, \mu)$ is separable.

Proof: use simple functions. More details later.

### 1.2 Linear spaces

Here the additional structure is algebraic, namely we allow for addition of two elements and
a somewhat limited notion of multiplication. Let $F$ be a field (for us this is $\mathbb{R}$ or $\mathbb{C}$), we
allow for multiplication of an element of the space with an element of $F$, the space is then
said to be linear over $F$.

**Examples:**
1. finite dimensional vector spaces (linear algebra).
2. Let $p \in (0, \infty)$. Then $L^p(\mathbb{R}) = \text{all (complex valued Borel) measurable functions on } \mathbb{R}$
such that
   \[
   \left( \int_{\mathbb{R}} |f(x)|^p dx \right)^{1/p} < \infty
   \]
   Similarly $L^\infty(\mathbb{R})$ is all measurable $f$ such that for some $M > 0$ finite the set $\{x : |f(x)| > M\}$
     has measure zero. Note that $f$ could be unbounded.

   $L^p$ spaces are strong type spaces. We could also define weak $L^p(\mathbb{R})$, denoted by $L^{p, \infty}(\mathbb{R})$
to be all measurable $f$ such that
   \[
   \sup_{\lambda > 0} \lambda \left( \left\{|f| > \lambda \right\} \right)^{1/p} < \infty
   \]
   For $p = \infty$, weak $L^\infty$ and $L^\infty$ are the same.

   One could further generalize this to $L^p$ on an arbitrary measure space. For instance we
could replace the Lebesgue measure with another Borel measure on $\mathbb{R}$.

3. $\ell^p(\mathbb{N}) = \text{all complex sequences } (a_1, a_2, \ldots)$ such that
   \[
   \left( \sum_{n \geq 1} |a_n|^p \right)^{1/p} < \infty
   \]
   Now $\ell^\infty = \text{all bounded sequences}$.

   **Exercise:** prove that these spaces are actually linear spaces, namely they are closed
under addition and scalar multiplication.
1.2. LINEAR SPACES

1.2.1 The Hahn-Banach extension theorem

Given a linear space \( X \) and \( p : X \to \mathbb{R} \), we say \( p \) is homogeneously convex if

- (positive homogeneity) \( p(ax) = ap(x) \) for \( a > 0 \) and \( x \in X \), and
- (subadditivity) \( p(x_1 + x_2) \leq p(x_1) + p(x_2) \) for \( x_1, x_2 \in X \).

Intuitively, the graph of \( p \) over \( X \) looks like a convex cone with top at the origin.

Note that this is equivalent to

\[
p(ax_1 + bx_2) \leq ap(x_1) + bp(x_2)
\]

for all \( a, b \geq 0 \) with \( a + b > 0 \) and \( x_1, x_2 \in X \).

The following is known as the HB extension theorem for linear spaces over \( \mathbb{R} \) (there is a version for \( \mathbb{C} \) discussed later). Note that it does not requires geometric structures on \( X \), all we need is linearity.

**Theorem 3.** Let \( X \) be a linear space over \( \mathbb{R} \) and let \( p : X \to \mathbb{R} \) be a homogeneously convex function. Let \( Y \subset X \) linear subspace and \( \ell : Y \to \mathbb{R} \) linear such that

\[
\ell(y) \leq p(y) \quad \text{for all } y \in Y.
\]

Then there exists \( L : X \to \mathbb{R} \) linear that agrees with \( \ell \) on \( Y \) such that

\[
L(x) \leq p(x) \quad \text{for all } x \in X.
\]

One could think of this theorem as an example of a **local to global** principle for linear maps: as long as the local constraint is reasonable we could preserve it globally. Here we need homogeneity and convexity.

Proof of the HB theorem in this generality uses Zorn’s lemma which is equivalent to the axiom of choice. (for more concrete \( X \) such as \( L^p, \ell^p \), one could avoid the axiom of choice.)

A relation \( R \) on a set \( X \) is a subset of \( X \times X \) (could be empty!). We say \( xRy \) if \( (x, y) \in R \). \( R \) is an equivalence relation if three conditions holds:

- (reflexive) \( xRx \) for all \( x \in X \),
- (symmetric) if \( xRy \) then \( yRx \), and
- (transitive) if \( xRy \) and \( yRz \) then \( xRz \).

We say \( R \) is a partial ordering if reflexive, transitive, and anti-symmetric (namely if \( xRy \) and \( yRx \) then \( x = y \)).

We say \( R \) is a linear ordering if \( x, y \in X \) with \( x \neq y \) then either \( xRy \) or \( yRx \). In this case we sometimes say that \( X \) is a linearly ordered chain. Note that the size a chain could be uncountable.
We say \( \alpha \in X \) is an upper bound for \( Y \subset X \) if \( yR\alpha \) for all \( y \in Y \).

Example: the inclusion order in \( \mathcal{P}(\mathbb{R}) \).

We say that an element \( m \in X \) is maximal if for every \( \alpha \in X \) if \( mR\alpha \) then \( \alpha = m \).

**Lemma 2** (Zorn’s lemma). \( X \neq \emptyset \), \( R \) is a partial ordering on \( X \) such that any linearly ordered chain \( Y \) inside \( X \) has an upper bound in \( X \). Then for each such \( Y \) there is a maximal \( m \in X \) such that \( m \) is an upper bound for \( Y \).

Note that it is very important that we could find an upper bound for a chain. Without that we have a counter example: \( X = \mathbb{Z}_+ \) is the set of all positive integers, and the chain \( 1 \leq 2 \leq \ldots \) does not have any (maximal) upper bound.

Technically speaking, this assumption is used in the proof of the Lemma (via say, transfinite induction plus the axiom of choice); details could be found in most introductory textbooks in mathematical logic.

**Proof of the HB theorem.** We divide the proof into two steps. In step 1, we show that if \( Y \) is a proper subspace of \( X \) then we can increase the dimension of \( Y \) while preserving the given hypothesis. In step 2, by induction on the dimension of \( Y \) (iterating step 1) and using Zorn’s lemma we obtain the whole space \( X \).

**Step 1:** Assume existence of \( z \in X \setminus Y \). Let \( \tilde{Y} = \text{span}(Y, z) = \{ay + bz, a, b \in \mathbb{R}, y \in Y\} \).

Want to define \( \ell_0 \) such that \( \ell_0(y) = \ell(y) \) and

\[
\ell_0(ay + bz) = a\ell_0(y) + b\ell_0(z)
\]

so only need \( \ell_0(z) \). We also want

\[
\ell_0(ay + bz) \leq p(ay + bz)
\]

thus

\[
b\ell_0(z) \leq p(ay + bz) - a\ell(y)
\]

all \( a, b \in \mathbb{R} \) and \( y \in Y \). Just need \( a = 1 \).

By considering \( b > 0 \) and \( b < 0 \) we end up needing

\[
\sup_{c > 0, y \in Y} \frac{1}{c} [p(y - cz) - \ell(y)]
\]

\[
\leq \inf_{c > 0, y \in Y} \frac{1}{c} [p(y + cz) - \ell(y)]
\]

After algebraic manipulations this follows from the given convexity assumption on \( p \), more specifically the following estimate, \( c_1, c_2 > 0 \),

\[
p\left( \frac{c_1}{c_1 + c_2}x + \frac{c_2}{c_1 + c_2}y \right) \leq \frac{c_1}{c_1 + c_2}p(x) + \frac{c_2}{c_1 + c_2}p(y)
\]
Step 2: Here we iterate step 1 and use Zorn’s lemma. If $X$ is finite dimensional then it is clear that the process will stop after finitely many iterations. If $X$ is infinitely dimensional (the dimension of $X$ could even be uncountable), we need to be able to go beyond any linearly ordered chain of these extension in order to apply Zorn’s lemma.

Formally, consider the set of all possible extensions of $\ell$ from $Y$ to bigger subspaces of $X$ while remain dominated by $p$. Define a partial order where $(Y_\alpha, \ell_\alpha) \leq (Y_\beta, \ell_\beta)$ if $Y_\alpha \subset Y_\beta$ and $\ell_\alpha$ agrees with $\ell_\beta$ on $Y_\alpha$. Now any linearly ordered chain $(Y_\alpha, \ell_\alpha)_{\alpha \in I}$ (here $I$ could be uncountable) has an upper bound $(Y', \ell')$ defined by

$$Y' = \bigcup_{\alpha \in I} Y_\alpha$$

and $\ell'$ is the same as $\ell_\alpha$ inside $Y_\alpha$. Now apply Zorn’s lemma. □

Corollary 1 (complex HB). $X$ is a complex vector space, $p : X \rightarrow \mathbb{R}_+$ such that

$$p(ax + by) \leq |a|p(x) + |b|p(y)$$

all $|a| + |b| > 0$ complex numbers. Assume that $\ell : Y \rightarrow \mathbb{C}$ linear functional such that $|\ell(y)| \leq p(y)$ all $y \in Y$. Then can extend $\ell$ linearly to $X$ so that it is still dominated by $p(y)$.

Note: the extension given in the HB theorems is not necessarily unique!

Proof. Idea of proof: $\ell_1(y) = Re(\ell(y))$ and $\ell_2(y) = Im(\ell(y))$ are real functionals bounded by $p$, and $\ell_2(y) = -i\ell_1(iy)$. So

$$\ell(y) = \ell_1(y) - i\ell_1(iy)$$

By real HB we could extend $\ell_1$ to $X$, thus we extend $\ell$ to. But still need to show that the extended $\ell$ satisfies

$$|\ell(x)| \leq p(x)$$

For any $x$, let $\alpha$ be such that $|\alpha| = 1$ and $\ell(x) = \alpha|\ell(x)|$, then the desired estimate follows:

$$|\ell(x)| = \ell(\alpha^{-1}x) = \ell_1(\alpha^{-1}x) \leq p(\alpha^{-1}x) = p(x)$$

□

We emphasize again that technically speaking the HB extension theorem applies to any linear spaces; although heuristically speaking the existence of $p$ implies some implicit geometric structures.

1.2.2 The HB theorem with symmetry constraints

One possible way to generalize the HB theorem is to introduce additional symmetries. More specifically, if the local space $Y$ and the linear map $\ell$ (on $Y$) and the convex function $p$ (defined globally on $X$) are invariant under some “compatible” set of linear transformations, we would also like to impose this on our extension $\ell$. 


More specifically, let $\mathcal{A}$ be a set of linear maps from $X$ to itself that commute, thus if $T_1, T_2 \in \mathcal{A}$ then $T_1 T_2 x = T_2 T_1 x$ for all $x \in X$. Assume that $p$ is invariant under elements of $\mathcal{A}$, thus $p(x) = p(Tx)$ if $T \in \mathcal{A}$ and $x \in X$. Assume that (on $Y$) $\ell$ is invariant under $\mathcal{A}$. Then

**Theorem 4.** $\ell$ could be extended to all of $X$ so that it remains invariant under $\mathcal{A}$ and still controlled by $p$.

Idea of proof: We could add more to $\mathcal{A}$ all products of its elements, thus we may assume that $\mathcal{A}$ contains 1 and is closed under composition (a semigroup). Let $\mathcal{B}$ contain all finite convex combination of elements of $\mathcal{A}$. By definition, (inside $Y$) the given $\ell$ is controlled by $p_0(x) = \inf_{T \in \mathcal{B}} p(Tx)$, which is a homogeneous convex function on $X$. Thus we could extend $\ell$ to all of $X$ while remains dominated by $p_0$ (thus is dominated by $p$). We need to show that $\ell$ is invariant under $\mathcal{A}$. Let $T \in \mathcal{A}$, it can be shown that

$$p_0(x - T x) \leq 0$$

for every $x \in X$. Indeed, by definition $p_0(x) \leq p_0((1 + \cdots + T^{n-1})x/n)$ for any $n \geq 1$, therefore

$$p_0(x - T x) \leq \frac{1}{n} p(x - T^n x) \to 0 \quad \text{as } n \to \infty.$$ 

Thus $\ell(x - T x) \leq p_0(x - T x)$ and thus $\ell(x) \leq \ell(T x)$ for all $x$. Since $\ell(-x) = -\ell(x)$, it follows that $\ell(x) = \ell(T x)$.

**An application:** The most typical applications of the HB extension theorem are hyperplane separation theorems which require some local convexity of the underlying space. We will revisit these applications later, here we discuss a concrete example.

### 1.3 Topological spaces

We recall the notion of locally compact Hausdorff spaces (LCH) and discuss related results.

#### 1.3.1 Compactness

$X$ is compact if for any open covering there is a finite subcover. A space is locally compact if every point has a compact neighborhood.

Properties: 1. (finite intersection property) For any family of closed set with nonempty intersection we could find a finite subfamily with nonempty intersection. In fact this is equivalent to compactness.

2. **Tychonoff’s theorem:** If $(X_\alpha)_{\alpha \in \mathcal{A}}$ is a family of compact spaces then the product $X = \prod_{\alpha \in \mathcal{A}} X_\alpha$ with the product topology is compact.

(The product topology is the minimal topology such that all coordinate projections $\pi_\alpha : X \to X_\alpha$ are continuous, equivalently speaking it is generated by $\pi_\alpha^{-1}(open \ sets)$. )
3. If there is some geometry (i.e. the topology is metrizable) then compactness is equivalent to sequential compactness, which states that for any sequence there is a subsequence that converges.

4. **Net convergence:** Without geometry, one needs to use more than sequences. A net \((x_\alpha)_{\alpha \in I}\) is a collection indexed by some **directed set** \(I\), i.e. a set with some partial ordering \(<\) so that any two elements has at least one common upper bound. We say this net converges to \(x\) if given any neighborhood of \(x\) there is \(\beta \in I\) such that \(x_\alpha \in P\) if \(\beta < \alpha\).

   (be careful, this limit is not necessary unique in general).

   **Bolzano–Weierstrass’s theorem:** \(X\) is compact iff every net has a convergent subnet.

### 1.3.2 Continuous functions on LCH spaces

Recall that a function \(f : X \to \mathbb{R}\) is continuous if for every open \(A \subset \mathbb{R}\) the set \(f^{-1}(A)\) is open in \(X\). Note that a continuous function maps a compact subset of \(X\) to a compact subset of \(\mathbb{R}\), thus continuous functions are always bounded on compact sets.

A space is Hausdorff if one could separate two points using two disjoint open sets. There are variants (both weaker/stronger) of this notion, but we won’t discuss them here. The most important property of a Hausdorff space is that limit is unique (if exists). Also a closed subset of a compact subset is compact.

**Basic questions:**

1. Are there (a lot of) nonconstant continuous functions on a topological space? (We are interested in bump functions, since this is related to the second question below.)

2. Can we extend continuously a given local continuous function (i.e. supported on a compact subset) to all of the space?

3. Can we construct partition of unity on such a space?

We will show that if the space is LCH then the answer is yes for all questions. (One could do better than LCH but we will not discuss that here.) **These confirmative answers lead to the study of \(C(X)\) the space of continuous functions on \(X\).**

**Urysohn’s lemma:** Let \(X\) be an LCH space. Let \(K\) be compact and \(U\) be open in \(X\), such that \(K \subset U\). Then there exists a continuous function \(f : X \to [0, 1]\) such that

- \(f = 1\) on \(K\) and

- \(f = 0\) outside a compact subset of \(U\).

**Tietze’s extension theorem:** Let \(X\) be LCH and \(K \subset X\) compact subset. Then any function \(f \in C(K)\) could be continuously extended to all of \(X\). Furthermore the extended function vanishes outside a compact set.

**Partition of unity:** this ia collection of nonnegative functions whose sum is 1 everywhere on the space, but locally at each point only finitely many of them are nonzero. One certainly could only do this partition for a compact subset of the space.
1.3.3 Proof of Urysohn’s lemma

Proof consists of two steps.

Step 1: First we consider a simpler setting when the space is actually compact Hausdorff. The idea is that a compact Hausdorff space has a nicer structure, namely one could separate two disjoint closed sets using two open sets. (also known as the “normal” property). Urysohn’s lemma works for normal spaces, here the assumptions would simply be $K$ is closed inside an open set $U$.

To see why compact Hausdorff implies normal, take two closed disjoint sets $C_1, C_2$, they are then compact. Given $x \in C_1$ and $y \in C_2$ we could use open sets $W_{x,y}$ and $V_{x,y}$ to separate them, now $(V_{x,y})$ covers $C_2$ so using compactness we could get a finite subcover $V_{x,y_k}$, thus we could separate $x$ from $C_2$ using two open sets $\cap W_{x,y_k}$ and $\cup V_{x,y_k}$. Repeat this argument to separate $C_1$ from $C_2$ using open sets.

Thus now $X$ is a normal space, $K$ closed, $U$ open. We construct a large family of open sets $(U_r)$ that interpolates $K$ and $U$. This family is indexed by dyadic rational numbers $r = \frac{m}{2^n}$ that are in $(0, 1)$, such that $K \subset U_r \subset U$ and $U_r \subset U_s$ if $r < s$. One then defines

$$g(x) = \inf \{ r : x \in U_r \}$$

for all $x \in X$. Clearly $g(x) = 0$ if $x \in K$ and $g(x) = 1$ if $x \in U^c$, thus we could define $f(x) = 1 - g(x)$ and $f = 1$ on $K$ and vanish outside $U$. If we want $f$ to vanish outside a compact subset $K'$ of $U$, we could apply this for $K$ and $U_{1/2}$, and notice that $\overline{U_{1/2}}$ is a compact subset of $U$.

Such $f$ is continuous: it suffices to show

$$f^{-1}((\alpha, \infty)) \text{ open} \quad \text{for all } \alpha \in \mathbb{R}.$$ 

Wlog assume $0 \leq \alpha \leq 1$. Then

$f(x) < \alpha \iff x \in U_r$ for some $r < \alpha$, thus

$$f^{-1}(-\infty, \alpha) = \bigcup_{r<\alpha} U_r$$

is open.

$f(x) > \alpha \iff x \not\in U_r$ for some $r < \alpha$. Using the inclusion assumption, this is equivalent to existence of $r > \alpha$ such that $x \not\in U_r$, so

$$f^{-1}(\alpha, \infty) = \bigcup_{r>\alpha} \overline{U_r^c}$$

is open.

Thus the remaining step is to construct the family. Here we use normality: to get $W_1$ open and $W_2$ open that surrounds $K$ and $U^c$ (disjoint closed sets). Thus $K \subset W_1 \subset W_2^c \subset U$, and we define $U_{1/2} = W_1$, now the closure of $U_{1/2}$ is inside $W_2^c$ so inside $U$. Thus

$$K \subset U_{1/2} \subset \overline{U_{1/2}} \subset U$$
1.3. TOPOLOGICAL SPACES

We repeat this with the new pairs \((K, U_{1/2})\) and \((\overline{U_{1/2}}, U)\) and so on, get the family.

**Step 2:** Reduction to compact setting. The idea is to show that for some \(V \subset X\) open it holds that \(\overline{V}\) is compact and \(K \subset V \subset \overline{V} \subset U\). Thus if the Lemma holds for compact Hausdorff, we simply restrict to the subspace \(\overline{V}\) and then extend the local function on \(\overline{V}\) to the whole of \(X\) by letting it be 0 outside \(\overline{V}\).

The existence of such a \(V\) can be done as follows: first we show that given each \(x \in K\) we could get a compact neighborhood \(N_x\) that remains inside \(U\). Then the family of interior of these \(N_x\) forms an open cover of \(K\), thus using a finite subcovering we easily get a open set containing \(K\) such that its closure is inside \(U\).

Now the existence of such a neighborhood follows from LCH property: the idea is for Hausdorff space we could actually separate point and a compact set.

Now given each \(x\) we could get a neighborhood of \(x\), called \(M_x\), that is compact but not necessarily inside \(U\), then the set \(P = M_x \cap U\) is compact and is a neighborhood of \(x\), it is almost contained inside \(U\). One uses Hausdorffness to separate \(x\) further from the compact boundary of this set.

### 1.3.4 Proof of Tietze’s theorem

Essentially, Urysohn’s lemma implies Tietze’s theorem, at least for compact Hausdorff spaces (or more generally normal space). We’ll focus on this, the extension to the locally compact case is similar to the last proof.

The idea is to approximate \(f\) by a sequence of continuous functions that have global extensions, and this sequence converges uniformly to a continuous function on \(X\).

To start, wlog assume \(0 \leq f \leq 1\) everywhere on \(K\). Then by Urysohn’s lemma there is a bump function \(0 \leq h_1 \leq 1\) such that \(h_1 = 1\) on the closed set \(f^{-1}([\frac{2}{3}, 1])\) and \(h_1 = 0\) outside the (bigger) open set \(f^{-1}(\frac{1}{3}, \infty)\). Clearly

\[
0 \leq f - \frac{1}{3} h_1 \leq \frac{2}{3}
\]

on \(K\).

Applying this argument to \(f_1 = \frac{3}{2}(f - h_1)\) which takes values in \([0, 1]\) on \(K\), we get a bump function \(h_2\) continuous on \(X\) such that locally on \(K\) it holds that \(\frac{3}{2}(f - \frac{1}{3} h_1) - \frac{1}{3} h_2 \leq \frac{2}{3}\), or equivalently

\[
0 \leq f - \frac{1}{3} h_1 - \frac{2}{9} h_2 \leq \frac{4}{9}
\]

Repeating this argument, we get continuous functions (globally on \(X\)) \(h_1, h_2, \ldots\) such that \(h_j\) takes value in \([0, 1]\) and

\[
0 \leq f - \frac{1}{3} h_1 - \frac{2}{9} h_2 \cdots - \frac{2^{n-1}}{3^n} h_n \leq \frac{2^n}{3^n}
\]

It is clear that the sequence \(\frac{1}{3} h_1, \frac{1}{3} h_1 + \frac{2}{5} h_2, \frac{1}{3} h_1 + \frac{2}{5} h_2 + \cdots + \frac{2^{n-1}}{3^n} h_n\) converges uniformly to a continuous function on the compact space \(X\).
1.3.5 Partition of unity

Given \( K \) compact subset and a finite open cover \((U_j)_{1 \leq j \leq n}\) can we find a partition of unity consisting of compactly supported bump functions such that for each \( j \) at least one such function is supported in \( U_j \)? Yes if LCH.

The idea is to find a compact subset \( V_j \) (could be empty) from each \( U_j \) such that they cover \( K \). This follows from the fact that each point in \( K \) has a compact neighborhood inside some \( U_j \), thus by compactness one could refine this and get a finite covering consisting of sets of this type, and let \( V_j \) be the union of those set that are strictly inside \( U_j \). (\( V_j \) covers \( K \) since any set in the covering has to be a subset of some \( V_j \).)

Then use Urysohn’s lemma to construct \( g_j \) bump functions that equal 1 in \( V_j \) and vanish outside \( U_j \). (If \( V_j = \emptyset \) simply take \( g_j \equiv 0 \).)

Clearly \( g := \sum_j g_j \geq 1 \) on \( K \) and continuous, but this will be zero outside a compact set, so we can’t simply divide everything by \( g \) to get the partition of unity. The idea is to use Urysohn’s again to get a bump function \( f \) that equals 1 in \( K \) and vanish outside \( \{g > 0\} \) (which is an open set). Now we could add to \( g \) the function \( 1 - f \), which does not change anything inside \( K \) but will be 1 as soon as \( g = 0 \). We get the partition of unity of \( K \) consisting of \( \frac{g_j}{g + 1 - f} \).
Chapter 2

Overview of normed linear spaces

Starting from this chapter, we begin examining linear spaces with at least one extra structure (topology or geometry). We assume linearity; this is a natural feature of functional spaces.

In this chapter, we start with spaces whose geometric and linear structures are compatible. More precisely, we assume that the metric is translation invariant \( d(x, y) = d(x + z, y + z) \) and homogeneous \( d(\lambda x, \lambda y) = |\lambda|d(x, y) \). We then define \( \|x\| = d(x, 0) \) the norm of \( x \), and it is clear that \( \|x\| \) satisfies the following three properties (below \( x, y \in X \) and \( \lambda \in \mathbb{C} \)):

- \( \|x\| \geq 0 \) and equality holds iff \( x = 0 \).

- \( \|x + y\| \leq \|x\| + \|y\| \).

- \( \|\lambda x\| = |\lambda|\|x\| \).

A linear space equipped with such a norm is called a normed linear space (NLS).

Conversely, if such a norm exists we could always define a metric \( d(x, y) = \|x - y\| \) that is translation invariant and homogeneous.

If the norm topology is separable then we say the NLS is separable.

A Banach space is a complete normed linear space, i.e. the induced metric space is complete. A Hilbert space has additional structures, where we could talk about “angle” through the scalar product. We’ll discuss these spaces further in separate chapters.

We recall that an incomplete metric space could be completed, and this applies to normed linear spaces too: the metric remains invariant under dilation and translation in the completed space thus it remains a norm.

Some examples:

1. Let \( (X, \mu) \) be a measure space. Let \( 1 \leq p \leq \infty \). Then \( L^p(X, \mu) \) are normed linear spaces, with

\[
\|f\|_p = (\int_X |f(x)|^p d\mu)^{1/p}
\]
In fact they are complete, this is a theorem of Riesz-Fisher. To see this, suffices to show that if \( \{f_n\} \) has \( \sum_n \|f_n\|_p < \infty \) then \( \sum_n f_n \) converges. \(^1\)

Now by the triangle inequality \( \int (\sum_n |f_n|)^p \, d\mu < \infty \), thus \( h(x) = \sum_n |f_n(x)| \) is finite for \( \mu \)-almost every \( x \). Thus \( \sum_n f(x) \) converges to some \( g(x) \) measurable, clearly \( \|g\|_p \leq \|h\|_p < \infty \) so \( g \in L^p \).

Finally, \( |(\sum_n f_n) - g|^p \) converges pointwise a.e. to 0 and dominated by \( (h(x) + |g(x)|)^p \) which are integrable, thus by dominated convergence \( \int |(\sum_n f_n) - g|^p \, d\mu \to 0 \), thus \( \sum f_n \to g \) in \( L^p \).

On the other hand, for \( p < \infty \) weak \( L^p(X, \mu) \) is not a normed linear space, since the triangle inequality fails

\[
\|f\|_{p, \infty} = \sup_{\lambda > 0} \lambda \mu(\{x \in X : |f| > \lambda\})^{1/p}
\]

However, we still say “the weak \( L^p \) norm” in practice.

Here is a counter example to the triangle inequality: use \( f(x) = x1_{0 \leq x \leq 1} \) and \( g(x) = (1 - x)1_{0 \leq x \leq 1} \).

2. If \( X \) is compact Hausdorff then \( C(X) \) is a complete normed linear space if we use the sup norm

\[
\|f\| = \sup_{x \in X} |f(x)|
\]

If \( X \) is not compact then the sup is not necessarily finite and so this is not even a norm.

For locally compact (Hausdorff) spaces we could instead look at \( C_c(X) \) consisting of compactly supported continuous functions, and this space is a normed linear space. However, \( C_c(X) \) is not complete in the noncompact case, infact its completion is \( C_0(X) \) the space of continuous functions on \( X \) that vanish at \( \infty \). (These are functions \( f \in C(X) \) such that for every \( M > 0 \) the set \( \{|f| \geq M\} \) is compact.)

To see this, we first show that \( C_0(X) \) is complete under the uniform norm. One way to show this is to use one-point compactification. Alternatively, consider a Cauchy sequence \( (f_n) \) in \( C_0(X) \), then for each \( x \in X \) \( f_n(x) \) is a Cauchy sequence of complex numbers, thus it converges pointwise to some \( f(x) \). Furthermore on any compact subset of \( X \) the uniform convergence \( f_n(x) \to f(x) \) implies continuity of \( f \). Since \( X \) is locally compact and continuity is a local property, it follows that \( f \in C(X) \), then using \( f_n \in C_0 \) it is not hard to see that \( f \in C_0 \). (One way to check this is to use the net convergence characterization of continuity).

Now, given a function \( f \in C_0(X) \) we could approximate it by a convergent sequence of compactly supported continuous functions on \( X \) using Urysohn’s lemma: let \( K_n = \{|f| \geq 1/n\} \) which is compact and sits inside \( U_n = \{|f| < 1/(n-1)\} \) an open set. Then by Urysohn’s lemma we may find \( \phi_n \) a bump function that vanishes outside \( U_n \) but equals 1 on \( K_n \); it is clear that \( \phi_n f \to f \) in the uniform metric.

3. \( L^2 \)-Sobolev spaces on \( \mathbb{R} \). If \( k \geq 0 \) integer we could define

\[
H^k := \{ f \in L^2 : f, \ldots, f^{(k)} \in L^2 \}
\]

\(^1\)For normed linear spaces, “completeness” is equivalent to “every absolutely summable sequence is summable”.
be the space of $L^2$ functions whose first $k$ derivatives exists almost everywhere and are $L^2$ integrable. $H^k$ is complete and we may alternatively define $H^k$ as the completion of the space of locally compacted $C^\infty$ functions on $\mathbb{R}$ under the norm

$$\|f\|_{H^k} = \|f\|_2 + \cdots + \|f^{(k)}\|_2$$

Using the Fourier transform, we could generalize this to allow for $k$ fractional and even negative, and we could also use $L^p$ instead of $L^2$. We’ll revisit this in the future if time permits.

4. If $Y$ is a closed subspace of $X$ then the quotient space $X/Y$ (the space of equivalent classes where $x_1 \sim x_2$ if $x_1 - x_2 \in Y$) is a normed linear space with norm

$$\|[x]\| = \inf_{y \in Y} \|x - y\|$$

An equivalent definition is $\|[x]\| = \inf_{z \in [x]} \|z\|$. Note that closedness of $Y$ is essential here to ensure that $\|[x]\| = 0$ iff $x \in Y$.

5. If $Y$ is a subspace of $X$ then the closure $\overline{Y}$ of $Y$ (with respect to the norm topology) is another subspace of $X$. If this closure is the same as $X$ then we say that $Y$ is dense in $X$. Note that this closure is not necessarily complete and $\overline{Y}$ is not the same as the completion of $Y$ under the norm. For instance we could take $X = Y$ incomplete NLS, then the closure of $Y$ under the norm topology of $X$ is the same as $X$, still incomplete.

2.1 Linear functionals and dual spaces

There are two basic notions of dual spaces:

- algebraic dual, which consists of all linear maps (aka linear functionals) $\ell : X \to \mathbb{R}$ (or $\mathbb{C}$); and

- continuous dual, which consists of all continuous linear functionals.

We will be only interested in continuous dual spaces, which will be implicitly understood whenever we refer to dual spaces.

Note that we could define continuous dual spaces even if $X$ is only a topological linear space, without any norm.

2.1.1 Linear functionals

For a normed linear space, we could talk about boundness of a linear functional. We say that a linear functional $\ell$ is bounded if there exists $C > 0$ such that for every $x \in X$ it holds that

$$|\ell(x)| \leq C\|x\|$$

Equivalently, this means $\sup_{x \in X: \|x\| \leq 1} |\ell(x)| < \infty$ (equivalently $\sup_{x \in X: \|x\| = 1} |\ell(x)| < \infty$) which sometimes is used as the definition. Note that this fact also holds for linear transformations from $X$ to a Banach spaces.
Theorem 5. On normed linear spaces, a linear functional is continuous iff it is bounded.

Proof. In one direction, clearly being bounded implies being continuous. For the other direction, we’d like to show that the image of the closed unit ball under \( \ell \) is a bounded set if \( \ell \) is linear continuous. Note that \( \ell \) maps compact sets to compact sets, but unfortunately as we’ll see the closed unit ball is not compact if \( X \) is infinite dimensional. So one has to exploit linearity of \( \ell \). Assume towards a contradiction that \( x_n \) is a sequence of unit vector s.t. \( |\ell(x_n)| > n \). Then \( x_n/n \) converges to 0 in the norm, but \( |\ell(x_n/n)| > 1 \) which violates continuity of \( \ell \).  

One of the most frequently/implicitly used theorems for bounded linear functionals on normed linear space is the so-called B.L.T. (bounded linear transformation) theorem. The theorem applies even for bounded linear maps from \( X \) to another Banach space.

Theorem 6. Let \( D \) be a dense subspace of the normed linear space \( X \). Let \( \ell : D \to Y \) where \( Y \) is a Banach space, and for some \( C > 0 \) it holds for all \( x \in D \) that

\[
\| \ell(x) \|_Y \leq C \| x \|_X
\]

Then \( \ell \) has an unique extension to a bounded linear map from \( X \) to \( Y \) and satisfies the above estimate for all \( x \in X \).

Proof. It is not hard to see that if such \( \ell \) exists it has to be unique. To define \( \ell \), fix \( x \in X \). Since \( D \) is dense in \( X \) there exists a sequence \( (x_n) \) in \( D \) that converges to \( x \). We then define \( \ell x = \lim x_n \). Note that this limit exists because \( \ell x_n \) is a Cauchy sequence in \( Y \) which is a complete space. One could easily show that the value of \( \ell x \) does not depend on the choice of the sequence \( x_n \). Linearity and boundedness could be easily checked.

While working with singular integral operators on \( L^p \) spaces we typically invoke the above theorem implicitly: these operators are explicitly defined only for a nicer dense subset of functions (say sufficiently smooth and with sufficient decay), and so the theorem says that as long as we could bound the operators on these dense subspace we could extend the operator to all of the corresponding \( L^p \) and get the same bound.

Example: the Hilbert transform \( Hf(x) = p.v. \int \frac{1}{y} f(x - y) dy \) is defined for smooth compactly supported functions on \( \mathbb{R} \) which is dense in \( L^p \). It turns out that \( \|Hf\|_p \leq C \|f\|_p \), \( 1 < p < \infty \), thus \( H \) extends to a bounded maps on \( L^p \).

2.1.2 Dual spaces

For a bounded linear functional we could define \( \|\ell\|_* = \sup_{x \in X : \|x\| = 1} |\ell(x)| \) and this defines a norm on the dual space of \( X \). By definition \( |\ell(x)| \leq \|x\| \|\ell\|_* \).

Theorem 7. The dual space of a normed linear space is a Banach space.

Proof. This is actually a special case of a more general fact: the space \( B(X,Y) \) of bounded operators from normed linear space \( X \) to a Banach space \( Y \)

\[
\|Tx\|_Y \leq \|T\| \|x\|_X
\]
is in turn another Banach space. (For us \( Y = \mathbb{R} \) (or \( \mathbb{C} \)) with the distance norm). It is not hard to see that this is a normed linear space, the main thing is to show completeness. Given any Cauchy sequence \((T_n)\) in \( B(X,Y) \) it is not hard to see that \( \sup_n \|T_n\| < \infty \). Now for any \( x \in X \) the sequence \( T_n x \) is a Cauchy sequence in \( Y \) therefore it converges (thanks to completeness of \( Y \)) in \( Y \), and we let \( T_\infty x \) to be this limit. It is clear that \( T_\infty \) is linear and bounded \( \|T_\infty\| \leq \sup_n \|T_n\| < \infty \). It remains to show that \( \lim_{n \to \infty} \|T_n - T\| = 0 \). □

Examples:
1. The dual space of a Hilbert space is itself.
2. If \( 1 < p < \infty \) then the dual space of \( L^p(X,\mu) \) is \( L^q(X,\mu) \) where \( q \) is the conjugate exponent \( \frac{1}{p} + \frac{1}{q} = 1 \). In particular, \( \|f\|_{L^p(X,\mu)} = \sup_{\|g\|_q = 1} |\int fg d\mu| \).

If \( \mu \) is a \( \sigma \)-finite measure (namely one could break the space into countably many subsets where \( \mu \) is finite) then the dual space of \( L^1(X,\mu) \) is \( L^\infty(X,\mu) \). This may not be true without the \( \sigma \)-finite assumption. ON the other hand, the dual space of \( L^\infty(X,\mu) \) generally speaking not \( L^1(X,\mu) \) (but there are examples when they are, say when \( X \) is a finite set with the counting measure).

3. If \( X \) is a locally compact Hausdorff space, then the dual of \( C_0(X) \) is the space of regular Borel measures with finite total mass. This result is one of Riesz’s representation theorems and we will prove them later in the course.

### 2.1.3 Reflexivity

If \( X \) is a normed linear space we let \( X^* \) denote its dual and \( X^{**} \) denote the dual of \( X^* \).

**Definition 3 (Reflexive spaces).** We say that \( X \) is reflexive if \( X = X^{**} \) (up to isomorphism).

In particular, a reflexive space has to be a Banach space to begin with, but certainly not all Banach spaces are reflexive.

The interest in reflexive spaces is natural, since we always can isometrically embed \( X \) into \( X^{**} \). To see this, fix \( x \in X \). Let \( \|\cdot\|_* \) and \( \|\cdot\|_{**} \) be the norms on \( X^* \) and \( X^{**} \) respectively. We could map \( x \in X \) to the following linear map \( \widehat{x} \) on \( X^* \):

\[
\widehat{x}(\ell) := \ell(x), \quad \ell \in X^*.
\]

Note that \( \widehat{x} \) is bounded on \( X^* \), since by definition \( |\widehat{x}(\ell)| \leq \|x\| \|\ell\|_* \). Thus \( \|\widehat{x}\|_{**} \leq \|x\| \).

On the other hand, using the Hahn Banach theorem it follows that we could find \( \ell \in X^* \) such that \( \ell(x) = \|x\| \) and \( |\ell(y)| \leq \|y\| \) for all \( y \in X \). It follows that \( \|\ell\|_* \leq 1 \), therefore \( |\widehat{x}(\ell)| = \|x\| \|\ell\|_* \), and consequently \( \|\widehat{x}\|_{**} \geq \|x\| \). Thus \( \|x\| = \|\widehat{x}\|_{**} \) and the embedding \( x \mapsto \widehat{x} \) is an isometric embedding of \( X \) into \( X^{**} \).

As we'll see, being reflexive helps in many situation; we'll explore that gradually.
Examples:
1. Hilbert spaces are reflexive.
2. $L^p$ are reflexive if $1 < p < \infty$. As discussed above, generally speaking both $L^1$ and $L^\infty$ are not reflexive. For examples they are not reflexive when the underlying space is $\mathbb{R}^n$ with Lebesgue measure or $\mathbb{Z}$ with counting measure.
3. A closed linear subspace of a reflexive space is also reflexive.
4. $C[\text{-}1,1]$ is not reflexive. Note that this space is separable but its dual is not (since the dual of this space contains in particular all Borel measures). As we’ll see later, for a reflexive space its dual space has to be separable too. One could prove this directly w/o using separability.

2.2 (Non)compactness of the unit ball

One of the key themes in analysis is existence of limit of a sequence (in some functional spaces) that are bounded uniformly in norm. By rescaling if necessary we may assume the bound is 1.

Sometimes, it suffices to get a convergent subsequence. In other words we are interested in the sequential compactness of the closed unit ball.

Since NLS is in particular a metric space, sequential compactness $=$ compactness. So the question is about compactness of the closed unitball in the norm topology.

For Euclidean spaces it is clear that for finite dimensional setting the closed unit ball is compact. For infinite dimensional setting, the answer is no, for instance consider the sequence $a_n = (0, \ldots, 0, 1, 0, \ldots)$ (the $n$th coordinate is 1). If there is a limit $x = (x_1, x_2, \ldots)$ for some subsequence $a_{n_k}$, i.e.

$$\|a_{n_k} - x\| \to 0$$

as $k \to \infty$. Then clearly $\|x\| = 1$. On the other hand, given any $j$ by taking $k$ large it is clear that $0 \leq |x_j| \leq \|a_{n_k} - x\| \to 0$, thus $x_1 = x_2 = \cdots = 0$, which is a contradiction.

This suggests that the answer should be similar for a general normed linear space: positive for spaces with finite dimensions and negative in the infinite dimensional setting (whether the space is complete or not).

In the finite dimension case, this follows from the fact that in finite dimensional linear spaces all norms are equivalent. In other words, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms then there exists $C_1, C_2 > 0$ such that $\frac{1}{C_2}\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1$ for all $x \in X$. To see this, for simplicity consider linear spaces over $\mathbb{R}$, and assume that the given space is spanned by $x_1, \ldots, x_m$, it suffices to show the conclusion for $\|\cdot\|_2$ being the Euclidean norm $\|a_1x_1 + \cdots + a_mx_m\|_2 = \sqrt{a_1^2 + \cdots + a_m^2}$ the Euclidean norm. Then by the triangle inequality

$$\|a_1x_1 + \cdots + a_mx_m\|_1 \leq (|a_1| + \cdots + |a_m|) \max_j \|x_j\|_1 \leq \sqrt{m} \sqrt{a_1^2 + \cdots + a_m^2} \max_j \|x_j\|_1$$

so we could take $C_1 = \sqrt{m} \max_j \|x_j\|_1$. To show existence of $C_2$, note that the identity map $\ell : (X, \|\cdot\|_2) \to (X, \|\cdot\|_1)$ (i.e. $\ell x = x$) is continuous since it is Lipschitz. The unit ball
2.2. **(NON)COMPACTNESS OF THE UNIT BALL**

in \((X, \|\cdot\|_2)\) is compact, thus its image under this continuous function is also compact, thus bounded in \(\|\cdot\|_1\), therefore for some \(C_2 > 0\) we have

\[
\sup \|x\|_1 \leq C_2
\]

the sup is over \(x\) with \(\|x\|_2 = 1\), which is equivalent to the desired estimate.

In the infinite dimensional case, this is a theorem of Riesz.

**Theorem 8** (Riesz). *If \(X\) is an infinite dimensional NLS then the closed unit ball is not compact with respect to the norm topology.*

**Proof.** Our plan is to find a sequence of elements of unit vectors in \(X\), \(x_1, x_2, \ldots\), such that the distance \(\|x_i - x_j\|\) between any two elements of the sequence is uniformly larger than 0, for instance \(\|x_i - x_j\| \geq \frac{1}{10}\) for all \(i \neq j\). If that could be constructed it is clear that no subsequence of \((x_n)\) is Cauchy, thus no subsequence of this sequence is convergent and therefore the closed unit ball is not (sequentially) compact.

To construct this sequence it suffices to show that if \(Y\) is a closed proper subspace of \(X\) then one could find \(x \in X \setminus Y\) with \(\|x\| = 1\) such that \(\text{dist}(x, Y) \geq \frac{1}{10}\). Once this is proved we could start with any point \(x_1\) of unit length and let \(Y\) be spanned by \(x_1\) (which is closed) and then select \(x_2\) of unit length as above (in particular \(\|x_2 - x_1\| \geq \frac{1}{10}\)), then reset \(Y\) to be spanned by \(x_1, x_2\) (a closed subspace because it has finite dimension) and select \(x_3\), etc.

Thus we only to show existence of \(x \in X\) of unit norm satisfying \(\text{dist}(x, Y) \geq \frac{1}{10}\) whenever \(Y\) is a closed proper subspace of \(X\). Let \(z \in X \setminus Y\) and let

\[
d := \text{dist}(z, Y) > 0.
\]

(if \(d = 0\) then the closedness of \(Y\) would imply that \(z \in Y\), contradiction.) Now for some \(y \in Y\) we have \(\|z - y\| < 10d\). Let \(x = z - y\), it follows that \(\text{dist}(x, Y) = \text{dist}(z, Y) = d\), thus

\[
\text{dist}(x, Y) > \|x\|/10
\]

now we just rescale \(x\) so that it has length 1. □

Because of this result, there is a natural question of determining a topology on \(X\) so that the closed unit ball is compact with respect to this topology. One wants a weaker topology (with fewer open sets, too many open sets is probably the reason why the ball is not compact).

There is a natural notion of weak-topology that is weaker than the norm topology. Here we wants just enough open sets so that all bounded linear functionals are continuous.

**Theorem 9.** *The closed unit ball is compact with respect to the weak-topology if and only if \(X\) is reflexive.*

Note that compactness of the closed unit ball is compact in this weak topology does not imply local compactness of \(X\).

We’ll prove this result later when discussing topologies on Banach spaces; as we’ll see this theorem is a consequence of the Banach Alaoglu theorem.
Exercises:
1. Prove that for a normed linear space, “completeness” of the norm is equivalent to “every absolutely summable sequence is summable”.
2. Prove that for every $x \in X$ a normed linear space it holds that $\|x\| = \sup_{\ell \in X^*: \|\ell\| = 1} |\ell(x)|$.
3. Prove that if $K$ is compact Hausdorff then $C(K)$ is complete with respect to the sup norm. Use this to complete the proof that $C_0(X)$ is complete if $X$ is locally compact Hausdorff without appealing to the one-point compactification trick.
4. Let $S$ be a subset of $X$ a normed linear space, and let $Y$ be the linear span of $S$ (consisting of all finite linear combination of elements of $S$). Let $L \subset X^*$ to be the set of all $\ell \in X^*$ that vanishes on $S$. Prove that $z \in \overline{Y}$ the closure of $Y$ if and only if $\ell(z) = 0$ for every $\ell \in L$. (Hint: one direction should be easier, for the other direction you should use the BLT theorem somewhere.)
5. Prove that all closed linear subspaces of a reflexive (Banach) space are reflexive. [Hint. Use problem 4 at some point.]
Chapter 3

Basic geometrical and topological properties of normed linear spaces and their duals

3.1 Topologies on normed linear spaces

Recall that the norm topology is induced from the norm and the weak topology is the minimal topology that makes all bounded linear functionals continuous.

If $X = Y^*$ is the dual of another normed linear space $Y$ (thus in particular $X$ is a Banach space) we could introduce another topology, called the weak* topology: this is the minimal topology such that all bounded linear functionals $\hat{y}, y \in Y$, are continuous. We could use the finite intersections of the following sets as a neighborhood base at 0 in this topology:

$$\{x \in X : |\hat{y}(x)| < \epsilon\}$$

Note that the maps $y \mapsto \hat{y}$ isometrically embedded $Y$ inside $X^*$, the weak* topology is smaller than or equal to the weak topology.

Certainly if $X$ is reflexive then the weak* and weak topologies are the same.

3.1.1 The Banach–Alaoglu theorem

**Theorem 10** (Banach–Alaoglu, 1940). Let $X$ be the dual of another normed linear space. Then the closed unit ball $B$ of $X$ is compact in the weak* topology.

**Proof.** Let $X = Y^*$. For simplicity assume that the spaces are over $\mathbb{R}$. For each $y \in Y$ let $I_y = [-\|y\|_Y, \|y\|_Y] \subset \mathbb{R}$ and consider

$$I = \prod_{y \in Y} I_y$$
consisting of tuples indexed by $Y$. By Tychonoff’s theorem, $I$ is compact in the product topology and it is also Hausdorff (tensor product of Hausdorff spaces with the product topology is Hausdorff). Now, for each elements $x$ in the unit ball $B$ of $X$ consider the map

$$x \mapsto T x := (x(y))_{y \in Y} \in I$$

(the fact that $Tx \in I$ follows because for $x \in B$ we have $|x(y)| \leq \|x\|_X \|y\|_Y = \|y\|_Y$). Note that elements of $TB$ are special because the coordinates of $Tx$ are actually related linearly.

Now, $T$ is clearly linear and one to one and it embeds $B$ into $I$, in fact it is not hard to see that if we use the weak* topology on $B$ and the inherited product topology on $TB$ then $T$ is indeed a topological isomorphism between $B$ and $TB$. We now show that $TB$ is a closed subset of $I$, which together with compactness of $I$ will then implies that $TB$ (hence $B$) is compact.

Let $z = (z_y)_{y \in Y} \in \overline{TB}$ where closure taken inside $I$ under the product topology. Then we want to show that $z$ is “linear and bounded by 1”, namely

$$z_{\alpha x + \beta y} = \alpha z_x + \beta z_y$$

Once that’s done it follows that the maps $y \mapsto z_y$ defines a linear map on $Y$ with norm at most 1 and so $z \in TB$ as desired.

To verify the above property, simply observe that there is a net in $TB$ that converge to $z$ (unique because of Hausdorffness), and linearity relationship between coordinates is actually preserved in the limit, so this property survives and $z$ has it. □

**Examples:**

1. If $X$ is the space of regular Borel measures on $\mathbb{R}$ with the norm = total mass, then its is the dual space of $C_0(X)$ which is a separable space. Therefore the closed unit ball in $X$ is sequentially compact. In particular, given any sequence $(\mu_n)$ of probability measures on $\mathbb{R}$ there is a subsequence $(\mu_{n_k})$ that converges vaguely to some Borel measures $\mu$, i.e. for every continuous function $f$ that vanishes at $\infty$ it holds that

$$\lim_{n \to \infty} \int f(x)d\mu_n(x) = \int f(x)d\mu(x)$$

Note that $\mu$ is not necessarily a probability measures, in particular it is possible that $\mu = 0$, which happens when the masses in $\mu_{n_k}$ escape to $\infty$ (for instance we could take $\mu_n$ to have the density function $1_{[n,n+1]}$). To avoid this there is a notion of tightness, which basically assume that given any $\epsilon > 0$ we could find a compact $K_\epsilon$ such that $\mu_n(\mathbb{R} \setminus K_\epsilon) < \epsilon$ for every $n$. With this assumption we could show that $\mu$ is a probability measure, and basically we have just shown the Helly selection theorem for probability measures on $\mathbb{R}$.

### 3.1.2 The sequential Banach-Alaoglu theorem

Recall that compactness means any net has a convergent subnet. In practice we are interested in the sequential variant of this theorem, which says that
3.1. TOPOLOGIES ON NORMED LINEAR SPACES

Theorem 11. If \( X = Y^* \) where \( Y \) is separable normed linear space, then the closed unit ball \( B \) in \( X \) is sequentially compact in the weak* topology. In other words given any sequence \( (x_n) \) in \( B \) there exists a subsequence \( x_{n_k} \) and \( x \in B \) such that

\[
\lim_{k \to \infty} x_{n_k}(y) = x(y)
\]

for every \( y \in Y \).

Namely, since \( Y \) is separable the weak* topology is metrizable (i.e. it arises from some metric on \( X \)), thus compactness and sequentially compactness are the same on this topology.

We could also prove this directly: let \( (y_k) \) be a dense subset of \( Y \), then from the sequence \( x_n \) we could select a subsequence \( x_{n_k} \) such that \( y_1(x_{n_k}) \) converges. Keep doing this and use a diagonal argument we get a sequence \( x_{n_k} \) such that for any fixed \( j \) it holds that \( \lim_{k \to \infty} y_j(x_{n_k}) \) exists. Using the fact that \( (y_j) \) is dense in \( Y \) it follows that for every \( y \in Y \) the limit \( \lim_{k \to \infty} y(x_{n_k}) \) exists, let this limit be \( z(y) \), it is clear that \( z \) is a linear functional on \( Y \) and also bounded with \( \|z\| \leq 1 \). This implies \( x_{n_k} \) converges weakly to \( z \) and element of \( B \).

3.1.3 Weak compactness of the closed unit ball in reflexive spaces

As a consequence of the Banach Alaoglu theorem, for any reflexive space the closed unit ball is compact in the weak topology. The reverse direction is true (Kakutani’s theorem) and is part of the homework.

If the given space is furthermore separable then its dual is also separable, so (from the sequential BA theorem) the closed unit ball is sequentially compact in this weak topology.

It is surprising that this sequential compactness property still holds even if we don’t assume separability on the given reflexive space!

Theorem 12. Let \( X \) be a Banach space. If \( X \) is reflexive then the closed unit ball in \( X \) is sequentially compact in the weak topology.

Notes: The reverse direction also holds, this is a result of Eberlein–Smulian: if every bounded sequence has a convergent subsequence then \( X \) is reflexive. In fact, ES showed that a subset of a reflexive Banach space is weakly compact if any only if it is sequentially compact, which implies the above statement.

Weak convergence of sequences: In a normed linear space, we say that a sequence converges weakly if it converges in the weak topology. More precisely, \( (x_n) \) converges weakly to \( x \) if for every \( \ell \in X^* \) it holds that \( \lim_{n \to \infty} \ell(x_n) = \ell(x) \). This is in contrast with strong convergence of the sequence \( x_n \), which means \( \|x_n - x\| \to 0 \). Clearly strong convergence implies weak convergence, but the opposite direction is not true in general.

It follows from the above theorem that in a reflexive Banach space (for instance \( L^p \) with \( 1 < p < \infty \)) any norm bounded sequence has a subsequence that converges weakly.

Proof. We’ll show that reflexivity implies sequentially compactness of the closed unit ball. Let \( (x_n) \) be the given sequence and let \( Y \) be the closure of the linear space spanned by \( (x_n) \).
Then $Y$ is a closed linear subspace of $X$ so is also reflexive, on the other hand $Y$ is separable, thus one could see that the closed unit ball inside $Y$ is sequentially compact with respect to the weak topology on $Y$. Thus there is a subsequence $x_{n_k}$ and $x_\infty \in Y$ such that for every bounded linear functionals $\ell$ on $Y$ we have

$$\ell(x_{n_k}) \to \ell(x)$$

Since every bounded linear functionals on $X$ is also a bounded linear functional on $Y$, this implies $x_{n_k}$ converges weakly to $x$. □

### 3.2 Uniform convexity

Uniform convexity is a geometric notion introduced by Clarkson (1936).

**Definition 4 (Uniformly convex).** We say that a normed linear space $X$ is uniformly convex if for each $0 < \epsilon \leq 2$ there exists $\delta = \delta(\epsilon) > 0$ such that the following holds: If $\|x\| = \|y\| = 1$ and $\|x - y\| \geq \epsilon$ then

$$\left\| \frac{x + y}{2} \right\| \leq 1 - \delta$$

Note that this is a property of the norm: there may be equivalent norms on the same space that is not uniformly convex.

Uniform convexity does not imply completeness. For instance we could take the rational numbers with the usual distance. Then $\|\frac{x+y}{2}\|^2 + \|\frac{x-y}{2}\|^2 = \|x\|^2 + \|y\|^2$ which implies uniform convexity.

**Theorem 13 (Milman–Pettis).** If $X$ is uniformly convex Banach then it is reflexive.

Note that there are reflexive Banach spaces where it is not even possible to replace the given norm with an equivalent norm that is uniformly convex, this is due to M. Day (BAMS, 1941). The idea is to use a vector valued Banach space with norm $\|\cdot\| = (\sum_n \|x_j\|^p)^{1/p}$ where $x_n$ belongs to a reflexive Banach space $B_n$ (which makes it reflexive); now for each $j$ one could still replace the norm on $B_j$ by an equivalent norm that is uniformly convex, but as $n \to \infty$ these replacement couldn’t be done uniformly because the underlying equivalence constants blow up if the $B_j$ are carefully chosen.

**Properties:**

1. Hilbert spaces are uniformly convex, since in Hilbert spaces we have the paralelloagram law $\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2$.
2. $C[-1, 1], L^1(\mathbb{R}), L^\infty(\mathbb{R})$ are not uniformly convex because they are not reflexive.
3. Clarkson(1936): if $1 < p < \infty$ then $L^p$ and $\ell_p$ spaces are uniformly convex.

For $p \geq 2$ this holds essentially some form of parallelogram law holds.

$$\left(\|x + y\|^p + \|x - y\|^p\right)^{1/p} \leq 2^{1/p}\left(\|x\|^{p'} + \|y\|^{p'}\right)^{1/p'}$$
where \(1/p + 1/p' = 1\). For \(p \in (1, 2)\) we have a variant \((\|x + y\|^{p'} + \|x - y\|^{p'})^{1/p'} \leq 2^{1/p'}(\|x\|^p + \|y\|^{p'})^{1/p'}\).  

4. (Radon–Riesz, aka property (H)) If \(x_n\) converges weakly to \(x\) in a uniformly convex Banach space \(X\), i.e. \(\ell(x_n) \to \ell(x)\) for all \(\ell \in X^\ast\), and \(\|x_n\| \to \|x\|\), then \(\|x_n - x\| \to 0\).  

5. A space is called uniformly smooth if its dual is uniformly convex.  

6. A normed linear space is uniformly convex if the following holds for every bounded sequences \((x_n)\) and \((y_n)\):  
   \[
   \lim_{n \to \infty} \frac{\|x_n\|^2 + \|y_n\|^2}{2} - \frac{\|x_n + y_n\|^2}{2} = 0
   \]
then \(\lim_{n \to \infty} \|x_n - y_n\| = 0\). Below are two notions weaker than uniformly convex:  

**Local uniform convexity:** A norm is locally uniformly convex if the following holds for every \(x \in X\) and \((x_n)\) in \(X\):  
   \[
   \lim_{n \to \infty} \frac{\|x\|^2 + \|x_n\|^2}{2} - \frac{\|x + x_n\|^2}{2} = 0
   \]
then \(\lim_{n \to \infty} \|x - x_n\| = 0\). It is clear that this is a consequence of uniform convexity.  

**Strict convexity:** A norm is strictly convex if the following holds for every \(x, y\): if  
   \[
   \frac{\|x\|^2 + \|y\|^2}{2} = \|\frac{x+y}{2}\|^2
   \]
then \(x = y\). It is clear that this is a consequence of local uniform convexity.  

It can be shown that if \(X\) is a separable normed linear space over \(\mathbb{R}\) then there exists an equivalent locally uniformly convex norm.  

### 3.3 Isometries between normed linear spaces  

We say that a map \(T : X \to Y\) between normed linear spaces \(X\) and \(Y\) is an isometry if  
   \[
   \|Tx_1 - Tx_2\|_Y = \|x_1 - x_2\|_X
   \]
for every \(x_1, x_2 \in X\). Note that we do not assume that \(T\) is affine (i.e. \(T(\alpha x + \beta y) = \alpha Tx + \beta Ty\), which is the same as linear if we impose \(T0 = 0\)).  

**The Banach–Mazur distance:** The distance is useful to compare two norms on \(\mathbb{R}^n\) - it is known that all norms are equivalent so the point is to be more quantitative.  

---

\(^1\)Alternatively, use Hanner’s inequality which says that if \(\|\cdot\|\) is the \(L^p\) or \(\ell^p\) norm then  
   \[
   \|f + g\|^p + \|f - g\|^p \geq (\|f\| + \|g\|)^p + \|f\| - \|g\|)^p
   \]
(3.1)  
if \(p \in [1, 2]\) and the reverse holds if \(p \in [2, \infty)\).  

\(^2\)To see this, it suffices to show the theorem for sequences with \(\|x_n\| = 1\). Hahn Banach gives us \(\ell \in X^\ast\) such that \(\ell(x) = 1\) and \(\|\ell\| \leq 1\). Then by weak convergence of \(x_n\) to \(x\) we obtain \(2 \geq \|x_n + x\| \geq \ell(x_n + x) \to 2\ell(x) = 2\) as \(\min(n, m) \to \infty\). This easily implies \(\|x_m + x_n\| \to 2\), from there and uniform convexity it follows that \((x_n)\) is a Cauchy sequence and converges to some \(y\), clearly \(\ell(y) = \lim \ell(x_n) = \ell(x)\) for every \(\ell \in X^\ast\), thus \(y = x\).
ANALYTIC: Let $X$ and $Y$ be normed linear spaces. The multiplicative BZ distance is defined to be

$$d(X, Y) = \inf \{ \| T \| \| T^{-1} \| \}$$

the infimum is over all isomorphisms $T : X \to Y$. Sometimes people also use $\log d(X, Y)$ which satisfies the usual (additive) triangle inequality.

GEOMETRIC: given two bounded convex sets with nonempty interiors (aka convex bodies) $K$ and $L$ in $\mathbb{R}^n$ that are symmetric (i.e. $K = -K$ and $L = -L$) the Banach–Mazur distance between them is

$$d(K, L) = \inf \{ C > 0 : L \subset TK \subset CL , \ T \text{ linear on } \mathbb{R}^n \}$$

For $\mathbb{R}^n$ these two notions are equivalent. Basically given any norm on $\mathbb{R}$ the unit ball is a symmetric convex body and conversely given any symmetric convex body $K$ we could construct a norm such that its unit ball is $K$, namely

$$\| x \| = \inf \{ t > 0 : x \in tK \}.$$ 

**Theorem 14** (F. John). Let $\ell_2^n$ denote $\mathbb{R}^n$ with the Euclidean norm. If $X$ is a $n$-dimensional NLS over $\mathbb{R}$ then

$$d(X, \ell_2^n) \leq \sqrt{n}$$

(This is actually sharp; equality holds for instance if $X = \ell_\infty^n$, for the geometric case take the unit ball of this space.)

It follows from the theorem and the (multiplicative) triangle inequality that given any two $n$-dimensional normed linear spaces over $\mathbb{R}$ it holds that

$$d(X, Y) \leq n$$

This result is also sharp up to a constant (Gluskin, 1981, a randomized construction).

**Existence of nonlinear isometry:** A natural question is whether non-affine isometry exists (or equivalently nonlinear if we assume the normalization $T0 = 0$)?

This is not true for complex NLS, even complex Banach spaces, see for instance the conjugation mapping $\mathbb{C}$ to itself $z \mapsto \bar{z}$. We would need surjectivity, since otherwise the map $T : \mathbb{R} \to \mathbb{R}^2$ mapping $x$ to $(x, \sin(x))$ is not linear but is an isometry if we equip $\mathbb{R}^2$ with the $L^\infty$ norm $\| (x, y) \| = \max(|x|, |y|)$ (and distance norm in $\mathbb{R}$).

It turns out that if the underlying normed linear spaces are over $\mathbb{R}$ then the answer is no.

**Theorem 15** (Mazur-Ulam, 1932). If $X$ and $Y$ are NLS over $\mathbb{R}$ and $T : X \to Y$ surjective and isometric with $T0 = 0$ then $T$ is linear.

If $T0 = 0$ is not given the conclusion could be equivalently changed to “$T$ is affine”, namely $T(\alpha x + (1 - \alpha)y) = \alpha T(x) + (1 - \alpha)T(y)$.

**Main ideas of proof:** Isometries are continuous so it suffices to show $T(kx + my) = kT(x) + mT(y)$ for all rational $m, k$, in fact suffices to show for $k = m = 1/2$. We construct a nested sequence of subsets $A_1 \supset A_2 \ldots$ each of them is symmetric around $\frac{x+y}{2}$ such that $\cap A_n = \{ \frac{x+y}{2} \}$. This construction will be invariant under surjective isometries, in particular
to get \( \frac{Tx + Ty}{2} \) we could use the nested sequence \( TA_1 \supset TA_2 \ldots \), which are symmetric around \( \frac{x + y}{2} \). When that’s done we could take intersections to get \( \frac{Tx + Ty}{2} = T(\frac{x + y}{2}) \) as desired.

Now, let \( A_1 \) contains all elements \( z \) of \( X \) such that \( \|x - z\| = \|y - z\| = \|x - y\| \). To construct \( A_{n+1} \) from \( A_n \), simply let \( A_{n+1} \) contains all \( z \in X \) such that for all \( w \in A_n \) it holds that \( \|w - z\| \leq \frac{1}{2} \text{diam}(A_n) \), one could check that \( A_{n+1} \) is still symmetric around \( \frac{x + y}{2} \) and \( \text{diam}(A_{n+1}) \leq \text{diam}(A_n)/2 \leq \ldots \leq \frac{\text{diam}(A_1)}{2^n} \to 0 \) as \( n \to \infty \). Note that we need \( TX = Y \) since we want \( TA_{n+1} \) contains all \( z \in Y = TX \) such that “so and so” holds.

**Exercises:**

1. For any \( 1 < p < \infty \) and any measure space \((X, \mu)\) prove that \( L^p(X, \mu) \) is uniformly convex (use the hints from the lecture notes).

2. Show that the Banach Mazur distance satisfies \( d(\ell_\infty^n, \ell_2^n) \geq \sqrt{n} \). [Hint: use the geometric formulation with corresponding convex bodies, and invoke a generalized parallelogram law.] (Note that together with F. John’s theorem this implies \( d(\ell_\infty^n, \ell_2^n) = \sqrt{n} \).) Then via Holder’s inequality and the analytic formulation prove that \( d(\ell_p^n, \ell_q^n) = n^{\frac{1}{p} - \frac{1}{q}} \) if \( 1 \leq p, q \leq 2 \) or \( 2 \leq p, q \leq \infty \).

3. Prove F. John’s theorem for \( n = 2 \). [Hint. Use the geometric formulation, you may want to use the fact that the Banach-Mazur distance is invariant under isomorphism of the plane.]

4. Let \( X \) be a normed linear space. Let \( B = \{ x \in X : \|x\| \leq 1 \} \) and \( B^{**} = \{ z \in X^{**} : \|z\|_{**} \leq 1 \} \). We know that there is a canonical map \( x \mapsto \hat{x} \) that embeds \( B \) isometrically linearly into \( B^{**} \). Prove that the image of \( B \) under this map is dense in \( B^{**} \) in the weak* topology of \( X^{**} \) (i.e. the minimal topology on \( X^{**} \) that makes all bounded linear functionals constructed from elements of \( X^* \) continuous on \( X^{**} \)). Use this fact to show that if \( B \) is weakly compact then \( B = B^{**} \) and therefore \( X \) is reflexive.

5. Find two examples demonstrating that ”compactness” and ”sequentially compactness” do not imply each other.

6. Let \( K \) be a convex subset of \( X \) a Banach space. Prove that if \( K \) is closed in the norm topology then it is closed in the weak topology.
Chapter 4

Basis on Hilbert spaces and Banach spaces

A Hilbert space is a complete normed linear space where the norm arises from an inner
product $\|u\| = \sqrt{\langle u, u \rangle}$, and the inner product $\langle u, v \rangle$ is a bilinear form that satisfies the
following properties (we state the properties for Hilbert spaces over $\mathbb{C}$):

(i) $\langle u, v \rangle \geq 0$, equality iff $u = 0$.
(ii) $\langle u, v \rangle = \langle v, u \rangle$.
(iii) $\langle u, v \rangle$ is linear in $u$ (hence conjugate linear in $v$).

Examples: $H = L^2(X, d\mu)$ with $\langle f, g \rangle = \int_X f(x)g(x)d\mu$; by Holder this is finite

$$|\int_X f(x)g(x)d\mu| \leq (\int_X |f|^2d\mu)^{1/2}(\int_X |g|^2d\mu)^{1/2}$$

4.1 Basic properties

Parallelogram law: $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

Schwarz’s inequality:

$$|\langle u, v \rangle| \leq \|u\|\|v\|$$

Proof. $\|x + y\| \leq \|x\| + \|y\|$, therefore

$$\|x\|^2 + \|y\|^2 + \langle x, y \rangle + \langle y, x \rangle = \langle x + y, x + y \rangle$$

$$\leq (\|x\| + \|y\|)^2 = \|x\|^2 + \|y\|^2 + 2\|x\|\|y\|$$

Therefore $Re \langle x, y \rangle \leq \|x\|\|y\|$ thus get desired estimate if $\langle x, y \rangle \in \mathbb{R}$. In general multiply $x$
with $\alpha$ unimodular such that $|\langle x, y \rangle| = |\alpha x, y \rangle \leq \|\alpha x\|\|y\| = \|x\|\|y\|$.

Orthogonality: We say $u \perp v$ if $\langle u, v \rangle = 0$. A set $E \subset H$ is orthonormal if $\|u\| = 1$ for
all $u \in E$ and $\langle u, v \rangle = 0$ if $u, v \in E$ with $u \neq v$.

Bessel’s inequality: If $\{x_1, \ldots, x_N\}$ orthonormal then $\sum_{n=1}^N |\langle x, x_n \rangle|^2 \leq \|x\|^2$. A
sequence $(x_n)$ is a Bessel sequence if $\sum_n |\langle x, x_n \rangle|^2 < \infty$ for all $x \in H$. 39
Theorem 16 (Riesz). For any \( \ell \in H^* \) there is \( h_\ell \in H \) such that \( \ell(x) = \langle x, h_\ell \rangle \) for all \( x \in H \), and the map \( \ell \mapsto h_\ell \) is a surjective isometry (conjugate linear). Consequently \( H^* \sim H \).

Proof. Clearly \( \text{ker}(\ell) \) is a closed linear subspace of \( H \), let \( Y = \{ y \in H : y \perp \text{ker}(\ell) \} \). Then any \( x \in H \) has a unique decomposition into \( y + z \) with \( y \in Y \) and \( z \in \text{ker}(\ell) \). Since \( \ell \) is nonzero on \( Y \) so there is \( y_0 \in Y \) such that \( \ell(y_0) = 1 \). We now decompose each \( x \in H \) using \( x = \ell(x)y_0 + (x - y_0\ell(x)) \), now the first element is clearly in \( Y \) and the second is in \( \text{ker}(\ell) \) since \( \ell(y_0) = 1 \). It follows that \( H = \text{span}(y_0, \text{ker}(\ell)) \) and we could let \( h_\ell = \frac{y_0}{\|y_0\|^2} \), then

\[
\ell(x) = \ell(\langle x, h_\ell \rangle y_0) = \langle x, h_\ell \rangle
\]

it is not hard to check isometry/(conjugate)linearity of the map \( \ell \mapsto h_\ell \). □

A generalization of Riesz’s theorem is Lax-Milgram’s theorem:

Theorem 17 (Lax-Milgram). Let \( B(x,y) : H^2 \to \mathbb{C} \) be a bounded bilinear map (conjugate linear in \( y \)), \( |B(x,y)| \leq C\|x\|\|y\| \) such that \( |B(x,x)| \geq c\|x\|^2 \). Then every \( \ell \in H^* \) has the form \( \ell(x) = B(x,y_\ell) \) and the map \( \ell \mapsto y_\ell \) is bounded invertible (conjugate) linear.

Proof. Since \( B(x,y) \) is bounded linear in \( x \) for each \( y \), we could find \( h_y \) such that \( B(x,y) = \langle x, T(y) \rangle \) with \( \|y\| \leq \|T(y)\| \leq \|y\| \), furthermore \( y \mapsto T(y) \) is linear. It follows that one could invert this map, now by Riesz representation theorem \( \ell(x) = \langle x, h_\ell \rangle = \langle x, T(T^{-1}(h_\ell)) \rangle = B(x, T^{-1}h_\ell) \), so we could let \( y_\ell = T^{-1}(h_\ell) \). □

4.2 Orthogonal basis for a Hilbert space

We say that a set \( E \subset H \) is an orthogonal basis if \( iE \) is an orthogonal set and whenever \( v \perp \text{span}(E) \) we must have \( v = 0 \). If furthermore every elements of \( E \) has unit norm then we say that \( E \) is an orthonormal basis.

Theorem 18. A Hilbert space is separable iff it has a countable orthonormal basis.

Proof. If there is a countable orthonormal basis \( \{v_n\} \) then we simply use finite linear combination of \( v_n \) with (complex) rational coefficients and get a dense countable subset. Conversely if \( H \) has a countable dense subset \( D \) we could apply Gram Schmidt to get a countable orthogonal set \( E \). Let \( v \perp \text{span}(E) \), we will show \( v = 0 \). If not, assume \( \|v\| = 1 \), then for all \( x \in E \), exploiting orthogonality of \( v \) and \( x \), we have

\[
\|v - x\|^2 = \|v\|^2 + \|x\|^2 \geq 1
\]

so \( \text{span}(E) \) is not dense in \( H \), contradiction. □

Notes:
1. If \( H \) is separable then we can extend any given basis to the full basis using Gram-Schmidt and avoid axiom of choice for Hahn-Banach.
2. If \( H \) is not separable and \( E \) is an orthogonal basis of \( H \) then for each \( x \in H \) the set of elements \( v \in E \) such that \( x \) has nonzero projection onto that element \( \langle x, v \rangle v \neq 0 \) is
4.3 Schauder basis

Schauder basis: A sequence \( E = (x_n) \) is a (Schauder) basis for a normed linear space \( X \) if for every \( x \in H \) there is a unique scalar sequence \( \alpha_n \) such that \( \lim_{n \to \infty} \| x - \sum_{k=0}^{n} \alpha_k x_k \| = 0 \).

Equivalently, \( (x_n) \) is a Schauder basis if two conditions hold: the closure of the linear span of \( \{x_n\} \) is the whole space and \( \sum_{n=1}^{\infty} a_n x_n = 0 \) iff \( a_n = 0 \).

Note that this is different from an algebraic basis (aka Hamel basis) which consists of linearly independent vectors and any \( x \in X \) could be written as a finite linear combination of these elements. Note that this could be defined on any linear spaces, topology (convergence) is not needed, as we don’t permit infinite sums. Also the order of the elements in a Hamel basis is not important.

A Schauder basis is said to be an unconditional basis if it remains a basis even after reordering. There are two other equivalent characterizations

**Theorem 19.** Let \( (x_n) \) be a (Schauder) basis in a Banach space (over \( \mathbb{R} \) or \( \mathbb{C} \)). The following are equivalent:

(i) \( (x_n) \) is an unconditional basis.

(ii) For any scalar sequence \( (\alpha_n) \), if \( \sum \alpha_n x_n \) converges then it converges unconditionally (i.e. it remains convergent even if we change the summation order).

(iii) There exists \( C > 0 \) finite such that

\[
\left\| \sum_{k=1}^{n} \epsilon_k \alpha_k x_k \right\| \leq C \left\| \sum_{k=1}^{n} \alpha_k x_k \right\|
\]

uniformly over \( n \) and all sequences \( \epsilon_n \) with \( |\epsilon_n| \leq 1 \) and all scalar coefficients \( \alpha_1, \ldots, \alpha_n \).

**Property:**

1. For NLS, existence of Schauder basis means the space is separable.
2. Every orthogonal basis in a separable Hilbert space is an unconditional Schauder basis.

**Examples:**

1. Fourier basis: \( L^2[0, 1] \) has \( (e^{2\pi inx}/\sqrt{2\pi})_{n \in \mathbb{Z}} \) as an unconditional Schauder basis. This remains a basis for \( L^p[0, 1] \) \( 1 < p < \infty \) but not unconditional for \( p \neq 2 \).

2. Haar basis: \( 1_{[0, 1]} \) and \( h_I(x) = \frac{1}{\sqrt{|I|}}(1_I - 1_{I_l}) \) indexed by dyadic subintervals \( I \subset [0, 1] \) is a basis for \( L^2 \) and all \( L^p[0, 1] \) with \( 1 \leq p < \infty \), note that \( p \) could be 1 here. This is an unconditional basis if \( 1 < p < \infty \).

3. Orthogonal polynomials: If \( w(x), x \in \mathbb{R} \) has sufficiently fast decay (say subexponential - Bernstein’s theorem) then the orthogonal polynomials with respect to \( d\mu = w dx \) is
a basis for $L^2(\mathbb{R}, \mu)$. Here $p_n(x) = a_{n,n}x^n + \cdots + a_{0,n}$ where $a_{n,n} > 0$ and 

$$
\int p_n(x)\overline{p_m(x)}d\mu = \delta_{mn}
$$

which is 1 if $n = m$ and zero otherwise.

**Proof of Theorem 19.** It is clear that (i) and (ii) are the same. So we only show equivalence to (iii).

We first observe that the following are equivalent for any given sequence $(x_n)$ in a Banach space $X$:

(a) the series $\sum_n x_n$ converges unconditionally;

(b) for any $\epsilon > 0$ there exists $n = n(\epsilon)$ such that $\|\sum_{k \in M} x_k\| < \epsilon$ for any finite subset $M \subset (n, \infty)$ of the integers.

Note that the direction (b) $\rightarrow$ (a) is a consequence of the completeness of the space. (Basically (b) implies that for any permutation of $\mathbb{N}$ the sequence $\sum_{k=1}^n x_{\sigma(k)}$ would be a Cauchy sequence).

For the other direction (a) $\rightarrow$ (b), assume towards a contradiction that there is an $\epsilon > 0$ such that no such $n$ could be found, thus we could find a sequence of sets $M_1$, $M_2$, ... such that $\max M_j < \min M_{j+1}$ and $\|\sum_{k \in M_j} x_k\| > \epsilon$. It is not hard to build a permutation $\sigma$ of $\mathbb{N}$ such that each $M_j$ appears as a consecutive block of $\sigma(1), \sigma(2), \ldots$, thus $\sum_{k=1}^n x_{\sigma(k)}$ is not a Cauchy sequence in $X$ so won't be convergent, contradiction.

Using this observation we claim that if a given series $\sum_n x_n$ is unconditionally convergent then $\sum_n x_{\sigma(n)}$ is the same as $\sum_n x_n$ for all permutation of $\mathbb{N}$. Using this fact it is not hard to see that (i) and (ii) are equivalent. To see this claim, just note that for any $\sigma$ there is some $M$ such that $\sigma(M+1), \cdots > n(\epsilon)$, therefore $\|\sum_{k>M} x_{\sigma(k)}\| \leq \epsilon$ and $\|\sum_{k>n(\epsilon)} x_k\| \leq \epsilon$, and by canceling out the common terms it follows that 

$$
\|\sum_{k \leq n(\epsilon)} x_k - \sum_{k \leq M} x_{\sigma(k)}\| = \|\sum_{k \leq M: \sigma(k)>n(\epsilon)} x_{\sigma(k)}\| < \epsilon
$$

thus $\|\sum_k x_{\sigma(k)} - \sum_k x_k\| < 3\epsilon$ for all $\epsilon > 0$, implying the desired claim.

Now we show that (ii) and (iii) are equivalent. In fact for simplicity we will only show the equivalence to the version with $\epsilon = \pm 1$.

It is clear that (iii) implies (ii): suppose that $\sum_k \alpha_k x_k$ converges and (iii) holds true. Then for any $\epsilon > 0$ there exists $n = n(\epsilon)$ such that $\|\sum_{k>n} \alpha_k x_k\| < \epsilon/(2C)$. Given any finite $A \subset (n, \infty)$ by choosing the sign sequence that equals 1 on $A$, equals 0 on $\{1, 2, \ldots, n\}$ and equals ±1 constantly on $\mathbb{N}\setminus A$ and using (iii) we obtain $\|\sum_{k \in A} \alpha_k x_k\| \leq 2C \sum_{k>n} \alpha_k x_k\| < \epsilon$. Thus by the observation above the series $\sum_k \alpha_k x_k$ converges unconditionally.

Now, to show that (ii) implies the sign sequence version of (iii), we will use the following lemma

**Lemma 3.** Let $P_n$ be the projection $P_n x = \sum_{k \leq n} \alpha_k x_k$ if $x = \sum_k \alpha_k x_k$ is the expansion into the Schauder basis $(x_n)$ for $x$. Then $\sup_n \|P_n\| < \infty$. 

We first prove the above implication using the lemma. The proof of the lemma will be discussed in the next class, after we have introduced several basic tools in Banach spaces, such as the principle of uniform boundedness and the open mapping theorem. Now, it suffices to show that there exists \( N > 0 \) and a scalar \( C_N > 0 \) such that for every sequence of scalar \( \alpha_j \) and signs \( \epsilon_j \) and \( n \geq N \) it holds that \( \| \sum_{j=N}^n \epsilon_j \alpha_j x_j \| < C_N \| \sum_{j=1}^n \alpha_j x_j \| \). Indeed, the desired claim would follow from the fact that given any \( N \) and any sign sequence \( \epsilon_1, \ldots \) it holds uniformly over \( n \geq N \) that

\[
\| \sum_{j=1}^N \epsilon_j \alpha_j x_j \| \lesssim_N \| \sum_{j=1}^n \alpha_j x_j \|
\]

To see this, simply use Lemma 3 to obtain \( \| \alpha_j x_j \| \leq \|(P_{j-1}-P_j)(\sum_{k=1}^n \alpha_k x_k)\| \leq C \| \sum_{k=1}^n \alpha_k c_k \| \) and then use the triangle inequality.

Now assume towards a contradiction that for every \( N > 0 \) one can not find such \( C_N \). We then construct a sequence \( n_1 < n_2 < \ldots \) and \( A_j \subset [n_j, n_{j+1} - 1] \) and \( \beta_1, \beta_2, \ldots \) recursively as follows, with the following property: for any \( j \) it holds that \( \| \sum_{k \in A_j} \beta_k a_k \| \geq j^2 \| \sum_{n_j \leq k \leq n_{j+1}-1} \beta_k a_k \| \).

Suppose that we have chosen \( n_1 < \cdots < n_j \) and \( A_1, \ldots, A_{j-1} \) and \( \beta_1, \ldots, \beta_{n_j-1} \) satisfying the above requirements. Then given \( C > 0 \) by the assumption there exists \( n_{j+1} \) and \( \beta_1, \ldots, \beta_{n_{j+1}-1} \) and \( B \subset [n_j, n_{j+1}) \) such that

\[
\| \sum_{k \in B} \beta_k x_k - \sum_{k \in [n_j, n_{j+1}) - B} \beta_k x_k \| \geq C \| \sum_{1 \leq k \leq n_{j+1}-1} \beta_k x_k \|
\]

On the other hand, by the triangle inequality

\[
\| \sum_{n_j \leq k \leq n_{j+1}-1} \beta_k x_k \| \leq \| \sum_{1 \leq k \leq n_{j+1}-1} \beta_k x_k \| + \| P_{n_j-1} ( \sum_{1 \leq k \leq n_{j+1}-1} \beta_k x_k ) \|
\]

\[
\lesssim_{n_j} \| \sum_{1 \leq k \leq n_{j+1}-1} \beta_k x_k \|
\]

therefore by choosing \( C \) large enough we could ensure that

\[
\| \sum_{k \in B} \beta_k x_k - \sum_{k \in [n_j, n_{j+1}) - B} \beta_k x_k \| \geq 2 j^2 \| \sum_{n_j \leq k \leq n_{j+1}-1} \beta_k x_k \|
\]

from here it is clear that either \( A_j = B \) or \( A_j = [n_j, n_{j+1}) - B \) will work, with \( \alpha_k = \beta_k \) for \( k \in [n_j, n_{j+1}) \). This completes the selection of \( A_j \) and \( n_j \) and \( \alpha_j \).

Now by rescaling \( \alpha_k \)'s we may assume that \( \| \sum_{k \in A_j} \alpha_k x_k \| = 1 \) while \( \| \sum_{k = n_j}^{n_{j+1}-1} \alpha_k x_k \| \leq 1/j^2 \). Now, given any \( m_1 < m_2 \) we let \( n_s < \cdots < n_t \) be the elements of \( (n_j) \) inside \( [m_1, m_2) \), then using the uniform boundedness of \( P_n \) we obtain

\[
\| \sum_{k=m_1}^{m_2} \alpha_k x_k \| \lesssim \sum_{j=s-1}^t \| \sum_{n_j \leq k < n_{j+1}} \alpha_k x_k \| \lesssim \frac{1}{s}
\]
and clearly if $m_1 \to \infty$ then so is $s$, thus $\sum_k \alpha_k x_k$ converges. On the other hand it does not converge unconditionally since it violates the second property of the initial observation: given any $n$ we could select $j$ such that $n_j > n$, and so $A_j \subset (n, \infty)$ and $\| \sum_{k \in A_j} \alpha_k x_k \| = 1$. □

**Exercise:**

1. Prove that $(e^{2\pi inx})_{n \in \mathbb{Z}}$ is an unconditional Schauder basis for $L^2[0, 1]$.

2. Prove that the Haar functions are orthogonal in $L^2[0, 1]$ and form a Schauder basis for $L^1[0, 1]$ but not unconditional in $L^1$.

3. (Radon–Nikodym) Let $\mu$ and $\nu$ two nonnegative $\sigma$-finite measures on the same $\sigma$ algebra $(X, \mathcal{A})$ such that the following holds: for every measurable $A$ if $\mu(A) = 0$ then $\nu(A) = 0$. Show that there exists $g$ nonnegative such that for every measurable $E$ it holds that $\nu(E) = \int_E g \, d\mu$. 
Chapter 5

Basic results about Banach Spaces

Here we will prove several basic results about bounded linear maps between Banach spaces: the principle of uniform boundedness, the open mapping theorem, and the closed graph theorem, and applications. We then focus on $L^p$ spaces.

5.1 The Baire category theorem

We first recall the Baire category theorem for metric spaces, which will be used in the proofs of these theorems.

Below we say that a set $A$ is nowhere dense if its closure $\overline{A}$ has empty interior.

**Theorem 20** (Baire). A complete metric space can not be written as a countable union of nowhere dense set.

**Proof:** If $X = \bigcup A_n$ we can construct a Cauchy sequence $\{x_n\}$ that does not have a limit as follows: Since $A_1$ is nowhere dense there is an open ball of small radius $B_1 = B(x_1, r_1)$ such that $\overline{B_1} \subset X \setminus \overline{A_1}$. Then again find an open ball $B_2 = B(x_2, r_2)$ such that $\overline{B_2} \subset B_1 \setminus \overline{A_2}$. Repeat this, make sure $r_k \leq r_{k-1}/2$, we get a Cauchy sequence which can not converge to anything in any $A_k$, so no limit inside $X$. □

5.2 The principle of uniform boundedness

Let $X$ be a Banach space and $T_\alpha, \alpha \in A$ is a family of bounded linear functionals from $X$ to some normed linear space $Y$.

**Theorem 21** (P.U.B, aka Banach-Steinhaus). If $\sup_{\alpha \in A} \|T_\alpha x\| < \infty$ for every $x \in X$ then $\sup_{\alpha \in A} \|T_\alpha\| < \infty$.

Remark: the opposite direction is clearly true.

**Example:** If $g_n \in L^\infty(\mathbb{R})$ such that $\sup_n |\int g_n(x)f(x)dx| < \infty$ for every $f \in L^1(\mathbb{R})$ then $\sup_n \|g_n\|_\infty < \infty$. 

45
CHAPTER 5. BOUNDED MAPS ON BANACH SPACES

Proof. We first show that for a linear operator \( T : X \to Y \) (only need normed linear spaces \( X, Y \)) \( T \) is bounded if any only if \( T^{-1}(\|y\|_Y \leq 1) \) has nonempty interior in \( X \).

Certainly if \( T \) is bounded then the set \( T^{-1}(\|y\|_Y \leq 1) \) contains \( \{\|x\|_X \leq 1/\|T\|\} \) so nonempty interior.

Conversely, if \( T^{-1}(\text{unit ball in } Y) \) contains an interior point \( x_0 \), then for some \( \epsilon > 0 \) it holds for every \( \|x\|_X \leq \epsilon \) that \( \|T(x_0 + x)\| \leq 1 \). It follows from the triangle inequality that \( \|Tx\| \leq 1 + \|Tx_0\| \) for such \( x \), thus \( \|T\| \leq \frac{1+\|Tx_0\|}{\epsilon} \).

Therefore in order to show P.U.B it suffices to show that the set

\[
B_1 = \bigcap_{\alpha \in A} T_\alpha^{-1}(\|y\| \leq 1)
\]

has an interior point. In fact it suffices to show that for some \( n > 0 \) the (closed) set \( B_n \) (replace the radius by \( n \)) has nonempty interior. But \( \bigcup B_n = X \) because \( \sup_{\alpha} \|T_\alpha x\| < \infty \). Thus the desired claim follows from the Baire category theorem. \( \Box \)

5.3 The open mapping theorem

Theorem 22 (Open mapping, aka Banach Schauder). Let \( T : X \to Y \) be a linear map between Banach spaces. If \( T \) is bounded and surjective then \( T \) maps open sets to open sets.

Remark: as a corollary, if \( T : X \to Y \) bounded bijective linear operators between Banach spaces then \( T^{-1} \) exists and is bounded linear.

Proof. Let \( B_r = \{\|x\| < r\} \subset X \).

By translation invariant if suffices to show that for every \( r > 0 \) the set \( T(B_r) \) contains a neighborhood of 0. By dilation, if this holds for one \( r > 0 \) then it holds for all \( r > 0 \).

Now this in turns follows if we could show that \( T(B_r) \) contains an interior point: indeed suppose that for some \( \epsilon > 0 \) and \( y_0 \) we have \( \{\|y - y_0\| < \epsilon\} \subset T(B_r) \), then by linearity \( T(B_{2r}) \) contains \( T(B_r) - T(B_r) \) which in turn contains \( \{\|y\| < \epsilon/2\} \) a neighborhood of 0.

To show that \( T(B_r) \) has nonempty interior, we first note that \( \overline{T(B_r)} \) has nonempty interior. This follows from surjectivity of \( T \) and the Baire category theorem \( \bigcup_n T(\|x\| < n) = Y \). Therefore it suffices to show that \( T(B_1) \) contains \( \overline{T(B_{r/10})} \).

To see this, we note that \( \overline{T(B_r)} \) contains a neighborhood of 0 via the set difference argument (as above). Assume that \( \{\|y\| \leq r/C\} \subset T(B_r) \) for some \( C > 0 \); by dilation invariant this \( C \) could be chosen uniformly over \( r \). Let \( y_0 \in \overline{T(B_{r/10})} \) then there is a point \( x_1 \in B_{r/10} \) such that \( \|y - Tx_1\| < \frac{r}{20C} \), thus \( y - Tx_1 \in \overline{T(B_{r/20})} \), so again there is \( x_2 \in B_{r/20} \) such that \( \|y - Tx_1 - Tx_2\| < \frac{r}{40C} \), thus by repeating this argument we obtain \( x_1, x_2, \ldots \) such that \( \|x_n\| \leq \frac{r}{2nC} \) and \( T(x_1 + \cdots + x_n) \to y \) in \( Y \) as \( n \to \infty \). Since \( X \) is compact it follows that \( x_1 + \cdots + x_n \to x \) in \( X \) as \( n \to \infty \), thus using continuity of \( T \) if follows that \( y = Tx \), and clearly \( \|x\| \leq \sum \|x_k\| < r \), thus \( y \in TB_r \) as desired. \( \Box \)
5.4 The closed graph theorem

Graph of a function \( f : X \to Y \) is defined to be \( \Gamma(f) = \{(x, f(x)), x \in X\} \) which is a subset of \( X \times Y \) equipped with the product topology. If \( X \) and \( Y \) are normed spaces then the product topology on \( X \times Y \) is the topology of the norm \( \|(x, y)\| = \|x\| + \|y\| \).

**Theorem 23** (Close graph). Let \( T : X \to Y \) be linear where \( X \) and \( Y \) are Banach spaces. Then \( T \) is bounded iff \( \Gamma(T) \) is closed.

**Remark:** Since \((X, Y)\) is a Banach space also, \( \Gamma(T) \) is closed iff whenever \((x_n, T(x_n))\) converges to \((x, y)\) one must have \( y = Tx \). Comparing with the usual continuity of \( T \) (which is equivalent to boundedness of \( T \)), which states that whenever \( x_n \to x \) we must have \( Tx_n \to Tx \), we see that the extra thing we could assume is the convergence of \( Tx_n \).

**Proof.** Clearly if \( T \) is bounded then \( T \) is continuous and the graph is closed. Assume now \( \Gamma(T) \) is closed, then \( \Gamma(T) \) is a Banach subspace of \( X \times Y \). Let \( \pi_1(x, y) := x \) and \( \pi_2(x, y) := y \) be the coordinate projections. By the open mapping theorem, the bounded bijective map \( \pi_1(x, y) : \Gamma(T) \to X \) is boundedly invertible, therefore \( T = \pi_2 \circ \pi_1^{-1} \) is bounded. \( \square \)

**Corollary 2** (Hellinger-Toeplitz). If \( A \) is linear operator on \( H \) a Hilbert space and \( \langle x, Ay \rangle = \langle Ax, y \rangle \) then \( A \) is bounded.

**Proof.** Clearly \( \Gamma(A) \) is closed: if \((x_n, Ax_n)\) converges to \((x, y)\) then for every \( z \in H \) we have \( \langle z, y \rangle = \lim \langle z, Ax_n \rangle = \lim \langle Az, x_n \rangle = \langle Az, x \rangle = \langle z, Ax \rangle \), thus \( y = Ax \). \( \square \)

5.5 An application: uniform boundedness of partial sum projections for Schauder basis

Here we prove Lemma 3 from the last Chapter.

Recall that \( P_n \) is the projection \( P_n x = \sum_{k \leq n} \alpha_k x_k \) if \( x = \sum \alpha_k x_k \) is the expansion into the Schauder basis \((x_n)\) for \( x \in X \) a Banach space.

We need to show that \( \sup_n \|P_n\| < \infty \).

By the principle of uniform boundedness it suffices to show that the linear operator \( P_n \) are bounded (the pointwise uniform boundedness of \( P_n \) follows from the fact that at each \( x \in X \) we have \( \|P_n x\| \to \|x\| \) and a convergent sequence is always bounded). Now to show that \( P_n \) are bounded, we consider the following norm on \( X \)

\[
\|x\|_1 = \sup_n \|P_n x\|
\]

it is not hard to see that this is a norm and \( \|x\| \leq \|x\|_1 \). We will show that \( P_n \) is bounded in \( (X, \|\cdot\|_1) \) and the norm \( \|\cdot\|_1 \) is actually equivalent to \( \|\cdot\| \), these two facts will take care of the lemma. Now the boundedness of \( P_n \) with respect to the new norm is clear

\[
\|P_n x\|_1 = \sup_m \|P_m P_n x\| = \sup_m \|P_{\min(m,n)} x\| \leq \|x\|_1
\]
Now we show that \((X, \|\cdot\|_1)\) is complete. Once we did that, the identity map from \((X, \|\cdot\|_1) \to (X, \|\cdot\|)\) is a bijective bounded linear map thus by the open mapping theorem it is boundedly invertible, thus the two norms are equivalent and completes the proof of the lemma. Now, let \(X'\) be the completion of \(X\) under \(\|\cdot\|_1\), then it is not hard to see that \(x_n\) is still a Schauder basis for \((X', \|\cdot\|_1)\) (see the homework). So if \(a \in X'\) we have \(a = \sum \alpha_j x_j\) convergence in \(\|\cdot\|_1\), but this series also converges in \(\|\cdot\|\) too because \(\|\sum_{m<j\leq n} \alpha_j x_j\| \leq \|\sum_{m\leq n} \alpha_j x_j\|_1 \to 0\) and \(X\) is complete. So, let \(b \in X\) be its limit under \(\|\cdot\|\). Now, it is clear that \(b = a\) because

\[
\|b - \sum_{j\leq n} \alpha_j x_j\|_1 = \sup_m \|P_m (b - \sum_{j\leq n} \alpha_j x_j)\| = \sup_{m\geq n} \|(P_m - P_n)b\| \to 0
\]

if \(n \to \infty\). □

5.6 \(L^p\) spaces

In this section we conduct a case study of \(L^p\) spaces, which are the most fundamental type of Banach spaces.

5.6.1 Separability

Recall that if \((X, \mathcal{A}, \mu)\) is separable then \(L^p(X, \mathcal{A}, \mu)\) is separable for \(1 \leq p < \infty\); this is part of the homework.

5.6.2 Duality

We’ll use Hilbert space techniques to show that

**Theorem 24.** For \(1 < p < \infty\) the dual of \(L^p(X, \mu)\) is \(L^q(X, \mu)\) where \(1/p + 1/q = 1\). If furthermore the measure \(\mu\) is \(\sigma\)-finite then the dual of \(L^1\) is \(L^\infty\).

**Proof.** We first consider the simpler case when \(\mu\) is \(\sigma\)-finite, in this case we will show the result for all \(1 \leq p < \infty\).

If \(\mu\) is not necessary \(\sigma\)-finite we will need \(p > 1\). Let \(\ell \in (L^p)^*\). Suppose that \(E \subset X\) measurable such that \(\mu|_E\) is \(\sigma\)-finite. Clearly \(\ell\) is also a bounded linear functional on \(L^p(E, \mu)\) whose norm does not exceed \(\|\ell\|\). Thus, by the \(\sigma\)-finite case, for every \(f \in L^p(E)\) there is \(g_E \in L^q(E, \mu)\) such that

\[
\ell(f) = \int_E fg_E d\mu
\]

\[
\|g_E\|_q \leq \|\ell\|
\]

Now if \(E \subset E'\) both measurable and \(\sigma\)-finite wrt \(\mu\) then it is clear that \(g_E = g_{E'}\) on \(E\).
5.6. $L^p$ SPACES

Now, using the fact that a countable union of $\sigma$-finite sets is $\sigma$-finite one could find $F$ $\sigma$-finite wrt $\mu$ such that $\|g_F\|_q = \sup_E \|g_E\|_q$. We will show that

$$\ell(f) = \int fg_F d\mu$$

for every $f \in L^p(X)$. Observe that $g|_{A \setminus F} = 0$ for every $A$ $\sigma$-finite, because otherwise $\int |g_{A \setminus F}|^q d\mu > 0$ and therefore $\|g_{A \cup F}\|_q > \|g\|_F$ contradiction (note that here is where we need $p > 1$ so that $q < \infty$). In particular let $A = \{|f| > 0\}$ which is $\sigma$-finite wrt $\mu$ and so

$$\ell(f) = \int f g_A d\mu = \int f g_F d\mu$$

We now focus on the $\sigma$-finite case. For simplicity we will assume $\mu(X) < \infty$, otherwise we could decompose $X$ into countably many subsets where $\mu$ is finite and apply this argument to each subspace and add things up: because of continuity of $\ell$, for each $f \in L^p(X)$ we have

$$\ell(f) = \lim_{n \to \infty} \sum_{k \leq n} \int_{X_k} f g_k d\mu$$

and use monotone convergence plus the fact that $\|\sum_{k \leq n} g_k\|_q = \|\sum_{k \leq n} \|g_k\|_q^{1/q}\|^{q} \leq \|\ell\| < \infty$ to interchange the sum and define $g = \sum_k g_k$. (Note that the $g_k$ have disjoint support).

Now we assume $\mu(X) < \infty$. In this case it can be shown that any $\ell$ could be written as a linear combination of finitely many positive linear functionals, i.e. $\ell(f) \geq 0$ if $f \geq 0$.

For each $E \subset X$ measurable construct a measure

$$\nu(E) = \ell(1_E)$$

it is clear that $\nu \ll \mu$, therefore by the Radon Nikodym theorem there exists $g \in L^1$, $g \geq 0$, such that

$$\ell(1_E) = \nu(E) = \int_E g d\mu$$

for any measurable $E \subset X$. Now given any $f \in L^p$ we will show that $\ell(f) = \int f g d\mu$. We may assume wlog that $f \geq 0$, now by the monotone convergence theorem it suffices to show this for $f$ being simple functions, for which the claim holds. It remains to show that $g \in L^q$ (note that since we assume $\mu(X) < \infty$ this is better than $g \in L^1$ since $L^1$ contains $L^q$ for all $q \geq 1$). Simply let $f = |g|^{p-1} 1_{|g| \leq K}$, then

$$\int |g|^{q} 1_{|g| \leq K} d\mu = \ell(f) \leq \|\ell\| \|f\|_p = \|\ell(\int |g|^{q} 1_{|g| \leq K} d\mu)^{1/p}\|
$$

thus

$$\left(\int |g|^{q} 1_{|g| \leq K} d\mu\right)^{1/p} \leq \|\ell\|$$

therefore by monotone convergence we obtain $\|g\|_q \leq \|\ell\|$. □
5.6.3 Basic facts about bounded operators on $L^p$

The most typical type of operator on $L^p$ spaces are integral operators. Let $(X, \mu)$ and $(Y, \nu)$ be two measure spaces. Let

$$Tf(y) = \int_Y K(x, y) f(x) d\mu(x)$$

mapping from measurable functions on $(X, \mu)$ to measurable functions on $(Y, \nu)$. At the beginning the operator $T$ may not be defined for all $f$, as the integral may not be convergent.

The adjoint of $T$ is

$$T^*g(x) = \int_X K(x, y) g(y) d\sigma(y)$$

Think of this as a generalization of the matrix in finite dimensional setting. (integration is like adding over indices and $K$ gives the matrix entries.) (This turns out to be the case if say $X = \{1, \ldots, n\}$ and $Y = \{1, \ldots, m\}$ with the counting measures.) The function $K$ is called the kernel (assumed measurable etc.).

Boundedness of integral operators

**Theorem 25.** For $1 \leq p \leq \infty$ it holds that

$$\|T\|_{L^p \to L^p} \leq \left[ \sup_x \int_X |K(x, y)| d\sigma(y) \right]^{1/p} \left[ \sup_y \int_X |K(x, y)| d\mu(x) \right]^{1/p'}$$

In the special case $p = 2$ one also has

$$\|T\|_{L^2 \to L^2} \lesssim \left( \int_{X \times X} |K(x, y)|^2 d\mu(x) d\sigma(y) \right)^{1/2}$$

Remarks: The dual estimates hold for $T^*$, namely we also have the same $L^2 \to L^2$ estimate and

$$\|T\|_{L^{p'} \to L^{p'}} \leq \left[ \sup_x \int_X |K(x, y)| d\sigma(y) \right]^{1/p} \left[ \sup_y \int_X |K(x, y)| d\mu(x) \right]^{1/p'}$$

In fact we will see that $\|T\|_{p \to p} = \|T^*\|_{p' \to p'}$.

Operators for which $(\int_{X \times X} |K(x, y)|^2 d\mu(x) d\sigma(y))^{1/2} < \infty$ are called Hilbert-Schmidt operators and the right hand side is called the Hilbert Schmidt norm. In the finite dimensional setting, for instance if $X = Y = \{1, \ldots, n\}$ with counting measures then this norm is the same as $(\sum_{1 \leq i, j \leq n} |K(i, j)|^2)^{1/2}$, or equivalently the $\ell^2$ norm of the sequence of eigenvalues (if diagonalizable).

For $1 < p < \infty$ to use the first estimate we need both factor to be finite. But if $p = 1$ or $p = \infty$ we only need one of them to be finite. More precisely

$$\|T\|_{L^1 \to L^1} \leq \sup_x \left[ \int_X |K(x, y)| d\sigma(y) \right]$$
5.6. $L^p$ SPACES

\[ \|T\|_{L^\infty \to L^\infty} \leq \sup_y \int_X |K(x, y)|d\mu(x) \]

These turn out to be the easiest cases. For the $L^1$ case this follows from

\[ \|Tf\|_{L^1(Y, \nu)} \leq \int_Y \int_X |K(x, y)|\|f(x)|d\mu(x)d\sigma(y) \]

\[ = \int_X |f(x)|\int_Y |K(x, y)|d\sigma(y)\] \[ \leq \|f\|_1 \sup_x \int_X |K(x, y)|d\sigma(y) \]

Similarly for the $L^\infty$ case we have, for each $y \in Y$

\[ |Tf(y)| \leq \int_X |K(x, y)||f(x)|d\mu(x) \]

\[ \leq \|f\|_\infty \sup_y \int_X |K(x, y)|d\mu(x) \]

To get the case $1 < p < \infty$ of the first estimate, we will need Riesz-Thorin interpolation theorem (aka convexity theorem).

**The Riesz-Thorin interpolation theorem**

**Theorem 26 (Riesz-Thorin).** Let $1 \leq p_0 \leq p_1 \leq \infty$ and $1 \leq q_0 \leq q_1 \leq \infty$. Let $T$ be a linear map from $(X, \mu)$ to $(Y, \nu)$ such that

\[ M_0 := \|T\|_{L^{p_0}(X, \mu) \to L^{q_0}(Y, \nu)} < \infty \]

\[ M_1 := \|T\|_{L^{p_1}(X, \mu) \to L^{q_1}(Y, \nu)} < \infty \]

Assume $1 < p, q < \infty$ be such that for some $0 < \alpha < 1$

\[ \left( \frac{1}{p}, \frac{1}{q} \right) = \left( 1 - \alpha \right) \left( \frac{1}{p_0}, \frac{1}{q_0} \right) + \alpha \left( \frac{1}{p_1}, \frac{1}{q_1} \right) \]

Then

\[ \|T\|_{L^p(X, \mu) \to L^q(Y, \nu)} \leq M_0^{1-\alpha} M_1^\alpha \]

**Proof.** We will use the Hadamard three lines theorem: let $h$ be analytic bounded in the strip \{Re$(z) \in [0, 1]$\}. For each $\alpha \in [0, 1]$ let $H(\alpha) = \sup_{\text{Re}(z)=\alpha} |h(z)|$. Then

\[ H(\alpha) \leq H(0)^{1-\alpha} H(1)^\alpha \]

The proof of this is an application of the maximum principle: if $H(0) = H(1) = 1$ then the claims follows by the maximum principle, in the general case modify $h$ by a suitable exponential factor to reduce to this setting.
We now set up the bounded analytic function \( h(z) \). For \( \Re(z) \in [0,1] \) let \( p_z \) and \( q_z \) be defined by
\[
\left( \frac{1}{p_z}, \frac{1}{q_z} \right) = (1 - z)(\frac{1}{p_0}, \frac{1}{q_0}) + z(\frac{1}{p_1}, \frac{1}{q_1})
\]
Then \( p_z \) and \( q_z \) are analytic functions. For any \( f \in L^p \) and \( g \in L^{q'} \) \((1/q' + 1/q = 1)\) we define
\[
h(z) = \int_Y T(f_z)g_z d\sigma
\]
where \( f_z(x) = |f(x)|^{\frac{2}{p}-1}f(x) \) and \( g_z(x) = |g(x)|^{\frac{q'}{q} - 1}g(x) \). It is not hard to see that \( h(z) \) is bounded analytic for \( \Re(z) \in [0,1] \) and \( h(\alpha) = \int Tfgd\sigma \), therefore using the duality characterization for \( L^p \) we obtain
\[
M = \sup_{f \in L^p, g \in L^{q'}} |H(\alpha)|
\]
Note that everything depends on \( f, g \) where \( f \in L^p \) and \( g \in L^{q'} \). We normalize \( \|f\|_p = \|g\|_{q'} = 1 \), then it is not hard to see that
\[
|H(0)| \leq M_0 \quad , \quad |H(1)| \leq M_1
\]
therefore the desired claims follows from Hadamard three lines lemma. \( \square \)

**Exercises:**

1. Let \( X \) be a Banach space. Prove that separability of \( X^* \) implies separability of \( X \).
2. Prove that the measure space \((X, A, \mu)\) is separable if and only if \( L^2(X, A, \mu) \) is separable. Is this true if we replace 2 by any fixed \( p \in (1, \infty) \)? Use this to construct a finite measure space \((X, \mu) \ (\mu(X) < \infty)\) such that \( L^2(X, \mu) \) is nonseparable.
3. Prove that if \( X \) is a compact metric space then \( C(X) \) with the sup norm is separable.
4. Let \( X \) be a normed linear space and assume that \((x_n)\) is a Schauder basis, furthermore the canonical projection \( P_n \) into the first \( n \) vectors is bounded for each \( n \geq 1 \). Show that \((x_n)\) remains a Schauder basis for the completion of \( X \). [Hint: show that \( P_n \) extends boundedly to the completed space and use \( P_n \) to find the linear expansion/proving required properties for Schauder basis.]
5. Let \( \ell \) be a bounded linear functional on \( L^p(X, \mu) \) where \( \mu(X) < \infty \). Show directly (without using the duality theorem) that \( \ell \) could be written as a linear combination of finitely many positive linear functionals on \( L^p(X, \mu) \). Here \( 1 \leq p < \infty \).
6. Show that every NLS is isomorphic to a subspace of the space of bounded continuous function on some complete metric space \( Y \). \( \square \)

---

1 if the given space is separable then one could take \( Y \) to be \([0,1]\) - this is the Banach-Mazur theorem.
Chapter 6

Bounded and continuous functions on a locally compact Hausdorff space and dual spaces

Recall that the dual space of a normed linear space is a Banach space, and the dual space of $L^p$ is $L^q$ where $1/p + 1/q = 1$ if $1 < p < \infty$. If the underlying measure space is $\sigma$-finite then the dual of $L^1$ is $L^\infty$. What about the dual of $L^\infty$? Or its subspace $C_0(X)$ and $C_c(X)$ where say $X$ is locally compact Hausdorff. In this chapter we discuss several representation theorems related to these themes.

6.1 Riesz representation theorems

6.1.1 The dual space of $L^\infty$

The dual space of $L^\infty(X, \Sigma, \mu)$ consists of bounded (i.e. the total measure of $X$ is finite) and finitely additive signed/complex measures on $\Sigma$ that is absolutely continuous with respect to $\mu$.

Clearly if $\sigma$ is a such a measure we may define a linear functional $\ell_\sigma(f) := \int f d\sigma$, it is not hard to see that this is well defined as a bounded linear functional on $L^\infty$. (The fact that $\sigma \ll \mu$ ensures that if $f$ and $f'$ equals almost everywhere with respect to $\mu$, i.e. they represent the same $L^\infty$ function, then the two integrals agree.) Note that it is possible to define integration with respect to a finitely additive measure: start with positive function and define the integral as the supremum over integration of simple functions dominated by the given positive function. Then define integral of signed functions and complex valued functions etc.

Conversely, given a bounded linear functional $\ell$, after several reductions we may assume that $\ell$ is nonnegative, namely if $f \geq 0$ then $\ell(f) \geq 0$. Then we may define $\sigma(E) = \ell(1_E)$, it is not hard to see that $\sigma$ is a finitely additive bounded measure, and if $\mu(E) = 0$ then $1_E \equiv 0$ in $L^\infty(\mu)$ thus $\sigma(E) = \ell(1_E) = \ell(0) = 0$. Boundedness of the measures follows from choosing
6.1.2 Measures on a locally compact Hausdorff spaces

There are several related and sometimes equivalent notions of measures on a given locally compact Hausdorff space. Here we list and compare some of them.

Regular Borel measures and Radon measures

Recall that a Borel set is a set in the $\sigma$-algebra generated by the open sets.

A regular Borel measure is a Borel measure (i.e. all Borel sets on $X$ are measurable) such that for any Borel set $E$ the following holds.

$$
\mu(E) = \inf \{ \mu(U) : U \text{ open, } E \subset U \}
$$

$$
\mu(E) = \sup \{ \mu(K) : K \text{ compact, } K \subset E \}
$$

A Radon measure is a locally finite regular Borel measure, i.e. in addition to being regular Borel it also assigns a finite value to all compact subsets of $X$.

Note that some textbooks define Radon measure as locally finite inner regular Borel measures, thus requiring only that measures of Borel sets could be approximated from inside using measures of compact subsets.

Baire sets and Baire measures

A Baire set is an element of the sigma algebra generated by all compact subsets of $X$ that are at the same time countable intersections of open sets.

A Baire measure is a measure on the Baire sigma algebra.

One could check that a Baire measure is also regular with respect to the Baire sigma algebra.

The Baire sigma algebra is the smallest sigma algebra such that all continuous functions are measurable.

It can be shown that any Baire measure extends uniquely to a Radon measure and vice versa.

6.1.3 The dual space of $C_o(X)$ and $C_c(X)$

Let $X$ be LCH. Since the closure of $C_c(X)$ under the uniform norm is $C_o(X)$, it suffices to consider the dual of $C_o(X)$. Note that if $X$ is not compact then $C(X)$ is not a normed linear space with the sup norm, and we have to look at its dual as a topological dual. More on this in later chapters.

We say that a signed measure is a signed Radon measure if it could be written as the difference of two Radon measures. We say that a complex valued measure is a complex valued Radon measure if its real and imaginary parts are signed Radon measures.
Theorem 27 (Riesz). Let $X$ be a LCH. Then the dual of $C_o(X)$ is $M(X)$ the space of complex valued Radon measures (locally finite regular Borel measures) on $X$. Namely the following map is an isometric isomorphism between this dual and $M(X)$

$$
\mu \mapsto \ell_{\mu}(f) = \int f d\mu
$$

and $\|\ell_{\mu}\| = \|\mu\|$ the total variation of $\mu$ defined by $\sup \sum_{j=1}^{m} |\mu(A_j)|$ supremum over all collection $A_1, \ldots, A_m$ of disjoint subsets of $X$.

**Step 1:** We first show that if $\ell$ is a bounded positive linear functionals on $C_o,\mathbb{R}(X)$, the space of real valued continuous functions on $X$ that vanish at $\infty$, then there is a Radon measure $\mu$ such that for any $f \in C_o,\mathbb{R}(X)$ it holds that

$$
\ell(f) = \int f d\mu
$$

To see this, we would like to define $\mu$ by $\mu(E) = \ell(1_E)$, however $1_E$ is not continuous therefore one can not apply $\ell$ to this function. To get around this we may define for each open set $U \subset X$

**Definition:** $\mu(U)$ is defined to be the supremum over $\ell(f)$ where $f \in C_o(\mathbb{R})$ and the support of $f$ is inside $U$ and $0 \leq f \leq 1$ pointwise.

(Note that we may not be able to do this if $U$ wasn’t open since such a continuous function may not exists; for open $U$ the existence of $f$ is due to Urysohn’s lemma).

Now we obtain a premeasure on open sets which is a ring of sets, thus by Caratheodory we may extend $\mu$ to the sigma algebra generated by these sets, which are exactly the Borel sets.

We want to show that $\ell(f) = \int f d\mu$ for every $f \in C_o(X)$. Without loss of generality assume $0 \leq f \leq 1$. Since $f$ vanishes at infinity the sets $A_k = \{ f \geq k/n \}$ are compact, and we may find continuous functions $f_k$ such that

$$
1_{A_k} \leq n f_k \leq 1_{A_{k-1}}
$$

and $f = \sum_{k=1}^{a} f_k$. We will use the following lemma

**Lemma 4.** If $E \subset F$ are compact subsets and $f$ is continuous on $X$ such that $1_E \leq f \leq 1_F$ then $\mu(E) \leq \ell(f) \leq \mu(F)$.

Using the lemma and positivity of $\ell$ it follows that

$$
\frac{1}{n} \sum_{k=1}^{n} \mu(A_k) \leq \ell(f) \leq \frac{1}{n} \sum_{k=1}^{n} \mu(A_{k-1})
$$

consequently

$$
|\ell(f) - \int f d\mu| \leq \frac{1}{n} \mu(A_0) \leq \frac{1}{n} \mu(\text{supp}(f)) = O\left(\frac{1}{n}\right)
$$
by sending $n \to \infty$ we obtain the desired claim.

We now show the lemma.

For the second estimate, recall that by definition $\mu$ is outer regular, thus $\mu(F) = \inf\{\mu(U) : F \subset U, U \text{ open}\}$. Now for every $U$ open containing $F$ we have $0 \leq \bar{f} \leq 1_U$ thus by definition of $\mu(U)$ it follows that $\ell(f) \leq \mu(U)$. Consequently $\mu(F) \geq \ell(f)$.

For the first estimate, it suffices to show that for every $\epsilon > 0$ we have $\mu(E) \leq (1 + \epsilon)\ell(f)$. Let $U_\epsilon = \{f > \frac{1}{1+\epsilon}\}$ which is an open set that contains $E$. Thus

$$\mu(E) \leq \mu(U_\epsilon) = \sup\{\ell(g) : 0 \leq g \leq 1, \supp(g) \subset U_\epsilon\}$$

$$\leq \sup\{\ell(g) : 0 \leq g \leq (1 + \epsilon)f\}$$

$$\leq (1 + \epsilon)\ell(f)$$

We now show regularity properties for $\mu$. (It is clear that $\mu$ is finite on compact set using Urysohn’s lemma.) Outer regularity of $\mu$ follows from construction using premeasure, and it remains to show inner regularity. Let $E$ be Borel, it suffices to show that

$$\mu(E) \leq \sup\{\mu(K) : K \subset E \text{ compact}\}$$

now using outer regularity it suffices to do this for $E$ open. Let $U$ be an open set, we have

$$\mu(E) = \sup\{\ell(f) : 0 \leq f \text{ continuous} \leq 1, \sup(f) \subset U\}$$

$$\leq \sup\{\mu(\sup(f)) \ldots\}$$

$$\leq \sup\{\mu(K) : K \subset E \text{ compact}\}$$

**Step 2:** We now reduce the desired result to the positive setting. This is done via writing bounded linear functionals on $C_0(X)$ as a linear combination of positive linear functionals. To see this note that without loss of generality it suffices to consider bounded linear functional on $C_{0,R}(X)$ real valued members of $C_0(X)$. Let $\ell_+$ be defined by

$$\ell_+(f) = \sup\{\ell(g) : 0 \leq g \leq f, g \in C_{0,R}(X)\}$$

and let $\ell_- = \ell_+ - \ell$, it is not hard to check that both $\ell_+$ and $\ell_-$ are positive linear functionals on $C_{0,R}(X)$.

### 6.2 The Stone-Weierstrass approximation theorem

Let $X$ be compact Hausdorff. Let $P$ be a subspace of $C_{R}(X)$ such that $P$ is also an algebra inside $C_{R}(X)$, i.e. $p_1p_2 \in P$ if both $p_1, p_2 \in P$.

**Theorem 28.** Assume that $P$ separates point, i.e. if $x_1 \neq x_2$ are elements of $X$ then there exists one element $p \in P$ such that $p(x_1) \neq p(x_2)$. Then one of the following holds: $\overline{P} = C_{R}(X)$ or there is some $x_0 \in X$ such that $\overline{P} = \{f \in C_{R}(X) : f(x_0) = 0\}$. 
Note that there are also versions for complex valued functions, in which case the same conclusion holds if we assume further that \( P \) is closed under conjugation \( p \in P \) then \( \overline{p} \in P \).

There are also versions for \( C_0(X) \) and \( C_{0,\mathbb{R}}(X) \) where \( X \) is locally compact Hausdorff.

The original Weierstrass approximation theorem is for \( X \) being compact intervals, which implies that polynomials are dense inside continuous functions. Applications in probability (moment problems etc.)

Proof: Since \( P \) is also an algebra that separate points, without loss of generality we may assume that \( P \) is closed. We divide the proof into two steps

Step 1: We will show that \( P \) is a lattice, namely if \( f, g \in P \) then so are \( \max(f, g) \) and \( \min(f, g) \).

Step 2: Using the lattice property of \( P \), we will show that if \( f \in C_\mathbb{R}(X) \) be such that for every \( x \neq y \) in \( X \) there exists \( h \in P \) such that \( h(x) = f(x) \) and \( h(y) = f(y) \), then \( f \in P \).

Combining these two steps, we prove the theorem as follows. Given any \( x \neq y \) consider the algebra \( \langle h(x), h(y) \rangle \), \( h \in P \). Clearly this is a subalgebra subspace of \( \mathbb{R}^2 \) and it can’t be \( \{0, 0\} \) since \( P \) separate points. Thus it could be either \( 0 \times \mathbb{R} \) or \( \mathbb{R} \times 0 \) or \( \mathbb{R}^2 \). Now in the last case by step 2 any \( f \in C_\mathbb{R}(X) \) must be in \( P \) as desired. In the first case for instance, \( h(x) = 0 \) for all \( h \in P \) and so we must have \( \{h(y), h \in P\} = \mathbb{R} \) for any other \( y \). So by Step 2 again it follows that any function \( f \) such that \( f(x) = 0 \) must be in \( P \) as desired.

Proof of step 1: since \( \max(f, g) \) and \( \min(f, g) \) could be written as linear combination of \( |f + g| \) and \( |f - g| \) and \( f \) and \( g \), therefore it suffices to show that if \( f \in P \) then so is \( |f| \). The idea is to approximate \( |f| \) with a polynomial of \( f \) uniformly. Indeed, clearly \( f \) is bounded therefore without loss of generality assume \( |f| \leq 1 \), it then suffices to show that \( |x| \) could be approximated uniformly by polynomials on \([-1, 1]\). To see this use the Taylor expansion of \( \sqrt{1 - t} = 1 - \frac{1}{2}t + \ldots \) which converges uniformly on \( 0 \leq t \leq 1 \), then write

\[
|x| = \sqrt{x^2} = \sqrt{1 - (1 - x^2)} = 1 - \frac{1}{2}(1 - x^2) + \ldots
\]

Proof of step 2: Let \( \epsilon > 0 \). It suffices to show that there exists \( g \in P \) such that \( |g(x) - f(x)| \leq \epsilon \) uniformly over \( x \in X \). Assume that for every \( x \in X \) there exists \( g_x \in P \) such that \( |g_x(x) - f(x)| < \epsilon \) and \( g_x(y) \leq f(y) + \epsilon \) for every \( y \in X \). Then the desired \( g \) could be constructed as follows. By continuity for each \( x \in X \) there exists \( U_x \subset X \) neighborhood of \( x \) such that \( \sup_{U_x} |g_x - f| < \epsilon \). By compactness of \( X \) one could refine the collection \( U_x, x \in X \) to a finite subcollection \( U_{x_1}, \ldots, U_{x_m} \), and we simply let

\[
g = \max(g_{x_1}, \ldots, g_{x_m})
\]

which is in \( P \) by the lattice property and clearly \( \sup_X |g - f| \leq \epsilon \).

Now to show the existence of such a \( g_x \) we fix \( x \) and notice that, using the given hypothesis, for each \( y \in Y \) we may find \( h_y \in P \) and two open neighborhoods of \( x \) and \( y \) respectively, denoted respectively by \( U_y \) and \( V_y \) (note that \( x \) is fixed so we ignore the dependence on \( x \)) such that

\[
\sup_{U_y} |h_y - f| < \epsilon, \quad \sup_{V_y} |h_y - f| < \epsilon
\]
By compactness of $X$ again we may refine the covering $V_y, y \in Y$ to a finite subcovering $V_{y_1}, \ldots, V_{y_n}$, and we simply let $g_x = \min(h_{y_1}, \ldots, h_{y_n})$ which has the desired property with $U_x := U_{y_1} \cap \cdots \cap U_{y_n}$. 
Chapter 7

Locally convex spaces, the hyperplane separation theorem, and the Krein-Milman theorem

Recall that $C(X)$ is not a normed linear space when $X$ is not compact. On the other hand we could use semi norms on $C(X)$: given any compact $K \subset X$ let $\| \cdot \|_K$ be the sup norm on $K$. This family of seminorms determine $C(X)$: together they provides a lot of properties so that various theorems (for NLS) remains valid here.

7.1 Two equivalent definitions of LCS

Let $X$ be a linear space over $\mathbb{R}$ (results over $\mathbb{C}$ are similar). We consider topologies on $X$ such that addition and scalar multiplication are continuous operations (which are assumed in the definition below), in which case we say that $X$ is a topological linear space.

Now there are two equivalent definitions of local convexity for a topological linear space. In one definition, we use seminorms. A seminorm $\rho$ on $X$ is essentially a norm except for the nondegenerate condition, namely it satisfies the triangle inequality $\rho(x + y) \leq \rho(x) + \rho(y)$, nonnegativity $\rho(x) \geq 0$, homogeneity $\rho(\lambda x) = |\lambda| \rho(x)$, however it is possible that $\rho(x) = 0$ when $x \neq 0$. A typical example is $\rho(x) = |\ell(x)|$ where $\ell$ is a linear functional on $X$. Now, a topological linear space $X$ is said to be locally convex if there exists a family of seminorms $\rho_\alpha$, $\alpha \in I$, such that the topology on $X$ is the minimal topology such that (addition and multiplication are continuous and) $\rho_\alpha$ are continuous.

Another definition uses convex sets. A subset $A$ of $X$ is called balanced if $\lambda x \in A$ whenever $x \in A$ and $|\lambda| \leq 1$. We also say that $A$ is absorbent if for every $x \in X$ there is some $t > 0$ such that $tA$ contains $x$. Then a topological linear space $X$ is said to be locally convex if there exists a neighborhood base at 0 consisting of only convex balanced absorbent (open) sets.

To see the equivalence of two definitions, we start with the seminorm definition. Then a
neighborhood base at 0 could be taken to consist of all open sets of the form

\[ U = \{ x \in X : \rho_{\alpha_1}(x) < \epsilon, \ldots, \rho_{\alpha_m}(x) < \epsilon \} \]

where \( \epsilon > 0 \) and \( \alpha_1, \ldots, \alpha_m \) are elements of \( I \). \( (m \geq 1 \) is arbitrary). It is clear that these open sets are convex and balanced and absorbent.

Conversely given a neighborhood base consisting of convex balanced absorbent sets we could use the Minkowski gauge functional to construct the seminorm. Namely if \( A \) is convex balance absorbent we let

\[ \rho_A(x) = \inf \{ t > 0 : x \in tA \} \]

(\( \rho_A \) is not hard to check that \( \rho_A \) is seminorm.)

### 7.1.1 Basic properties

A family of seminorm is said to be separated if whenever \( \rho_\alpha(x) = 0 \) for every \( \alpha \in I \) it must follow that \( x = 0 \). This is actually equivalent to the Hausdorffness of the topology.

If \( (x_i)_{i \in D} \) is a net in \( X \) then \( (x_i) \to x \) if and only if \( \rho_\alpha(x_i - x) \to 0 \) (as a net in \( \mathbb{R} \)) for all \( \alpha \in A \).

If there is a lot of seminorms one expects that the topology is very rich; on the other hand if there are fewer seminorms one expects a more well-structured topology. In particular if there are only countably many seminorms involved then the topology is equivalent to the topology of some pseudo-metric, i.e., a distance notion that resembles the metric notion except for the fact that two distinct points could have distance 0. To see this, enumerate the seminorms \( \rho_1, \ldots, \), and let

\[ d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_n(x - y)}{1 + \rho_n(x - y)} \]

If one assume that the (countable) family is separated then the above pseudo metric is actually a metric. If this metric is furthermore complete then the given space is called a Frechet space.

**Examples:** Recall that \( C(X) \) is a locally convex space with seminorms given by \( \rho_K(f) = \sup_{x \in K} |f(x)| \) where \( K \subset X \) is compact. Other examples are

(i) the space of Schwartz functions on \( \mathbb{R}^n \): These are functions that are \( C^\infty \) and their derivatives decay faster than any polynomial. One could use \( \rho_{\alpha, \beta}(f) = |x|^\alpha |D^\beta f(x)| \) where \( \alpha \) and \( \beta \) are nonnegative integer multi-indices.

(ii) if \( \ell_\alpha, \alpha_A \) is a family of linear functionals on \( X \) then the minimal topology such that \( \ell_\alpha \) are continuous and linear operations (addition/scalar multiplication) are continuous is called the weak topology induced by this set of linear functionals. Examples of these (for normed linear spaces) are the weak topology and the weak* topology. It can be shown that if \( \ell \) is a linear functional that is continuous with respect to this topology it must be a finite linear combination of the given linear functionals.
(iii) Given any open set $\Omega \subset \mathbb{R}^n$ the space $D(\Omega)$ consisting of compactly supported infinitely smooth functions on $U$ is also a locally convex space and actually complete (recall that the version without infinite smoothness, i.e. $C_c(U)$, is not complete).

### 7.1.2 Linear maps

Let $X$ and $Y$ be locally convex spaces with seminorms $\rho_\alpha$ and $\rho'_\beta$ where $\alpha \in I$ and $\beta \in J$ index sets. Then it can be shown that a linear map $\ell : X \to Y$ is continuous if and only if given any $\beta \in J$ there exists $\alpha_1, \ldots, \alpha_m \in I$ and $C > 0$ such that the following holds for all $x \in X$:

$$\rho'_\beta(\ell(x)) \leq M(\rho_{\alpha_1}(x) + \cdots + \rho_{\alpha_m}(x))$$

(Note that this generalizes the usual equivalence of continuity and boundedness of linear maps between normed linear spaces.)

### 7.2 The Hyperplane separation theorem

Let $K$ be a convex subset of a topological linear space $X$ over $\mathbb{R}$ and let $y \in X \setminus K$. Assume that $K$ contains at least one interior point.

**Theorem 29.** There exists a continuous linear functional $\ell$ on $X$ such that $\sup_{x \in K} \ell(x) \leq \ell(y)$. If furthermore $K$ is closed then this inequality could be taken strict.

**Proof.** By translation invariant we may assume that 0 is an interior point of $K$. Consider the Minkowski gauge functional

$$\rho_K(x) = \inf\{t > 0 : x \in tK\}$$

since 0 is an interior point of $K$ it is clear that $\rho_K(x) < \infty$ for any $x \in X$. Furthermore $\rho_K$ is convex and positive homogeneous, and since $y \not\in K$ we have $\rho_K(y) \geq 1 \geq \rho_K(x)$ for every $x \in K$. Let $\ell$ be defined on the one dimensional subspace spanned by $y$ using $\ell(y) = 1$. Then $|\ell(z)| \leq \rho_K(z)$ inside this subspace, so by Hahn Banach we may extend $\ell$ to all of $X$ such that $|\ell(x)| \leq \rho_K(x)$ for all $x$. In particular $\ell(x)$ is continuous since $\rho_K$ is continuous at 0.

Now, if $K$ is closed then one could see that $\rho_K(y) > 1 \geq \sup_{x \in K} \rho_K(x)$ and we could therefore obtain a strict inequality. \(\square\)

For a locally convex Hausdorff space over $\mathbb{R}$, it follows from the above theorem that for any two distinct points $x \neq y$ there is a continuous linear functional $\ell$ on $X$ such that $\ell(x) \neq \ell(y)$. To see this, by Hausdorffness and local convexity we could find a convex open set $A$ such that $x \in A$ while $y \in X \setminus A$. Since $A$ is closed convex we could apply the hyperplane separation theorem and get a continuous linear function $\ell$ such that $\sup_{z \in \overline{A}} \ell(z) < \ell(y)$. In particular $\ell(x) < \ell(y)$. 


7.3 The Krein-Milman theorem

An extreme point of a convex set $K$ is a point $x$ such that if $x$ is a convex combination of $x_1, x_2 \in K$ then $x_1 = x_2 = x$.

Theorem 30 (Krein-Milman). Let $X$ be locally convex Hausdorff and $K$ is a nonempty compact subset of $X$. Then

(i) $K$ has at least one extreme point.

(ii) $K$ is the closure of the convex hull of its extreme points.

Remark: If $X = \mathbb{R}^n$ for some finite $n$ then we don’t need to take the closure in (ii), in fact Caratheodory showed that one could get any point of $X$ from a convex combination of at most $n + 1$ extreme points. For infinite dimensional space the closure is essential.

Proof:

We first generalize the notion of extreme points to extreme subsets of $K$. A subset $A$ of $K$ is said to be extreme if it is nonempty convex and furthermore if any $x \in A$ is a convex combination of two points $x_1$ and $x_2$ in $K$ then we must have $x_1, x_2 \in A$. It is clear that if a family of extreme subsets of $K$ has nonempty intersection then this intersection is also extreme. In particular $K$ is an extreme subset of itself.

Now, consider the collection $\mathcal{A}$ of all closed extreme subsets of $K$, which is nonempty since it contains $K$, and we may order this collection partially using set inclusion, namely $A \leq B$ if $A \supseteq B$.

(i) We first show that there exists a maximal element in $\mathcal{A}$, which we will show to be a point later. In order to show existence of the maximal element we plan to use Zorn’s lemma, and what is needed here is the fact that any chain (i.e. a totally ordered subcollection of $\mathcal{A}$) has an upper bound. The idea is to take the intersection of the elements in this chain, and what needs to be shown is the fact that this intersection is not empty. Assume towards a contradiction that this intersection is empty, it follows from compactness of $K$ that there exists a finite subcollection that has empty intersection, but this is a contradiction because the intersection of a finite chain is simply the smallest element, which is nonempty.

Now, let $A$ be a close extreme subset of $K$ and assume towards a contradiction that it is not a point. Say $x \neq y$ are two elements of $A$, then there is a continuous linear functional on $X$ that separates $x$ and $y$. We’ll show that the subset $B$ of $A$ where $\ell$ achieves maximum is an extreme subset of $K$, which would violate maximality of $A$. (Clearly $B \neq A$). Since $A$ is extreme in $K$ it suffices to show that $B$ is extreme in $A$. Now extremality of $B$ in $A$ follows easily from linearity of $\ell$.

(ii) We now show that the closure of the convex hull $E$ of extremal points of $K$ is $K$, i.e. $K = \overline{E}$. Suppose that $z \in K \setminus \overline{E}$. Then by the hyperplane separation theorem there is a continuous linear functional $\ell$ such that

$$\sup_{x \in \overline{E}} \ell(x) < \ell(z)$$

Again the set of maximum of $\ell$ on $K$ is a proper subset of $K$ and also an extreme subset, and
this set is also closed and disjoint from $\overline{E}$. So by repeating the above argument one could find one extremal point of $K$ inside this set, which is therefore not inside $E$, a contradiction.

Examples: Let $X$ be a compact Hausdorff space and consider $C_\mathbb{R}(X)$ (we could do locally compact Hausdorff too with $C_{\mathbb{R},0}(X)$) and let $\mathcal{A}$ consists of all positive linear functional on $C(X)$. Let $A$ be the subset of $\mathcal{A}$ with $\ell(1) = 1$, it is not hard to see that $A$ is convex and its extreme points are the point evaluation linear functional $e_x f = f(x)$.

7.4 Inductive limit and weak solutions

Let $\Omega$ be a domain inside $\mathbb{R}^n$. As mentioned before the space $D(\Omega)$ consists of $C^\infty$ functions whose supports are compact subsets of $\Omega$. One way to construct a locally convex topology on this space is to use inductive limit of topologies. Let $X_1, X_2, \ldots, X_n, \ldots,$ be linear spaces such that $X_1 \subset X_2 \subset \cdots \subset X = \bigcup X_n$. Each $X_n$ has a locally convex topology that is consistent with the topologies on other $X_m$ in following sense: the topology of $X_n$ is the induced topology from $X_{n+1}$. If that is the case one could construct a limiting topology on $X$ which should be thought of as $\lim_{n \to \infty} X_n$.

In our context we let $K_1 \subset K_2 \ldots$ be compact subsets of $\Omega$ such that $\bigcup K_j = \Omega$. Then let $X_n$ be the space of $C^\infty$ functions on $\mathbb{R}^n$ whose support are subsets of $K_n$. Note that $X_n$ is a complete metrizable locally convex space. We then define the topology on $D(\Omega)$ to be the inductive limit of the topologies of $X_n$.

The dual space of $D(\Omega)$ is called the space of (tempered) distribution on $\Omega$ denoted by $D'(\Omega)$. This space contains $D(\Omega)$ as a dense subspace. We may define the action of a linear differential operator $D^j$ (here $j$ is a multiindex and the sum is over some finite collection) on any $\ell \in D'$ by defining for each $\phi \in D(\Omega)$

$$D^j \ell(\phi) = (-1)^j \ell(D^j \phi)$$

A weak solution to a PDE is a distributional solution in the above sense. If this distribution arise from some sufficiently smooth functions then we say it is a classical solution.
Part II

Spectral analysis for linear operators
Chapter 8

Elementary spectral theory

Let $B$ denote the space of all bounded linear operators on some given Banach space $X$ over $\mathbb{C}$. The analysis here works for more general settings, say $B$ could be a unital Banach subalgebra of this space (i.e. an associative subalgebra that contains a unit and is a complete subspace with respect to the operator norm).

8.1 Spectrum and resolvent set

We say $M \in B$ is invertible if its inverse exists and is in $B$. Observe that if $M$ is invertible then so is all $N$ with $\|N - M\| < \frac{1}{\|M^{-1}\|}$. To see this by writing $N = M + (N - M) = M(I + M^{-1}(N - M))$ it suffices to consider the case $M = I$. In this case (i.e. $M = I$) use geometric series to show that the inverse is exactly

$$I + (N - I) + (N - I)^2 1 + \ldots$$

**Definition:** The resolvent set of $M$ in $B$ consists of all complex number $\lambda$ such that $\lambda I - M$ is invertible in $B$. The resolvent set is denoted by $\rho(M)$ and the spectrum of $M$ is $\sigma(M) := \mathbb{C} \setminus \rho(M)$.

Properties:
1. $\rho(M)$ is open in $\mathbb{C}$. *(This is a consequence of the above observation.)*
2. The resolvent function $(\lambda I - M)^{-1}$ is an analytic function of $\lambda$ on $\rho(M)$. *(This is because the above observation also shows that the resolvent could be expanded as a power series around each point $\lambda$ in the resolvent set $\rho(M)$)*

$$\frac{1}{\lambda - h - M} = \sum_{j=0}^{\infty} (\lambda - M)^{n-1} h^n$$

*which has positive radius of convergence $|h| < \|(\lambda - M)^{-1}\|$.*

3. $\|(\lambda - M)^{-2}\| \geq 1/\text{dist}(\lambda, \sigma(M))$. 
CHAPTER 8. ELEMENTARY SPECTRAL THEORY

This is a corollary of the above analysis.

4. (Gelfand) The spectrum \( \sigma(M) \) is a bounded nonempty closed subset of \( \mathbb{C} \). Furthermore, the spectral radius of \( M \), defined by \( |\sigma(M)| = \max |\lambda| \) over \( \lambda \in \sigma(M) \) satisfies

\[
\sigma(M) = \lim_{k \to \infty} \|M^k\|^{1/k}
\]

Closedness is clear, boundedness follows from the fact that if \( \lambda \) is large enough in modulus then \( \lambda^{-1}M \) has small norm and so \( (1 - \lambda^{-1}M) \) is invertible and so \( \lambda \in \rho(M) \). To show that \( \sigma(M) \) is nonempty assume towards a contradiction that it is. Then consider the contour integral along \( C_R = \{ |\xi| = R \} \)

\[
\int_{C_R} (\lambda - M)^{-1} d\lambda
\]

clearly analyticity of the resolvent operator implies that this would be 0. On the other hand we know that \((\lambda - M)^{-1}\) has the Laurent series expansion at \( \infty \)

\[
(\lambda - M)^{-1} = \sum_{n=0}^{\infty} M^n \xi^{-(n+1)}
\]

therefore the contour integral will gives the residue term i.e. the term with \( n = 0 \) and we get \( I \), contradiction.

To show the identity for the spectral radius, by elementary complex analysis it follows that \( |\sigma(M)| \geq \limsup \|M^k\|^{1/k} \). So it suffices to show that \( |\sigma(M)| \leq |M^k|^{1/k} \) for all \( k \geq 1 \). To see this, observe that

\[
\| \sum_{n \geq 0} M^n \xi^{-(n+1)} \| \leq \left[ \sum_{j=0}^{k-1} \|M^j\| |\xi|^{-(j+1)} \right] \left[ \sum_{m \geq 0} (\|M^k\||\xi|^{-k})^m \right]
\]

this follows from \( \|M^{mk+j}\| \leq \|M^j\||M^k||^m \). It follows that if \( |\xi| > \|M^k\| \) then the series converges absolutely and therefore \( \xi \in \rho(M) \); this completes the proof that \( |\sigma(M)| = \lim \|M^k\|^{1/k} \).

8.2 Functional calculus and spectral mapping

We clearly could define a polynomial of a given element \( M \in B \); in fact via power series this works for analytic functions provided that the spectral norm of \( M \) is smaller than the radius of convergence at 0. In particular it works if the given function is analytic in some domain \( \Omega \) that contains \( \sigma(M) \). We demonstrate that this later fact would be enough. Indeed, let \( C \) be a contour in the intersection of \( \rho(M) \) and the domain of analyticity of \( f \) that winds once around every point in \( \sigma(M) \) but zero time around \( \Omega^c \). Then define

\[
f(M) = \int_C (\xi - M)^{-1} f(\xi) d\xi
\]
note that by the Cauchy theorem this is independent of the choice of the contour. It turns out that this is consistent with the usual polynomial case and furthermore
\[ \sigma(f(M)) = f(\sigma(M)) \]
(also known as the spectral mapping property). We also have the resolvent identity
\[
(\xi_1 - \xi_2)[(\xi_2 - M)^{-1} - (\xi_1 - M)^{-1}] = (\xi_1 - M)^{-1}(\xi_2 - M)^{-1}
\]
which is useful to show that the functional calculus maps the algebra of analytic functions on open sets containing \( \sigma(M) \) into \( B \) is a homomorphism.

Also, if \( g \) is analytic on an open set containing \( f(\sigma(M)) \) and \( h = g \circ f \) then \( h(M) = g(f(M)) \).

**Spectral Projections:** Assume the spectrum \( \sigma(M) \) could be decomposed into an union of \( n \) disjoint closed components \( \sigma_1 \cup \ldots \cup \sigma_n \).

Let \( C_j \) be a contour in \( \rho(M) \) that winds once around each point of \( \sigma_j \) but not other components, and \( P_j = \int_{C_j} (\xi - M)^{-1}d\xi \).

Then \( P_j \) are disjoint projections \( P_j^2 = P_j \) and \( P_jP_k = 0 \) if \( j \neq k \), and
\[
\sum_j P_j = I
\]
and if \( \sigma_j \neq \emptyset \) then \( P_j \neq 0 \).

### 8.3 Examples of operators and their spectra

1. **Shifts**
   Let \( X = \ell^2 \) consiting of \( x = (a_0, a_1, \ldots) \) such that \( \sum |a_j|^2 < \infty \). The right and left shifts \( R \) and \( L \)
   \[
   Rx = (0, a_0, a_1, \ldots), \quad Lx = (a_1, a_2, \ldots)
   \]
Now \( LR = I \) but \( RL \neq I \) so these linear bounded maps are not invertible. The spectrum of \( R \) and \( L \) consists of all points in the unit disk \( \{ \lambda \in \mathbb{C} : |\lambda| \leq 1 \} \). Now \( R' = L \) and \( L' = R \) as adjoint of each other.

The idea is the spectral radius of \( L \) is 1 so spectrum contained inside the disk.

2. **The Fourier transform** \( Ff(\xi) = \int f(x)e^{-2\pi i \xi}dx \). Note that \( F^4 = I \) therefore the spectrum of \( F \) is a subset of \( \{ x \in \mathbb{C} : x^4 = 1 \} \) which consists of \( \pm 1, \pm \sqrt{-1} \). Let \( H_n \) denotes the orthogonal polynomials wrt to \( e^{-x^2}dx \)
\[
\int H_mH_ne^{-x^2}dx = \sqrt{\pi}2^n n! \delta_{mn}
\]
then it could be shown that $F[e^{-x^2/2}H_n(x)] = (-\sqrt{-1})^n e^{-\xi^2/2}H_n(\xi)$. This follows from
\[
\exp(-x^2/2 + 2xt - t^2) = \sum_{n\geq 0} e^{-x^2/2}H_n(x)t^n/n!
\]

3. **Volterra integral operator** Let $X = C([0,1])$ and let $Tf(x) = \int_0^x K(x,t)f(t)dt$, also known as the Volterra integral operator, here $K$ is continuous. The spectrum of $T$ consists of only one point $\lambda = 0$. This follows by computing $T^nf$, for instance if $K \equiv 1$ then $T^n f(x) = 1/(n-1)! \int_0^x (x-t)^{n-1} f(t)dt$ and show that $\|T^n f\| \leq \|f\|/n!$ in the sup norm, and use the spectral radius identity.

4. **Diagonal multiplication** Let $X = \ell^p(\mathbb{Z}_+)$ and $M$ acts diagonally $(Mx)_n = \lambda_n x_n$, then the spectrum of $M$ is the closure of $\{\lambda_n\}$.

8.4 **Adjoint operators and spectrum**

For convenience of notation let $\langle x, \ell \rangle := \ell(x)$ for every $\ell \in X^*$ and $x \in X$.

Let $T : X \to X$ be a bounded operator on a Banach space $X$. Then its adjoint operator $T^*$ is defined to be the bounded operator from $X^*$ to $X^*$ such that for every $x \in X$ and $\ell \in X^*$ it holds that
\[
\langle x, T^* \ell \rangle = \langle Tx, \ell \rangle
\]

It is not hard to see that the adjoint operator exists (and is linear and bounded), and $\|T^*\| \leq \|T\|$.

Note that this definition differs the adjoint notion of an operator on a Hilbert space by a conjugate. Thus some texts use $T'$ for the Banach space conjugate of $T$.

**Theorem 31** (Phillips). *If $T : X \to X$ bounded linear on a complex Banach space $X$ then $\sigma(T) = \sigma(T^*)$ and for every $\lambda \in \rho(T)$ it holds that $([T - \lambda]^{-1})^* = (T^* - \lambda)^{-1}$.*

As a corollary, for Hilbert space we have $\sigma(T) = \sigma(T^*)$.

We will prove the first part, the second part follows as a by product.

**Step 1:** $\rho(T) \subset \rho(T^*)$. Take $\lambda \in \rho(T)$. It suffices to show the following lemma (then apply it to $T - \lambda$)

**Lemma 5.** Let $T : X \to X$ be bounded linear and boundedly invertible. Then $(T^*)^{-1}$ exists and $(T^*)^{-1} = (T^{-1})^*$.

**Proof.** It suffices to show, using the open mapping theorem, that $T^* : X^* \to X^*$ bounded and injective and onto.

**Injectivity:** If $T^* \ell = 0$ for some $\ell \in X^*$ then we will show that $\ell = 0$.
\[
\langle Tx, \ell \rangle = \langle x, T^* \ell \rangle = 0
\]
for every $x$, so using the fact that $T$ is onto we obtain $\ell(X) = 0$ so $\ell = 0$. 
Onto: If $\ell \in X^*$ then we want some $\ell_0 \in X^*$ such that $T^*\ell_0 = \ell$. We equivalent transform this equation into

$$\langle x, T^*\ell_0 \rangle = \langle x, \ell \rangle \quad \forall x \in X$$

or equivalently $\langle Tx, \ell_0 \rangle = \langle x, \ell \rangle$ for all $x \in X$. Since $T$ is invertible this is the same as $\langle y, \ell_0 \rangle = (T^{-1}y, \ell)$ for every $y \in X$, which is the same as $\ell_0 = (T^{-1})^*\ell$. □

**Step 2:** Here we show that $\rho(T^*) \subset \rho(T)$. We will use the following Lemma

**Lemma 6.** Let $T : X \to X$ be bounded linear on a complex Banach space. Then

(i) If range($T^*$) is dense in $X^*$ (in the weak* topology) then $T$ is injective.

(ii) If $T^*$ is injective then range($T$) is dense.

We first use the lemma to complete this step. If $\lambda \in \rho(T^*)$ we will show $\lambda \in \rho(T)$. By the lemma it follows that $T - \lambda$ is injective and has dense range. It remains to show that $(T - \lambda)^{-1}$ is bounded on its range (then we could invoke the BLT theorem). To see this for every $y \in \text{range}(T - \lambda)$, i.e. $y = (T - \lambda)z$ for $z \in X$, and $\ell \in X^*$ we have

$$\langle y, \ell \rangle = \langle z, (T - \lambda)^*\ell \rangle$$

and so

$$\langle z, \ell \rangle = \langle y, [T^* - -\lambda]^{-1}\ell \rangle$$

$$|\langle z, \ell \rangle| \leq ||y|| ||\ell|| [T^* - -\lambda]^{-1}||$$

since this is true for all $\ell$ we obtain $\|z\| \leq ||[T^* - -\lambda]^{-1}|| ||y||$, as desired.

It remains to show the lemma. For (i) note that if $Tx = 0$ then $(Tx, \ell) = 0$ for all $\ell \in X^*$ and therefore $\langle x, T^*\ell \rangle = 0$, but the denseness of the range of $T^*$ implies that $\langle x, \ell \rangle = 0$ for all $\ell \in X^*$, and consequently $x = 0$. For (ii), assume that range($T$) is a strict subspace of $X$, then we could find, using Hahn Banach, a nonzero bounded linear functional such that $\ell(Tx) = 0$ for all $x \in X$. Then $\langle x, T^*\ell \rangle = 0$ for all $x$ therefore $T^*\ell = 0$ but injectivity of $T^*$ implies that $\ell = 0$ contradiction.
Chapter 9

Spectral theory for compact operators on Banach spaces

Recall that a subset $S$ of a metric space $X$ is precompact if its closure is compact, or equivalently every sequence contains a Cauchy subsequence. Another characterization is that $S$ is totally bounded, namely for any $\epsilon > 0$ one could cover $S$ by finitely many $\epsilon$-balls.

If $X$ is a normed linear space we can add/multiply, and we have the following basic properties:

(i) If $S$ is precompact then so is $\alpha S$ for any $\alpha$ scalar.
(ii) If $S_1$ and $S_2$ are precompact then $S_1 + S_2 := \{s_1 + s_2, s_1 \in S_1, s_2 \in S_2\}$ is precompact.
(iii) If $S$ is precompact then so is the convex hull of $S$.
(iv) Let $T : X \to Y$ where $X$ and $Y$ are say normed linear spaces. If $S \subset X$ is precompact then so is $TS \subset Y$.

9.1 Compact operators

Let $T : X \to Y$ a bounded linear map between Banach spaces $X$ and $Y$. Let $B_1$ be the unit ball in $X$. We say that $T$ is pre-compact if $TB_1$ is a precompact subset of $Y$.

Properties:

(i) A finite rank operator (namely the range of the operator is finite dimensional) is compact.
(ii) If $T_1 : X \to Y$ and $T_1 : X \to Y$ are compact operators then $\alpha_1 T_1 + \alpha_2 T_2$ is also compact for any scalar $\alpha_1, \alpha_2$.
(iii) If $T : X \to Y$ is compact and $M : U \to X$ and $N : Y \to V$ are bounded linear maps between Banach spaces then $NTM : U \to V$ is compact.
(iv) If $T : X \to Y$ is compact then it maps a weakly convergent sequence in $X$ to a strongly convergent sequence in $Y$. Such operator is called completely continuous, so we could say that compactness implies complete continuity.\[1\]

[1]If $T : X \to X$ is completely continuous and $X$ is reflexive then $T$ is compact, this is left as an exercise.
To see property (iv), first note that \( Tx_n \) converges weakly to \( Tx \). To see this, let \( \ell \in X^* \), then \( \ell(Tx_n) = (T^*\ell)x_n \) converges to \( T^*\ell x \) which is the same as \( \ell(Tx) \). It then follows that any strongly convergent subsequence of \( Tx_n \) has to converge to \( Tx \). Now since \( x_n \) converges weakly to \( x \) in \( X \) it follows that \( x_n \) is a family of bounded operator on \( X^* \) that is uniformly bounded pointwise, consequently by the principle of uniform boundedness \( \sup_n \|x_n\| < \infty \), thus \( (Tx_n) \) is precompact, so any subsequent contains a Cauchy subsequence; thus by a routine argument it follows that \( Tx_n \) is convergent to \( Tx \).

(v) If \( T_n \) is a sequence of compact operators from \( X \) to \( Y \) and \( T_n - T \| \to 0 \) as \( n \to \infty \) for some \( T : X \to Y \) bounded linear, then \( T \) is compact.

To see property (v), note that for any \( \epsilon > 0 \) one could choose \( N \) large such that \( \|T_N - T\| < \epsilon/2 \). Now, we could use finitely many \( \epsilon/2 \) balls to cover \( T_N B_1 \). The \( \epsilon \) balls with the same centers will then cover \( TB_1 \). (Recall that \( B_1 \) is the unit ball in \( X \).)

As a corollary, we have

Corollary 3. If \( T : X \to Y \) is the limit of a sequence of finite rank operators in the norm topology then \( T \) is compact.

The converse direction of this corollary is not true in general (for Banach spaces \( X \) and \( Y \)), the first construction of counter examples is due to P. Enflo ('73). However, if \( Y \) is a Hilbert space then the converse is true.

Theorem 32. Any compact operator \( T : X \to Y \) where \( X \) is Banach and \( Y \) is Hilbert can be approximated by a sequence of finite rank operators.

Proof. We sketch the main ideas of the proof. For every \( n \) the set \( TB_1 \) is covered by \( \bigcup_{k=1}^{M_n} B_Y(y_k, 1/n) \). We may assume that \( M_n \) is an increasing sequence. Let \( P_n \) be the projection onto the span of \( y_1, \ldots, y_{M_n} \), which is clearly a finite rank operator, thus \( P_n T \) is also finite rank. It remains to show that \( \|T - P_n T\| = O(1/n) \). We observe that \( \|P_n y - y_k\| \leq \|y - y_k\| \) for every \( 1 \leq k \leq M_n \) and every \( y \in Y \), since projection are contractions. It follows that

\[
\|P_nTy - Ty\| \leq \|P_nTy - y_k\| + \|Ty - y_k\| \leq 2\|Ty - y_k\|
\]

and clearly given any \( y \in B_1 \) there is a \( k \) such that \( \|Ty - y_k\| \leq 1/n \), therefore \( \|P_nTy - Ty\| \leq 2/n \) for every \( y \in B_1 \), thus \( \|P_nT - T\| \leq 2/n \) as desired. \( \square \)

Theorem 33 (Schauder). Let \( T : X \to Y \) be a bounded linear operator between Banach spaces \( X \) and \( Y \). Then \( T \) is compact if and only if \( T^* : Y^* \to X^* \) is compact.

Proof. It suffices to show the forward direction, namely if \( T \) is compact then \( T^* \) is also compact. For the other direction, apply the forward direction it follows that \( T^{**} \) is compact from \( Y^{**} \) to \( X^{**} \), and by restricting \( T^{**} \) to \( X \) we obtain \( T \) therefore \( T \) is also compact.

Now, assume that \( T \) is compact. Given any sequence \( \ell_n \in X^* \) with \( \|\ell_n\| \leq 1 \) we will show that \( (T^*\ell_n) \) has a Cauchy subsequence \( T^*\ell_{n_j} \), in other words given any \( \epsilon > 0 \) we have

\[
\sup_{\|x\| \leq 1} \|\ell_{n_j}(Tx) - \ell_{n_k}(Tx)\| \leq \epsilon
\]
if $j$ and $k$ are large enough.

Let $B$ be the unit ball in $X$ and let $K = \overline{TB}$ which is a compact subset of $X$. Then the above estimate is a consequence of $\sup_{y \in K} \| \ell_n y - \ell_k y \| \leq \epsilon$. We may view $\ell_n$ as a sequence of continuous functions on $K$, which are uniformly bounded pointwise and equicontinuous on $K$:

$$\sup_n \| \ell_n y \| \leq \| y \|$$

$$\sup_n \| \ell_n y_1 - \ell_n y_2 \| \leq \| y_1 - y_2 \|$$

Thus by the Arzela Ascoli theorem the sequence $\ell_n$ has an uniformly convergent subsequence, as desired. □

### 9.2 Compactness of integral operators

We now discuss compactness of the integral operator $T$

$$Tf(y) = \int_U K(x, y) f(x) d\mu(x), \quad y \in V$$

where $U$ and $V$ are say compact metric spaces, viewing $T$ as operator on different function spaces.

Viewing $T$ as a map from $L^2(U, d\mu)$ to $L^2(V, d\nu)$ for some measure $\nu$ (note that both spaces are separable Hilbert spaces), we know that one sufficient condition that guarantee boundedness of $T$ is $\int_{U \times V} |K|^2 d\mu d\nu < \infty$. It turns out that this would also imply compactness of $T$. To see this, for each $x$ we expand $K(x, y)$ into the (countable) orthogonal basis of $L^2(Y, d\nu)$, which we may denote by $\phi_1, \phi_2, \ldots$

$$K(x, y) = \sum_{j=1}^{\infty} K_j(x) \phi_j(y)$$

note that for almost every $x \in X$ the function $K(x, y)$ is $L^2(Y, d\nu)$ integrable in $y$ and so

$$\int \int |K(x, y)|^2 d\mu(x) d\nu(y) = \sum_j \int_X |K_j(x)|^2 d\mu(x)$$

thus we may approximate $T$ with the finite rank operator

$$T_n f(y) = \sum_{j \leq n} \int K_j(x) f(x) d\mu(x) u_j(y)$$

so $T$ is compact.
9.3 Spectral properties of compact operators

9.3.1 Riesz’s theorem

One of the main results about compact operators is the following fact: if $T : X \to X$ is a compact operator on a Banach space $X$ and $1 - T$ is injective, then $1 - T$ is boundedly invertible. Note that it is possible for $\|T\|$ to be large, so the basic theory about small perturbation of 1 does not applies here. The key to showing this fact is the following theorem of Riesz.

**Theorem 34 (Riesz).** Let $T : X \to X$ be a compact operator on a Banach space $X$. Then range of $1 - T$ is a closed subspace of $X$ and furthermore $\dim(\ker(1 - T))$ and $\dim(1 - T)$ are finite and equal to each other.

Part of the proof of this theorem is the following lemma.

**Lemma 7.** Let $T : X \to X$ be a compact operator on a Banach space $X$. Then

(i) $\ker(1 - T)$ is finite dimensional.

(ii) There exists some $k \geq 1$ such that $\ker((1 - T)^m) = \ker((1 - T)^{m + 1})$ for every $m \geq k$.

(iii) range$(1 - T)$ is a closed subspace of $X$.

**Proof of Lemma 7** (i) If $y \in \ker(1 - T)$ then $Ty = y$. Assume towards a contradiction that $\ker(1 - T)$ is infinite dimensional. Then by another result of Riesz we could find an infinite sequence $(y_n)$ in this kernel such that $\|y_n\| = 1$ and $\|y_n - y_m\| \geq 1/2$ for all $m \neq n$. This implies the set $\{Ty_n, n \geq 1\}$ is not precompact, contradiction.

(ii) Note that $\ker([1 - T]^k) \subset \ker([1 - T]^{k+1})$ for all $k$. Now, if $\ker([1 - T]^k) = \ker([1 - T]^{k+1})$ then $\ker([1 - T]^m) = \ker([1 - T]^{m+1})$ for all $m \geq k$. Assume towards a contradiction that we have the strict inclusion $\ker([1 - T]^k) \subset \ker([1 - T]^{k+1})$ for all $k \geq 1$. Since $\ker([1 - T]^k)$ are closed, by Riesz lemma we may find $x_n \in \ker([1 - T]^n)$ such that $\|x_n\| = 1$ and $\text{dist}(x_n, \ker([1 - T]^{n-1})) \geq 1/2$. We will show that $\|Tx_n - Tx_m\| \geq 1/2$ for all $m \neq n$, which will contradict compactness of $T$. Now, without loss of generality assume that $m < n$, then

$$Tx_n - Tx_m = x_n - (1 - T)x_n - Tx_m$$

and $(1 - T)x_n \in \ker([1 - T]^{n-1})$ and $Tx_m \in \ker([1 - T]^{n-1})$ (since $T$ commutes with $(1 - T)^{n-1}$). Therefore by choice of $x_n$ we obtain

$$\|Tx_n - Tx_m\| \geq \text{dist}(x_n, \ker([1 - T]^{n-1})) \geq 1/2$$

(iii) Assume that $y_k = (1 - T)x_k$ is a sequence in range$(1 - T)$ that converges to some $y \in Y$, we will show that for some $x \in X$ we have $y = Tx$. Certainly if $x_k$ has a convergent subsequence we could simply take $x$ to be the corresponding limit. Now, $x_k = y_k + Tx_k$ so it suffices to obtain convergence of some subsequence of $Tx_k$, and using compactness of $T$ this would follow if $x_k$ is uniformly bounded. Unfortunately it is possible for $x_k$ to be unbounded, but we could correct this by modifying $x_k$ by an appropriate term in $\ker(1 - T)$ to make it...
bounded. Note that this correction does not change $y_k$ and hence does not change the goal of this part. Let

$$d_k = \text{dist}(x_k, \text{ker}(1 - T))$$

By modifying $x_k$ by an amount inside $\text{ker}(1 - T)$ we may assume that $d_k \leq \|x_k\| \leq 2d_k$. Thus it suffices to show that

$$\sup_k d_k < \infty$$

Assume towards a contradiction that some subsequence of $d_k$ converges to $\infty$. Without loss of generality we may assume $\lim d_k = \infty$. We then have

$$\frac{y_k}{d_k} = (1 - T)\left(\frac{x_k}{d_k}\right)$$

now $y_k/d_k \to 0$ and $x_k/d_k$ is uniformly bounded, so using compactness of $T$ it follows that some subsequence of $T(x_k/d_k)$ converges, which in turn implies that some subsequence of $x_k/d_k$ converges to some $x \in X$. We obtain $0 = (1 - T)x$ so $x \in \text{ker}(1 - T)$, on the other hand it is clear that $\text{dist}(x_k/d_k, \text{ker}(1 - T)) \geq 1$, contradiction. \(\square\)

We now prove Riesz's theorem using the lemma. It remains to show that $\text{ker}(1 - T)$ has the same dimension as the codimension of $1 - T$. We first reduce the proof to the case when $\text{ker}(1 - T)$ is trivial. Let $k$ be the index given by part (ii) of the lemma and let $Y = \text{ker}[(1 - T)^{k+1}]$ a closed subspace of $X$. Since $(1 - T)Y \subset \text{ker}[(1 - T)^k] = Y$ by choice of $k$, it follows that $TY \subset Y$, thus $T$ induces an operator $T_Z$ on $Z := X/Y$, which is a Banach space with the induced norm. It follows immediately that $T_Z$ is compact on $Z$, and $1 - T_Z$ is also injective on $Z$. Now, the dimension of $\text{ker}(1 - T)$ is the same as the dimension of the kernel of $1 - T$ viewing as an operator on $Y$, which by standard linear algebra is the same as the codimension of the range of $1 - T$ viewing as an operator on $Y$. Thus it remains to show that the codimension of $1 - T$ on $Z$ is 0. One could see that we have reduced the proof to the case of the compact operator $T_Z$ on the Banach space $Z$ which is also injective.

Now, if $(1 - T_Z)Z$ is a strict subspace of $Z$ it follows that $(1 - T_Z)^nZ$ is a strictly decreasing sequence of closed subspaces of $Z$, and again we may find $z_n \in (1 - T_Z)^nZ$ such that $\|z_n\| = 1$ and $\text{dist}(z_n, (1 - T_Z)^{n+1}Z) \geq 1/2$. Then for $n < m$ we have $\|T_Z z_n - T_Z z_m\| = \|z_n - (1 - T_Z)z_n - T_Z z_m\| \geq \text{dist}(z_n, (1 - T_Z)^{n+1}Z) \geq 1/2$, violating the compactness of $T_Z$. It follows that the codimension of $1 - T_Z$ is 0 as desired.

### 9.3.2 Spectral properties

**Theorem 35.** Let $T$ be a compact operator on a Banach space $X$. Then

(i) its spectrum $\sigma(T)$ consists of at most countably many elements, all of them are eigenvalues with finite dimensional eigenspace.

(ii) 0 is the only possible accumulation point for the elements of $\sigma(T)$ if such an accumulation point exists, and 0 will belong to $\sigma(T)$ if the dimension of $X$ is infinite.

Note that if $X$ is a (separable) Hilbert space then we could furthermore diagonalize $T$ using the eigenvectors.
We now prove Theorem 35. First, given any \( \lambda \in \sigma(T) \) such that \( \lambda \neq 0 \) we show that it is an eigenvalue with finite dimensional eigenspace. Clearly \( \frac{1}{\lambda}T \) is compact. Thus, if \( ker(1 - \frac{1}{\lambda}T) \) is trivial then by Riesz’s theorem it is also onto, therefore by the open mapping theorem \( 1 - \frac{1}{\lambda}T \) is boundedly invertible and therefore \( \lambda \notin \sigma(T) \). Thus \( ker(1 - \frac{1}{\lambda}T) \) is nontrivial and also finite dimensional by Riesz’ theorem, and so \( \lambda \) is an eigenvalue with finite dim eigenspace.

Now, we will show that 0 is the only possible accumulation point, which also implies that \( \sigma(T) \) is at most countable. It suffices to show that given any \( \epsilon > 0 \) there is some \( C = C(T, \epsilon) \) finite such that at most \( C(T, \epsilon) \) elements of \( \sigma(T) \) would be outside \([-\epsilon, \epsilon]\). Let \( \lambda_1, \ldots, \lambda_m \) be a finite collection of distinct elements of \( \sigma(T) \) with \( |\lambda_j| > \epsilon \), it suffices to show that \( m < O_{T, \epsilon}(1) \).

The idea is to let \( Y_n \) be the eigenspace associated with \( \lambda_n \) and observe that for any \( n \) it holds that \( Y_n \cap span\{Y_m, m < n\} \) is trivial. Let \( Y_{<n} = span\{Y_m, m < n\} \) which is now a strictly nested sequence. Then by Riesz’s lemma one could choose \( y_n \in Y_n \) such that \( \|y_n\| = 1 \) and \( dist(y_n, Y_{<n}) \geq 1/2 \). We will show that

\[
\|Ty_n - Ty_m\| \geq \epsilon/2
\]

for any \( n \neq m \). Indeed, without loss of generality assume \( n > m \), then using the fact that \( y_n \in Y_n \) and the fact that \( T \) leaves \( Y_m \) invariant we have

\[
\|Ty_n - Ty_m\| = \|Ty_m\| = \|\lambda_n y_n - Ty_m\| \geq |\lambda_n| dist(y_n, Y_{<n}) \geq \frac{|\lambda_n|}{2} \geq \epsilon/2
\]

Consequently, \( m \) is bounded above by the maximum number of points in \( TB \) where \( B \) is the unit ball inside \( X \) that are \( \epsilon/2 \) apart. Since \( TB \) is precompact this is finite and depends only on \( T \) and \( \epsilon \) (and certainly \( X \)), and independent of the sequence \( (\lambda_k) \). \( \square \)
Chapter 10

Spectral theorems for bounded self-adjoint operators on a Hilbert space

Let \( H \) be a Hilbert space. For a bounded operator \( A : H \to H \) its Hilbert space adjoint is an operator \( A^* : H \to H \) such that \( \langle Ax, y \rangle = \langle x, A^*y \rangle \) for all \( x, y \in H \). We say that \( A \) is bounded self adjoint if \( A = A^* \).

In this chapter we discussed several results about the spectrum of a bounded self adjoint operator on a Hilbert space. We emphasize that in this chapter \( A \) is bounded, there is also a notion of unbounded self adjoint operator which we will discuss in subsequent chapters.

10.1 Diagonalization form

The first result says that \( A \) could be diagonalized using some change of basis.

**Theorem 36.** Let \( A : H \to H \) be a bounded self-adjoint operator on a Hilbert space \( H \). Then there exists some \( L^2(X, \mu) \) and \( U : L^2(X, \mu) \to H \) isometric isomorphism such that for some bounded \( M \) on \( (X, \mu) \) and every \( f \in L^2(X, \mu) \) it holds that

\[
(U^{-1}AU)f(x) = M(x)f(x) , \quad x \in X
\]

**Proof.** We note that if \( H \) is the direct sum of subspaces \( H_1 \) and \( H_2 \) such that \( H_1 \) and \( H_2 \) are invariant under \( A \) then it suffices to prove the theorem for the restriction of \( A \) to each subspace. This applies to direct sums indexed by larger index sets.

Now it is not hard to see that \( H \) could be written as an orthogonal direct sum of subspaces of the form \( \text{span}(A^n\xi, n \geq 0) \) where \( \xi \in H \). (The proof uses Zorn’s lemma.) Note that these subspaces are invariant under \( A \), therefore it suffices to show the theorem when \( H = \text{span}(A^n\xi, n \geq 0) \) for some fixed \( \xi \).

Now, recall from spectral calculus for bounded operators on a Banach space that if \( f \) is analytic on a domain containing \( \sigma(A) \) then we could define \( f(A) \) and furthermore \( \sigma(f(A)) = \)
f(σ(A)). For polynomials we could do this directly using factorization into linear factors for polynomials.

In the case of bounded self adjoint operator we will show below that σ(A) ⊂ ℜ. (Note that this fact also holds for unbounded case, but we will not discuss that in this section.)

**Lemma 8.** Let A be bounded self adjoint on a complex Hilbert space H. Then σ(A) ⊂ ℜ.

To see this lemma, we will show that if λ ∈ ℂ has nonzero imaginary part then λ ∈ ρ(A). To do this, we will show that

\[ |⟨x, (λ − A)x⟩| ≥ c∥x∥^2 > 0 \]

for some c depending on λ. This would imply that λ − A is invertible using an application of the Lax-Milgram theorem: consider the bilinear form \( B(x, y) = ⟨x, (λ − A)y⟩ \), which is nondegenerate once we proved the above estimate, thus given any z we could find y such that \( ⟨x, z⟩ = B(x, y) \) for all \( x ∈ H \), which implies that \( z = (λ − A)y \), thus \( λ − A \) is bijective on \( H \) and so is boundedly invertible and so \( λ ∈ ρ(A) \) as desired.

To show the above estimate, simply write

\[ λ = a + ib \]

where \( a, b ∈ ℜ \) and \( b ≠ 0 \), then using the self-adjoint property of \( A \) it follows that \( ⟨x, (a − A)x⟩ \) is a real number, therefore

\[ |⟨x, (a + ib − A)x⟩| = |⟨x, (a − A)x⟩ + ib∥x∥^2| ≥ b∥x∥^2 \]

as desired. This completes the proof of the above lemma.

We now discuss functional calculus for bounded self adjoint operators. Note that since \( σ(A) \) is bounded and closed it will follow that \( σ(A) \) is a compact subset of ℜ, and thus we could define \( f(A) \) even if \( f \) is merely continuous (which would be weaker than the analytic assumption required by the complex method generally applied to all bounded operators). The idea is to use the Weierstrass theorem and define \( f(A) \) to be the limit in operator norm of \( p_n(A) \) where \( (p_n) \) is a sequence of polynomials that approximates \( f \). To see that this could be done, it suffices to show that if \( g \) is a polynomial then

\[ ∥g(A)∥ = \sup_{x ∈ σ(A)} |g(x)| \]

To see this last claim, we first show it for real polynomial. In fact we will consider \( g(x) = x \). Then as we proved before

\[ \sup_{x ∈ σ(A)} |g(x)| = \lim_n ∥A^n∥^{1/n} ≤ ∥A∥ \]

while \( ∥A∥ ≤ \sup_{x ∈ σ(A)} |g(x)| \) using either the spectral theorem, or by elementary methods.\(^1\)

The real polynomial case then follows from the spectral mapping theorem \( σ(g(A)) = g(σ(A)) \)

---

\(^1\)Here we could easily see that \( ∥Ax∥^2 = ⟨A^2x, x⟩ \leq ∥A^2∥∥x∥^2 \) thus by repeating we obtain \( ∥A^{2^n}∥^{2^{-n}} ≤ ∥A∥ \) as desired.
and the fact that \( g(A) \) which is self adjoint when \( g \) is real polynomial. To allow for complex polynomial \( p \), simply write
\[
\|p(A)\|^2 = \|p(A)^*p(A)\| = \|(p\bar{p})(A)\| = \sup_{\lambda \in \sigma(A)} (p\bar{p})(\lambda) = \sup_{\lambda \in \sigma(A)} |p(\lambda)|^2
\]

We note that as a consequence of the definition we also have

**Corollary 4.** For all continuous \( g \) on \( \sigma(A) \) it holds that \( \|g(A)\| = \sup_{x \in \sigma(A)} |g(x)| \).

Now, we may define a linear functional on the space of polynomials on \( \sigma(A) \) as follows: given such a polynomial \( p \), let \( L(p) = \langle p(A)\xi, \xi \rangle \), it is clear that \( \|L\| \leq \sup_{x \in \sigma(A)} |p(x)| \). Using Weierstrass’s theorem it follows that we could extend \( L \) to the space of continuous functions on \( \sigma(A) \). Using self-adjointness of \( A \) it is clear that \( L \) is positive: if \( f \geq 0 \) then write \( f = g^2 \) and \( g = \lim p_n \) limit of polynomials then \( L(f) = L(g^2) = \lim (p_n(A)^2\xi, \xi) = \lim \|p_n(A)\xi\|^2 \geq 0 \). Thus by the Riesz representation theorem we may write \( L(f) = \int f d\mu \) where \( \mu \) is some finite Borel measures on \( \sigma(A) \).

We now construct the operator \( U \), initially from the space of continuous functions on \( \sigma(A) \) to \( H \). If \( q \) is a polynomial then let \( Uq = q(A)\xi \), clearly
\[
\int_{\sigma(A)} |q(x)|^2 d\mu = \langle q\bar{q} \rangle - \langle (q\bar{q})(A)\xi, \xi \rangle = \|q(A)\xi\|^2 = \|Uq\|^2
\]
thus the restriction of \( U \) to the polynomials is an isometry and the image of the polynomials under \( U \) is clearly dense inside \( H \) by the given assumption. Thus \( U \) extends to an isometric isomorphism from \( L^2(\sigma(A), d\mu) \) to \( H \).

Finally we will show that \( U^{-1}AUf(x) = xf(x) \) for all \( f \in L^2 \), note that this equality is understood in the almost everywhere since with respect to \( \mu \), furthermore \( m(x) := x \) is bounded in \( \sigma(A) \) thanks to compactness of \( \sigma(A) \). Thanks to Weierstrass’s theorem again it suffices to show this equality for \( f \) being polynomials. In that case let \( g(x) = xf(x) \) also a polynomial, we then have
\[
U^{-1}AUf = U^{-1}(Af(A)\xi) = U^{-1}(g(A)\xi) = g(x) = xf(x)
\]
\[\square\]

### 10.2 Projection-valued measure and spectral projection

Recall that \( P : H \to H \) is a projection is \( P^2 = P \). We say that it is an orthogonal projection if \( \ker(P) \) and \( \text{range}(P) \) are orthogonal subspaces, which would be the case if and only if \( P = P^* \) (the underlying assumption is that \( P \) is bounded). Recall a basic fact:

**Theorem 37.** Given any closed subspace \( K \) of \( H \) there is an orthogonal projection \( P_K \) onto \( K \): \( \text{range}(P) = K \) and \( \ker(P) \) is exactly \( K^\perp \).
In this section we take a closer look at the spectral representation of $A$.

We say that a family $P_{\Omega}$ indexed by Borel subsets $\Omega \subset \mathbb{R}$ if a (compactly supported) projection-valued measure if

(i) $P_{\Omega}$ is an orthogonal projection on $H$
(ii) $P = 0$ (and $P_{[-M,M]} = I$ for some $M$ sufficiently large).
(iii) If $\Omega = \bigcup_{n \geq 1} \Omega_n$ disjoint union then $P_{\Omega} \xi = \sum P_{\Omega_n} \xi$ convergence in the norm.
(iv) If $\Omega_1$ and $\Omega_2$ are Borel sets then $P_{\Omega_1} P_{\Omega_2} = P_{\Omega_1 \cap \Omega_2}$.

We note that property (iv) is a corollary of the first three properties.

We first note that we could construct a measure out of $(P_{\Omega})$ when testing the projections on some vector $\phi \in H$. Namely let $\mu_{\phi_1,\phi_2}(\Omega) = \langle P_{\Omega} \phi_1, \phi_2 \rangle$ then this defines a compactly supported complex Borel measure on $\mathbb{R}$ with finite total mass, in fact it is not hard to see that

$$\| \mu_{\phi_1,\phi_2} \| \leq \| P_\mathbb{R} \| = 1$$

therefore we could integrate any bounded Borel measurable function $f$ and obtained a bounded bilinear form

$$T_f(\phi_1, \phi_2) = \int f(\lambda) d\mu_{\phi_1,\phi_2}(\lambda)$$

Then by Riesz representation theorem we may write $T_f(\phi_1, \phi_2) = \langle B \phi_1, \phi_2 \rangle$ where $B$ is a bounded linear operator on $H$. If we approximate $f$ using simple functions, it can be seen that $B$ is the limit in the weak operator topology of the corresponding linear combinations of $P_\Omega$ (note that this implies $B$ is the limit in the strong operator topology too). We will let $\int f(\lambda) dP_\lambda$ denote this operator $B$ and we think of this as the integration of $f$ over the measure induced by the family $(P_{\Omega})$. In particular, $\int \lambda dP_\lambda$ is a bounded self adjoint operator on $H$.

It turns out that the converse of this is also true:

**Theorem 38.** Given a bounded self adjoint operator $A$ on a complex Hilbert space $H$ let $P_{\Omega} = 1_\Omega(A)$ then $(P_{\Omega})$ is a compactly supported projection valued measure, and $\int f(\lambda) dP_\lambda$ converges to $f(A)$ (as defined by spectral caculus) in the strong operator norm ($T_j \to T$ means $\| T_j x - T x \| \to 0$ for all $x$). Furthermore this is the unique projection valued operator with this property.

To check that $\int f(\lambda) dP_\lambda$ converges to $f(A)$ in strong operator norm topology, it suffices to show convergence in the weak operator norm, then it suffices to check that if $f = 1_\Omega$ then $\langle \int 1_\Omega dP_\lambda \phi_1, \phi_2 \rangle = \langle 1_\Omega(A) \phi_1, \phi_2 \rangle$ for all $\phi_1, \phi_2 \in H$. By definition the left hand side the same as $\int 1_\Omega dP_{\phi_1,\phi_2} = \langle P_{\Omega} \phi_1, \phi_2 \rangle$ which is the same as the right hand side.

### 10.3 Spectral representation and decomposition

**Absolutely continuous, singular, and point spectral**

Recall that $H$ has the spectral representation $L^2(X, \mu)$ which comes from direct summing the spectral representations $L^2(\mu_j)$ of $K_j$. We may decompose $\mu_j$ into three parts (absolutely...
continuous, point, and singular parts) using the Radon Nikodym theorem. This leads to decomposition of \( K \) and also decomposition of \( H \) into the absolutely continuous, singular, and point spectrum \( H^{(p)}, H^{(s)}, \) and \( H^{(c)} \). Note that these are orthogonal subspaces and the corresponding spectral measure of the cyclic subspace has the inherited properties. Say if \( x \in H^{(c)} \) then the spectral measure of \( A \) on \( \text{span}(A^n x) \) is absolutely continuous with respect to the Lebesgue measure.

**Uniqueness of the spectrum** Let \( A \) be bounded self adjoint on \( H \) and assume that \( H \) be separable. Recall from prior sections that there is a decomposition of \( H \) into orthogonal direct sum of \( K_1, K_2, \ldots \) where \( K_j \) are orthogonal closed subspaces of \( H \) and each of them is invariant under \( A \), furthermore there is a linear isometry \( U_j \) mapping some \( L^2(\sigma(A), \mu_j) \) into \( K_j \) such that \( U_j^{-2} A U_j \) acts on this \( L^2 \) space by multiplication, in fact \( U_j^{-1} f(A) U_j g(x) = f(x) g(x) \) for all bounded measureable \( f \) and \( g \in L^2(\mu_j) \). One could certainly modify \( d\mu_j \) multiplicatively using a bounded positive function, this would affect the isometry part of the map \( U \) but the \( L^2 \) space remains the same and the corresponding action of \( f(A) \) is still pointwise multiplication.

Now if \( H \) has another decomposition into an orthogonal direct sum of closed subspaces \( L_1, L_2, \ldots \) with spectral representations \( L^2(T_j, \nu_j) \) for \( L_j \), then upto a set of (spectral) measure 0 we have \( \bigcup S_j = \bigcup T_k \), namely we will show that for any \( k \) the set difference \( T_k \setminus \bigcup S_j \) has zero measure under \( \nu_k \). Note that if we assume furthermore that these measures are absolutely continuous with respect to the Lebesgue measure then \( \bigcup S_j = \bigcup T_k \) up to set of Lebesgue measure 0.

To see this take \( f \in L^2(T_k, \nu_k) \), and let \( F \) be any bounded Borel measurable function. Note that \( S_j \) and \( T_k \) are invariant under \( F(A) \), and \( F \) coresponds to some \( h_f \in H \). We now decompose \( h_f \) into \( h_j \) where \( h_j \in K_j \), and \( h_j \) in turn corresponds to \( f_j \in L^2(S_j, \mu_j) \). Then

\[
\int |F(x) f(x)|^2 d\nu_k = \| h_f \|^2 = \| h_1 \|^2 + \| h_2 \|^2 + \cdots = \int |F(x)|^2 |f_1|^2 d\mu_1 + \int |F(x)|^2 |f_2(x)|^2 d\mu_2 + \cdots
\]

therefore if \( T_k \setminus \bigcup S_j \) has positive \( \mu_k \) measure we may choose \( F \) to be the characteristic function of this set and \( f \equiv 1 \), clearly the left hand side is now positive while the right hand side is 0, contradiction.

**Spectral multiplicity**

Consider the spectral representation of \( H \), assumed separable, using the orthogonal direct sum of \( K_1, \ldots \) with \( K_j = L^2(S_j, \mu_j) \) where \( S_j \) is the support of \( \mu_j \).

Given \( \lambda \in \mathbb{R} \) we define its spectral multiplicity with respect to this representation as the number of \( j \) such that \( \lambda \in S_j \).

Assume for simplicity that the spectrum of \( A \) is absolutely continuous with respect to the Lebesgue measure. Then it can be shown that the spectral multiplicity is independent of the spectral representation of \( H \); this is a theorem of Hellinger.
Chapter 11
Spectral theory for unbounded self-adjoint operator

Let $A$ be a linear operator on some dense subspace $D$ of a Hilbert space $H$.

Let $A^*$ be defined as follows: for every $\omega \in H$, if there exists $z \in H$ such that for every $x \in D$ it holds that $\langle Ax, \omega \rangle = \langle x, z \rangle$ then $\omega \in D^*$ the domain of $A^*$ and $A^* \omega := z$ (it is clear that such $z$ is unique if exists).

We say that $\lambda$ is in the resolvent set of $A$ if $A - z$ is bijective from $D$ to $H$, and $\sigma(A) = \mathbb{C} \setminus \rho(A)$ as before.

**Theorem 39.** $A$ be self-adjoint on $H$ with dense domain $D \subset H$. Then $\sigma(A) \subset \mathbb{R}$ and there is an orthogonal projection valued measure $P_{\Omega}$ such that $P_{\Omega}$ commutes with $A$ on the domain $D$ of $A$, which in turn consists of all vector $\phi$ such that

$$\int t^2 d\langle P_t \phi, \phi \rangle < \infty$$

and

$$A\phi = \int t dP_t \phi$$

We first show that $\sigma(A) \subset \mathbb{R}$. To do this we will show that for every $z \not\in \mathbb{R}$ the range of $A - z$ is closed and then we use the denseness of $D$ to show that $A - z$ is onto, as a by product of the proof we will also have $A - z$ is injective, in fact $(A - z)^{-1}$ is bounded from $H$ to $D$.

We now show the spectral theorem. To do that we first show that

(i) $R(z) = (A - z)^{-1}$ is a complex differentiable function of $z \in \mathbb{C}_+ := \{Im(z) > 0\}$ in the strong topology, namely the topology on the space of bounded operators induced by the seminorms $A\phi$, $\phi \in H$. In other words, $R(z)u$ is complex differentiable for every $u \in H$.

(ii) the adjoint of $R(z)$ is $R(z)$.

To see (i), first note that $R(z)$ is a bounded operator on $H$ for those $z$, let $u(z) = (A - z)^{-1}u$ for any $u \in H$. Then by simple algebra

$$u(z + \omega) - u(z) = \omega R(z) R(z + \omega)u$$
so \( u(z) \) is complex differentiable for all \( u \), thus \( R(z) \) is complex differentiable and therefore is analytic in the strong operator topology.

To see (ii), note that \( \langle u, R(\overline{z})v \rangle = \langle (A - z)R(z)u, R(\overline{z})v \rangle = \langle R(z)u, (A - \overline{z})R(\overline{z})v \rangle = \langle R(\overline{z})u, v \rangle \).

As a consequence, we know that \( F_u(z) := \langle R(z)u, u \rangle \) is complex valued analytic function on the upper half plane and \( F_u(z) \lesssim \frac{1}{\text{Im}(z)} \). Furthermore using the second property it follows that \( F(z) = \overline{F(\overline{z})} \).

We now show that \( F_u(z) \) has nonnegative imaginary part for any \( u \in H \) and \( z \) in the upper half plane and

\[
\lim_{y \to \infty} y \text{Im} F_u(iy) = \|u\|^2
\]

To see these points, let \( v = R(z)u \), then \( F_u(z) = \langle v, (A - z)v \rangle = \langle v, Av \rangle - \overline{z} \langle v, v \rangle \), note that self adjointness of \( A \) implies that \( \langle v, Av \rangle \in \mathbb{R} \), therefore \( \text{Im} F_u(z) = \text{Im}(z)\|v\|^2 > 0 \) as desired. Let \( z = iy \) where \( y \in \mathbb{R} \), then it follows that

\[
y \text{Im} F_u(iy) = y^2\|v\|^2
\]

on the other hand

\[
\|u\|^2 = \langle (A - iy)v, (A - iy)v \rangle = \|Av\|^2 + y^2\|v\|^2
\]

thus as \( y \to \infty \) we have \( y^2\|v\|^2 \to \|u\|^2 \) as desired once we show that \( Av = AR(iy)u \) converges to 0 as \( y \to \infty \) for any fixed \( u \in H \). Note that this holds if \( u \in D \): we could commute \( A \) and \( R(iy) \) and then \( \|AR(iy)u\| \lesssim \|R(iy)\|\|Au\| = O(1/y) \). If \( u \not\in D \) we could approximate it using some thing in \( D \), this could be done uniformly as long as we could show that \( AR(iy) \) is uniformly bounded. Infact we will show that

\[
\|AR(iy)\| \leq 2
\]

To see this, simply write

\[
\|AR(iy)u\| = \|u + iyR(iy)u\| \leq \|u\| + y\|R(iy)\|\|u\| \leq 2\|u\|
\]

Now one could think of \( F_u(z) \) as an analytic function that maps the upper half plane into itself, thus using Poisson integral and the Riesz representation theorem we could show that there is some nonnegative measure \( \mu_u \) and some linear function \( l(z) \) of \( z \) such that

\[
F_u(z) = l(z) + \int \frac{d\mu_u(t)}{t - z}
\]

(this is basically Herglotz’s theorem) for all \( z \) in the upper half plane. Then combining with the fact that \( yF_u(iy) \to \|u\|^2 \) it follows that the linear part is 0 and \( \mu_u(\mathbb{R}) = \|u\|^2 \) (this is basically a result of Nevanlinna.)
We note that the values of $F_u$ in the upper half plane determines its values in the lower half plane via the relationship $F_u(\bar{z}) = F_u(z)$. Therefore the equation
\[ F_u(z) = \int \frac{d\mu_u(t)}{t-z} \]
holds for all $z \in \mathbb{C} \setminus \mathbb{R}$. Then via algebras it follows that for all $u, v \in H$
\[ \langle R(z)u, v \rangle = \int \frac{d\mu_{u,v}(t)}{t-z} \]
where $\mu_{u,v}$ is a signed measure, basically a linear combination of $\mu_{u\pm v}, \mu_{u\pm iv}$.

It is not hard to check the following properties
(i) $\mu_{u,u} = \mu_u$ is a nonnegative measure with total mass $\|u\|^2$;
(ii) $\mu_{u,v}$ is linear in $u$ and conjugate linear in $v$ and $\mu_{u,v} = \overline{\mu_{v,u}}$;
(iii) the total variation $\|\mu_{u,v}\|$ is controlled by $\|u\||v||$. (use Cauchy Schwartz)

Then via repeated application of the Riesz representation theorem there exists a bounded self adjoint operator $P_\Omega$ for each Borel $\Omega \subset R$ such that $\mu_{u,v}(\Omega) = \langle P_\Omega u, v \rangle$ for all $u, v \in H$.

We will check that $P_\Omega$ indeeds define a projection valued measure, and it will then follow that
\[ \langle R(z)u, v \rangle = \int \frac{1}{t-z}d \langle P_tu, v \rangle \]

Check:
(i) Clearly $P_=0$ since $\mu_{u,v}(\emptyset) = 0$ for all $u, v$, also $P_\mathbb{R} = 1$ since $\|u\|^2 = \mu_U(\mathbb{R}) = \int \langle P_tu, u \rangle = \langle P_\mathbb{R}u, u \rangle$ for all $u \in H$.
(ii) We will show that $P_\Omega$ commutes with $A$. First we show that $P_\Omega$ commutes with $R(z)$.

Note that $R(z_1)$ and $R(z_2)$ commutes with each other for all $z_1$ and $z_2$ non real. Using the representation $R(z)u, v = \int (t-z)^{-1}d \langle E_t u, v \rangle$ we have
\[ \int (t-z_1)^{-1}d \langle R(z_2)P_t u, v \rangle = \int (t-z_1)^{-1}d \langle P_t u, R(z_2)v \rangle = \]
\[ = \langle R(z_1)u, R(z_2)v \rangle = \langle R(z_2)R(z_1)u, v \rangle = \langle R(z_1)R(z_2)u, v \rangle = \ldots \]
\[ = \int (t-z_1)^{-1}d \langle P_t R(z_2)u, v \rangle \]

Using the fact that the Cauchy transform determines uniquely the measure, it follows that $R(z_2)P_\Omega = P_\Omega R(z_2)$ for any Borel $\Omega$.

Let $u \in H$, then $AR(z)u = u + zR(z)u$, thus for any $\Omega$ we have
\[ P_\Omega A(R(z)u) = P_\Omega u + zP_\Omega(R(z)u) = P_\Omega u + zR(z)P_\Omega u = AR(z)P_\Omega u = AP_\Omega R(z)u \]
since the range of $R(z)$ is $D$ it follows that $P_\Omega Au = AP_\Omega u$ for all $u \in D$.

(iii) We will show that $P_{\Omega_1}P_{\Omega_2} = P_{\Omega_1 \cap \Omega_2}$. This is a consequence of the resolvent identity $R(z)R(w) = (z-w)R(z)R(w)$. Indeed, this identity implies
\[
\langle R(z)R(w)u, v \rangle = \frac{1}{z-w} \int (\frac{1}{t-z} - \frac{1}{t-w}) d \langle Pu, v \rangle = \int \frac{1}{(t-z)(t-w)} d \langle Pu, v \rangle
\]

On the other hand, as before
\[
\langle R(z)R(w)u, v \rangle = \int (t-z)^{-1} d \langle R(w)Pu, v \rangle
\]

therefore for any Borel $\Omega_1 \subset \mathbb{R}$ and any $w \not\in \mathbb{R}$ we have
\[
\int \frac{1}{t-w} d \langle 1_{\Omega_1} Pu, v \rangle = \langle R(w)Pu, v \rangle = \int \frac{1}{t-w} d \langle PuP_{\Omega_1}u, v \rangle
\]

Using again the fact that the Cauchy transform determines the measure again, it follows that given any $\Omega_2 \subset \mathbb{R}$ we have
\[
\langle P_{\Omega_2 \cap \Omega_1} u, v \rangle = \langle P_{\Omega_2}P_{\Omega_1} u, v \rangle
\]
as desired.

(iv) Finally, assume $\phi \in D$. Then $\phi = R(z)u$ for some $u \in H$, and $z$ is arbitrary in the upper half plane. We then have
\[
\langle P_{\Omega_2} \phi, \phi \rangle = \langle R(z)R(\overline{z})P_{\Omega_1}u, u \rangle = \int \frac{1}{t^2 + |z|^2} d \langle Pu, u \rangle
\]

therefore $d \langle P_t \phi, \phi \rangle = \frac{1}{t^2 + |z|^2} d \langle Pu, u \rangle$ as measures. In particular,
\[
\int t^2 d \langle P_t \phi, \phi \rangle \leq \int d \langle Pu, u \rangle < \infty
\]

Using orthgonality of $P_\Omega$ on disjoint $\Omega$’s one could construct $\int tdPu$ for all $u \in D$ using simple functions (the integral exists in the strong operator topology) and we could show that this is the same as $Au$. To see this, let $v \in D$ and $u = (A-z)v$ then similarly we could show for any $\phi \in H$:
\[
d \langle P_t v, \phi \rangle = \frac{1}{t-z} d \langle Pu, \phi \rangle
\]

\[
(t-z)d \langle P_t v, \phi \rangle = d \langle P_t(A-z)v, \phi \rangle = d \langle P_tA v, \phi \rangle - zd \langle P_t v, \phi \rangle
\]

\[
\int t d \langle P_t v, \phi \rangle = \langle Av, \phi \rangle
\]

therefore $\int t d \langle P_t v, \phi \rangle = \langle Av, \phi \rangle$ for every $\phi$, thus for any $v \in D$ we have
\[
Av = \int t dP_t v
Chapter 12
Fredholm determinant

12.1 Motivation

Fredholm determinant started in Fredholm’s investigation of the integral equation

\[(1 + K)u = f\]

where \( K \) is the integral operator \( Kf(y) = \int_Y K(x, y)f(y)dy \) mapping functions on \( Y \) to functions on \( X \). (Here \( X, Y \) are compact metric spaces.) We know that \( K \) is compact from \( L^2(Y) \) to \( L^2(X) \) if the kernel \( K \) is square integrable in \( X \times Y \). The setting originally considered is for continuous functions, for which we have the following results:

(i) If \( K \) is a continuous function of \( x \) and \( y \) then \( K \) is compact from \( L^1(Y) \) to \( C(X) \).
(ii) If \( K \) is a continuous function of \( x \) in the \( L^1 \) norm wrt to \( y \), i.e. \( \|K(x,.)\|_{L_1^y} \) is continuous wrt \( x \), then \( K \) is compact from \( C(Y) \) to \( C(X) \).

Proof of these facts uses the Arzela Ascoli criteria: \( S \) hausdorff compact, functions equicontinuous and pointwise uniformly bounded, then the family is precompact in the sup norm on \( S \).

For simplicity, assume \( K : [0,1]^2 \to \mathbb{C} \) is continuous and \( f \in C[0,1] \). Note that this is not a Hilbert space. By Riesz’s theorem the integral equation

\[ u(x) + \int_0^1 K(x, y)u(y)dy = f(x) \]

is solvable on \( C[0,1] \) (i.e. given \( f \in C[0,1] \) one could find \( u \in C[0,1] \)) iff the integral operator \( K \) is injective on this space, or equivalently its range is the whole space. It can be shown that a vector is in the range of \( K \) iff it is orthogonal to the null space of \( K^* \) which has kernel \( K^*(x, y) = K(y, x) \).

Fredholm investigated the above equation by discretizing the equation and appeal to linear algebra, and then taking a limit at the end. This give rises to a number called the Fredholm determinant of \( 1 + K \) (we simply say the Fredholm determinant for \( K \)), which determines whether the given integral equation is solvable or not. The determinant concept
CHAPTER 12. FREDHOLM DETERMINANT

has been extended to other settings, most commonly for $K$ being a trace class operator on some separable Hilbert space.

We first discuss Fredholm’s original approach for $C[0,1]$ (modulo some simplifications), and then we’ll discuss the Hilbert space theory later.

12.2 Fredholm’s approach for integral operators

We now investigate Fredholm’s approach towards solving the above equation. Fredholm’s idea is to use linear algebra, approximating the integral using discretized sums and taking an appropriate limit at the end.

$$u_i + h \sum K_{ij} u_j = f_i, \quad i = 1, \ldots, n,$$

where $h = 1/n$, $f_i = f(ih)$ and $K_{ij} = K(ih, jh)$ and $u_i = u(ih)$. The determinant of the matrix acting on the vector $(u_1, \ldots, u_n)$ is denoted by $D(h)$

$$D(h) = \det(I + h(K_{ij}))$$

which is clearly a polynomial in $h$ of degree $n$

$$D(h) = \sum_{m=0}^{n} c_m h^m$$

$$c_m = \frac{1}{m!} \left. \frac{d}{dh} \right|_{h=0}^\infty \sum_{k=0}^\infty K(x_1, \ldots, x_k)$$

We then use the product rule, which says that if $C = \det(C_1, \ldots, C_n)$ is the determinant of a matrix with columns $C_1, \ldots, C_n$ then by multilinearity $\frac{d}{dh} C = \sum_k \det(C_1, \ldots, \frac{d}{dh} C_k, \ldots, C_n)$. In our case each column is linear in $h$, and $C_j(0) = (0, \ldots, 1, \ldots, 0)$ the unit is in the $j$th position. Therefore

$$D(h) = 1 + h \sum_j K_{jj} + \frac{h^2}{2} \sum_{i,j} \det \begin{pmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{pmatrix} + \ldots$$

For convenience let $K\left(\begin{array}{c} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{array}\right)$ denote the determinant of the matrix $K(x_i, y_j)$, then letting $h = 1/n$ and send $n \to \infty$ we obtain

$$D = \sum_{k=0}^\infty \frac{1}{k!} \int \ldots \int K\left(\begin{array}{c} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{array}\right) dx_1 \ldots dx_k$$

This is called the Fredholm determinant of the integral operator $K$. Note that this is a complex number.
12.2. FREDHOLM’S APPROACH FOR INTEGRAL OPERATORS

12.2.1 Convergence and Continuity of the determinant

We’ll show that the sum defining $D$ converges. We’ll use Hadamard’s inequality: for column vectors $v_1, \ldots, v_n \in \mathbb{R}^n$, it holds that

$$|\det(v_1|v_2| \ldots |v_n)| \leq n^{n/2} \prod_{i=1}^n |v_i|_\infty$$

This follows from the fact that the volume of the parallelopipde is at most the product of the side lengths.

Now, the given assumption on $K$ implies that $\sup_{x,y} |K(x, y)| \leq M$ for some finite $M$. So by Hadamard inequality we have

$$K\left(\begin{array}{c} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{array}\right) \leq (Mk^{1/2})^k \quad (12.1)$$

which implies the desired convergence.

Let $\|\|$ denote the sup norm on $[0,1]^2$ below, we obtain

Lemma 9. Let $F$ and $G$ be two functions on $[0,1]^2$. Then

$$|\det(F(x_i, x_j)) - \det(G(x_i, x_j))| \leq n^{1+\frac{2}{n}} \|F - G\|_{sup} \max(\|F\|, \|G\|)^{n-1}$$

here all matrices are $n \times n$.

Proof. For convenience, let $F$ be the $n \times n$ matrix with entries $F(x_i, x_j)$ and $G$ be the $n \times n$ matrix with entries $G(x_i, x_j)$. It is clear that

$$\det(F) - \det(G) = \det M_1 + \det M_2 + \cdots + \det M_n$$

where $M_k$ is the matrix whose first $k - 1$ rows are the same as $G$, and the $k$th row is the same as $F - G$, and the last $n - k$ rows are the same as $F$.

Apply Hadamard inequality we obtain the first desired estimate. □

We now discuss how to solve the integral equation using Fredholm’s determinant. The idea is to solve it at the discrete level and then send $h \to 0$ via $n \to \infty$. Via this heuristic we obtain

$$(I + K)(I + L) = (I + L)(I + K) = I$$

where

$$Lf = \int L(x, y) f(y) dy$$

$$L(x, y) = -D^{-1} \sum_{k \geq 0} \frac{1}{k!} \int \cdots \int K\left(\begin{array}{c} x, x_1, \ldots, x_k \\ y, y_1, \ldots, y_k \end{array}\right) d\xi_1 \ldots d\xi_k$$

We now ready to prove
Theorem 40. Let $K$ acts on $C[0,1]$. Let $K$ be a continuous kernel.

(i) If $D = 0$ then the operator $I + K$ has a nontrivial null space and therefore is not invertible.

(ii) Conversely if $D \neq 0$ then the operator $I + K$ is invertible, furthermore its inverse is given by $I + L$ where $L$ is defined above.

Proof: We first note that if $K_1$ and $K_2$ are two integral operators with kernel $K_1(x,y)$ and $K_2(x,y)$ respectively then $(1 + K_1)(1 + K_2) = 1 + K_3$ where $K_3$ is another integral operator whose kernel is given by

$$K_3(x,y) = K_1(x,y) + K_2(x,y) + \int K_1(x,z)K_2(z,y)dz$$

(i) Now, assume that $D \neq 0$. We need to show

$$K(x,y) + L(x,y) + \int K(x,z)L(z,y)dz = 0$$

$$L(x,y) + K(x,y) + \int L(x,z)K(z,y)dz = 0$$

We will show the first equality. For convenience of notation let

$$R(x,y) := \sum_{k \geq 0} \frac{1}{k!} \int \ldots \int K\left(\begin{array}{c} x, & x_1, & \ldots, & x_k \\ y, & y_1, & \ldots, & y_k \end{array}\right) d\xi_1 \ldots d\xi_k$$

Then $L = -\frac{1}{D} R$.

Now, computing the determinant $K\left(\begin{array}{c} x, & x_1, & \ldots, & x_k \\ y, & y_1, & \ldots, & y_k \end{array}\right)$ using its first row, we obtain

$$K\left(\begin{array}{c} x, & x_1, & \ldots, & x_k \\ y, & y_1, & \ldots, & y_k \end{array}\right) = K(x,y)K\left(\begin{array}{c} x_1, & x_2, & \ldots, & x_k \\ y_1, & y_2, & \ldots, & y_k \end{array}\right) +$$

$$+ \sum_{j=1}^{k} (-1)^j K(x,y)\left(\begin{array}{c} x_1, & x_2, & \ldots, & x_k \\ y, & y_1, & \ldots, (y_j), & y_k \end{array}\right)$$

here $(y_j)$ means one omit $y_j$.

Now let $y_1 = x_1, \ldots, y_k = x_k$ and integrate the above over $x_1, \ldots, x_k \in [0,1]$. Then the integrals of the last $k$ terms in the above expansion are actually the same, one could prove this by simple change of variable. It follows that

$$\int \ldots \int K\left(\begin{array}{c} x_1, & x_2, & \ldots, & x_k \\ x, & x_1, & \ldots, & x_k \end{array}\right) dx_1 \ldots dx_k$$

$$= K(x,y) \int \ldots \int K\left(\begin{array}{c} x_1, & x_2, & \ldots, & x_k \\ x_1, & x_2, & \ldots, & x_k \end{array}\right) dx_1 \ldots dx_k$$
12.2. FREDHOLM’S APPROACH FOR INTEGRAL OPERATORS

\[ + kK(x, x_1) \begin{pmatrix} x_1, x_2, \ldots, x_k \\ y, x_2, \ldots, x_k \end{pmatrix} dx_1 \ldots dx_k \]

Dividing by \( k! \) and summing over \( k \geq 0 \), we obtain

\[ R(x, y) = K(x, y)D - \int K(x, x_1)R(x_1, y)dx_1 \]

which implies the desired equality \( K(x, y) + L(x, y) + \int K(x, z)L(z, y)dz = 0. \)

For the second equality, one argues similarly, computing the determinant using the first column.

(ii) Since \( D = 0 \), by part (i) we have

\[ R(x, y) + \int K(x, z)R(z, y)dz = 0 \]

for every \( x, y \). If \( R \not\equiv 0 \) then one could find one \( y \) such that \( g(.) = R(. , y) \) is not the zero function (note that it is continuous), and it satisfies \( g + Kg = 0 \), therefore \( 1 + K \) is not injective.

It is however possible that \( R \equiv 0 \). If so, consider the following functions

\[ D(z) = \sum_{k \geq 0} \frac{z^k}{k!} \int \ldots \int K \begin{pmatrix} x_1, \ldots, x_k \\ x_1, \ldots, x_k \end{pmatrix} dx_1 \ldots dx_k \]

\[ R(x, y, z) := \sum_{k \geq 0} \frac{z^{k+1}}{k!} \int \ldots \int K \begin{pmatrix} x, x_1, \ldots, x_k \\ y, y_1, \ldots, y_k \end{pmatrix} d\xi_1 \ldots d\xi_k \]

One could think of \( D(z) \) and \( R(x, y, z) \) as the version of \( D \) and \( R \) with \( zK \) instead of \( K \). It is clear that \( D \) and \( R \) are entire functions. Since \( D(1) = 0 \) the entire function \( D \) has a zero of finite order \( n \geq 1 \) at \( z = 1 \). By algebraic manipulation we have

\[ \int R(x, x, z)dx = zD'(z) \]

therefore there is some \( (x, y) \) such that \( R(x, y, z) \) can not vanish at \( z = 1 \) with order more than \( n - 1 \). Then for some \( 1 \leq \ell < n \) it holds that

\[ R(x, y, z) = (z - 1)^\ell g(x, y) + O((z - 1)^{\ell+1}) \]

where \( g(x, y) \not\equiv 0 \), and \( g \) is continuous in \( x, y \) (to see this note that \( g \) is the uniform limit of a sequence of continuous functions on \([0, 1]^2\)). We recall that \( R(x, y, z) = zK(x, y)D(z) - z \int K(x, x_1)R(x_1, y, z)dx_1 \) thus by dividing everything by \( (z - 1)^\ell \) and then send \( z \to 1 \) we obtain

\[ g(x, y) = - \int K(x, x_1)g(x_1, y)dx_1 \]

and since \( g \not\equiv 0 \) continuous it follows that \( 1 + K \) is not injective. □
Theorem 41. Assume that $K$ is Holder $c$-continuous where $c > 1/2$. Then the nonzero eigenvalues of $K$ on $C[0, 1]$ (only countability many of them since $K$ is compact) satisfies $\sum_j |\lambda_j| < \infty$ and for every $z \in \mathbb{C}$ we have

$$D(z) = \prod_j (1 + z\lambda_j)$$

and we also have the trace formula

$$\int K(x, x)dx = \sum_j \lambda_j$$

Proof. We recall Hadamard’s factorization theorem (or may be a consequence of this theorem): Let $f$ be an entire function such that for some finite positive $C_1, C_2$ and $\rho \in [0, 1)$ it holds for every complex number $z$

$$|f(z)| \leq C_1 \exp(C_1|z|^\rho)$$

Assume that $f(0) \neq 0$. Then $f$ has at most a countable number of zeros, furthermore

$$\sum_{z: f(z)=0} \frac{1}{|z|} < \infty$$

and for every complex number $\lambda$ it holds that

$$f(\lambda) = f(0) \prod_{z: f(z)=0} (1 - \frac{\lambda}{z})$$

We plan to use this theorem to show the first part of the theorem. We note that the second part of the theorem, i.e. the trace formula would then follows from

$$\int K(x, x)dx = D'(z)|_{z=0}$$

(which is part of the definition of $D$) and the absolute convergence of the product $\prod_j (1 + z\lambda_j)$ (viewed as an infinite power series for $z$).

First, we will show that for $z \neq 0$ it holds that: $D(z) = 0$ if and only if $-1/z$ is an eigenvalue of $K$. This is simply a consequence of our last theorem applied to $D = \text{det}(1 + zK)$.

Now, we will show that if $K$ is Holder $c$-continuous then

$$|D(z)| \lesssim \exp(O(|z|^{2c}))$$

To see this, using Stirling’s formula it suffices to show that, for some $C$ finite,

$$|D(z)| \lesssim \sum_{n \geq 0} C^n |z|^{2n/(1+2c)} \frac{1}{n^n}$$
which is equivalent to showing that, for some $C$ finite,

$$|D(z)| \lesssim \sum_{k \geq 0} \frac{C^k |z|^k}{k^{(1+c)k}}$$

Using the definition of $D$ it suffices to show that given any $x_1, \ldots, x_k, y_1, \ldots, y_k$ it holds that for some $C > 0$ finite

$$|K(x_1, x_2, \ldots, x_k) - z K(y_1, y_2, \ldots, y_k)| \leq (Ck^{1/2})^k k^{-c}$$

Note that incomparision with (12.1) we gain a factor of $k^{-c}$. To see this it suffices to show that

$$|K(x_1, x_2, \ldots, x_k) - z K(y_1, y_2, \ldots, y_k)| \leq (Ck^{1/2})^k \prod_{j=1}^{k-1} |y_{j+1} - y_j|^c$$

(note that the constant $C$ may be different in different display). Indeed, without loss of generality we may assume $y_1 \leq y_2 \cdots \leq y_k$, then by the geometric arithmetical mean inequality we have $\prod_{j=1}^{k-1} |y_{j+1} - y_j| \leq (\frac{1}{k-1})^{k-1}$, as desired.

Now, note that if $c_1, \ldots, c_k$ are $k$ column vectors $(k \times 1)$ then

$$|\det(c_1 \ldots |c_k)| = |\det(c_1, c_2 - c_1, \ldots, c_k - c_{k-1})|$$

therefore using Hadamard’s inequality we obtain

$$|K(x_1, x_2, \ldots, x_k) - z K(y_1, y_2, \ldots, y_k)| \leq \prod_{j=1}^k \left(\sum_{n=1}^k |K(x_n, y_{j+1}) - K(x_n, y_j)|^2\right)^{1/2}$$

which implies the desired estimate using Holder continuity of $K$. □

### 12.3 Fredholm determinant for operators on Hilbert spaces

Let $H$ be a separable Hilbert space over $\mathbb{C}$. The Fredholm determinant $\det(1 + K)$ could be defined for trace class operators $K$ on a Hilbert space $H$. Note that this is not the same as the setting considered in the last section since $C[0,1]$ with the sup norm is not a Hilbert space.

We first define the singular values of a compact operator $T$ on $H$. Let $T^*$ be its adjoint, clearly $T^*T$ is a nonnegative self adjoint operator on $H$, so one could define its square root $A = \sqrt{T^*T}$ using functional calculus. Note that $T^*T$ and $A$ are both compact operators with nonnegative eigenvalues.

The singular values of $T$ are defined to be the positive eigenvalues of $A$, counting with multiplicity.
**Definition:** We say that $T$ is a trace class operator if the sum of its singular values is finite (counting with multiplicity). In that case the trace norm of $T$ is defined to be this sum, denoted by $\|T\|_{tr}$.

**Properties:** $T$ and $T^*$ has the same trace norm, and if $T$ is trace class then so is $TB$ and $BT$ where $B$ is any bounded operator on $H$, and $\|\cdot\|_{tr}$ satisfies the triangle inequality.

$$\|T_1 + T_2\|_{tr} \leq \|T_1\|_{tr} + \|T_2\|_{tr}$$

This inequality is an immediate consequence of the following equivalent characterization of the trace norm:

**Lemma 10.** For any trace class operator $T$

$$\|T\|_{tr} = \sup_{f_n,e_n} \sum_n |\langle Tf_n,e_n\rangle|$$

where the sup is taken over all $(f_n)$ and $(e_n)$ orthonormal bases of $H$.

To show this characterization we will first derive the **polar factorization of $T$:** namely there is a partial unitary operator $U$ such that $T = UA$, here unitary of $U$ simply means that $U^*U$ when acting on the range of $A$ is the identity operator. This should be thought of as the operator analogue of the usual polar factorization of a complex number.

To define $U$, note that $\|Au\| = \|Tu\|$ for all $u \in H$, therefore we may define an isometry $U$ from $\text{range}(A)$ to $\text{range}(T)$ by mapping $Au$ to $Tu$.

One then extends $U$ to $H$ by letting $U$ to be zero on the orthogonal complement of $\text{range}(A)$. It follows immediately that $U^*U \subset \overline{\text{range}(A)}$: one simply notice that for every $z$ in the orthogonal complement of $\text{range}(A)$ and $z' \in H$ it holds that $\langle z,U^*z'\rangle = \langle Uz,z'\rangle = 0$.

Now let $z$ in the closure of the range of $A$. We want to show that $(U^*U - 1)z = 0$, which is the desired local unitary property of $U$. Using the isometric property of $U$ on $\overline{\text{range}(A)}$, for every $z'$ in the closure of the range of $A$ we have

$$\langle z',z\rangle = \langle Uz',Uz\rangle = \langle z',U^*Uz\rangle$$

therefore $(U^*U - 1)z$ belongs to the orthogonal complement of $\overline{\text{range}(A)}$. But $U^*U - 1$ leaves $\overline{\text{range}(A)}$ invariant since the range of $U^*$ is inside $\text{range}(A)$. This contradiction completes the proof of the local unitary property of $U$.

We are now back to proving the above equivalent characterization of the trace class norm of $T$. Let $(F_n)$ be a complete set of normalized eigenvectors of $A = \sqrt{T^*T}$. (Note that since $A^* = A$ we could find such a set.) Let $G_n = UF_n$. (Note that $\|F_n\| = \|G_n\| = 1$. ) Then

$$\sum_{n} |\langle TF_n,G_n\rangle| = \sum_{n} |\langle UAF_n,UF_n\rangle| = \sum_{n} |\langle AF_n,F_n\rangle| = \|T\|_{tr}$$

therefore it remains to show that

$$\sum_{n} |\langle Tf_n,e_n\rangle| \leq \|T\|_{tr}$$
for any pair of orthogonal bases \((f_n)\) and \((e_n)\). Let \(s_n\) be the singular values of \(T\), namely
\[ AF_n = s_n F_n. \]
Then expand \(f_n\) into \((F_n)\) we have
\[ f_n = \sum_k \langle f_n, F_k \rangle F_k \]
and by an application of Fubini’s theorem
\[
\sum_n |\langle Tf_n, e_n \rangle| \leq \sum_k |s_k| \sum_n |\langle f_n, F_k \rangle \langle G_k, e_n \rangle| 
\]
(it will be clear from the proof that the double sum is absolutely summable)
\[
\leq \sum_k |s_k| (\sum_n |\langle f_n, F_k \rangle|^2)^{1/2} (\sum_n |\langle G_k, e_n \rangle|^2)^{1/2} 
\]
\[
\leq \sum_k |s_k|\|F_k\|\|G_k\| = \sum_k |s_k| = \|T\|_{tr} 
\]
This completes the proof of the characterization. \(\Box\)

**Trace:** Given a trace class operator one may define a linear functional, namely the trace of \(T\)
\[
\text{Trace}(T) = \sum_n \langle Tf_n, f_n \rangle 
\]
where \((f_n)\) is any orthonormal basis of \(H\). This definition is independent of the choice of the basis. Let \((g_n)\) be another orthonormal basis, then by expanding \(f_n\) into this new basis we have
\[
\sum_n \langle Tf_n, f_n \rangle = \sum_n \sum_k \langle f_n, g_k \rangle \langle Tg_k, f_n \rangle 
\]
it is not hard to see that this double sum is abs convergence (using Cauchy Schwartz). Then by Fubini
\[
\sum_n \langle Tf_n, f_n \rangle = \sum_k (\sum_n \langle f_n, g_k \rangle \langle Tg_k, f_n \rangle) 
\]
using orthogonality of \(f_n\) and normalization of \(f_n\) we obtain
\[
= \sum_k (\sum_n \sum_j \langle f_n, g_k \rangle \langle f_j, f_n \rangle \langle Tg_k, f_j \rangle) 
\]
\[
= \sum_k \langle Tg_k, g_k \rangle 
\]
By definition it is clear that \(|\text{Trace}(T)| \leq \|T\|_{tr}\) and the sum defining the trace is absolutely convergent.

Now, Lidskii’s trace formula says that
CHAPTER 12. FREDHOLM DETERMINANT

**Theorem 42** (Lidskii). If $T$ is trace class on a separable Hilbert space $H$ then

$$\text{Trace}(T) = \sum_j \lambda_j$$

where $\lambda_j$ are the eigenvalues of $T$.

Note that $T$ is compact so it has a countable set of nonzero eigenvalues. It is clear that the sum of the eigenvalues is bounded above by the trace norm.

**Proof.** The proof of this formula could be divided into two steps: first consider the cases when $T$ does not have any nonzero eigenvalues, and then in the second step we reduce the general case to this setting.

Let’s assume for now that $T$ does not have any nonzero eigenvalues, then we want to show that $\text{Trace}(T) = 0$. Let $s_1, s_2, \ldots$ be the singular values of $T$. Then we’ll show that for every $\lambda > 0$

$$e^{\lambda|\text{Trace}(T)|} \leq O(1 + |\lambda|)^M) e^{\lambda \sum_{j>M} s_j} \quad (12.2)$$

(the implicit constant is independent of $\lambda$), from here by sending $\lambda \to \infty$ we obtain $|\text{Trace}(T)| \leq \sum_{j>M} s_j$; and the desired estimate now follows from the fact that $\sum_j s_j = \|T\|_tr < \infty$. To show the above claim, we approximate $T$ by finite rank operators, say $T_n = P_n TP_n \to T$ where $P_n$ is the projection into the first $n$ basis vectors of $H$, here one may fix any orthonormal basis. Let $D_n(\lambda) = \det(1 + \lambda T_n)$, here the determinant is defined using linear algebra, in other words if $\Lambda_n$ is the set of eigenvalues of $T_n$ then $D_n(\lambda) = \prod_{\alpha \in \Lambda_n} (1 + \lambda \alpha)$.

We will show that uniformly on any compact subsets of the complex plane it holds that

$$e^{\lambda \text{Trace}(T)} = \lim_{n \to \infty} D_n(\lambda) \quad (12.3)$$

Note that by definition we have $\|T_n - T\| \to 0$ (in operator norm) and $\text{Trace}(T_n) \to \text{Trace}(T)$. Since the spectral radius of $T$ is 0 it is clear that the spectral radius $\sigma(T_n)$ of $T_n$ converges to 0 too. In particular given any bounded set of the complex plane we may choose $n$ large such that this bounded set is contained inside the ball of radius $1/\sigma(T_n)$. Furthermore one could also show that $s_j(T_n) \leq s_j(T)$ if the singular values of $T$ and $T_n$ are ordered in decreasing order.

Now, it is not hard to see that

$$\frac{D_n'(\lambda)}{D_n(\lambda)} = \sum_{\alpha \in \Lambda_n} \frac{\alpha}{1 + \lambda \alpha}$$

$$= \text{Trace}(T_n) + O\left(\sum_{k \geq 2} (|\lambda| \sigma(T_n))^{k-1} \|T_n\|_{tr}\right)$$

therefore uniformly over $\lambda$ in a bounded subset of $\mathbb{C}$ it holds that

$$\lim_{n \to \infty} \frac{D_n'(\lambda)}{D_n(\lambda)} - \text{Trace}(T) = 0$$
which implies the desired limiting equality (12.3).

Now using (12.3), for any \( \lambda > 0 \) we have

\[
e^{\lambda |\text{Trace}(T)|} \leq \liminf_{n \to \infty} \prod_{\alpha \in \Lambda_n} (1 + \lambda |\alpha|)
\]

We will show that the last limit is bounded above by \( \prod_{j=1}^{\infty} (1 + \lambda s_j(T)) \) where \( s_j(T) \) are the singular values of \( T \), which easily implies the desired estimate (12.2). To show this, using \( s_j(T_n) \leq s_j(T) \) it suffices to show that

\[
\sum_{\alpha \in \Lambda_n} \log(1 + \lambda |\alpha|) \leq \sum_{j=1}^{n} \log(1 + \lambda s_j(T_n))
\]

Thanks to convexity of \( \log \), one could show that this inequality is a consequence of the fact that

\[
\prod_{\alpha \in \Lambda_n} |\alpha| \leq \prod_{j=1}^{N} s_j(T_n)
\]

which in turn could be easily verified (equality holds if \( T_n \) is nonsingular).

Now to reduce the general case to the above setting, we consider the subspace \( K_1 \) of \( H \) spanned by the eigenfunctions and generalized eigenfunctions of \( T \). Let \( K_2 \) be the orthogonal complement of \( K_1 \). Using linear algebra it is clear that the trace of the restriction of \( T \) to \( K_1 \) is exactly the sum of the eigenvalues of \( T \). More precisely if \( (g_n) \) is a basis for \( K_2 \) and \( (f_m) \) is a basis for \( K_1 \) then we may take the union of the two as a basis for \( H \) and

\[
\text{Trace}(T) = \sum_{m} \langle Tf_m, f_m \rangle + \sum_{n} \langle Tg_n, g_n \rangle
\]

\[
= \sum_{\alpha \in \Lambda(T)} \alpha + \sum_{n} \langle T^*g_n, g_n \rangle
\]

thus using the previous argument it suffices to show that \( T^* \) leaves \( K_2 \) invariant and the only eigenvalue of \( T^* \) on \( K_2 \) is 0 (it is clear that \( T_2 \) is also compact and trace class). These properties could be easily checked.

\[\square\]

**Determinant** We now discuss \( \det(1 + T) \) for trace class operators \( T \) on a separable Hilbert space \( H \).

Define an inner product on \( H^k \) by

\[
\langle (w_1, \ldots, w_k), (v_1, \ldots, v_k) \rangle = \det(\langle w_i, v_j \rangle)_{i,j}
\]

Then \( T \) extends to a trace class operator \( T_k \) on \( H^k \), defined by \( T_k(w_1, \ldots, w_k) = (Tw_1, \ldots, Tw_k) \), with \( ||T_k||_tr \leq ||T||^k_{tr} \). Then define

\[
\det(1 + T) = \sum_{k \geq 0} \text{Trace}(T_k)
\]
CHAPTER 12. FREDHOLM DETERMINANT

One could check that if \( T \) is finite rank then this determinant is the same as what we would obtain from linear algebra, and \( \det(1 + .) \) is locally Lipschitz wrt to the trace class norm. Approximating \( T \) by a sequence of finite rank operators (for which \( \det(1 + T) \) could be defined using standard linear algebra), we could show that \( \det(1 + T) \) is the limit of the corresponding determinants, and thus this limit is independent of the choice of the sequence. Now, one could use the polar decomposition \( T = UA \) and approximate \( A \) by projections into finite dimensional subspaces of \( H \) spanned by eigenfunctions of \( A \). By this approximation scheme it can be shown that

\[
\det(1 + T) = \prod (1 + \lambda_j)
\]

where \( \lambda_j \) are the eigenvalues of \( T \) and

\[
\det[(1 + T_1)(1 + T_2)] = \det(1 + T_1) \det(1 + T_2)
\]

for any two trace class operators \( T_1 \) and \( T_2 \).

Using \( \exp(T) = 1 + T + \cdots + \frac{T^n}{n!} + \cdots \) we could also define \( \det(\exp T) \) too, and in fact \( \det(\exp T) = e^{\text{Trace}(T)} \).

As an example, if \( T \) is an integral operator on some \( L^2(X) \) with kernel \( K(x, y) \) with mild assumptions on \( K \) and \( X \), then it could be shown that

\[
\text{Tr}(T_k) = \frac{1}{k!} \int \cdots \int K \left( \begin{array}{c} x_1, x_2, \ldots, x_k \\ x_1, x_2, \ldots, x_k \end{array} \right) dx_1 \cdots dx_k
\]

and the definition of the determinant coincides with the previous section.