Functional Analysis I

Term 1, 2009–2010

Vassili Gelfreich

Contents

Preface

These notes follow the lectures on Functional Analysis given in the Autumn Term of 2009. If you find a mistake or misprint please inform the author by sending an e-mail to `v.gelfreich@warwick.ac.uk`. The author thanks James Robinson for his set of notes and selection of exercises which significantly facilitated the preparation of the lectures. The author also thanks all students who helped with proofreading the notes.

1 Vector spaces

1.1 Definition.

A vector space V over a field $\mathbb{K}$ is a set equipped with two binary operations called vector addition and multiplication by scalars. Elements of V are called vectors and elements of $\mathbb{K}$ are called scalars. The sum of two vectors $x, y \in V$ is denoted $x+y$, the product of a scalar $\alpha \in \mathbb{K}$ and vector $x \in V$ is denoted αx.

It is possible to consider vector spaces over an arbitrary field $\mathbb{K}$, but we will consider the fields $\mathbb{R}$ and $\mathbb{C}$ only. So we will always assume that $\mathbb{K}$ denotes either $\mathbb{R}$ or $\mathbb{C}$ and refer to V as a real or complex vector space respectively.

In a vector space, addition and multiplication have to satisfy the following set of axioms: Let x, y, z be arbitrary vectors in V, and α, β be arbitrary scalars in $\mathbb{K}$, then

- Associativity of addition: $x+(y+z) = (x+y)+z$.
- Commutativity of addition: $x+y = y+z$.
- There exists an element $0 \in V$, called the *zero vector*, such that $x+0 = x$ for all $x \in V$.
- For all $x \in V$, there exists an element $y \in V$, called the additive inverse of x, such that $x+y = 0$. The additive inverse is denoted $-x$.
- "Associativity" of multiplication:[1] $\alpha(\beta x) = (\alpha\beta)x$.
- Distributivity:

$$\alpha(x+y) = \alpha x + \alpha y \qquad \text{and} \qquad (\alpha+\beta)x = \alpha x + \beta x.$$

- There is an element $1 \in \mathbb{K}$ such that $1x = x$ for all $x \in V$. This element is called *the multiplicative identity* in $\mathbb{K}$.

[1]The purist would not use the word "associativity" for this property as it includes two different operations: $\alpha\beta$ is a product of two scalars and βx involves a vector and scalar.

It is convenient to define two additional operations: subtraction of two vectors and division by a (non-zero) scalar are defined by

$$\begin{aligned} x-y &= x+(-y), \\ x/\alpha &= (1/\alpha)x. \end{aligned}$$

1.2 Examples of vector spaces

1. $\mathbb{R}^n$ is a real vector space.
2. $\mathbb{C}^n$ is a complex vector space.
3. $\mathbb{C}^n$ is a real vector space.
4. The set of all polynomials P is a vector space:

$$P = \left\{ \sum_{k=0}^{n} \alpha_k x^k : \ \alpha_k \in \mathbb{K}, \ n \in \mathbb{N} \right\}.$$

5. The set of all bounded sequences $\ell^\infty(\mathbb{K})$ is a vector space:

$$\ell^\infty(\mathbb{K}) = \left\{ (x_1, x_2, \ldots) : \ x_k \in \mathbb{K} \text{ for all } k \in \mathbb{N}, \ \sup_{k\in\mathbb{N}} |x_k| < \infty \right\}.$$

For two sequences $x, y \in \ell^\infty(\mathbb{K})$, we define $x+y$ by

$$x+y = (x_1+y_1, x_2+y_2, \ldots).$$

For $\alpha \in \mathbb{K}$, we set

$$\alpha x = (\alpha x_1, \alpha x_2, \ldots).$$

We will always use these definitions of addition and multiplication by scalars for sequences.

6. Let $1 \le p < \infty$. The set $\ell^p(\mathbb{K})$ of all p^{th} power summable sequences is a vector space:

$$\ell^p(\mathbb{K}) = \left\{ (x_1, x_2, \ldots) : \ x_k \in \mathbb{K}, \ \sum_{k=1}^{\infty} |x_k|^p < \infty \right\}.$$

The definition of the multiplication by scalars and vector addition is the same as in the previous example. Let us check that the sum $x+y \in \ell^p(\mathbb{K})$ for any $x, y \in \ell^p(\mathbb{K})$. Indeed,

$$\begin{aligned} \sum_{k=1}^{\infty} |x_k+y_k|^p &\le \sum_{k=1}^{\infty} (|x_k|+|y_k|)^p \le \sum_{k=1}^{\infty} 2^p(|x_k|^p+|y_k|^p) \\ &\le 2^p \sum_{k=1}^{\infty} |x_k|^p + 2^p \sum_{k=1}^{\infty} |y_k|^p < \infty. \end{aligned}$$

7. The space $C[0,1]$ of all real-valued continuous functions on the closed interval $[0,1]$ is a vector space. The addition and multiplication by scalars are defined naturally: for $f,g \in C[0,1]$ and $\alpha \in \mathbb{R}$ we defined by $f+g$ the function whose values are given by

$$(f+g)(t) = f(t) + g(t)\,, \qquad t \in [0,1]\,,$$

and αf is the function whose values are

$$(\alpha f)(t) = \alpha f(t)\,, \qquad t \in [0,1]\,,$$

We will always use these definitions on all spaces of functions to be considered later.

8. The set $\tilde{L}^1(0,1)$ of all real-valued continuous functions f on the open interval $(0,1)$ for which

$$\int_0^1 |f(t)|\,dt < \infty$$

is a vector space.

If $f \in C[0,1]$ then $f \in \tilde{L}^1(0,1)$. Indeed, since $[0,1]$ is compact f is bounded (and attains its lower and upper bounds). Then

$$\int_0^1 |f(t)|\,dt \le \max_{t\in[0,1]} |f(t)| < \infty,$$

i.e. $f \in \tilde{L}^1(0,1)$.

We note that $\tilde{L}^1(0,1)$ contains some functions which do not belong to $C[0,1]$. For example, $f(t) = t^{-1/2}$ is not continuous on $[0,1]$ but it is continuous on $(0,1)$ and

$$\int_0^1 |f(t)|\,dt = \int_0^1 |t|^{-1/2} dx = 2t^{1/2}\Big|_0^1 = 2 < \infty,$$

so $f \in \tilde{L}^1(0,1)$.

We conclude that $C[0,1]$ is a strict subset of $f \in \tilde{L}^1(0,1)$.

1.3 Hamel bases

Definition 1.1 *The linear span of a subset E of a vector space V is the collection of all finite linear combinations of elements of E:*

$$\mathrm{Span}(E) = \left\{ x \in V : x = \sum_{j=1}^{n} \alpha_j e_j,\ n \in \mathbb{N},\ \alpha_j \in \mathbb{K},\ e_j \in E \right\}.$$

We say that E spans *V if $V = \mathrm{Span}(E)$,* i.e. *every element of V can be written as a finite linear combination of elements of E.*

Definition 1.2 *A set E is* linearly independent *if any finite collection of elements of E is linearly independent:*

$$\sum_{j=1}^{n} \alpha_j e_j = 0 \qquad \Longrightarrow \qquad \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

for any choice of $n \in \mathbb{N}$, $e_j \in E$ and $\alpha_j \in \mathbb{K}$.

Definition 1.3 *A* Hamel basis E for V *is a linearly independent subset of V which spans V.*

Examples:

1. Any basis in $\mathbb{R}^n$ is a Hamel basis.
2. The set $E = \{ 1, x, x^2, \dots \}$ is a Hamel basis in the space of all polynomials.

Lemma 1.4 *If E is a Hamel basis for a vector space V then any element $x \in V$ can be uniquely written in the form*

$$x = \sum_{j=1}^{n} \alpha_j e_j$$

where $n \in \mathbb{N}$, $\alpha_j \in \mathbb{K}$ and $e_j \in E$.

Exercise: Prove the lemma.

Definition 1.5 *We say that a set is* finite *if it consists of a finite number of elements.*

Theorem 1.6 *If V has a finite Hamel basis then every Hamel basis for V has the same number of elements.*

Definition 1.7 *If V has a finite basis E then the* dimension of V *(denoted* $\dim V$*) is the number of elements in E. If V has no finite basis then we say that V is* infinite-dimensional.

Example: In $\mathbb{R}^n$ any basis consists of n vectors. Therefore $\dim \mathbb{R}^n = n$.

Let V and W be two vector spaces over $\mathbb{K}$.

Definition 1.8 *A map $L : V \to W$ is called* linear *if for any $x, y \in V$ and any $\alpha \in \mathbb{K}$*

$$L(x + \alpha y) = L(x) + \alpha L(y) .$$

Definition 1.9 *If a linear map $L : V \to W$ is a bijection, then L is called a linear* isomorphism. *We say that V and W are linearly isomorphic if there is a bijective linear map $L : V \to W$.*

Proposition 1.10 *Any n-dimensional vector space over $\mathbb{K}$ is linearly isomorphic to $\mathbb{K}^n$.*

Proof: If $E = \{ e_j : 1 \le j \le n \}$ is a basis in V, then every element $x \in V$ is represented uniquely in the form

$$x = \sum_{j=1}^{n} \alpha_j e_j \,.$$

The map $L : x \mapsto (\alpha_1, \dots, \alpha_n)$ is a linear bijection $V \to \mathbb{K}^n$. Therefore V is linearly isomorphic to $\mathbb{K}^n$. □

In order to show that a vector space is infinite-dimensional it is sufficient to find an infinite linearly independent subset. Let's consider the following examples:

1. $\ell^p(\mathbb{K})$ is infinite-dimensional ($1 \le p \le \infty$).

   Proof. The set

   $$E = \{ (1,0,0,0,\dots),\ (0,1,0,0,\dots),\ (0,0,1,0,\dots),\ \dots \}$$

   is not finite and linearly independent. Therefore $\dim \ell^p(\mathbb{K}) = \infty$.

   Remark: This linearly independent subset is not a Hamel basis. Indeed, the sequence $x = (x_1, x_2, x_3, \dots)$ with $x_k = e^{-k}$ belongs to $\ell^p(\mathbb{K})$ for any $p \ge 1$ but cannot be represented as a sum of finitely many elements of the set E.

2. $C[0,1]$ is infinite-dimensional.

   Proof: The set $E = \{ x^k : k \in \mathbb{N} \}$ is linearly independent subset of $C^0[0,1]$: Indeed, suppose

   $$p(x) = \sum_{k=1}^{n} \alpha_k x^k = 0 \qquad \text{for all } x \in [0,1].$$

   Differentiating the equality n times we get $p^{(n)}(x) = n!\alpha_n = 0$. Which implies $\alpha_n = 0$. Therefore $p(x) \equiv 0$ implies $\alpha_k = 0$ for all k.

The linearly independent sets provided in the last two examples are not Hamel bases. This is not a coincidence: we will see later that $\ell^p(\mathbb{K})$ and $C[0,1]$ (as well as many other functional spaces) do not have a *countable* Hamel basis.

Theorem 1.11 *Every vector space has a Hamel basis.*

The proof of this theorem is based on Zorn's Lemma.

We note that in many interesting vector spaces (called normed spaces), a very large number of elements should be included into a Hamel basis in order to enable representation of every element in the form of a finite sum. Then the basis is too large to be useful for the study of the original vector space. A natural idea would be to allow infinite sums in the definition of a basis. In order to use infinite sums we need to define convergence which cannot be done using the axioms of vector spaces only. An additional structure on the vector space should be defined.

2 Normed spaces

2.1 Norms

Definition 2.1 *A norm on a vector space V is a map $\|\cdot\| : V \to \mathbb{R}$ such that for any $x, y \in V$ and any $\alpha \in \mathbb{K}$:*

1. $\|x\| \geq 0$, *and* $\|x\| = 0 \Leftrightarrow x = 0$ *(positive definiteness);*
2. $\|\alpha x\| = |\alpha| \, \|x\|$ *(positive homogeneity);*
3. $\|x + y\| \leq \|x\| + \|y\|$ *(triangle inequality).*

The pair $(V, \|\cdot\|)$ is called a normed space.

In other words, a normed space is a vector space equipped with a norm.

Examples:

1. $\mathbb{R}^n$ with each of the following norms is a normed space:

   (a) $\|x\| = \sqrt{\sum_{k=1}^{n} |x_k|^2}$

   (b) $\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p}$, $1 \leq p < \infty$

   (c) $\|x\|_\infty = \max_{1 \leq k \leq n} |x_k|$.

2. $\ell^p(\mathbb{K})$ is a vector space with the following norm ($1 \leq p < \infty$)

$$\|x\|_{\ell^p} = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p} .$$

3. $\ell^\infty(\mathbb{K})$ is a vector space with the following norm

$$\|x\|_{\ell^\infty} = \sup_{k \in \mathbb{N}} |x_k| .$$

In order to prove the triangle inequality for the ℓ^p norm, we will state and prove several inequalities.

2.2 Four famous inequalities

Lemma 2.2 (Young's inequality) *If* $a,b>0$, $1<p,q<\infty$, $\frac{1}{p}+\frac{1}{q}=1$, *then*

$$ab \le \frac{a^p}{p}+\frac{b^q}{q}\,.$$

Proof: Consider the function $f(t)=\frac{t^p}{p}-t+\frac{1}{q}$ defined for $t\ge 0$. Since $f'(t)=t^{p-1}-1$ vanishes at $t=1$ only, and $f''(t)=(p-1)t^{p-2}\ge 0$, the point $t=1$ is a global minimum for f. Consequently, $f(t)\ge f(1)=0$ for all $t\ge 0$. Now substitute $t=ab^{-q/p}$:

$$f(ab^{-q/p})=\frac{a^p b^{-q}}{p}-ab^{-q/p}+\frac{1}{q}\ge 0\,.$$

Multiplying the inequality by b^q yields Young's inequality. □

Lemma 2.3 (Hölder's inequality) *If* $1\le p,q\le\infty$, $\frac{1}{p}+\frac{1}{q}=1$, $x\in\ell^p(\mathbb{K})$, $y\in\ell^q(\mathbb{K})$, *then*

$$\sum_{j=1}^{\infty}|x_j y_j|\le \|x\|_{\ell^p}\|y\|_{\ell^q}\,.$$

Proof. If $1<p,q<\infty$, we use Young's inequality to get that for any $n\in\mathbb{N}$

$$\sum_{j=1}^{n}\frac{|x_j|}{\|x\|_{\ell^p}}\,\frac{|y_j|}{\|y\|_{\ell^q}}\le\sum_{j=1}^{n}\left(\frac{1}{p}\,\frac{|x_j|^p}{\|x\|_{\ell^p}^p}+\frac{1}{q}\,\frac{|y_j|^q}{\|y\|_{\ell^q}^q}\right)\le\frac{1}{p}+\frac{1}{q}=1$$

Therefore for any $n\in\mathbb{N}$

$$\sum_{j=1}^{n}|x_j y_j|\le \|x\|_{\ell^p}\|y\|_{\ell^q}\,.$$

Since the partial sums are monotonically increasing and bounded above, the series converge and Hölder's inequality follows by taking the limit as $n\to\infty$.

If $p=1$ and $q=\infty$:

$$\sum_{j=1}^{n}|x_j y_j|\le\max_{1\le j\le n}|y_j|\,\sum_{j=1}^{n}|x_j|\le\|x\|_{\ell^1}\|y\|_{\ell^\infty}\,.$$

Therefore the series converges and Hölder's inequality follows by taking the limit as $n\to\infty$. □

Lemma 2.4 (Cauchy-Schwartz inequality) *If* $x,y\in\ell^2(\mathbb{K})$ *then*

$$\sum_{j=1}^{\infty}|x_j y_j|\le\left(\sum_{j=1}^{\infty}|x_j|^2\right)^{1/2}\left(\sum_{j=1}^{\infty}|y_j|^2\right)^{1/2}\,.$$

Proof: This inequality coincides with Hölder's inequality with $p = q = 2$. $\square$

Now we state and prove the triangle inequality for the ℓ^p norm.

Lemma 2.5 (Minkowski's inequality) *If* $x, y \in \ell^p(\mathbb{K})$ *for* $1 \le p \le \infty$ *then* $x + y \in \ell^p(\mathbb{K})$ *and*

$$\|x+y\|_{\ell^p} \le \|x\|_{\ell^p} + \|y\|_{\ell^p} .$$

Proof: If $1 < p < \infty$, define q from $\frac{1}{p} + \frac{1}{q} = 1$. Then using Hölder's inequality (finite sequences belong to ℓ^p with any p) we get[2]

$$\begin{aligned}
\sum_{j=1}^{n} |x_j + y_j|^p &= \sum_{j=1}^{n} |x_j + y_j|^{p-1} |x_j + y_j| \\
&\le \sum_{j=1}^{n} |x_j + y_j|^{p-1} |x_j| + \sum_{j=1}^{n} |x_j + y_j|^{p-1} |y_j| \\
&\le \left(\sum_{j=1}^{n} |x_j + y_j|^{(p-1)q} \right)^{1/q} \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p} \quad \text{(Hölder's inequality)} \\
&\quad + \left(\sum_{j=1}^{n} |x_j + y_j|^{(p-1)q} \right)^{1/q} \left(\sum_{j=1}^{n} |y_j|^p \right)^{1/p} .
\end{aligned}$$

Dividing the inequality by $\left(\sum_{j=1}^{n} |x_j + y_j|^p \right)^{1/q}$ and using that $(p-1)q = p$ and $1 - \frac{1}{q} = \frac{1}{p}$, we get for all n

$$\left(\sum_{j=1}^{n} |x_j + y_j|^p \right)^{1/p} \le \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p} + \left(\sum_{j=1}^{n} |y_j|^p \right)^{1/p} .$$

The series on the right hand side converge to $\|x\|_{\ell^p} + \|y\|_{\ell^p}$. Consequently the series on the left hand side also converge, $x + y \in \ell^p(\mathbb{K})$, and Minkowski's inequality follows by taking the limit as $n \to \infty$.

Exercise: Prove Minkowski's inequality for $p = 1$ and $p = \infty$. $\square$

2.3 Examples of norms on a space of functions

Each of the following formulae defines a norm on $C[0,1]$, the space of all continuous functions on $[0,1]$:

[2] We do not start directly with $n = \infty$ because a priori we do not know convergence for some of the series involved in the proof.

1. the "sup(remum) norm"
$$\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|\,;$$
2. the "L^1 norm"
$$\|f\|_{L^1} = \int_0^1 |f(t)|\,dt\,;$$
3. the "L^2 norm"
$$\|f\|_{L^2} = \left(\int_0^1 |f(t)|^2\,dt\right)^{1/2}.$$

Exercise: Check that each of these formulae defines a norm. For the case of the L^2 norm, you will need a Cauchy-Schwartz inequality for integrals.

Example: Let $k \in \mathbb{N}$. The space $C^k[0,1]$ consists of all continuous real-valued functions which have continuous derivatives up to order k. The norm on $C^k[0,1]$ is defined by
$$\|f\|_{C^k} = \sum_{j=0}^{k} \sup_{t\in[0,1]} |f^{(j)}(t)|\,,$$
where $f^{(j)}$ denotes the derivative of order j.

2.4 Equivalence of norms

We have seen that various different norms can be introduced on a vector space. In order to compare different norms it is convenient to introduce the following equivalence relation.

Definition 2.6 *Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space V are* equivalent *if there are constants $c_1, c_2 > 0$ such that*
$$c_1\|x\|_1 \le \|x\|_2 \le c_2\|x\|_1 \qquad \textit{for all } x \in V.$$
In this case we write $\|\cdot\|_1 \sim \|\cdot\|_2$.

Theorem 2.7 *Any two norms on $\mathbb{R}^n$ are equivalent.*[3]

Example: The norms $\|\cdot\|_{L^1}$ and $\|\cdot\|_\infty$ on $C[0,1]$ are not equivalent.

[3] You already saw this statement in Analysis III and Differentiation in Year 2. The proof is based on the observation that the unit sphere $S \subset \mathbb{R}^n$ is sequentially compact. Then we checked that $f(x) = \|x\|_2/\|x\|_1$ is continuous on S and consequently it is bounded and attains its lower and upper bounds on S. We set $c_1 = \min_S f$ and $c_2 = \max_S f$.

Proof: Consider the sequence of functions $f_n(t) = t^n$ with $n \in \mathbb{N}$. Obviously $f_n \in C[0,1]$. We see that

$$\begin{aligned} \|f_n\|_\infty &= \max_{t\in[0,1]} |t|^n = 1\,, \\ \|f_n\|_{L^1} &= \int_0^1 t^n\,dt = \frac{1}{n+1}\,. \end{aligned}$$

Suppose the norms are equivalent. Then there is a constant $c_2 > 0$ such that for all f_n:

$$\frac{\|f_n\|_\infty}{\|f_n\|_{L^1}} = n+1 \le c_2\,,$$

which is not possible for all n. This contradiction implies the norms are not equivalent.

□

2.5 Linear Isometries

Suppose V and W are normed spaces.

Definition 2.8 *If a linear map $L : V \to W$ preserves norms, i.e. $\|L(x)\| = \|x\|$ for all $x \in V$, it is called a* linear isometry.

This definition implies L is injective, i.e., $L : V \to L(V)$ is bijective, but it does not imply $L(V) = W$, i.e., L is not necessarily invertible. Note that sometimes the invertibility property is included into the definition of the isometry. Finally, in Metric Spaces the word "isometry" is used to denote distance-preserving transformations.

Definition 2.9 *We say that two normed spaces are* isometrically isomorphic *(or simply* isometric*), if there is an invertible linear isometry between them.*

A linear invertible map can be used to "pull back" a norm as follows.

Proposition 2.10 *Let $(V, \|\cdot\|_V)$ be a normed space, W a vector space, and $L : W \to V$ a linear isomorphism. Then*

$$\|x\|_W := \|L(x)\|_V$$

defines a norm on W.

Proof: For any $x, y \in V$ and any $\alpha \in \mathbb{K}$ we have:

$$\begin{aligned} \|x\|_W &= \|L(x)\|_V \ge 0\,, \\ \|\alpha x\|_W &= \|L(\alpha x)\|_V = |\alpha|\,\|L(x)\|_V = |\alpha|\,\|x\|_W\,. \end{aligned}$$

If $\|x\|_W = \|L(x)\|_V = 0$, then $L(x) = 0$ due to non-degeneracy of the norm $\|\cdot\|_V$. Since L is invertible, we get $x = 0$. Therefore $\|\cdot\|_W$ is non-degenerate.

Finally, the triangle inequality follows from the triangle inequality for $\|\cdot\|_V$:

$$\|x+y\|_W = \|L(x)+L(y)\|_V \leq \|L(x)\|_V + \|L(y)\|_V = \|x\|_W + \|y\|_W .$$

Therefore, $\|\cdot\|_W$ is a norm. □

Note that in the proposition the new norm is introduced in such a way that $L : (W, \|\cdot\|_W) \to (V, \|\cdot\|_V)$ is a linear isometry.

Let V be a finite dimensional vector space and $n = \dim V$. We have seen that V is linearly isomorphic to $\mathbb{K}^n$. Then the proposition implies the following statements.

Corollary 2.11 *Any finite dimensional vector space V can be equipped with a norm.*

Corollary 2.12 *Any n-dimensional normed space V is isometrically isomorphic to $\mathbb{K}^n$ equipped with a suitable norm.*

Since any two norms on $\mathbb{R}^n$ (and therefore on $\mathbb{C}^n$) are equivalent we also get the following statement.

Theorem 2.13 *Let V be a finite-dimensional vector space. Then all norms on V are equivalent.*

3 Convergence in a normed space

3.1 Definition and examples

The norm on a vector space V can be used to measure distances between points $x, y \in V$. So we can define the limit of a sequence.

Definition 3.1 *A sequence $(x_n)_{n=1}^{\infty}$, $x_n \in V$, $n \in \mathbb{N}$,* converges to a limit *$x \in V$ if for any $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that*

$$\|x_n - x\| < \varepsilon \qquad \text{for all } n > N.$$

Then we write $x_n \to x$.

We note that the sequence of vectors $x_n \to x$ if and only if the sequence of non-negative real numbers $\|x_n - x\| \to 0$.

Proposition 3.2 *The limit of a convergent sequence is unique.*

Exercise: Prove it.

Proposition 3.3 *Any convergent sequence is bounded.*

Exercise: Prove it.

Proposition 3.4 *If x_n converges to x, then $\|x_n\| \to \|x\|$.*

Exercise: Prove it.

It is possible to check convergence of a sequence of real numbers without actually finding its limit: it is sufficient to check that it satisfies the following definition:

Definition 3.5 (Cauchy sequence) *A sequence $(x_n)_{n=1}^{\infty}$ in a normed space V is Cauchy if for any $\varepsilon > 0$ there is an N such that*

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

Theorem 3.6 *A sequence of real numbers converges* iff *it is Cauchy.*

Proposition 3.7 *Any convergent sequence is Cauchy.*

Exercise: Prove it.

Proposition 3.8 *Any Cauchy sequence is bounded.*

Exercise: Prove it.

Example: Consider the sequence $f_n \in C[0,1]$ defined by $f_n(t) = t^n$.

1. $f_n \to 0$ in the L^1 norm.

   Proof: We have already computed the norms:

   $$\|f_n\|_{L^1} = \frac{1}{n+1} \to 0\,.$$

   Consequently, $f_n \to 0$. □

2. f_n does not converge in the sup norm.

   Proof: If $m > 2n \geq 1$ then

   $$f_n(2^{-1/n}) - f_m(2^{-1/n}) = \frac{1}{2} - \frac{1}{2^{m/n}} \geq \frac{1}{4}\,.$$

   Consequently (f_n) is not Cauchy in the sup norm and hence not convergent.

This example shows that the convergence in the L^1 norm does not imply the pointwise convergence and, as a results, does not imply the convergence in the sup norm (often called *the uniform convergence*). Note that in contrast to the uniform and L^1 convergences the notion of pointwise convergence is not based on a norm on the space of continuous function.

Exercise: The pointwise convergence does not imply the L^1 convergence.

Hint: Construct f_n with a very small support but make the maximum of f_n very large to ensure that $\|f_n\|_{L^1} > n$. Therefore f_n is not bounded in the L_1 norm, hence not convergent.

We can also make $\operatorname{supp} f_n \cap \operatorname{supp} f_m = \emptyset$ for all m, n such that $n \neq m$. Then for any t there is at most one n such that $f_n(t) \neq 0$. The last property guarantees pointwise convergence: $f_n(t) \to 0$ for any t. □

Proposition 3.9 *If $f_n \in C[0,1]$ for all $n \in \mathbb{N}$ and $f_n \to f$ in the sup norm, then $f_n \to f$ in the L^1 norm, i.e.,*

$$\|f_n - f\|_\infty \to 0 \qquad \Longrightarrow \qquad \|f_n - f\|_{L^1} \to 0\,.$$

Proof:

$$0 \leq \|f_n - f\|_{L^1} = \int_0^1 |f_n(t) - f(t)|\,dt \leq \sup_{0\leq t\leq 1} |f_n(t) - f(t)| = \|f_n - f\|_\infty \to 0\,.$$

Therefore $\|f_n - f\|_{L^1} \to 0$. □

We have seen that different norms may lead to different conclusions about convergence of a given sequence but sometime convergence in one norm implies convergence in another one. The following lemma shows that equivalent norms give rise to the same notion of convergence.

Lemma 3.10 *Suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms on a vector space V. Then for any sequence (x_n):*

$$\|x_n - x\|_1 \to 0 \qquad \Leftrightarrow \qquad \|x_n - x\|_2 \to 0\,.$$

Proof: Since the norms are equivalent, there are constant $c_1, c_2 > 0$ such that

$$0 \le c_1\|x_n - x\|_1 \le \|x_n - x\|_2 \le c_2\|x_n - x\|_1$$

for all n. Then $\|x_n - x\|_2 \to 0$ implies $\|x_n - x\|_1 \to 0$, and vice versa. □

3.2 Topology on a normed space

We say that a collection $\mathscr{T}$ of subsets of V is a *topology on V* if it satisfies the following properties:

1. $\emptyset, V \in \mathscr{T}$;
2. any finite intersection of elements of $\mathscr{T}$ belongs to $\mathscr{T}$;
3. any union of elements of $\mathscr{T}$ belongs to $\mathscr{T}$.

A set equipped with a topology is called a *topological space*. The elements of $\mathscr{T}$ are called *open sets*. The topology can be used to define a convergent sequence and continuous function.

We note that each norm on V can be used to define a topology on V, i.e., to define the notion of an open set.

Definition 3.11 *A subset $X \subset V$ is* open*, if for any $x \in X$ there is $\varepsilon > 0$ such that the ball of radius ε centred around x belongs to X:*

$$B(x,\varepsilon) = \{y \in V : \|y - x\| < \varepsilon\} \subset X\,.$$

Example: In any normed space V:

1. The unit ball centred around the zero, $B_0 = \{x : \|x\| < 1\}$, is open.
2. Any open ball $B(x,\varepsilon)$ is open.
3. V is open.
4. The empty set is open.

It is not too difficult to check that the collection of open sets defines a topology on V. You can easily check from the definition that equivalent norms generate the same topology, i.e., open sets are exactly the same. The notion of convergence can be defined in terms of the topology.

Definition 3.12 *An* open neighbourhood of *x is an open set which contains x.*

Lemma 3.13 *A sequence $x_n \to x$ if and only if for any open neighbourhood X of x there is $N \in \mathbb{N}$ such that $x_n \in X$ for all $n > N$.*

Proof: ($\Longrightarrow$). Let $x_n \to x$. Take any open X such that $x \in X$. Then there is $\varepsilon > 0$ such that $B(x, \varepsilon) \subset X$. Since the sequence converges there is N such that $\|x_n - x\| < \varepsilon$ for all $n > N$. Then $x_n \in B(x, \varepsilon) \subset X$ for the same values of n.

($\Longleftarrow$). Take any $\varepsilon > 0$. The ball $B(x, \varepsilon)$ is open, therefore there is N such that $x_n \in B(x, \varepsilon)$ for all $n > N$. Hence $\|x_n - x\| < \varepsilon$ and $x_n \to x$. □

3.3 Closed sets

Definition 3.14 *A set $X \subset V$ is closed if its complement $V \setminus X$ is open.*

Example: In any normed space V:

1. The unit sphere $S = \{ x : \|x\| = 1 \}$ is closed
2. Any closed ball
$$\overline{B}(x, \varepsilon) = \{ y \in V : \|y - x\| \leq \varepsilon \}$$
is closed.
3. V is closed.
4. The empty set is closed.

Lemma 3.15 *A subset $X \subset V$ is closed if and only if any convergent sequence in X has its limit in X.*

Proof: The proof literally repeats the proof given in Differentiation. □

Definition 3.16 *We say that a subset $L \subset V$ is a* linear subspace*, if it is a vector space itself, i.e., if $x_1, x_2 \in L$ and $\lambda \in \mathbb{K}$ imply $x_1 + \lambda x_2 \in L$.*

Proposition 3.17 *Any finite dimensional linear subspace L of a normed space V is closed.*

Proof: Since L is finite-dimensional, it has a finite Hamel basis
$$E = \{ e_1, e_2, \ldots, e_n \} \subset L$$
such that $L = \mathrm{Span}(E)$. Suppose L is not closed, then by Lemma 3.15 there is a convergent sequence $x_k \to x^*$, $x_k \in L$ but $x^* \in V \setminus L$. Then x^* is linearly independent from E (otherwise it would belong to L). Consequently
$$\tilde{E} = \{ e_1, e_2, \ldots, e_n, x^* \}$$

is a Hamel basis in $\tilde{L} = \mathrm{Span}(\tilde{E})$. In this basis, the components of x_k are given by $(\alpha_1^k, \ldots, \alpha_n^k, 0)$ and x^* corresponds to the vector $(0, \ldots, 0, 1)$. We get in the limit as $k \to \infty$

$$(\alpha_1^k, \ldots, \alpha_n^k, 0) \to (0, \ldots, 0, 1),$$

which is obviously impossible. Therefore L is closed.[4] □

Example: The subspace of polynomial functions is linear but not closed in $C[0,1]$ equipped with the sup norm.

3.4 Compactness

Definition 3.18 (sequential compactness) *A subset K of a normed space $(V, \|\cdot\|_V)$ is (sequentially) compact if any sequence $(x_n)_{n=1}^\infty$ with $x_n \in K$ has a convergent subsequence $x_{n_j} \to x^*$ with $x^* \in K$.*

Proposition 3.19 *A compact set is closed and bounded.*

Theorem 3.20 *A subset of $\mathbb{R}^n$ is compact iff it is closed and bounded.*

Corollary 3.21 *A subset of a finite-dimensional vector space is compact iff it is closed and bounded.*

Example: The unit sphere in $\ell^p(\mathbb{K})$ is closed, bounded but not compact.

Proof: Take the sequence e_j such that

$$e_j = (0, \ldots, 0, \underbrace{1}_{j^{\text{th}} \text{ place}}, 0, \ldots).$$

We note that $\|e_j - e_k\|_{\ell^p} = 2^{1/p}$ for all $j \neq k$. Consequently, $(e_j)_{j=1}^\infty$ does not have any convergent subsequence, hence S is not compact. □

Lemma 3.22 (Riesz' Lemma) *Let X be a normed vector space and Y be a closed linear subspace of X such that $Y \neq X$ and $\alpha \in \mathbb{R}$, $0 < \alpha < 1$. Then there is $x_\alpha \in X$ such that $\|x_\alpha\| = 1$ and $\|x_\alpha - y\| > \alpha$ for all $y \in Y$.*

Proof: Since $Y \subset X$ and $Y \neq X$ there is $x \in X \setminus Y$. Since Y is closed, $X \setminus Y$ is open and therefore

$$d := \inf\{\, \|x - y\| : y \in Y \,\} > 0\,.$$

[4] This proof implicitly uses equivalence of norms on $\mathbb{R}^{n+1}$ to establish that convergence in the norm obtained by restricting the original norm $\|\cdot\|_V$ onto $\tilde{L}$ implies convergence of the components of the vectors.

Since $\alpha^{-1} > 1$ there is a point $z \in Y$ such that $\|x - z\| < d\alpha^{-1}$. Let $x_\alpha = \frac{x-z}{\|x-z\|}$. Then $\|x_\alpha\| = 1$ and for any $y \in Y$,

$$\|x_\alpha - y\| = \left\| \frac{x-z}{\|x-z\|} - y \right\| = \frac{\big\|x - (z + \|x-z\|y)\big\|}{\|x-z\|} > \frac{d}{d\alpha^{-1}} = \alpha\,,$$

as $z + \|x - z\|y \in Y$ because Y is a linear subspace. □

Theorem 3.23 *A normed space is finite dimensional iff the unit sphere is compact.*

Proof: Bolzano-Weierstrass Theorem and Lemma 3.15 imply that in a finite dimensional normed space the unit sphere is compact (the unit sphere is bounded and closed).

So we only need to show that if the unit sphere $S \subset B$ is sequentially compact, then the normed space V is finite dimensional. Indeed, if V is infinite dimensional, then Riesz' Lemma can be used to construct an infinite sequence of $x_n \in S$ such that $\|x_n - x_m\| > \alpha > 0$ for all $m \neq n$. This sequence does not have a convergent subsequence (none of the subsequences is Cauchy) and therefore S is not compact.

We construct x_n inductively. Fix $\alpha \in (0,1)$ and take any $x_1 \in S$.

Suppose that for some $n \geq 1$ we have found $E_n = \{x_1, \ldots, x_n\}$ such that $x_k \in S$ and $\|x_l - x_k\| > \alpha$ for all $1 \leq k, l \leq n$, $k \neq l$ (note that the second property is automatically fulfilled for $n = 1$). The linear subspace $Y_n = \mathrm{Span}(E_n)$ is n-dimensional and hence closed (see Proposition 3.17). Since X is infinite dimensional $Y_n \neq X$. Then Riesz' Lemma implies that there is $x_{n+1} \in S$ such that $\|x_{n+1} - x_k\| > \alpha$ for all $1 \leq k \leq n$.

Repeating this argument we generate x_n for all $n \in \mathbb{N}$. □

4 Banach spaces

4.1 Completeness: Definition and examples

Definition 4.1 (Banach space) *A normed space V is called* complete *if any Cauchy sequence in V converges to a limit in V. A complete normed space is called a* Banach space.

Theorem 4.2 *Every finite-dimensional normed space is complete.*

Proof: Theorem 3.6 implies that $\mathbb{R}$ is complete, i.e., every Cauchy sequence of numbers has a limit.

Now let V be a real vector space, $\dim V = n < \infty$. Take any basis in V. Then a sequence of vectors in V converges iff each component of the vectors converges, and a sequence of vectors is Cauchy iff each component is Cauchy. Therefore each component has a limit, and those limits constitute the limit vector for the original sequence. Hence V is complete.

Considering $\mathbb{C}$ as a real vector space we conclude that it is also complete. Therefore, any finite-dimensional complex vector space V is also complete. □

In particular, $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete.

Theorem 4.3 (ℓ^p is a Banach space) *The space $\ell^p(\mathbb{K})$ equipped with the standard ℓ^p norm is complete.*

Proof: Suppose that $x^k = (x_1^k, x_2^k, \ldots) \in \ell^p(\mathbb{K})$ is Cauchy. Then for every $\varepsilon > 0$ there is N such that

$$\|x^m - x^n\|_{\ell^p} = \sum_{j=1}^{\infty} |x_j^m - x_j^n|^p < \varepsilon$$

for all $m,n > N$. Consequently, for each $j \in \mathbb{N}$ the sequence x_j^k is Cauchy, and the completeness of $\mathbb{K}$ implies that there is $a_j \in \mathbb{K}$ such that

$$x_j^k \to a_j$$

as $k \to \infty$. Let $a = (a_1, a_2, \ldots)$. First we note that for any $M \geq 1$ and $m,n > N$:

$$\sum_{j=1}^{M} |x_j^m - x_j^n|^p \leq \sum_{j=1}^{\infty} |x_j^m - x_j^n|^p < \varepsilon\,.$$

Taking the limit as $n \to \infty$ we get

$$\sum_{j=1}^{M} |x_j^m - a_j|^p \leq \varepsilon\,.$$

This holds for any M, so we can take the limit as $M \to \infty$:

$$\sum_{j=1}^{\infty} |x_j^m - a_j|^p \le \varepsilon .$$

We conclude that $x^m - a \in \ell^p(\mathbb{K})$. Since $\ell^p(\mathbb{K})$ is a vector space and $x^m \in \ell^p(\mathbb{K})$, then $a \in \ell^p(\mathbb{K})$. Moreover, $\|x^m - a\|_{\ell^p} < \varepsilon$ for all $m > N$. Consequently $x^m \to a$ in $\ell^p(\mathbb{K})$ with the standard norm, and so $\ell^p(\mathbb{K})$ is complete. □

Theorem 4.4 **(C is a Banach space)** *The space $C[0,1]$ equipped with the sup norm is complete.*

Proof: Let f_k be a Cauchy sequence. Then for any $\varepsilon > 0$ there is N such that

$$\sup_{t\in[0,1]} |f_n(t) - f_m(t)| < \varepsilon$$

for all $m,n > N$. In particular, $f_n(t)$ is Cauchy for any fixed t and consequently has a limit. Set

$$f(t) = \lim_{n\to\infty} f_n(t) .$$

Let's prove that $f_n(t) \to f(t)$ uniformly in t. Indeed, we already know that

$$|f_n(t) - f_m(t)| < \varepsilon$$

for all $n,m > N$ and all $t \in [0,1]$. Taking the limit as $m \to \infty$ we get

$$|f_n(t) - f(t)| \le \varepsilon$$

for all $n > N$ and all $t \in [0,1]$. Therefore f_n converges uniformly:

$$\|f_n - f\|_\infty = \sup_{t\in[0,1]} |f_n(t) - f(t)| < \varepsilon .$$

for all $n > N$. The uniform limit of a sequence of continuous functions is continuous. Consequently, $f \in C[0,1]$ which completes the proof of completeness. □

Example: The space $C[0,2]$ equipped with the L^1 norm is not complete.

Proof: Consider the following sequence of functions:

$$f_n(t) = \begin{cases} t^n & \text{for } 0 \le t \le 1, \\ 1 & \text{for } 1 \le t \le 2 . \end{cases}$$

This is a Cauchy sequence in the L^1 norm. Indeed for any $n < m$:

$$\|f_n - f_m\|_{L^1} = \int_0^1 (t^n - t^m)\, dt = \frac{1}{n+1} - \frac{1}{m+1} < \frac{1}{n+1} ,$$

and consequently for any $m,n > N$

$$\|f_n - f_m\|_{L^1} < \frac{1}{N}.$$

Now let us show that f_n do not converge to a continuous function in the L^1 norm. Indeed, suppose such a limit exists and call it f. Then

$$\|f_n - f\|_{L^1} = \int_0^1 |t^n - f(t)|\,dt + \int_1^2 |1 - f(t)|\,dt \to 0.$$

Since

$$|f(t)| - |t^n| \le |t^n - f(t)| \le |f(t)| + |t^n|$$

implies that

$$\int_0^1 |f(t)|\,dt - \int_0^1 t^n\,dt \le \int_0^1 |t^n - f(t)|\,dt \le \int_0^1 |f(t)|\,dt + \int_0^1 t^n\,dt,$$

we have $\int_0^1 |t^n - f(t)|\,dt \to \int_0^1 |f(t)|\,dt$ as $n \to \infty$ and consequently

$$\int_0^1 |f(t)|\,dt + \int_1^2 |1 - f(t)|\,dt = 0.$$

As f is assumed to be continuous, it follows

$$f(t) = \begin{cases} 0, & 0 < t < 1, \\ 1, & 1 < t < 2. \end{cases}$$

We see that the limit function f cannot be continuous. This contradiction implies that $C[0,2]$ is not complete with respect to the L^1 norm.

4.2 The completion of a normed space

A normed space may be incomplete. However, every normed space X can be considered as a subset of a larger Banach space $\hat{X}$. The minimal among these spaces[5] is called the *completion of* X.

Informally we can say that $\hat{X}$ consists of limit points of all Cauchy sequences in X. Of course, every point $x \in X$ is a limit point of the constant sequence ($x_n = x$ for all $n \in \mathbb{N}$) and therefore $X \subset \hat{X}$. If X is not complete, some of the limit points are not in X, so $\hat{X}$ is larger then the original set X.

Definition 4.5 (dense set) *We say that a subset $X \subset V$ is* dense *in V if for any $v \in V$ and any $\varepsilon > 0$ there is $x \in X$ such that $\|x - v\| < \varepsilon$.*

[5] In this context "minimal" means that if any other space $\tilde{X}$ has the same property, then the minimal $\hat{X}$ is isometric to a subspace of $\tilde{X}$. It turns out that this property can be achieved by requiring X to be dense in $\hat{X}$.

Note that X is dense in V iff for every point $v \in V$ there is a sequence $x_n \in X$ such that $x_n \to v$.

Theorem 4.6 *Let $(X, \|\cdot\|_X)$ be a normed space. Then there is a complete normed space $(\mathscr{X}, \|\cdot\|_{\mathscr{X}})$ and a linear map $i : X \to \mathscr{X}$ such that i is an isometrical isomorphism between $(X, \|\cdot\|_X)$ and $(i(X), \|\cdot\|_{\mathscr{X}})$, and $i(X)$ is dense in $\mathscr{X}$.*

Moreover, $\mathscr{X}$ is unique up to isometry, i.e., if there is another complete normed space $(\tilde{\mathscr{X}}, \|\cdot\|_{\tilde{\mathscr{X}}})$ with these properties, then $\mathscr{X}$ and $\tilde{\mathscr{X}}$ are isometrically isomorphic.

Proof: The proof is relatively long so we break it into a sequence of steps.

Construction of $\mathscr{X}$. Let $\mathscr{Y}$ be the set of all Cauchy sequences in X. We say that two Cauchy sequences $\mathbf{x} = (x_n)_{n=1}^\infty$, $x_n \in X$, and $\mathbf{y} = (y_n)_{n=1}^\infty$, $y_n \in X$, are equivalent, and write $\mathbf{x} \sim \mathbf{y}$, if

$$\lim_{n\to\infty} \|x_n - y_n\|_X = 0 .$$

Let $\mathscr{X}$ be the space of all equivalence classes in $\mathscr{Y}$, i.e., it is the factor space: $\mathscr{X} = \mathscr{Y}/\sim$. The elements of $\mathscr{X}$ are collections of equivalent Cauchy sequences from X. We will use $[\mathbf{x}]$ to denote the equivalence class of $\mathbf{x}$.

Exercises: Show that $\mathscr{X}$ is a vector space.

Norm on $\mathscr{X}$. For an $\eta \in \mathscr{X}$ take any representative $\mathbf{x} = (x_n)_{n=1}^\infty$, $x_n \in X$, of the equivalence class η. Then the equation

$$\|\eta\|_{\mathscr{X}} = \lim_{n\to\infty} \|x_n\|_X . \tag{4.1}$$

defines a norm on $\mathscr{X}$. Indeed:

1. Equation (4.1) defines a function $\mathscr{X} \to \mathbb{R}$, i.e., for any $\eta \in \mathscr{X}$ and any representative $\mathbf{x} \in \eta$ the limit exists and is independent from the choice of the representative. (Exercise)

2. The function defined by (4.1) satisfies the axioms of norm. (Exercise)

Definition of $i : X \to \mathscr{X}$. For any $x \in X$ let

$$i(x) = [(x, x, x, x, \dots)]$$

(the equivalence class of the constant sequence). Obviously, i is a linear isometry, and it is a bijection $X \to i(X)$. Therefore the spaces X and $i(X)$ are isometrically isomorphic.

Completeness of $\mathscr{X}$. Let $\left(\eta^{(k)}\right)_{k=1}^\infty$ be a Cauchy sequence in $(\mathscr{X}, \|\cdot\|_{\mathscr{X}})$. For every $k \in \mathbb{N}$ take a representative $\mathbf{x}^{(k)} \in \eta^{(k)}$. Note that $\mathbf{x}^{(k)} \in \mathscr{Y}$ is a Cauchy sequence in the space $(X, \|\cdot\|_X)$. Then there is a strictly monotone sequence of integers n_k such that

$$\left\|x_j^{(k)} - x_l^{(k)}\right\|_X \le \frac{1}{k} \qquad \text{for all } j, l \ge n_k . \tag{4.2}$$

Now consider the sequence $\mathbf{x}^*$ defined by

$$\mathbf{x}^* = \left(x_{n_k}^{(k)}\right)_{k=1}^{\infty}.$$

Next we will check that $\mathbf{x}^*$ is Cauchy, and consider its equivalence class $\eta^* = [\mathbf{x}^*] \in \mathscr{X}$. Then we will prove that $\eta^{(k)} \to \eta^*$ in $(\mathscr{X}, \|\cdot\|_{\mathscr{X}})$.

The sequence $\mathbf{x}^*$ *is Cauchy.* Since the sequence of $\eta^{(k)}$ is Cauchy, for any $\varepsilon > 0$ there is M_ε such that

$$\lim_{n\to\infty} \|x_n^{(k)} - x_n^{(l)}\|_X = \|\eta^{(k)} - \eta^{(l)}\|_{\mathscr{X}} < \varepsilon \qquad \text{for all } k,l > M_\varepsilon.$$

Consequently, for every $k,l > M_\varepsilon$ there is $N_\varepsilon^{k,l}$ such that

$$\|x_n^{(k)} - x_n^{(l)}\|_X < \varepsilon \qquad \text{for all } n > N_\varepsilon^{k,l}. \tag{4.3}$$

Then fix any $\varepsilon > 0$. If $j,l > \frac{3}{\varepsilon}$ and $m > \max\{n_j, n_l, N_{\varepsilon/3}^{k,l}\}$ we have

$$\begin{aligned}
\|x_j^* - x_l^*\|_X &= \|x_{n_j}^{(j)} - x_{n_l}^{(l)}\|_X \\
&\le \|x_{n_j}^{(j)} - x_m^{(j)}\|_X + \|x_m^{(j)} - x_m^{(l)}\|_X + \|x_m^{(l)} - x_{n_l}^{(l)}\|_X \\
&< \frac{1}{j} + \frac{\varepsilon}{3} + \frac{1}{l} < \varepsilon
\end{aligned}$$

where we used (4.3) and (4.2). Therefore $\mathbf{x}^*$ is Cauchy and $\eta = [x^*] \in \mathscr{X}$.

The sequence $\eta^{(k)} \to [\mathbf{x}^*]$. Indeed, take any $\varepsilon > 0$ and $k > 3\varepsilon^{-1}$, then

$$\begin{aligned}
\|\eta^{(k)} - \eta^*\|_{\mathscr{X}} &= \lim_{j\to\infty} \|x_j^{(k)} - x_j^*\|_X = \lim_{j\to\infty} \|x_j^{(k)} - x_{n_j}^{(j)}\|_X \\
&\le \lim_{j\to\infty} \left(\|x_j^{(k)} - x_{n_k}^{(k)}\|_X + \|x_{n_k}^{(k)} - x_{n_j}^{(j)}\|_X\right) \\
&\le \frac{1}{k} + \varepsilon < 2\varepsilon
\end{aligned}$$

Therefore $\eta^{(k)} \to \eta^*$.

We have proved that any Cauchy sequence in $\mathscr{X}$ has a limit in $\mathscr{X}$, so $\mathscr{X}$ is complete.

Density of $i(X)$ in $\mathscr{X}$. Take an $\eta \in \mathscr{X}$ and let $\mathbf{x} \in \eta$. Take any $\varepsilon > 0$. Since $\mathbf{x}$ is Cauchy, there is N_ε such that $\|x_m - x_k\|_X < \varepsilon$ for all $k,m > N_\varepsilon$. Then

$$\|\eta - i(x_k)\|_{\mathscr{X}} = \lim_{m\to\infty} \|x_m - x_k\|_X \le \varepsilon.$$

Therefore $i(X)$ is dense in $\mathscr{X}$.

Uniqueness of $\mathscr{X}$ up to isometry. I will not show you the details of the proof. The proof uses the fact that $i(X)$ is isometrically isomorphic to $\tilde{i}(X)$ (since each one is isometrically isomorphic to X). Moreover, $i(X)$ is dense in $\mathscr{X}$ and $\tilde{i}(X)$ is dense in $\tilde{\mathscr{X}}$. In order to complete the proof one has to show that an isometry between two dense subsets can be extended to an isometry between the sets. □

4.3 Examples

The theorem provides an explicit construction for the completion of a normed space. Often this description is not sufficiently convenient and a more direct description is desirable.

1. **Example:** Consider the space $P[0,1]$ of all polynomial functions restricted to the interval $[0,1]$ and equip this space with the sup norm. This space is not complete. On the other hand any polynomial is continuous, and therefore $P[0,1]$ can be considered as a subspace of $C[0,1]$ which is complete. The Weierstrass approximation theorem states that any continuous function on $[0,1]$ can be uniformly approximated by polynomials. In other words, the polynomials are dense in $C[0,1]$. Then Theorem 4.6 implies that the completion of $P[0,1]$ is isometrically isomorphic to $C[0,1]$ equipped with the sup norm.

2. **Example:** Let $\ell_f(\mathbb{K})$ be the space of all sequences which have only a finite number of non-zero elements. This space is not complete in the ℓ^p norm. The completion of $\ell_f(\mathbb{K})$ in the ℓ^p norm is isometric to $\ell^p(\mathbb{K})$.

   Indeed, we have already seen that $\ell^p(\mathbb{K})$ is complete. So in order to prove the claim you only need to check that $\ell_f(\mathbb{K})$ is dense in $\ell^p(\mathbb{K})$.

   We see that the completion of a space depends both on the space and on the norm.

3. **Example:** $L^1(0,1)$ is the completion of $C[0,1]$ in the L^1 norm.

   According to this definition any $f \in L^1(0,1)$ is an equivalence class of a Cauchy sequence $f_n \in C[0,1]$ equipped with the L^1 norm. The norm of f is defined by

$$\|f\|_{L^1} = \lim_{n\to\infty} \|f_n\|_{L^1} = \lim_{n\to\infty} \int_0^1 |f_n(t)|\, dt\,.$$

   In spite of the fact that f can be considered as a limit of the sequence of continuous functions f_n in the L^1 norm, we cannot define the value of $f(t)$ for a given t as the limit of $f_n(t)$, because the limit may not exist or may depend on the choice of the representative in the equivalence class.

   In spite of that we can define the integral of f by setting

$$\int_0^1 f(t)\, dt := \lim_{n\to\infty} \int_0^1 f_n(t)\, dt\,. \tag{4.4}$$

   Indeed, the limit exists: the sequence of the integrals is Cauchy as

$$\begin{aligned} \left| \int_0^1 f_n(t)\, dt - \int_0^1 f_m(t)\, dt \right| &= \left| \int_0^1 \big(f_n(t) - f_m(t)\big)\, dt \right| \\ &\le \int_0^1 |f_n(t) - f_m(t)|\, dt = \|f_n - f_m\|_{L^1}\,, \end{aligned}$$

hence convergent. It is not difficult to check that the limit does not depend from the choice of a representative in the equivalence class f.

Obviously, if f_n is a constant sequence, i.e., $f_n(t) = f_0(t)$, $t \in [0,1]$, for all $n \in \mathbb{N}$, then $\int_0^1 f(t) = \int_0^1 f_0(t)\,dt$. Therefore this definition can be considered as an extension of the classical Riemann's integral used for continuous functions.

Taking into account the new definition of the integral we can write

$$\|f\|_{L^1} = \int_0^1 |f(t)|\,dt\,.$$

Then we recall that the norm is non-degenerate, therefore $f = g$ for $f, g \in L^1(0,1)$ if and only if $\|f-g\|_{L^1} = 0$, i.e.,

$$\int_0^1 |f(t)-g(t)|\,dt = 0\,,$$

where the integral should be interpreted in the sense of the new definition. Note that f and g are not continuous and consequently the equality above cannot be used to deduce that $f(t) = g(t)$ for all t.

Nevertheless a more direct description of $L^1(0,1)$ is possible: The space $L^1(0,1)$ is isometrically isomorphic to the space of equivalence classes of Lebesgue integrable functions on $(0,1)$: the functions f and g are equivalent if

$$\int_0^1 |f(t)-g(t)|\,dt = 0\,.$$

Here the integral should be considered as the Lebesgue integral. We note that it coincides with the definition provided above but its construction is more direct. Since the notion of the Lebesgue integration is very important for the functional analysis and its applications we will discuss it in more details in the next few lectures. A more detailed study of this topics is a part of MA359 Measure Theory module.

5 Lebesgue spaces

The exposition of the Lebesgue integral is based on the book H.A.Priesly, *Introduction to integration*, Oxford Sc.Publ., 1997, 306 p.

5.1 Integrable functions

Integrals of step functions

We say that $\varphi : \mathbb{R} \to \mathbb{R}$ is a *step function* if it is piecewise constant on a finite number of intervals, i.e., it can be represented as a finite sum

$$\varphi(x) = \sum_{j=1}^{n} c_j \chi_{I_j}(x)\,,$$

where $c_j \in \mathbb{R}$, $I_j \subset \mathbb{R}$ is an interval, $I_j \cap I_k = \emptyset$ if $k \neq j$, and χ_I is the characteristic function of I:

$$\chi_I(x) = \begin{cases} 1, & x \in I\,, \\ 0, & x \notin I\,. \end{cases}$$

We note that the intervals are allowed to be of any of the four possible types (e.g. (a,b), $[a,b)$, $(a,b]$ or $[a,b]$).

We define the integral of a step function φ by

$$\int \varphi := \sum_{j=1}^{n} c_j |I_j|,$$

where $|I_j|$ is the length of I_j. We note that this sum equals to the Riemann integral which you studied in Year 1, i.e., $\int \varphi$ is the "algebraic" area under the graph of the step function φ (the area is counted negative on those intervals where $\varphi(x) < 0$).

Sets of measure zero

Definition 5.1 *We say that a set $A \subset \mathbb{R}$ has measure zero if for any $\varepsilon > 0$ there is an (at most countable) collection of intervals that cover A and whose total length is less than ε:*

$$A \subset \bigcup_{j=1}^{\infty} [a_j, b_j] \qquad \text{and} \qquad \sum_{j=1}^{\infty} (b_j - a_j) < \varepsilon\,.$$

Exercise: Show that a countable union of measure zero sets has measure zero. Hint: for A_n choose a cover with $\varepsilon_n = \varepsilon/2^n$.

Examples. The set $\mathbb{Q}$ of all rational numbers has measure zero. The Cantor set has measure zero.

Definition 5.2 *A property is said to hold "almost everywhere" or "for almost every x" (and abbreviated to "a.e.") if the set of points at which the property does not hold has measure zero.*

Almost everywhere convergence

Theorem 5.3 *Let $(\varphi_n(x))_{n=1}^{\infty}$ be an increasing sequence of step functions ($\varphi_{n+1}(x) \geq \varphi_n(x)$ for all x) such that $\int \varphi_n < K$. Then $\varphi_n(x)$ converges for a.e. x.*

Proof: First note that an increasing sequence of numbers has a limit if and only if it is bounded from above. So in order to prove the theorem it is sufficient to show that the set

$$E = \{ x \in \mathbb{R} : \varphi_n(x) \to +\infty \}$$

has measure zero.

Without loosing in generality, we can assume that $\varphi_n(x) \geq 0$.[6] Let us define the set

$$E_{n,m} = \{ x : \varphi_n(x) > m \} .$$

This set is a finite union of intervals. Indeed, $\varphi_n(x) = \sum_j c_j^{(n)} \chi_{I_j^{(n)}}(x)$ is a step function. Let $\mathscr{I}_{n,m} = \{ j : c_j^{(n)} > m \}$. Then

$$E_{n,m} = \bigcup_{j \in \mathscr{I}_{n,m}} I_j^{(n)} .$$

The total length of those intervals is less than K/m. Indeed, since $c_j^{(n)} \geq 0$ for all j and $c_j^{(n)} > n$ for $j \in \mathscr{I}_n$,

$$K > \int \varphi_n = \sum_j c_j^{(n)} |I_j^{(n)}| > m \sum_{j \in \mathscr{I}_{n,m}} |I_j^{(n)}| .$$

Finally, $E \subset E_m = \bigcup_{l=1}^{\infty} E_{l,m}$ for every m. Since the sequence φ_n is increasing, $E_{n,m} \subset E_{n+1,m}$. Moreover $E_{n+1,m} \setminus E_{n,m}$ consists of a finite number of intervals. Then

$$E_m = \bigcup_{l=1}^{\infty} E_{l,m} \setminus E_{l-1,m}$$

is at most countable union of intervals (we denote $E_{0,m} = \emptyset$). Since $E_{n,m} = \bigcup_{l=1}^{n} E_{l,m} \setminus E_{l-1,m}$ and the total length of the intervals in $E_{n,m}$ is less than K/m for all n, the total length of the intervals in E_m is not larger than K/m.

Since E_m is at most countable union of intervals whose total length is less than K/m and $E \subset E_m$ for all m, the set E has measure zero. □

[6] Otherwise replace φ_n by $\varphi_n - \varphi_1$ which is non-negative. The new sequence satisfies the assumption of the theorem (possibly with a different constant K). Moreover, the convergence of $\varphi_n(x) - \varphi_1(x)$ is equivalent to the convergence of $\varphi_n(x)$.

Lemma 5.4 *If φ_n and ψ_n are two increasing sequences of step functions which respectively tend to f and g a.e. and $f(x) \geq g(x)$ a.e., then*

$$\lim_{n\to\infty} \int \varphi_n \geq \lim_{n\to\infty} \int \psi_n \,.$$

Proof: The following sets

$$\begin{aligned} E_1 &= \Big\{ x : \lim \varphi_n(x) \neq f(x) \Big\}, \\ E_2 &= \Big\{ x : \lim \psi_n(x) \neq g(x) \Big\}, \\ E_3 &= \Big\{ x : f(x) < g(x) \Big\} \end{aligned}$$

have measure zero. Let $E = E_1 \cup E_2 \cup E_3$. Assume $x \notin E$.

Fix arbitrary $k \in \mathbb{N}$. The sequence $\psi_k(x) - \varphi_n(x)$ is decreasing in n. Then as $n \to \infty$

$$\psi_k(x) - \varphi_n(x) \to \psi_k(x) - f(x) \leq g(x) - f(x) \leq 0 \,.$$

Consequently

$$(\psi_k - \varphi_n)^+(x) = \max\{\, \psi_k(x) - \varphi_n(x), 0 \,\} \to 0$$

Since it is a decreasing sequence of non-negative step functions which converges to 0 a.e.,[7]

$$\int (\psi_k - \varphi_n)^+ \to 0 \,.$$

Since ψ_k and φ_n are step functions

$$\int \psi_k - \int \varphi_n = \int (\psi_k - \varphi_n) \leq \int (\psi_k - \varphi_n)^+ \,.$$

Then taking the limit as $n \to \infty$ we get

$$\int \psi_k \leq \lim_{n\to\infty} \int \varphi_n$$

Finally, take the limit as $k \to \infty$ to obtain the desired inequality. □

Corollary 5.5 *If φ_n and ψ_n are two increasing sequences of step functions which tend to a function f a.e., then*

$$\lim_{n\to\infty} \int \varphi_n = \lim_{n\to\infty} \int \psi_n \,.$$

Proof: Use the previous lemma with $g = f$.

[7] This statement is provided without a proof.

Lebesgue integrable functions

Definition 5.6 *If a function $f : \mathbb{R} \to \mathbb{R}$ can be represented as an a.e. limit of an increasing sequence of step functions φ_n, then the integral of f is given by*

$$\int f = \lim_{n\to\infty} \int \varphi_n .$$

If the limit is finite we write $f \in L^{\text{inc}}(\mathbb{R})$.

Note that the limit is independent of the choice of the increasing sequence of step functions.

Unfortunately $L^{\text{inc}}(\mathbb{R})$ is not a vector space as $f \in L^{\text{inc}}(\mathbb{R})$ does not imply $-f \in L^{\text{inc}}(\mathbb{R})$. Indeed, f is bounded from below by φ_1 but not necessarily bounded from above. Then $-f$ is not bounded from below and therefore $-f \notin L^{\text{inc}}(\mathbb{R})$. For example,

$$f(x) = \begin{cases} \frac{1}{\sqrt{|x|}}, & x \neq 0, \\ 0, & x = 0, \end{cases} \qquad \text{or} \qquad f = \sum_{k=1}^{\infty} k^{1/2} \chi_{[(k+1)^{-1},k^{-1}]} .$$

Definition 5.7 (Lebesgue integral) *A function $f : \mathbb{R} \to \mathbb{R}$ is* (Lebesgue) integrable *if $f = g - h$ where $g, h \in L^{\text{inc}}(\mathbb{R})$ and*

$$\int f := \int g - \int h .$$

This integral is called the Lebesgue integral.[8]

Of course in the definition the choice of g and h is not unique. So we will have to check that the value of the integral does not depend from this freedom.

5.2 Properties of Lebesgue integrals

The properties of the Lebesgue integral are based on the properties of the integrals for the functions from $L^{\text{inc}}(\mathbb{R})$.

Proposition 5.8 *If $f, g \in L^{\text{inc}}(\mathbb{R})$ and $\lambda \in \mathbb{R}$, $\lambda \geq 0$, then*

1. $f + g$, λf, $\max\{f, g\}$, $\min\{f, g\} \in L^{\text{inc}}(\mathbb{R})$.
2. $\int (f + g) = \int f + \int g$.
3. *If additionally $f(x) \geq g(x)$ a.e., then $\int f \geq \int g$.*

[8]This construction is different (but equivalent) to the original definition of the Lebesgue integral. For an alternative approach see Measure Theory module.

Exercise: Prove the proposition.

First we state the main elementary properties of the Lebesgue integration.

Theorem 5.9 *If f, f_1, f_2 are integrable and $\lambda \in \mathbb{R}$, then*

1. *$f_1 + \lambda f_2$ is also integrable and $\int (f_1 + \lambda f_2) = \int f_1 + \lambda \int f_2$.*
2. *$|f(x)|$ is also integrable and $|\int f| \le \int |f|$.*
3. *If additionally $f(x) \ge 0$ a.e., then $\int f \ge 0$.*

Proof:

1. Since f_1, f_2 are integrable, there are functions $g_1, g_2, h_1, h_2 \in L^{\mathrm{inc}}(\mathbb{R})$ such that $f_k = g_k - h_k$, $k = 1, 2$. If $\lambda \ge 0$ then $g_1 + \lambda g_2, h_1 + \lambda h_2 \in L^{\mathrm{inc}}(\mathbb{R})$ and

$$\begin{aligned} \int (f_1 + f_2) &= \int (g_1 + \lambda g_2) - \int (h_1 + \lambda h_2) \\ &= \int g_1 + \lambda \int g_2 - \int h_1 - \lambda \int h_2 = \int f_1 + \lambda \int f_2 . \end{aligned}$$

   The case of $\lambda < 0$ can be reduced to the previous one. Indeed, in this case we can write $f_1 + \lambda f_2 = f_1 + (-\lambda)(-f_2)$ and observe that $-f_2 = h_2 - g_2$ and consequently is also integrable. Linearity is proved.

2. Since f is integrable, $f = g - h$ with $g, h \in L^{\mathrm{inc}}(\mathbb{R})$. Obviously for every x

$$|f(x)| = \max\{\, g(x), h(x)\,\} - \min\{\, g(x), h(x)\,\},$$

   then Proposition 5.8 implies that the maximum and the minimum belong to $L^{\mathrm{inc}}(\mathbb{R})$ and hence $|f|$ is integrable. The inequality

$$g(x) + \min\{\, g(x), h(x)\,\} \le h(x) + \max\{\, g(x), h(x)\,\}$$

   is valid for every x and in combination with Proposition 5.8 implies

$$\int g + \int \min\{\, g, h\,\} \le \int h + \int \max\{\, g, h\,\}.$$

   Consequently

$$\int f = \int g - \int h \le \int \max\{\, g, h\,\} - \int \min\{\, g, h\,\} = \int |f|.$$

   Applying this result for f replaced by $-f$ we conclude

$$-\int f = \int (-f) \le \int |-f| = \int |f|.$$

   Consequently, $|\int f| \le \int |f|$.

3. Since $f(x) = g(x) - h(x)$ with $g, h \in L^{\text{inc}}(\mathbb{R})$ and $f(x) \geq 0$ a.e., we conclude $g(x) \geq h(x)$ a.e.. Then Proposition 5.8 implies that $\int g \geq \int h$. Consequently
$$\int f = \int g - \int h \geq 0. \qquad \square$$

We note that $|f|$ is integrable does not imply that f is integrable.[9]

Exercise: If f is integrable than $f_+ = \max\{f, 0\}$ and $f_- = \min\{f, 0\}$ are integrable. (Hint: $f_+ = (f + |f|)/2$ and $f_- = (f - |f|)/2$.)

Integrals and limits

You should be careful when swapping lim and $\int$:

Examples:

$$\lim_{n\to\infty} \int n\chi_{\left(0,\frac{1}{n}\right)} = 1 \quad \neq \quad \int \lim_{n\to\infty} n\chi_{\left(0,\frac{1}{n}\right)} = 0.$$
$$\lim_{n\to\infty} \int \frac{1}{n}\chi_{(0,n)} = 1 \quad \neq \quad \int \lim_{n\to\infty} \frac{1}{n}\chi_{(0,n)} = 0.$$

The following two theorems establish conditions which allow swapping the limit and integration. They play the fundamental role in the theory of Lebesgue integrals.

Theorem 5.10 (Monotone Convergence Theorem) *Suppose that f_n are integrable, $f_n(x) \leq f_{n+1}(x)$ a.e., and $\int f_n < K$ for some constant independent of n. Then there is an integrable function g such that $f_n(x) \to g(x)$ a.e. and*
$$\int g = \lim_{n\to\infty} \int f_n.$$

Corollary 5.11 *If f is integrable and $\int |f| = 0$, then $f(x) = 0$ a.e.*

Proof: Let $f_n(x) = n|f(x)|$. This sequence satisfies MCT (integrable, increasing and $\int f_n = 0 < 1$), consequently there is an integrable $g(x)$ such that $f_n(x) \to g(x)$ for a.e. x. Since the sequence is increasing, $f_n(x) \leq g(x)$ a.e. which implies $|f(x)| \leq g(x)/n$ for all n and a.e. x. Consequently $f(x) = 0$ a.e. $\square$

Theorem 5.12 (Dominated Convergence Theorem) *Suppose that f_n are integrable functions and $f_n(x) \to f(x)$ for a.e. x.. If there is an integrable function g such that $|f_n(x)| \leq g(x)$ for every n and a.e. x, then f is integrable and*
$$\int f = \lim_{n\to\infty} \int f_n.$$

[9] Indeed, we can sketch an example. It is based on partitioning the interval $[0,1]$ into two very nasty subsets. So let $f(x) = 0$ outside $[0,1]$, for $x \in [0,1]$ let $f(x) = 1$ if x belongs to the Vitali set and $f(x) = -1$ otherwise. Then $|f| = \chi_{[0,1]}$ but f is not integrable.

It is also possible to integrate complex valued functions: $f : \mathbb{R} \to \mathbb{C}$ is integrable if its real and imaginary parts are both integrable, and

$$\int f := \int \operatorname{Re} f + i \int \operatorname{Im} f.$$

The MCT has no meaning for complex valued functions. The DCT is valid without modifications (and indeed follows easily from the real version).

5.3 Lebesgue space $L^1(\mathbb{R})$

Definition 5.13 *The Lebesgue space $L^1(\mathbb{R})$ is the space of integrable functions modulo the following equivalence relation: $f \sim g$ iff $f(x) = g(x)$ a.e. The Lebesgue space is equipped with the L^1 norm:*

$$\|f\|_{L^1} = \int |f|.$$

It is convenient to think about elements of $L^1(\mathbb{R})$ as functions $\mathbb{R} \to \mathbb{R}$ interpreting the equality $f = g$ as $f(x) = g(x)$ a.e.

From the viewpoint of Functional Analysis, the equivalence relation is introduced to ensure non-degeneracy of the L^1 norm. Indeed, in the space of integrable functions $\int |f| = 0$ is equivalent to $f(x) = 0$ a.e. and therefore does not imply the latter equality for all x.

Theorem 5.14 *$L^1(\mathbb{R})$ is a Banach space.*

Revise the properties of the Lebesgue integral which imply that $L^1(\mathbb{R})$ is a normed space. The completeness of $L^1(\mathbb{R})$ follows from the combination of the following two statements: The first lemma gives a criterion for completeness of a normed space, and the second one implies that the assumptions of the first lemma are satisfied for $X = L^1(\mathbb{R})$.

Lemma 5.15 *If $(X, \|\cdot\|_X)$ is a normed space in which*

$$\sum_{j=1}^{\infty} \|y_j\|_X < \infty$$

implies the series $\sum_{j=1}^{\infty} y_j$ converges, then X is complete.

Proof: Let $x_j \in X$ be a Cauchy sequence. Then there is a monotone increasing sequence $n_k \in \mathbb{N}$ such that for every $k \in \mathbb{N}$

$$\|x_j - x_l\|_X < 2^{-k} \qquad \text{for all } k, l \geq n_k.$$

Let $y_1 = x_{n_1}$ and $y_k = x_{n_k} - x_{n_{k-1}}$ for $k \geq 2$. Since $\|y_k\|_X \leq 2^{1-k}$ for $k \geq 2$,

$$\sum_{k=1}^{\infty} \|y_k\|_X \leq \|y_1\|_X + \sum_{k=1}^{\infty} 2^{1-k} = \|y_1\|_X + 2 < \infty .$$

Consequently there is $x^* \in X$ such that

$$x^* = \sum_{j=1}^{\infty} y_j .$$

On the other hand

$$\sum_{j=1}^{k} y_j = x_{n_1} + \sum_{j=2}^{k} (x_{n_j} - x_{n_{j-1}}) = x_{n_k}$$

and therefore $x_{n_k} \to x^*$. Consequently $x_k \to x^*$ and the space X is complete. □

Lemma 5.16 *If $(f_k)_{k=1}^{\infty}$ is a sequence of integrable functions such that $\sum_{k=1}^{\infty} \|f_k\|_{L^1} < \infty$, then*

1. *$\sum_{k=1}^{\infty} |f_k(x)|$ converges a.e. to an integrable function,*
2. *$\sum_{k=1}^{\infty} f_k(x)$ converges a.e. to an integrable function.*

Proof: The first statement follows from MCT applied to the sequence $g_n = \sum_{k=1}^{n} |f_k|$ and $K = \sum_{k=1}^{\infty} \|f_k\|_{L^1}$. So there is an integrable function $g(x)$ such that

$$g(x) = \sum_{k=1}^{\infty} |f_k(x)|$$

for almost all x. For these values of x the partial sums $h_n(x) = \sum_{k=1}^{n} f_k(x)$ obviously converge, so let

$$h(x) = \sum_{k=1}^{\infty} f_k(x).$$

Moreover

$$|h_n(x)| = \left| \sum_{k=1}^{n} f_k(x) \right| \leq \sum_{k=1}^{n} |f_k(x)| \leq \sum_{k=1}^{\infty} |f_k(x)| = g(x) .$$

Therefore the partial sums h_n satisfy DCT and the second statement follows. □

In addition to $L^1(\mathbb{R})$ we will sometimes consider the Lebesgue spaces $L^1(I)$ where I is an interval. We say that $f \in L^1(I)$ if $\chi_I f \in L^1(\mathbb{R})$, i.e., we extend the function by zero outside its original domain.

Proposition 5.17 *The space $C[0,1]$ is dense in $L^1(0,1)$.*

Proof: First show that step functions are dense in $L^1(0,1)$. Then check that every step function can be approximated by a piecewise linear continuous function. □

Consequently the space we constructed in this section is isometrically isomorphic to the completion of $C[0,1]$ in the L^1 norm.

5.4 L^p spaces

Another important class of Lebesgue spaces consists of L^p spaces for $1 \leq p < \infty$, among those the L^2 space is the most remarkable (it is also a Hilbert space, see the next chapter for details). In this section we will sketch the main definitions of those spaces noting that the full discussion requires more knowledge of Measure Theory than we can fit into this module).

If $I = (a,b)$ is an interval, then $L^p(I)$ can be defined in terms of the integration procedure developed earlier in this chapter. This definition is equivalent to the standard one which will be given a bit later.

The Lebesgue space $L^p(I)$ is the space of all integrable functions such that

$$\|f\|_{L^p}^p = \int_I |f|^p < \infty$$

modulo the equivalence relation: $f = g$ iff $f(x) = g(x)$ a.e. We note that in this case $L^p(I) \subset L^1(I)$. The definition of $L^p(\mathbb{R})$ is slightly different. We say that $f \in L^p(\mathbb{R})$ if f is locally integrable (i.e., $f \in L^1(I)$ for any interval I)[10] and its p^{th} power is integrable. The norm is defined by the same formula:

$$\|f\|_{L^p}^p = \int_{\mathbb{R}} |f|^p < \infty .$$

We note that although $L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \neq \emptyset$ (e.g. both spaces contain all step functions) none of those spaces is a subset of the other one. For example, $f(x) = 1/(1+|x|)$ belongs to $L^2(\mathbb{R})$ but not to $L^1(\mathbb{R})$. Indeed, $\int f^2 < \infty$ but $\int f = \infty$ so it is not integrable on $\mathbb{R}$. On the other hand

$$g(x) = \frac{\chi_{(0,1)}(x)}{|x|^{1/2}}$$

belongs to $L^1(\mathbb{R})$ but not to $L^2(\mathbb{R})$.

Theorem 5.18 *$L^p(\mathbb{R})$ and $L^p(I)$ are Banach spaces for $p \geq 1$ and any interval I.*

We will not give a complete proof but sketch the main ideas instead.

Let $\frac{1}{p} + \frac{1}{q} = 1$ and $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$. Then the Hölder inequality states[11] that

$$\int |fg| \leq \|f\|_p \|g\|_q .$$

[10] This is an important requirement. It is not sufficient to define L^p as a set of all functions such that f^p is integrable: this space would not be a vector space. Indeed, let $p = 2$, $g = \chi_{[0,1]}$ and f be the non-integrable function from the footnote[9]. Then $f^2 = g^2 = \chi_{[0,1]}$ is integrable. But $(f+g)^2 = f^2 + 2fg + g^2 = 2 + 2f$ is not integrable. Therefore if we followed this definition $f, g \in L^2$ would not imply $f + g \in L^2$.

[11] We will not discuss the proof of this inequality in these lectures.

Note that the characteristic function $\chi_I \in L^q(\mathbb{R})$ for any interval I and any $q \geq 1$, moreover $\|\chi_I\|_{L^q} = |I|^{\frac{1}{q}}$ where $|I| = b - a$ is the length of I. The Hölder inequality with $g = \chi_I$ implies that

$$\int \chi_I |f| = \int_I |f| \leq |I|^{1/q} \|f\|_p .$$

The left hand side of this inequality is the norm of f in $L^1(I)$:

$$\|f\|_{L^1(I)} \leq |I|^{1/q} \|f\|_{L^p(I)} .$$

Consequently any Cauchy sequence in $L^p(I)$ is automatically a Cauchy sequence in $L^1(I)$. Since L^1 is complete the Cauchy sequence converges to a limit in $L^1(I)$ (and consequently converges a.e.). In order to proof completeness of L^p it is sufficient to show that the p^{th} power of this limit is integrable. This can be done on the basis of the Dominated Convergence Theorem.

Exercise: The next two exercises show that $L^2(\mathbb{R})$ is complete (compare with the proof of completeness for $L^1(\mathbb{R})$).

1. Let $(f_k)_{k=1}^{\infty}$ be a sequence in $L^2(\mathbb{R})$ such that

$$\sum_{k=1}^{\infty} \|f_k\|_{L^2} < \infty .$$

   Applying the MCT to the sequence

$$g_n = \left(\sum_{k=1}^{n} |f_k| \right)^2$$

   show that $\sum_k f_k$ converges to a function f with integrable f^2.

2. Now use the DCT applied to $h_n = \left| f - \sum_{k=1}^{n} f_k \right|^2$ to deduce that $\sum_k f_k$ converges in the L^2 norm to a function in L^2.

□

If you look into a textbook, you will probably see a differently looking definition of the Lebesgue spaces. Traditionally a function f is asked to be *measurable* instead of *locally integrable*. Local integrability is a stronger property: every locally integrable function is measurable but there are measurable functions which are not locally integrable, e.g. x^{-2} is measurable but not locally integrable since $\int_{-1}^{1} x^{-2} = +\infty$. Nevertheless the two alternative definitions of the Lebesgue space are equivalent.

Let us discuss the notion of a measurable function from the perspective of our definitions.

First we need to define the measure, which can be considered as a generalisation of the length of an interval. We say that a subset $A \subset \mathbb{R}$ has finite Lebesgue measure $\mu(A)$ if the characteristic function χ_A is Lebesgue integrable. Then

$$\mu(A) := \int \chi_A \geq 0\,.$$

Obviously $\mu([a,b]) = b - a$ for an interval $[a,b]$ and consequently its Lebesgue measure coincides with the length.

In order to study large sets (like $\mathbb{R}$) we need to extend this definition to allow measuring sets with infinitely large measures. We say that $A \subset \mathbb{R}$ is measurable if χ_A is locally integrable. In particular, if A is measurable then $A_n = A \cap [-n,n]$ has finite measure for each n. Since $\mu(A_n)$ is an increasing sequence the following limit exists (but can be $+\infty$)

$$\mu(A) = \lim_{n\to\infty} \mu(A_n) \leq +\infty\,.$$

Note that if $\int \chi_A < \infty$ MCT implies that the limit coincides with the previous definition of $\mu(A)$.

For example $\mathbb{R}$ is measurable and $\mu(\mathbb{R}) = +\infty$.

Definition 5.19 *A function $f : \mathbb{R} \to \mathbb{R}$ is* measurable *if preimage of any interval is measurable.*

We note that sums, products and pointwise limits of measurable functions are measurable.

Consider the set of all measurable functions from $\mathbb{R}$ to $\mathbb{R}$ (or $\mathbb{C}$) whose absolute value raised to the p^{th} power has a finite Lebesgue integral, i.e.,

$$\|f\|_{L^p} := \left(\int |f|^p \right)^{1/p} < \infty.$$

This space modulo the equivalence relation "$f = g$ iff $f(x) = g(x)$ a.e." is called the Lebesgue space $L^p(\mathbb{R})$.

6 Hilbert spaces

6.1 Inner product spaces

You have already seen the inner product on $\mathbb{R}^n$.

Definition 6.1 *An inner product on a vector space V is a map $(\cdot,\cdot): V \times V \to \mathbb{K}$ such that for all $x,y,z \in V$ and for all $\lambda \in \mathbb{K}$:*

(i) $(x,x) \geq 0$*, and* $(x,x) = 0$ *iff* $x = 0$*;*

(ii) $(x+y,z) = (x,z) + (y,z)$*;*

(iii) $(\lambda x, y) = \lambda (x,y)$*;*

(iv) $(x,y) = \overline{(y,x)}$.

A vector space equipped with an inner product is called an inner product space.

- In a real vector space the complex conjugate in (iv) is not necessary.
- If $\mathbb{K} = \mathbb{C}$, then (iv) with $y = x$ implies that (x,x) is real and therefore the requirement $(x,x) \geq 0$ make sense.
- (iii) and (iv) imply that $(x, \lambda y) = \overline{\lambda}(x.y)$.

1. **Example:** $\mathbb{R}^n$ is an inner product space

$$(x,y) = \sum_{k=1}^{n} x_k y_k \,.$$

2. **Example:** $\mathbb{C}^n$ is an inner product space

$$(x,y) = \sum_{k=1}^{n} x_k \overline{y}_k \,.$$

3. **Example:** $\ell^2(\mathbb{K})$ is an inner product space

$$(x,y) = \sum_{k=1}^{\infty} x_k \overline{y}_k \,.$$

Note that the sum converges because $\sum_k |x_k y_k| \leq \frac{1}{2} \sum_k \left(|x_k|^2 + |y_k|^2\right)$.

4. **Example:** $L^2(a,b)$ is an inner product space

$$(f,g) = \int_a^b f(x)\overline{g}(x)\,dx \,.$$

6.2 Natural norms

Every inner product space is a normed space as well.

Proposition 6.2 *If V is an inner product space, then*

$$\|v\| = \sqrt{(v,v)}$$

defines a norm on V.

Definition 6.3 *We say that $\|x\| = \sqrt{(x,x)}$ is* the natural norm induced by the inner product.

Lemma 6.4 (Cauchy-Schwartz inequality) *If V is an inner product space and $\|v\| = \sqrt{(v,v)}$ for all $v \in V$, then*

$$|(x,y)| \leq \|x\| \, \|y\| \qquad \textit{for all } x,y \in V.$$

Proof of the lemma: The inequality is obvious if $y = 0$. So suppose that $y \neq 0$. Then for any $\lambda \in \mathbb{K}$:

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

Then substitute $\lambda = (x,y)/\|y\|^2$:

$$0 \leq (x,x) - 2\frac{|(x,y)|}{\|y\|^2} + \frac{|(x,y)|}{\|y\|^2} = \|x\|^2 - \frac{|(x,y)|}{\|y\|^2},$$

which implies the desired inequality. □

Proof of the proposition: Now we can complete the proof of Proposition 6.2. We note that positive definiteness and homogeneity of $\|\cdot\|$ easily follow from (i), and (iii), (iv) in the definition of the inner product. In order to establish the triangle inequality we use the Cauchy-Schwartz inequality. Let $x,y \in V$. Then

$$\begin{aligned} \|x+y\|^2 &= (x+y,x+y) = (x,x) + (x,y) + (y,x) + (y,y) \\ &\leq \|x\|^2 + 2\|x\| \, \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2, \end{aligned}$$

and the triangle inequality follows by taking the square root.

Therefore $\|\cdot\|$ is a norm. □

We have already proved the Cauchy-Schwartz inequality for $\ell^2(\mathbb{K})$ using a different strategy (see Lemma 2.4).

The Cauchy-Schwartz inequality in $L^2(a,b)$ takes the form

$$\left| \int_a^b f(x)g(x)\,dx \right| \leq \left(\int_a^b |f(x)|^2 dx \right)^{1/2} \left(\int_a^b |g(x)|^2 dx \right)^{1/2}.$$

In particular, it states that $f,g \in L^2(a,b)$ implies $fg \in L^1(a,b)$.

Lemma 6.5 *If V is an inner product space equipped with the natural norm, then $x_n \to x$ and $y_n \to y$ imply that*

$$(x_n, y_n) \to (x, y)\,.$$

Proof: Since any convergent sequence is bounded, the inequality

$$\begin{aligned} |(x_n,y_n)-(x,y)| &= |(x_n-x,y_n)+(x,y_n-y)| \\ &\le |(x_n-x,y_n)|+|(x,y_n-y)| \\ &\le \|x_n-x\|\,\|y_n\|+\|x\|\,\|y_n-y\| \end{aligned}$$

implies that $(x_n,y_n)\to(x,y)$. □

The lemma implies that we can swap inner products and limits.

6.3 Parallelogram law and polarisation identity

Natural norms have some special properties.

Lemma 6.6 (Parallelogram law) *If V is an inner product space with the natural norm $\|\cdot\|$, then*

$$\|x+y\|^2+\|x-y\|^2 = 2\big(\|x\|^2+\|y\|^2\big) \qquad \textit{for all } x,y\in V.$$

Proof: The linearity of the inner product implies that for any $x,y\in V$

$$\begin{aligned} \|x+y\|^2+\|x-y\|^2 &= (x+y,x+y)+(x-y,x-y) \\ &= (x,x)+(x,y)+(y,x)+(y,y) \\ &\quad +(x,x)-(x,y)-(y,x)+(y,y) \\ &= 2\big(\|x\|^2+\|y\|^2\big) \end{aligned}$$

□

Example (some norms are not induced by an inner product): There is no inner product which induces the following norms on $C[0,1]$:

$$\|f\|_\infty = \sup_{t\in[0,1]} |f(t)| \qquad \text{or} \qquad \|f\|_{L^1} = \int_0^1 |f(t)|\,dt\,.$$

Indeed, these norms do not satisfy the parallelogram law, e.g., take $f(x)=x$ and $g(x)=1-x$, obviously $f,g\in C[0,1]$ and

$$\|f\|_\infty = \|g\|_\infty = \|f-g\|_\infty = \|f+g\|_\infty = 1\,,$$

substituting these numbers into the parallelogram law we see $2\neq 4$.

Exercise: Is the parallelogram law for the L^1 norm satisfied for these f,g?

Lemma 6.7 (Polarisation identity) *Let V be an inner product space with the natural norm $\|\cdot\|$. Then*

1. *If V is real*
$$4(x,y) = \|x+y\|^2 - \|x-y\|^2 ;$$
2. *If V is complex*
$$4(x,y) = \|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2 .$$

Proof: Plug in the definition of the natural norm into the right hand side and use linearity of the inner product. □

Lemma 6.7 shows that the inner product can be restored from its natural norm. Although the right hand sides of the polarisation identities is meaningful for any norm, we ***should not*** rush to the conclusion that any normed space is automatically an inner product space. Indeed, the example above implies that for some norms these formulae cannot define an inner product. Nevertheless, if the norm satisfy the parallelogram law, we indeed get an inner product:

Proposition 6.8 *Let V be a real normed space with the norm $\|\cdot\|$ satisfying the parallelogram law, then*

$$(x,y) = \frac{\|x+y\|^2 - \|x-y\|^2}{4} = \frac{\|x+y\|^2 - \|x\|^2 - \|y\|^2}{2}$$

defines an inner product on V.[12]

Proof: Let us check that (x,y) satisfy the axioms of inner product. Positivity and symmetry are straightforward (Exercise). The linearity:

$$\begin{aligned}
4(x,y)+4(z,y) &= \|x+y\|^2 - \|x-y\|^2 + \|z+y\|^2 - \|z-y\|^2 \\
&= \tfrac{1}{2}(\|x+2y+z\|^2 + \|x-z\|^2) - \tfrac{1}{2}(\|x-2y+z\|^2 + \|x-z\|^2) \\
&= \tfrac{1}{2}\|x+2y+z\|^2 - \tfrac{1}{2}\|x-2y+z\|^2 \\
&= \tfrac{1}{2}(2\|x+y+z\|^2 + 2\|y\|^2 - \|x+z\|^2) \\
&\quad -\tfrac{1}{2}(2\|x-y+z\|^2 + 2\|y\|^2 - \|x+z\|^2) \\
&= \|x+y+z\|^2 - \|x-y+z\|^2 = 4(x+z,y) .
\end{aligned}$$

We have proved that
$$(x,y)+(z,y) = (x+z,y) .$$
Applying this identity several times and setting $z = x/m$ we obtain

$$n(x/m,y) = (nx/m,y) \qquad \text{and} \qquad m(x/m,y) = (x,y)$$

[12] Can you find a simpler proof?

for any $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. Consequently, for any rational $\lambda = \frac{n}{m}$

$$(\lambda x, y) = \lambda (x, y).$$

We note that the right hand side of the definition involves the norms only, which commute with the limits. Any real number is a limit of rational numbers and therefore the linearity holds for all $\lambda \in \mathbb{R}$. □

6.4 Hilbert spaces: Definition and examples

Definition 6.9 A Hilbert space *is a complete inner product space (equipped with the natural norm).*

Of course, any Hilbert space is a Banach space.

1. **Example:** $\mathbb{R}^n$ is a Hilbert space

$$(x, y) = \sum_{k=1}^{n} x_k y_k, \qquad \|x\| = \left(\sum_{k=1}^{n} x_k^2 \right)^{1/2}.$$

2. **Example:** $\mathbb{C}^n$ is a Hilbert space

$$(x, y) = \sum_{k=1}^{n} x_k \bar{y}_k, \qquad \|x\| = \left(\sum_{k=1}^{n} |x_k|^2 \right)^{1/2}.$$

3. **Example:** $\ell^2(\mathbb{K})$ is a Hilbert space

$$(x, y) = \sum_{k=1}^{\infty} x_k \bar{y}_k, \qquad \|x\| = \left(\sum_{k=1}^{\infty} |x_k|^2 \right)^{1/2}.$$

4. **Example:** $L^2(a, b)$ is a Hilbert space

$$(f, g) = \int_a^b f(x) \overline{g}(x)\, dx, \qquad \|x\| = \left(\int_a^b |f(x)|^2 dx \right)^{1/2}.$$

7 Orthonormal bases in Hilbert spaces

The goal of this section is to discuss properties of orthonormal bases in a Hilbert space H. Unlike Hamel bases, the orthonormal ones involve a countable number of elements: i.e. a vector x is represented in the form of an infinite sum

$$x = \sum_{k=1}^{\infty} \alpha_k e_k$$

for some $\alpha_k \in \mathbb{K}$.

We will mainly consider complex spaces with $\mathbb{K} = \mathbb{C}$. The real case $\mathbb{K} = \mathbb{R}$ is not very different. We will use $(\cdot,\cdot)$ to denote an inner product on H, and $\|\cdot\|$ will stand for the natural norm induced by the inner product.

7.1 Orthonormal sets

Definition 7.1 *Two vectors $x, y \in H$ are called* orthogonal *if $(x, y) = 0$. Then we write $x \perp y$.*

Theorem 7.2 (Pythagoras theorem) *If $x \perp y$ then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.*

Proof: Since $(x,y) = 0$

$$\|x+y\|^2 = (x+y, x+y) = (x,x) + (x,y) + (y,x) + (y,y) = \|x\|^2 + \|y\|^2 . \qquad \square$$

Definition 7.3 *A set E is orthonormal if $\|e\| = 1$ for all $e \in E$ and $(e_1, e_2) = 0$ for all $e_1, e_2 \in E$ such that $e_1 \neq e_2$.*

Note that this definition does not require the set E to be countable.

Exercise: Any orthonormal set is linearly independent.

Indeed, suppose $\sum_{k=1}^{n} \alpha_k e_k = 0$ with $e_k \in E$ and $\alpha_k \in \mathbb{K}$. Multiplying this equality by e_j we get

$$0 = \left(\sum_{k=1}^{n} \alpha_k e_k, e_j \right) = \sum_{k=1}^{n} \alpha_k (e_k, e_j) = \alpha_j .$$

Since $\alpha_j = 0$ for all j, we conclude that the set E is linearly independent.

Definition 7.4 (Kronecker delta) *The* Kronecker delta *is the function defined by*

$$\delta_{jk} = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{if } j \neq k . \end{cases}$$

Example: For every $j \in \mathbb{N}$, let $e_j = (\delta_{jk})_{k=1}^{\infty}$ (it is an infinite sequence of zeros with 1 at the j^{th} position). The set $E = \{ e_j : j \in \mathbb{N} \}$ is orthonormal in ℓ^2. Indeed, from the definition of the scalar product in ℓ^2 we see that $(e_j, e_k) = \delta_{jk}$ for all $j,k \in \mathbb{N}$.

Example: The set[13]

$$E = \left\{ f_k = \frac{e^{ikx}}{\sqrt{2\pi}} : k \in \mathbb{Z} \right\}$$

is an orthonormal set in $L^2(-\pi,\pi)$. Indeed, since $|f_k(x)| = \frac{1}{\sqrt{2\pi}}$ for all x:

$$\|f_k\|_{L^2}^2 = \int_{-\pi}^{\pi} |f_k(x)|^2 dx = 1\,,$$

and if $j \neq k$

$$(f_k, f_j) = \int_{-\pi}^{\pi} f_k(x)\overline{f}_j(x)\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(k-j)x} dx = \left.\frac{e^{i(k-j)x}}{i(k-j)}\right|_{x=-\pi}^{x=\pi} = 0\,.$$

Lemma 7.5 *If $\{e_1, \ldots, e_n\}$ is an orthonormal set in an inner product space V, then for any $\alpha_j \in \mathbb{K}$*

$$\left\| \sum_{j=1}^{n} \alpha_j e_j \right\|^2 = \sum_{j=1}^{n} |\alpha_j|^2\,.$$

Proof: The following computation is straightforward:

$$\begin{aligned} \left\| \sum_{j=1}^{n} \alpha_j e_j \right\|^2 &= \left(\sum_{j=1}^{n} \alpha_j e_j, \sum_{l=1}^{n} \alpha_l e_l \right) = \sum_{j=1}^{n}\sum_{l=1}^{n} \alpha_j \overline{\alpha}_l (e_j, e_l) \\ &= \sum_{j=1}^{n}\sum_{l=1}^{n} \alpha_j \overline{\alpha}_l \delta_{jl} = \sum_{j=1}^{n} \alpha_j \overline{\alpha}_j\,. \end{aligned}$$ □

7.2 Gram-Schmidt orthonormalisation

Lemma 7.6 (Gram-Schmidt orthonormalisation) *Let V be an inner product space and (v_k) be a sequence of linearly independent vectors in V (finite or infinite). Then there is an orthonormal sequence (e_k) such that*

$$\mathrm{Span}\{ v_1, \ldots, v_k \} = \mathrm{Span}\{ e_1, \ldots, e_k \} \qquad \textit{for all } k.$$

[13] Remember that for any $x \in \mathbb{R}$ and any $k \in \mathbb{Z}$: $e^{ikx} = \cos kx + i \sin kx$. Then $|e^{ikx}| = 1$ and $e^{\pm ik\pi} = \cos k\pi \pm i \sin k\pi = (-1)^k$.

Proof: Let $e_1 = \frac{v_1}{\|v_1\|}$. Then

$$\mathrm{Span}\{\, v_1 \,\} = \mathrm{Span}\{\, e_1 \,\}$$

and the statement is true for $n = 1$ as the set $E_1 = \{\, e_1 \,\}$ is obviously orthonormal.

Then we continue inductively.[14] Suppose that for some $k \geq 2$ we have found an orthonormal set $E_{k-1} = \{\, e_1, \dots, e_{k-1} \,\}$ such that its span coincides with the span of $\{\, v_1, \dots, v_{k-1} \,\}$. Then set

$$\tilde{e}_k = v_k - \sum_{j=1}^{k-1} (v_k, e_j) e_j \,.$$

Since $\sum_{j=1}^{k-1} (v_k, e_j) e_j \in \mathrm{Span}(E_{k-1}) = \mathrm{Span}\{\, v_1, \dots, v_{k-1} \,\}$ and $v_1, \dots, v_k$ are linearly independent, we conclude that $\tilde{e}_k \neq 0$. For every $j < k$

$$(\tilde{e}_k, e_j) = (v_k, e_l) - \sum_{j=1}^{k-1} (v_k, e_j)(e_j, e_l) = (v_k, e_l) - (v_k, e_l) = 0$$

which implies that $\tilde{e}_k \perp e_j$. Finally let $e_k = \tilde{e}_k / \|\tilde{e}_k\|$. Then $\{\, e_1, \dots, e_k \,\}$ is an orthonormal set such that

$$\mathrm{Span}\{\, e_1, \dots, e_k \,\} = \mathrm{Span}\{\, v_1, \dots, v_k \,\} \,.$$

If the original sequence is finite, the orthonormalisation procedure will stop after a finite number of steps. Otherwise, we get an infinite sequence of e_k. □

Corollary 7.7 *Any infinite-dimensional inner product space contains a countable orthonormal sequence.*

Corollary 7.8 *Any finite-dimensional inner product space has an orthonormal basis.*

Proposition 7.9 *Any finite dimensional inner product space is isometric to $\mathbb{C}^n$ (or $\mathbb{R}^n$ if the space is real) equipped with the standard inner product.*

Proof: Let $n = \dim V$ and e_j, $j = 1, \dots, n$ be an orthonormal basis in V. Note that $(e_k, e_j) = \delta_{kj}$. Any two vectors $x, y \in V$ can be written as

$$x = \sum_{k=1}^{n} x_k e_k \qquad \text{and} \qquad y = \sum_{j=1}^{n} y_j e_j \,.$$

Then

$$(x, y) = \left(\sum_{k=1}^{n} x_k e_k, \sum_{j=1}^{n} y_j e_j \right) = \sum_{k=1}^{n} \sum_{j=1}^{n} x_k \bar{y}_j (e_k, e_j) = \sum_{k=1}^{n} x_k \bar{y}_k \,.$$

Therefore the map $x \mapsto (x_1, \dots, x_n)$ is an isometry. □

We see that an arbitrary inner product, when written in orthonormal coordinates, takes the form of the "canonical" inner product on $\mathbb{C}^n$ (or $\mathbb{R}^n$ if the original space is real).

[14]For example, let $k = 2$. We define $\tilde{e}_2 = v_2 - (v_2, e_1) e_1$. Then $(\tilde{e}_2, e_1) = (v_2, e_1) - (v_2, e_1)(e_1, e_1) = 0$. Since v_1, v_2 are linearly independent $\tilde{e}_2 \neq 0$. So we can define $e_2 = \frac{\tilde{e}_2}{\|\tilde{e}_2\|}$.

7.3 Bessel's inequality

Lemma 7.10 (Bessel's inequality) *If V is an inner product space and $E = (e_k)_{k=1}^{\infty}$ is an orthonormal sequence, then for every $x \in V$*

$$\sum_{k=1}^{\infty} |(x, e_k)|^2 \leq \|x\|^2 .$$

Proof: We note that for any $n \in \mathbb{N}$:

$$\begin{aligned} \left\| x - \sum_{k=1}^{n} (x, e_k) e_k \right\|^2 &= \left(x - \sum_{k=1}^{n} (x, e_k) e_k, x - \sum_{k=1}^{n} (x, e_k) e_k \right) \\ &= \|x\|^2 - 2 \sum_{k=1}^{n} |(x, e_k)|^2 + \sum_{k=1}^{n} |(x, e_k)|^2 \\ &= \|x\|^2 - \sum_{k=1}^{n} |(x, e_k)|^2 . \end{aligned}$$

Since the left hand side is not negative,

$$\sum_{k=1}^{n} |(x, e_k)|^2 \leq \|x\|^2$$

and the lemma follows by taking the limit as $n \longrightarrow \infty$. □

Corollary 7.11 *If E is an orthonormal set in an inner product space V, then for any $x \in V$ the set*

$$\mathscr{E}_x = \{ e \in E : (x, e) \neq 0 \}$$

is at most countable.

Proof: For any $m \in \mathbb{N}$ the set $E_m = \{ e : |(x, e)| > \frac{1}{m} \}$ has a finite number of elements. Otherwise there would be an infinite sequence $(e_k)_{k=1}^{\infty}$ with $e_k \in E_m$, then the series $\sum_{k=1}^{\infty} |(x, e_k)|^2 = +\infty$ which contradicts to Bessel's inequality. Therefore $\mathscr{E}_x = \cup_{m=1}^{\infty} E_m$ is a countable union of finite sets and hence at most countable. □

7.4 Convergence

In this section we will discuss convergence of series which involve elements from an orthonormal set.

Lemma 7.12 *Let H be a Hilbert space and $E = (e_k)_{k=1}^{\infty}$ an orthonormal sequence. The series $\sum_{k=1}^{\infty} \alpha_k e_k$ converges iff $\sum_{k=1}^{\infty} |\alpha_k|^2 < +\infty$. Then*

$$\left\| \sum_{k=1}^{\infty} \alpha_k e_k \right\|^2 = \sum_{k=1}^{\infty} |\alpha_k|^2 . \tag{7.1}$$

Proof: Let $x_n = \sum_{k=1}^{n} \alpha_k e_k$ and $\beta_n = \sum_{k=1}^{n} |\alpha_k|^2$. Lemma 7.5 implies that $\|x_n\|^2 = \beta_n$ and that for any $n > m$

$$\|x_n - x_m\|^2 = \left\| \sum_{k=m+1}^{n} \alpha_k e_k \right\|^2 = \sum_{k=m+1}^{n} |\alpha_k|^2 = \beta_n - \beta_m .$$

Consequently, x_n is a Cauchy sequence in H iff β_n is Cauchy in $\mathbb{R}$. Since both spaces are complete, the sequences converge or diverge simultaneously.

If they converge, we take the limit as $n \to \infty$ in the equality $\|x_n\|^2 = \beta_n$ to get (7.1) (the limit commutes with $\|\cdot\|^2$). □

Definition 7.13 *A series $\sum_{n=1}^{\infty} x_n$ in a Banach space X is unconditionally convergent if for every permutation $\sigma : \mathbb{N} \to \mathbb{N}$ the series $\sum_{n=1}^{\infty} x_{\sigma(n)}$ converges.*

Every absolutely convergent series is unconditionally convergent, but the converse implication does not hold in general. In $\mathbb{R}^n$ a series is unconditionally convergent if and only if it is absolutely convergent.

Lemma 7.12 and Bessel's inequality imply:

Corollary 7.14 *If H is a Hilbert space and $E = (e_k)_{k=1}^{\infty}$ is an orthonormal sequence, then for every $x \in H$ the sequence*

$$\sum_{k=1}^{\infty} (x, e_k) e_k$$

converges unconditionally.

Lemma 7.15 *Let H be a Hilbert space, $E = (e_k)_{k=1}^{\infty}$ an orthonormal sequence and $x \in H$. If $x = \sum_{k=1}^{\infty} \alpha_k e_k$, then*

$$\alpha_k = (x, e_k) \qquad \textit{for all } k \in \mathbb{N}.$$

Proof: Exercise. □

7.5 Orthonormal basis in a Hilbert space

Definition 7.16 *A set E is a* basis *for H if every $x \in H$ can be written uniquely in the form*

$$x = \sum_{k=1}^{\infty} \alpha_k e_k$$

for some $\alpha_k \in \mathbb{K}$ and $e_k \in E$. If additionally E is an orthonormal set, then E is an orthonormal basis.

If E is a basis, then it is a linearly independent set. Indeed, if $\sum_k^n \alpha_k e_k = 0$ then $\alpha_k = 0$ due to the uniqueness.

Note that in this definition the uniqueness is a delicate point. Indeed, the sum $\sum_{k=1}^{\infty} \alpha_k e_k$ is defined as a limit of partial sums $x_n = \sum_{k=1}^{n} \alpha_k e_k$. A permutation of e_k changes the partial sums and may lead to a different limit. In general, we cannot even guarantee that after a permutation the series remains convergent.

If E is countable, we can assume that the sum involves all elements of the basis (some α_k can be zero) and that the summation is taken following the order of a selected enumeration of E. The situation is more difficult if E is uncountable since in this case there is no natural way of numbering the elements.

The situation is much simpler if E is orthonormal as in this case the series converge unconditionally and the order of summations is not important.

Proposition 7.17 *Let $E = \{e_j : j \in \mathbb{N}\}$ be an orthonormal set in a Hilbert space H. Then the following statements are equivalent:*

(a) *E is a basis in H;*

(b) $x = \sum_{k=1}^{\infty} (x, e_k) e_k$ *for all* $x \in H$;

(c) $\|x\|^2 = \sum_{k=1}^{\infty} |(x, e_k)|^2$;

(d) $(x, e_n) = 0$ *for all* $n \in \mathbb{N}$ *implies* $x = 0$;

(e) *the linear span of E is dense in H.*

Proof:
$(a) \iff (b)$: use Lemma 7.15.

$(b) \implies (c)$: use Lemma 7.12.

$(c) \implies (d)$: Let $(x, e_k) = 0$ for all k, then (c) implies that $\|x\| = 0$ hence $x = 0$.

$(d) \implies (b)$: let $y = x - \sum_{k=1}^{\infty} (x, e_k) e_k$. Corollary 7.14 implies that the series converges. Then Lemma 6.5 implies we can swap the limit and the inner product to get for every n

$$\begin{aligned} (y, e_n) &= \left(x - \sum_{k=1}^{\infty} (x, e_k) e_k, e_n \right) \\ &= (x, e_n) - \sum_{k=1}^{\infty} (x, e_k)(e_k, e_n) = (x, e_n) - (x, e_n) = 0 . \end{aligned}$$

Since $(y, e_n) = 0$ for all n, then (d) implies that $y = 0$ which is equivalent to $x = \sum_{k=1}^{\infty} (x, e_k) e_k$ as required.

$(e) \implies (d)$: since $\mathrm{Span}(E)$ is dense in H for any $x \in H$ there is a sequence $x_n \in \mathrm{Span}(E)$ such that $x_n \to x$. Take x such that $(x, e_n) = 0$ for all n. Then $(x_n, x) = 0$ and consequently

$$\|x\|^2 = \left(\lim_{n\to\infty} x_n, x\right) = \lim_{n\to\infty} (x_n, x) = 0\,.$$

Therefore $x = 0$.

$(a) \implies (e)$: Since E is a basis any $x = \lim_{n\to\infty} x_n$ with $x_n = \sum_{k=1}^{n} \alpha_k e_k \in \mathrm{Span}(E)$.
□

Example: The orthonormal sets from examples of Section 7.1 are also examples of orthonormal bases.

8 Closest points and approximations

8.1 Closest points in convex subsets

Definition 8.1 *A subset A of a vector space V is* convex *if $\lambda x + (1-\lambda)y \in A$ for any two vectors $x, y \in V$ and any $\lambda \in [0,1]$.*

Lemma 8.2 *Let A be a non-empty closed convex subspace of a Hilbert space H and $x \in H$. Then there is a unique $a^* \in A$ such that*

$$\|x - a^*\| = \inf_{a \in A} \|x - a\| .$$

Proof: The parallelogram rule implies:

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

Then

$$\|u-v\|^2 = 2\|x-u\|^2 + 2\|x-v\|^2 - 4\|x - \tfrac{1}{2}(u+v)\|^2 .$$

Let $d = \inf_{a \in A} \|x-a\|$. Since A is convex, $\frac{1}{2}(u+v) \in A$ for any $u, v \in A$, and consequently $\|x - \frac{1}{2}(u+v)\| \ge d$. Then

$$\|u-v\|^2 \le 2\|x-u\|^2 + 2\|x-v\|^2 - 4d^2 . \tag{8.1}$$

Since d is the infinum, for any n there is $a_n \in A$ such that $\|x - a_n\|^2 < d^2 + \frac{1}{n}$. Then equation (8.1) implies that

$$\|a_n - a_m\| \le 2d^2 + \frac{2}{n} + 2d^2 + \frac{2}{m} - 4d^2 = \frac{2}{n} + \frac{2}{m} .$$

Consequently (a_n) is Cauchy and, since H is complete, it converges to some a^*. Since A is closed, $a^* \in A$. Then

$$\|x - a^*\|^2 = \lim_{n \to \infty} \|x - a_n\|^2 = d^2 .$$

Therefore a^* is the point closest to x. Now suppose that there is another point $\tilde{a} \in A$ such that $\|x - \tilde{a}\| = d$, then (8.1) implies

$$\|a^* - \tilde{a}\| \le 2\|x - a^*\|^2 + 2\|x - \tilde{a}\|^2 - 4d^2 = 2d^2 + 2d^2 - 4d^2 = 0 .$$

So $\tilde{a} = a^*$ and a^* is unique. □

8.2 Orthogonal complements

In an infinite dimensional space a linear subspace does not need to be closed. For example the space ℓ_f of all sequences with only a finite number of non-zero elements is a linear subspace of ℓ^2 but it not closed in ℓ^2 (e.g. consider the sequence $x_n = (1, 2^{-1}, 2^{-2}, \ldots, 2^{-n}, 0, 0, \ldots)$).

Definition 8.3 *Let $X \subseteq H$. The orthogonal complement of X in H is the set*

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

Proposition 8.4 *If $X \subseteq H$, then $X^\perp$ is a closed linear subspace of H.*

Proof: If $u, v \in X^\perp$ and $\alpha \in K$ then

$$(u + \alpha v, x) = (u,x) + \alpha (v,x) = 0$$

for all $x \in X$. Therefore $X^\perp$ is a linear subspace. Now suppose that $u_n \in X^\perp$ and $u_n \to u \in H$. Then for all $x \in X$

$$(u,x) = (\lim_{n\to\infty} u_n, x) = \lim_{n\to\infty} (u_n, x) = 0 .$$

Consequently, $u \in X^\perp$ and so $X^\perp$ is closed. □

Exercises:

1. If E is a basis in H, then $E^\perp = \{ 0 \}$.
2. If $Y \subseteq X$, then $X^\perp \subseteq Y^\perp$.
3. $X \subseteq (X^\perp)^\perp$
4. If X is a closed linear subspace in H, then $X = (X^\perp)^\perp$

Definition 8.5 The closed linear span of $E \subset H$ *is a minimal closed set which contains* Span(E)*:*

$$\overline{\mathrm{Span}}(E) = \{ u \in H : \forall \varepsilon > 0 \ \exists x \in \mathrm{Span}(E) \text{ such that } \|x - u\| < \varepsilon \} .$$

Proposition 8.6 *If $E \subset H$ then $E^\perp = (\mathrm{Span}(E))^\perp = (\overline{\mathrm{Span}}(E))^\perp$.*

Proof: Since $E \subseteq \mathrm{Span}(E) \subseteq \overline{\mathrm{Span}}(E)$ we have $(\overline{\mathrm{Span}}(E))^\perp \subseteq (\mathrm{Span}(E))^\perp \subseteq E^\perp$. So we need to prove the inverse inclusion. Take $u \in E^\perp$ and $x \in \overline{\mathrm{Span}}(E)$. Then there is $x_n \in \mathrm{Span}(E)$ such that $x_n \to x$. Then

$$(x,u) = (\lim_{n\to\infty} x_n, u) = \lim_{n\to\infty} (x_n, u) = 0 .$$

Consequently, $u \in (\overline{\mathrm{Span}}(E))^\perp$ and we proved $E^\perp \subseteq (\overline{\mathrm{Span}}(E))^\perp$. □

Theorem 8.7 *If U is a closed linear subspace of a Hilbert space H then*

1. *any $x \in H$ can be written uniquely in the form $x = u + v$ with $u \in U$ and $v \in U^\perp$.*
2. *u is the closest point to x in U.*
3. *The map $P_U : H \to U$ defined by $P_U x = u$ is linear and satisfies*

$$P_U^2 x = P_U x \qquad \textit{and} \qquad \|P_U(x)\| \le \|x\| \qquad \textit{for all } x \in H\,.$$

Definition 8.8 *The map P_U is called* the orthogonal projector onto U.

Proof: Any linear subspace is obviously convex. Then Lemma 8.2 implies that there is a unique $u \in U$ such that

$$\|x - u\| = \inf_{a \in U} \|x - a\|\,.$$

Let $v = x - u$. Let us show that $v \in U^\perp$. Indeed, take any $y \in U$ and consider the function $\Delta : \mathbb{C} \to \mathbb{R}$ defined by

$$\Delta(t) = \|v + ty\|^2 = \|x - (u - ty)\|^2\,.$$

Since the definition of u together with $u - ty \in U$ imply that $\Delta(t) \ge \Delta(0) = \|x - u\|^2$, the function Δ has a minimum at $t = 0$. On the other hand

$$\begin{aligned} \Delta(t) = \|v + ty\|^2 &= (v + ty, v + ty) \\ &= (v, v) + t(y, v) + \bar{t}(v, y) + |t|^2 (y, y)\,. \end{aligned}$$

First suppose that t is real. Then $\bar{t} = t$ and $\frac{d\Delta}{dt}(0) = 0$ implies

$$(y, v) + (v, y) = 0\,.$$

Then suppose that t is purely imaginary, Then $\bar{t} = -t$ and $\frac{d\Delta}{dt}(0) = 0$ implies

$$(y, v) - (v, y) = 0\,.$$

Taking the sum of these two equalities we conclude

$$(y, v) = 0 \qquad \text{for every } y \in U.$$

Therefore $v \in U^\perp$.

In order to prove the uniqueness of the representation suppose $x = u_1 + v_1 = u + v$ with $u_1, u \in U$ and $v_1, v \in U^\perp$. Then $u_1 - u = v - v_1$. Since $u - u_1 \in U$ and $v - v_1 \in U^\perp$,

$$\|v - v_1\|^2 = (v - v_1, v - v_1) = (v - v_1, u_1 - u) = 0\,.$$

Therefore u and v are unique.

Finally $x = u + v$ with $u \perp v$ implies $\|x\|^2 = \|u\|^2 + \|v\|^2$. Consequently $\|P_U(x)\| = \|u\| \le \|x\|$. We also note that $P_U(u) = u$ for any $u \in U$. So $P_U^2(x) = P_U(x)$ as $P_U(x) \in U$ $\square$

Corollary 8.9 *If U is a closed linear subspace in a Hilbert space H and $x \in H$, then $P_U(x)$ is the closest point to x in U.*

8.3 Best approximations

Theorem 8.10 *Let E be an orthonormal sequence: $E = \{e_j : j \in \mathscr{J}\}$ where $\mathscr{J}$ is either finite or countable set. Then for any $x \in H$, the closest point to x in $\overline{\mathrm{Span}}(E)$ is given by*

$$y = \sum_{j \in \mathscr{J}} (x, e_j) e_j \,.$$

Corollary 8.11 *If E is an orthonormal basis in a closed subspace $U \subset H$, then the orthogonal projection onto U is given by*

$$P_U(x) = \sum_{j \in \mathscr{J}} (x, e_j) e_j \,.$$

Proof: Corollary 7.14 implies that $u = \sum_{j \in \mathscr{J}} (x, e_j) e_j$ converges. Then obviously $u \in \overline{\mathrm{Span}}(E)$ which is a closed linear subset. Let $v = x - u$. Since $(v, e_k) = (x, e_k) - (u, e_k) = 0$ for all $k \in$ J, we conclude $v \in E^{\perp} = (\overline{\mathrm{Span}}(E))^{\perp}$ (Lemma 8.6). Theorem 8.7 implies that u is the closest point. □

Example: The best approximation of an element $x \in \ell^2$ in terms of the elements of the standard basis $(e_j)_{j=1}^n$ is given by

$$\sum_{k=1}^{n} (x, e_j) e_j = (x_1, x_2, \ldots, x_n, 0, 0, \ldots) \,.$$

Example: Let $(e_j)_{j=1}^{\infty}$ be an orthonormal basis in H. The best approximation of an element $x \in H$ in terms of the first n elements of the orthonormal basis is given by

$$\sum_{k=1}^{n} (x, e_j) e_j \,.$$

Now suppose that the set E is not orthonormal. If the set E is finite or countable we can use the Gram-Schmidt orthonormalisation procedure to construct an orthonormal basis in $\overline{\mathrm{Span}}(E)$. After that the theorem above gives us an explicit expression for the best approximation. Let's consider some examples.

Example: Find the best approximation of a function $f \in L^2(-1,1)$ with polynomials of degree up to n. In other words, let $E = \{1, x, x^2, \ldots, x^n\}$. We need to find $u \in \mathrm{Span}(E)$ such that

$$\|f - u\|_{L^2} = \inf_{p \in \mathrm{Span}(E)} \|f - p\|_{L^2} \,.$$

The set E is not orthonormal. Let's apply the Gram-Schmidt orthonormalisation procedure to construct an orthonormal basis in $\mathrm{Span}(E)$. For the sake of shortness, let's write $\|\cdot\| = \|\cdot\|_{L^2(-1,1)}$.

First note that $\|1\| = \sqrt{2}$ and let

$$e_1 = \frac{1}{\sqrt{2}}.$$

Then $(1,x) = \int_{-1}^{1} x\,dx = 0$ and $\|x\|^2 = \int_{-1}^{1} |x|^2\,dx = \frac{2}{3}$ so let

$$e_2 = \sqrt{\frac{3}{2}}x.$$

Then

$$\begin{aligned}
\tilde{e}_3 &= x^2 - (x^2, e_2)e_2 - (x^2, e_1)e_1 \\
&= x^2 - \sqrt{\frac{3}{2}}x \int_{-1}^{1} t^2 \sqrt{\frac{3}{2}}t\,dt - \frac{1}{\sqrt{2}} \int_{-1}^{1} t^2 \frac{1}{\sqrt{2}}\,dt \\
&= x^2 - \frac{1}{2}\int_{-1}^{1} t^2\,dt = x^2 - \frac{1}{3}.
\end{aligned}$$

Taking into account that

$$\|\tilde{e}_3\|^2 = \int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2 dt = \frac{8}{45}$$

we obtain

$$e_3 = \frac{\tilde{e}_3}{\|\tilde{e}_3\|} = \sqrt{\frac{5}{8}}\,(3x^2 - 1).$$

Exercise: Show that $e_4 = \sqrt{\frac{7}{8}}(5x^3 - 3x)$ is orthogonal to e_1, e_2 and e_3.

The best approximation of any function $f \in L^2(-1,1)$ by a polynomial of third degree is given by

$$\begin{aligned}
&\frac{7}{8}(5x^3 - 3x)\int_{-1}^{1} f(t)(5t^3 - 3t)dt + \frac{5}{8}(3x^2 - 1)\int_{-1}^{1} f(t)(3t^2 - 1)dt \\
&\quad + \frac{3}{2}x \int_{-1}^{1} t f(t)\,dt + \frac{1}{2}\int_{-1}^{1} f(t)\,dt
\end{aligned}$$

For example, if $f(x) = |x|$ its best approximation by a third degree polynomial is

$$p_3 = \frac{15x^2 + 3}{16}.$$

We can check (after computing the corresponding integral);

$$\|f - p_3\|^2 = \frac{3}{16}.$$

Note that the best approximation in the L^2 norm is not necessarily the best approximation in the sup norm. Indeed, for example,

$$\sup_{x\in[-1,1]}\left||x|-\frac{15x^2+3}{16}\right|>\frac{3}{16}$$

(the supremum is larger than the values at $x=0$). At the same time

$$\sup_{x\in[-1,1]}\left||x|-\left(x^2+\frac{1}{8}\right)\right|=\frac{1}{8}.$$

8.4 Weierstrass Approximation Theorem

In this section we will prove an approximation theorem which is independent from the discussions of the previous lectures. This theorem implies that polynomials are dense in the space of continuous functions on an interval.

Theorem 8.12 *If $f:[0,1]\to\mathbb{R}$ is continuous on $[0,1]$ then the sequence of polynomials*

$$P_n(x)=\sum_{p=0}^{n}\binom{n}{p}f(p/n)x^p(1-x)^{n-p}$$

uniformly converges to f on $[0,1]$.

Proof: The binomial theorem states that

$$(x+y)^n=\sum_{p=0}^{n}\binom{n}{p}x^py^{n-p}.$$

Differentiating with respect to x and multiplying by x we get

$$nx(x+y)^{n-1}=\sum_{p=0}^{n}p\binom{n}{p}x^py^{n-p}.$$

Differentiating the original identity twice with respect to x and multiplying by x^2 we get

$$n(n-1)x^2(x+y)^{n-2}=\sum_{p=0}^{n}p(p-1)\binom{n}{p}x^py^{n-p}.$$

Now substitute $y=1-x$ and denote

$$r_p(x)=\binom{n}{p}x^p(1-x)^{n-p}.$$

We get

$$\begin{aligned}\sum_{p=0}^{n} r_p(x) &= 1\,,\\ \sum_{p=0}^{n} p r_p(x) &= nx\,.\\ \sum_{p=0}^{n} p(p-1) r_p(x) &= n(n-1)x^2\,.\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{p=0}^{n} (p-nx)^2 r_p(x) &= \sum_{p=0}^{n} p^2 r_p(x) - 2nx \sum_{p=0}^{n} p r_p(x) + n^2x^2 \sum_{p=0}^{n} r_p(x)\\ &= n(n-1)x^2 + nx - 2(nx)^2 + n^2x^2 = nx(1-x)\,.\end{aligned}$$

Let $M = \sup_{x\in[0,1]} |f(x)|$. Note that f is uniformly continuous, i.e., for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$|x-y| < \delta \qquad \Longrightarrow \qquad |f(x) - f(y)| < \varepsilon\,.$$

Now we can estimate

$$\begin{aligned}|f(x) - P_n(x)| &= \left| f(x) - \sum_{p=0}^{n} f(p/n) r_p(x) \right|\\ &= \left| \sum_{p=0}^{n} (f(x) - f(p/n)) r_p(x) \right|\\ &\leq \left| \sum_{|x-p/n|<\delta} (f(x) - f(p/n)) r_p(x) \right|\\ &\qquad + \left| \sum_{|x-p/n|>\delta} (f(x) - f(p/n)) r_p(x) \right|\end{aligned}$$

The first sum is bounded by

$$\left| \sum_{|x-p/n|<\delta} (f(x) - f(p/n)) r_p(x) \right| \leq \varepsilon \sum_{|x-p/n|<\delta} r_p(x) < \varepsilon$$

The second sum is bounded by

$$
\begin{aligned}
\left|\sum_{|x-p/n|>\delta}(f(x)-f(p/n))r_p(x)\right| &\leq 2M\sum_{|nx-p|>n\delta} r_p(x) \\
&\leq \sum_{p=0}^{n}\frac{(p-nx)^2}{n^2\delta^2}r_p(x) \\
&= \frac{2Mx(1-x)}{n\delta^2} \leq \frac{2M}{n\delta^2}
\end{aligned}
$$

which is less than ε for any $n > \frac{2M}{n\delta^2\varepsilon}$. Therefore for these values of n

$$|f(x)-P_n(x)| < 2\varepsilon\,.$$

Consequently,

$$\sup_{x\in[0,1]} |f(x)-P_n(x)| \to 0 \qquad \text{as } n\to\infty. \qquad \square$$

Corollary 8.13 *The set of polynomials is dense in $C[0,1]$ equipped with the supremum norm.*

9 Separable Hilbert spaces

9.1 Definition and examples

Definition 9.1 *A normed space is* separable *if it contains a countable dense subset.*

In other words, a space H is separable if there is a countable set $\{x_n \in H : n \in \mathbb{N}\}$ such that for any $u \in H$ and any $\varepsilon > 0$ there is $n \in \mathbb{N}$ such that

$$\|x_n - u\| < \varepsilon .$$

Examples: $\mathbb{R}$ is separable ($\mathbb{Q}$ is dense). $\mathbb{R}^n$ is separable ($\mathbb{Q}^n$ is dense), $\mathbb{C}^n$ is separable ($\mathbb{Q}^n + i\mathbb{Q}^n$ is dense).

Example: ℓ^2 is separable. Indeed, the set of sequences $(x_1, x_2, \ldots, x_n, 0, 0, 0, \ldots)$ with $x_j \in \mathbb{Q}$ is dense and countable.

Example: The space $C[0,1]$ is separable. Indeed, the Weierstrass approximation theorem states that every continuous function can be approximated (in the sup norm) by a polynomial. The dense countable set is given by polynomials with rational coefficients.

Example: $L^2(0,1)$ is separable. Indeed, continuous functions are dense in $L^2(0,1)$ (in the L^2-norm). The polynomials are dense in $C[0,1]$ (in the supremum norm and therefore in the L^2 norm as well). The set of polynomials with rational coefficients is dense in the set of all polynomials and, consequently, it is also dense in $L^2[0,1]$ (in the L^2 norm).

9.2 Isometry to ℓ^2

If H is a Hilbert space, then its separability is equivalent to existence of a countable orthonormal basis.

Proposition 9.2 *An infinite-dimensional Hilbert space is separable iff it has a countable orthonormal basis.*

Proof: If a Hilbert space has a countable basis, then we can construct a countable dense set by taking finite linear combinations of the basis elements with rational coefficients. Therefore the space is separable.

If H is separable, then it contains a countable dense subset $V = \{x_n : n \in \mathbb{N}\}$. Obviously, the closed linear span of V coincides with H. First we construct a linear independent set $\tilde{V}$ which has the same linear span as V by eliminating from V those x_n which are not linearly independent from $\{x_1, \ldots, x_{n-1}\}$. Then the Gram-Schmidt process gives an orthonormal sequence with the same closed linear span, i.e., it is a basis by characterisation (e) of Proposition 7.17. $\square$

The following theorem shows that all infinite dimensional separable spaces are isometric. So in some sense ℓ^2 is essentially the "only" separable infinite-dimensional space.

Theorem 9.3 *Any infinite-dimensional Hilbert space is isometric to* ℓ^2.

Proof: Let $\{e_j\}$ be an orthonormal basis in H. The map $A : H \to \ell^2$ defined by

$$A : u \to ((u,e_1),(u,e_2),(u,e_3),\dots\}$$

is invertible. Indeed, the image of A is in ℓ^2 due to Lemma 7.12, and the inverse map is given by

$$A^{-1} : (x_k)_{k=1}^{\infty} \mapsto \sum_{k=1}^{\infty} x_k e_k\,.$$

The characterisation of a basis in Proposition 7.17 implies that $\|u\|_H = \|A(u)\|_{\ell^2}$. $\square$

Note that there are Hilbert spaces which are not separable.

Example: Let $\mathscr{J}$ be uncountable. The space of all functions $f : \mathscr{J} \to \mathbb{R}$ such that

$$\sum_{j\in\mathscr{J}} |f(j)|^2 < \infty$$

is a Hilbert space. It is not separable.[15]

[15] How do we define the sum over an uncountable set? For any $n \in \mathbb{N}$ the set $\mathscr{J}_n = \{ j \in \mathscr{J} : |f(j)| > \frac{1}{n}\}$ is finite (otherwise the sum is obviously infinite). Consequently, the set $\mathscr{J}(f) := \{ j \in \mathscr{J} : |f(j)| > 0\}$ is countable because it is a countable union of finite sets: $\mathscr{J}(f) = \cup_{n=1}^{\infty} \mathscr{J}_n$. Therefore, the number of non-zero terms in the sum is countable and the usual definition of an infinite sum can be used.

10 Linear maps between Banach spaces

A linear map on a vector space is traditionally called a linear operator. All linear functions defined on a finite-dimensional space are continuous. This statement is no longer true in the case of an infinite dimensional space.

We will begin our study with continuous operators: this class has a rich theory and numerous applications. We will only slightly touch some of them (the most remarkable examples will be the shift operators on ℓ^2, and integral operators and multiplication operators on L^2).

Of course many interesting linear maps are not continuous, i.e., the differential operator $A : f \mapsto f'$ on the space of continuously differentiable functions. More accurately, let $D(A) = C^1[0,1] \subset L^2(0,1)$ be the domain of A Obviously $A : D(A) \to L^2(0,1)$ is linear but not continuous. Indeed, consider the sequence $x_n(t) = n^{-1}\sin(nt)$. Obviously $\|x_n\|_{L^2} \le n^{-1}$ so $x_n \to 0$, but $A(x_n) = \cos(nt)$ does not converge to $A(0) = 0$ in the L^2 norm so A is not continuous.

Some definitions and properties from the theory of continuous linear operators can be literally extended onto unbounded ones, but sometimes subtle differences appear: e.g., we will see that a bounded operator is self-adjoint iff it is symmetric, which is no longer true for unbounded operators. In a study of unbounded operators a special attention should be paid to their domains.

10.1 Continuous linear maps

Let U and V be vector spaces over $\mathbb{K}$.

Definition 10.1 *A function $A : U \to V$ is called a* linear operator *if*

$$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y) \qquad \text{for all } x, y \in U \text{ and } \alpha, \beta \in \mathbb{K}.$$

We will often write Ax to denote $A(x)$.

The collection of all linear operators from U to V is a vector space. If $A, B : U \to V$ are linear operators and $\alpha, \beta \in \mathbb{K}$ then we define

$$(\alpha A + \beta B)(x) = \alpha Ax + \beta Bx\,.$$

Obviously, $\alpha A + \beta B$ is also linear.

Definition 10.2 *A linear operator $A : U \to V$ is* bounded *if there is a constant M such that*

$$\|Ax\|_V \le M\|x\|_U \qquad \text{for all } x \in U. \tag{10.1}$$

If an operator is bounded, then the image of a bounded set is also bounded.

Lemma 10.3 *A linear operator $A : U \to V$ is continuous iff it is bounded.*

Proof: Suppose A is bounded. Then there is $M > 0$ such that

$$\|A(x) - A(y)\| = \|A(x-y)\| \leq M\|x-y\|$$

for all $x, y \in V$ and consequently A is continuous.

Now suppose A is continuous. Obviously $A(0) = 0$. Then for $\varepsilon = 1$ there is $\delta > 0$ such that $\|A(x)\| < \varepsilon = 1$ for all $\|x\| < \delta$. For any $u \in U$, $u \neq 0$,

$$A(u) = \frac{2\|u\|}{\delta} A\left(\frac{\delta}{2\|u\|}\, u\right).$$

Since $\left\|\frac{\delta}{2\|u\|}\, u\right\| = \frac{\delta}{2} < \delta$ we get $\|A(u)\| \leq \frac{2\|u\|}{\delta}$ and consequently A is bounded. □

The space of all bounded linear operators from U to V is denoted by $B(U,V)$.

Definition 10.4 *The* operator norm of $A : U \to V$ *is*

$$\|A\|_{B(U,V)} = \sup_{x \neq 0} \frac{\|A(x)\|_V}{\|x\|_U}.$$

We will often write $\|A\|_{\mathrm{op}}$ *instead of* $\|A\|_{B(U,V)}$.

Since A is linear

$$\|A\|_{B(U,V)} = \sup_{\|x\|_U = 1} \|A(x)\|_V.$$

We note that $\|A\|_{B(U,V)}$ is the smallest M such that (10.1) holds: indeed, it is easy to see that the definition of operator norm implies

$$\|A(x)\|_V \leq \|A\|_{B(U,V)}\, \|x\|_U$$

and (10.1) holds with $M = \|A\|_{B(U,V)}$. On the other hand, (10.1) implies $M \geq \frac{\|Ax\|_V}{\|x\|_U}$ for any $x \neq 0$ and consequently $M \geq \|A\|_{B(U,V)}$.

Theorem 10.5 *Let U be a normed space and V be a Banach space. Then $B(U,V)$ is a Banach space.*

Proof: Let $(A_n)_{n=1}^\infty$ be a Cauchy sequence in $B(U,V)$. Take a vector $u \in U$. The sequence $v_n = A_n(u)$ is a Cauchy sequence in V:

$$\|v_n - v_m\| = \|A_n(u) - A_m(u)\| = \|(A_n - A_m)(u)\| \leq \|A_n - A_m\|_{\mathrm{op}}\|u\|.$$

Since V is complete there is $v \in V$ such that $v_n \to v$. Let $A(u) = v$.

The operator A is linear. Indeed,

$$\begin{aligned} A(\alpha_1 u_1 + \alpha_2 u_2) &= \lim_{n\to\infty} A_n(\alpha_1 u_1 + \alpha_2 u_2) = \lim_{n\to\infty} (\alpha_1 A_n(u_1) + \alpha_2 A_n(u_2)) \\ &= \alpha_1 \lim_{n\to\infty} A_n u_1 + \alpha_2 \lim_{n\to\infty} A_n u_2 = \alpha_1 A u_1 + \alpha_2 A u_2. \end{aligned}$$

The operator A is bounded. Indeed, A_n is Cauchy and hence bounded: there is constant $M \in \mathbb{R}$ such that $\|A_n\|_{\text{op}} < M$ for all n. Taking the limit in the inequality $\|A_n u\| \le M\|u\|$ implies $\|Au\| \le M\|u\|$. Therefore $A \in B(U,V)$.

Finally, $A_n \to A$ in the operator norm. Indeed, Since A_n is Cauchy, for any $\varepsilon > 0$ there is N such that $\|A_n - A_m\|_{\text{op}} < \varepsilon$ or

$$\|A_n(u) - A_m(u)\| \le \varepsilon \|u\| \qquad \text{for all } m, n > N.$$

Taking the limit as $m \to \infty$

$$\|A_n(u) - A(u)\| \le \varepsilon \|u\| \qquad \text{for all } n > N.$$

Consequently $\|A_n - A\| \le \varepsilon$ and so $A_n \to A$. Therefore $B(U,V)$ is complete. □

10.2 Examples

1. **Example:** Shift operator: $T_l, T_r : \ell^2 \to \ell^2$:

$$T_r(x) = (0, x_1, x_2, x_3, \ldots) \qquad \text{and} \qquad T_l(x) = (x_2, x_3, x_4, \ldots)\,.$$

Both operators are obviously linear. Moreover,

$$\|T_r(x)\|_{\ell^2}^2 = \sum_{k=1}^{\infty} |x_k|^2 = \|x\|_{\ell^2}\,.$$

Consequently, $\|T_r\|_{\text{op}} = 1$. We also have

$$\|T_l(x)\|_{\ell^2}^2 = \sum_{k=2}^{\infty} |x_k|^2 \le \|x\|_{\ell^2}\,.$$

Consequently, $\|T_l\|_{\text{op}} \le 1$. However, if $x = (0, x_2, x_3, x_4, \ldots)$ then $\|T_l(x)\|_{\ell^2} = \|x\|_{\ell^2}$. Therefore $\|T_l\|_{\text{op}} = 1$.

2. **Example:** Multiplication operator: Let f be a continuous function on $[a,b]$. The equation

$$(Ax)(t) = f(t)x(t)$$

defines a bounded linear operator $A : L^2[a,b] \to L^2[a,b]$. Indeed, A is obviously linear. It is bounded since

$$\|Ax\|^2 \;=\; \int_a^b |f(t)x(t)|^2\,dt \le \|f\|_\infty^2 \int_a^b |x(t)|\,dt = \|f\|_\infty^2 \|x\|_{L^2}\,.$$

Consequently $\|A\|_{\text{op}} \le \|f\|_\infty$. Now let t_0 be a maximum of f. If $t_0 \ne b$, consider the characteristic function

$$x_\varepsilon = \chi_{[t_0, t_0+\varepsilon]}\,.$$

(If $t_0 = b$ let $x_\varepsilon = \chi_{[t_0-\varepsilon,t_0]}$.) Since f is continuous,

$$\frac{\|Ax_\varepsilon\|}{\|x_\varepsilon\|} = \frac{1}{\varepsilon}\int_{t_0}^{t_0+\varepsilon} |f(t)|^2 dt \to |f(t)|^2 \qquad \text{as } \varepsilon \to 0.$$

Therefore $\|A\|_{\text{op}} = \|f\|_\infty$.

3. **Example:** Integral operator on $L^2(a,b)$:

$$(Ax)(t) = \int_a^b K(t,s)x(s)\,ds \qquad \text{for all } t \in [a,b],$$

where

$$\int_a^b \int_a^b |K(s,t)|\,ds\,dt < +\infty\,.$$

Let us estimate the norm of A:

$$\begin{aligned}
\|Ax\|^2 &= \int_a^b \left|\int_a^b K(t,s)x(s)ds\right|^2 dt \\
&\le \int_a^b \left(\int_a^b |K(t,s)|^2 ds \int_a^b |x(s)|^2 ds\right) dt \qquad \text{(Cauchy-Schwartz)} \\
&= \int_a^b \int_a^b |K(t,s)|^2 ds dt\, \|x\|^2 .
\end{aligned}$$

Consequently

$$\|A\|_{\text{op}}^2 \le \int_a^b \int_a^b |K(t,s)|^2 ds dt\,.$$

Note that this example requires a bit more from the theory of Lebesgue integrals than we discussed in Section 5. If you are not taking Measure Theory and feel uncomfortable with these integrals, you may assume that x, y and K are continuous functions.

10.3 Kernel and range

Definition 10.6 *Kernel of A:*

$$\operatorname{Ker} A = \{x \in U : Ax = 0\}$$

Range of A:

$$\operatorname{Range} A = \{y \in V : \exists x \in U \text{ such that } y = Ax\}$$

We note that $0 \in \operatorname{Ker} A$ for any linear operator A. We say that $\operatorname{Ker} A$ is trivial if $\operatorname{Ker} A = \{0\}$.

Proposition 10.7 *If $A \in B(U,V)$ then* $\operatorname{Ker} A$ *is a closed linear subspace of U.*

Proof: If $x, y \in \mathrm{Ker} A$ and $\alpha, \beta \in \mathbb{K}$, then

$$K(\alpha x + \beta y) = \alpha K(x) + \beta K(y) = 0\,.$$

Consequently $\alpha x + \beta y \in \mathrm{Ker} A$ and it is a linear subspace. Furthermore if $x_n \to x$ and $A(x_n) = 0$ for all n, then $A(x) = 0$ due to continuity of A. □

Note that the range is a linear subspace but not necessarily closed (see Examples 3).

11 Linear functionals

11.1 Definition and examples

Definition 11.1 *If U is a vector space then a linear map $U \to \mathbb{K}$ is called* a linear functional on U,

Definition 11.2 *The space of all continuous functionals on a normed space U is called* the dual space*, i.e., $U^* = B(U, \mathbb{K})$.*

The dual space equipped with the operator norm is Banach. Indeed, $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ which are both complete. Then Theorem 10.5 implies that U^* is Banach.

1. **Example:** $\delta_x(f) = f(x)$, $x \in [a,b]$, is a bounded linear functional on $C[a,b]$.

2. **Example:** Let $\phi \in C[a,b]$ (or $\phi \in L^2(a,b)$). Then $\ell_\phi(x) = \int_a^b \phi(t)x(t)\,dt$ is a bounded linear functional on $L^2(a,b)$.

11.2 Riesz representation theorem

Let H be a Hilbert space. Then for any $y \in H$

$$\ell_y(x) = (x, y)$$

is a bounded functional $\ell_y : H \to \mathbb{K}$ (by the Cauchy-Schwartz inequality). The following theorem is one of the fundamental results of Functional Analysis: it states that the map $y \mapsto \ell_y$ is an isometry between H and its dual space H^*.

Theorem 11.3 (Riesz Representation Theorem) *Let H be a Hilbert space. For any bounded linear functional $f : H \to \mathbb{K}$ there is a unique $y \in H$ such that*

$$f(x) = (x, y) \qquad \text{for all } x \in H.$$

Moreover, $\|f\|_{H^*} = \|y\|_H$.

Proof: Let $K = \operatorname{Ker} f$. It is a closed linear subspace of H. If $K = H$ then $f(x) = 0$ for all x and the statement of the theorem is true with $y = 0$. Otherwise $K^\perp \neq \{0\}$ and there is a vector $z \in K^\perp$ with $\|z\|_H = 1$.

Now we show that $\dim K^\perp = 1$. Indeed, let $u \in K^\perp$. Since $K^\perp$ is a linear subspace $v = f(z)u - f(u)z \in K^\perp$. On the other hand

$$f(v) = f\left(f(z)u - f(u)z\right) = f(z)f(u) - f(u)f(z) = 0$$

and so $v \in K$. For any linear subspace $K \cap K^\perp = \{0\}$, and so $v = 0$. Then $f(z)u - f(u)z = v = 0$, i.e. $u = \frac{f(u)}{f(z)}z$. Consequently $\{z\}$ is the basis in $K^\perp$.

Theorem 8.7 implies that every vector $x \in H$ can be written uniquely in the form

$$x = u + v \qquad \text{where } u \in K \text{ and } v \in K^{\perp}.$$

Since $K^{\perp}$ is one dimensional and $\|z\| = 1$, $u = (x,z)z$. Moreover,

$$f(x) = f(u) + f(v) = f(u) = (x,z)f(z) = (x, \overline{f(z)}\,z)\,.$$

Set $y = \overline{f(z)}\,z$ to get the desired equality:

$$f(x) = (x,y)\,.$$

If there is another $y' \in H$ such that $f(x) = (x,y')$ for all $x \in H$, then $(x,y) = (x,y')$ for all x, i.e., $(x, y-y') = 0$. Setting $x = y - y'$ we conclude $\|y-y'\|^2 = 0$, i.e. $y = y'$ is unique.

Finally, the Cauchy-Schwartz inequality implies

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

i.e., $\|f\|_{H^*} = \|f\|_{\text{op}} \leq \|y\|$. On the other hand,

$$\|f\|_{\text{op}} \geq \frac{|f(y)|}{\|y\|} = \frac{|(y,y)|}{\|y\|} = \|y\|\,.$$

Consequently, $\|f\|_{H^*} = \|y\|_H$. □

12 Linear operators on Hilbert spaces

12.1 Complexification

In the next lectures we will discuss the spectral theory of linear operators. The spectral theory looks more natural in complex spaces. For example, a part of the theory studies eigenvalues and eigenvectors of linear maps (i.e. non-zero solutions of the equation $Ax = \lambda x$). In the finite-dimensional space a linear operator can be describe by a matrix. You already know that a matrix (even a real one) can have complex eigenvalues. Fortunately a real Hilbert space can always be considered as a part of a complex one due to the "complexification" procedure.

Definition 12.1 *Let H be a Hilbert space of $\mathbb{R}$. The complexification of H is the vector space*

$$H_{\mathbb{C}} = \{ x + iy : \ x, y \in H \}$$

where the addition and multiplication are respectively defined by

$$\begin{aligned} (x+iy)+(u+iw) &= (x+u)+i(y+w) \\ (\alpha+i\beta)(x+iy) &= (\alpha x-\beta y)+i(\alpha y+\beta x) . \end{aligned}$$

The inner product is defined by

$$(x+iy, u+iw) = (x,u) - i(x,w) + i(y,u) + (y,w) .$$

Exercise: Show that $H_{\mathbb{C}}$ is a Hilbert space.

Exercise: Show that $\|x+iy\|^2_{H_{\mathbb{C}}} = \|x\|^2 + \|y\|^2$ for all $x, y \in H$.

The following lemma states that any bounded operator on H can be extended to a bounded operator on $H_{\mathbb{C}}$.

Lemma 12.2 *Let H be a real Hilbert space and $A : H \to H$ be a bounded operator. Then*

$$A_{\mathbb{C}}(x+iy) = A(x) + iA(y)$$

is a bounded operator $H_{\mathbb{C}} \to H_{\mathbb{C}}$.

Exercise: Prove the lemma.

12.2 Adjoint operators

Theorem 12.3 *If $A : H \to H$ is a bounded linear operator on a Hilbert space H, then there is a unique bounded operator $A^* : H \to H$ such that*

$$(Ax, y) = (x, A^*y) \qquad \textit{for all } x, y \in H .$$

Moreover, $\|A^*\|_{\mathrm{op}} \le \|A\|_{\mathrm{op}}$.

Proof: Let $y \in H$ and $f(x) = (Ax, y)$ for all $x \in H$. The map $f : H \to \mathbb{K}$ is linear and

$$|f(x)| = |(Ax,y)| \leq \|Ax\|\,\|y\| \leq \|A\|_{\rm op}\|x\|\,\|y\|$$

where we have used the Cauchy-Schwartz inequality. Consequently, f is a bounded functional on H. The Riesz representation theorem implies that there is a unique $z \in H$ such that

$$(Ax,y) = (x,z) \qquad \text{for all } x \in H.$$

Define the function $A^* : H \to H$ by $A^* y = z$. Then

$$(Ax,y) = (x, A^* y) \qquad \text{for all } x, y \in H.$$

First, A^* is linear since for any $x, y_1, y_2 \in H$ and $\alpha_1, \alpha_2 \in \mathbb{K}$

$$\begin{aligned} (x, A^*(\alpha_1 y_1 + \alpha_2 y_2)) &= (Ax, \alpha_1 y_1 + \alpha_2 y_2) = \bar\alpha_1 (Ax, y_1) + \bar\alpha_2 (Ax, y_2) \\ &= \bar\alpha_1 (x, A^* y_1) + \bar\alpha_2 (x, A^* y_2) = (x, \alpha_1 A^* y_1 + \alpha_2 A^* y_2)\,. \end{aligned}$$

Since the equality is valid for all $x \in H$, it implies

$$A^*(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 A^* y_1 + \alpha_2 A^* y_2\,.$$

Second, A^* is bounded since

$$\|A^* y\|^2 = (A^* y, A^* y) = (AA^* y, y) \leq \|AA^* y\|\,\|y\| \leq \|A\|_{\rm op}\|A^* y\|\,\|y\|\,.$$

Dividing by $\|A^* y\|$ (do not forget to consider the case $A^* y = 0$ separately), we conclude that

$$\|A^* y\| \leq \|A\|_{\rm op}\|y\|\,.$$

Therefore A^* is bounded and $\|A^*\|_{\rm op} \leq \|A\|_{\rm op}$. □

Definition 12.4 *The operator A^* from Theorem 12.3 is called the* adjoint operator.

1. **Example:** If $A : \mathbb{C}^n \to \mathbb{C}^n$, then A^* is the Hermitian conjugate of A, i.e. if $A^* = \bar{A}^T$ (the complex conjugate of the transposed matrix).

2. **Example:** Integral operator on $L^2(0,1)$

   $$(Ax)(t) = \int_0^1 K(t,s)x(s)ds$$

   The adjoint operator

   $$(A^* y)(s) = \int_0^1 \overline{K(t,s)}y(t)dt$$

Indeed, for any $x, y \in L^2(0,1)$:

$$\begin{aligned}
(Ax,y) &= \int_0^1 \left(\int_0^1 K(t,s)x(s)\,ds \right) \bar{y}(t)\,dt \\
&= \int_0^1 \int_0^1 K(t,s)x(s)\bar{y}(t)\,ds\,dt \\
&= \int_0^1 x(s) \left(\overline{\int_0^1 \overline{K(t,s)} y(t)\,dt} \right) ds \\
&= (x, A^* y)\,.
\end{aligned}$$

Note that we used Fubini's Theorem to change the order of integration.

3. **Example:** Shift operators: $T_l^* = T_r$ and $T_r^* = T_l$. Indeed,

$$(T_r x, y) = \sum_{k=1}^{\infty} x_k \bar{y}_{k+1} = (x, T_l y)\,.$$

The following lemma states some elementary properties of adjoint operators.

Lemma 12.5 *If $A, B : H \to H$ are bounded operators on a Hilbert space H and $\alpha, \beta \in \mathbb{C}$, then*

1. $(\alpha A + \beta B)^* = \bar{\alpha} A^* + \bar{\beta} B^*$
2. $(AB)^* = B^* A^*$
3. $(A^*)^* = A$
4. $\|A^*\| = \|A\|$
5. $\|A^* A\| = \|AA^*\| = \|A\|^2$

Proof: Statements 1—3 follow directly from the definition of an adjoint operator (Exercise). Statement 4 follows from 3 and the estimate of Theorem 12.3: indeed,

$$\|A^*\| \le \|A\| = \|(A^*)^*\| \le \|A^*\|.$$

Finally since

$$\|Ax\|^2 = (AX, Ax) = (x, A^* Ax) \le \|x\|\,\|A^* Ax\| \le \|A^* A\|\,\|x\|^2$$

implies $\|A\|^2 \le \|AA^*\|$ and on the other hand $\|A^* A\| \le \|A^*\|\,\|A\| = \|A\|^2$, it follows that $\|A^* A\| = \|A\|^2$. □

12.3 Self-adjoint operators

Definition 12.6 *A linear operator A is* self-adjoint*, if $A^* = A$.*

Lemma 12.7 *An operator $A \in B(H,H)$ is self-adjoint iff*

$$(x, Ay) = (Ax, y) \qquad \text{for all } x, y \in H.$$

1. **Example:** $H = \mathbb{R}^n$, a linear map defined by a symmetric matrix is self-adjoint.
2. **Example:** $H = \mathbb{C}^n$, a linear map defined by a Hermitian matrix is self-adjoint.
3. **Example:** $A : L^2(0,1) \to L^2(0,1)$

$$Af(t) = \int_0^1 K(t,s) f(s)\, ds$$

with real symmetric K, $K(t,s) = K(s,t)$, is self-adjoint.

Theorem 12.8 *Let A be a self-adjoint operator on a Hilbert space H. Then all eigenvalues of A are real and the eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof: Suppose $Ax = \lambda x$ with $x \neq 0$. Then

$$\lambda \|x\|^2 = (\lambda x, x) = (Ax, x) = (x, A^* x) = (x, Ax) = (x, \lambda x) = \bar{\lambda} \|x\|^2 .$$

Consequently, λ is real.

Now if λ_1 and λ_2 are distinct eigenvalues and $Ax_1 = \lambda_1 x_1$, $Ax_2 = \lambda_2 x_2$, then

$$0 = (Ax_1, x_2) - (x_1, Ax_2) = (\lambda_1 x_1, x_2) - (x_1, \lambda_2 x_2) = (\lambda_1 - \lambda_2)(x_1, x_2).$$

Since $\lambda_1 - \lambda_2 \neq 0$, we conclude $(x_1, x_2) = 0$. □

Exercise: Let A be a self-adjoint operator on a real space. Show that the complexification of A is also self-adjoint and has the same eigenvalues as the original operator A.

Theorem 12.9 *If A is a bounded self-adjoint operator then*

1. *(Ax, x) is real for all $x \in H$*
2. $\|A\|_{\text{op}} = \sup_{\|x\|=1} |(Ax, x)|$

Proof: For any $x \in H$

$$(Ax,x) = (x,Ax) = \overline{(Ax,x)}$$

which implies (Ax,x) is real. Now let

$$M = \sup_{\|x\|=1} |(Ax,x)|\,.$$

The Cauchy-Schwartz inequality implies

$$|(Ax,x)| \le \|Ax\|\,\|x\| \le \|A\|_{\text{op}}\|x\|^2 = \|A\|_{\text{op}}$$

for all $x \in H$ such that $\|x\| = 1$. Consequently $M \le \|A\|_{\text{op}}$. On the other hand, for any $u,v \in H$ we have

$$\begin{aligned} 4\,\text{Re}(Au,v) &= (A(u+v),u+v) - (A(u-v),u-v) \\ &\le M\left(\|u+v\|^2 + \|u-v\|^2\right) \\ &= 2M\left(\|u\|^2 + \|v\|^2\right) \end{aligned}$$

using the parallelogram law. If $Au \neq 0$ let

$$v = \frac{\|u\|}{\|Au\|}Au$$

to obtain, since $\|u\| = \|v\|$, that

$$\|u\|\,\|Au\| \le M\|u\|^2\,.$$

Consequently $\|Au\| \le M\|u\|$ (for all u, including those with $Au = 0$) and $\|A\|_{\text{op}} \le M$. Therefore $\|A\|_{\text{op}} = M$. $\square$

Unbounded operators and their adjoint operators[16]

The notions of adjoint and self-adjoint operators play an important role in the general theory of linear operators. If an operator is not bounded a special care should be taken in the consideration of its domain of definition.

Let $D(A)$ be a linear subspace of a Hilbert space H, and $A : D(A) \to H$ be a linear operator. If $D(A)$ is dense in H we say that A is densely defined.

Example: Consider the operator $A(f) = \frac{df}{dt}$ on the set of all continuously differentiable functions, i.e., $D(A) = C^1[0,1] \subset L^2(0,1)$. This operator is densely defined.

Given a densely defined linear operator A on H, its adjoint A^* is defined as follows:

- $D(A^*)$, the domain of A^*, consists of all vectors $x \in H$ such that

  $$y \mapsto (x, Ay)$$

  is a continuous linear functional $D(A) \to \mathbb{K}$. By continuity and density of $D(A)$, it extends to a unique continuous linear functional on all of H.

- By the Riesz representation theorem, if $x \in D(A^*)$, there is a unique vector $z \in H$ such that

  $$(x, Ay) = (z, y) \qquad \text{for all } y \in D(A).$$

  This vector z is defined to be A^*x.

It can be shown that the dependence of $A^* : D(A^*) \to H$ is linear.

Note that two properties play a key role in this definition: the density of the domain of A in H, and the uniqueness part of the Riesz representation theorem.

A linear operator is symmetric if $(Ax, y) = (x, Ay)$ for all $x, y \in D(A)$. From this definition we see that $D(A) \subseteq D(A^*)$ and symmetric A coincides with the restriction of A^* onto $D(A)$. An operator is self adjoint if $A = A^*$, i.e., it is symmetric and $D(A) = D(A^*)$. In general, the condition for a linear operator on a Hilbert space to be self-adjoint is stronger than to be symmetric. If an operator is bounded then it is normally assumed that $D(A) = D(A^*) = H$ and therefore a symmetric operator is self-adjoint.

The Hellinger-Toeplitz theorem states that an everywhere defined symmetric operator on a Hilbert space is bounded.

[16]Optional topic

13 Introduction to Spectral Theory

13.1 Point spectrum

Let H be a complex Hilbert space and $A : H \to H$ a linear operator. If $Ax = \lambda x$ for some $x \in H$, $x \neq 0$, and $\lambda \in \mathbb{C}$, then λ is an eigenvalue of A and x is an eigenvector. The space

$$E_\lambda = \{ x \in H : Ax = \lambda x \}$$

is called the eigenspace.

Exercise: Prove the following: If $A \in B(H,H)$ and λ is an eigenvalue of A, then E_λ is a closed linear subspace in H. Moreover, E_λ is invariant, i.e., $A(E_\lambda) = E_\lambda$.

Definition 13.1 The point spectrum of A *consists of all eigenvalues of A:*

$$\sigma_p(A) = \{ \lambda \in \mathbb{C} : Ax = \lambda x \text{ for some } x \in H,\ x \neq 0 \}.$$

Proposition 13.2 *If $A : H \to H$ is bounded and λ is its eigenvalue then*

$$\|\lambda\| \leq \|A\|_{op}.$$

Proof: If $Ax = \lambda x$ with $x \neq 0$, then

$$\|A\|_{\text{op}} = \sup_{y \neq 0} \frac{\|Ay\|}{\|y\|} \geq \frac{\|Ax\|}{\|x\|} = |\lambda|. \qquad \square$$

Examples:

1. A linear map on an n-dimensional vector space has at least one and at most n different eigenvalues.

2. The right shift $T_r : \ell^2 \to \ell^2$ has no eigenvalues, i.e., the point spectrum is empty. Indeed, suppose $T_r x = \lambda x$, then

$$(0, x_1, x_2, x_3, x_4, \ldots) = \lambda (x_1, x_2, x_3, x_4, \ldots)$$

implies $0 = \lambda x_1$, $x_1 = \lambda x_2$, $x_2 = \lambda x_3, \ldots$ If $\lambda \neq 0$, we divide by λ and conclude $x_1 = x_2 = \cdots = 0$. If $\lambda = 0$ we also get $x = 0$. Consequently

$$\sigma_p(T_r) = \emptyset.$$

3. The point spectrum of the left shift $T_l : \ell^2 \to \ell^2$ is the open unit disk. Indeed, suppose $T_l x = \lambda x$ with $\lambda \in \mathbb{C}$. Then

$$(x_2, x_3, x_4, \ldots) = \lambda (x_1, x_2, x_3, x_4, \ldots)$$

is equivalent to $x_2 = \lambda x_1$, $x_3 = \lambda x_2$, $x_4 = \lambda x_3, \ldots$ Consequently, $x = (x_k)_{k=1}^\infty$ with $x_k = \lambda^{k-1} x_1$ for all $k \geq 2$. This sequence belongs to ℓ^2 if and only if $\sum_{k=1}^\infty |x_k|^2 = \sum_{k=1}^\infty |x_1| \, |\lambda|^{2k}$ converges or equivalently $|\lambda| < 1$. Therefore

$$\sigma_p(T_l) = \{ \lambda \in \mathbb{C} : |\lambda| < 1 \}.$$

13.2 Invertible operators

Let us discuss the concept of an inverse operator.

Definition 13.3 (injective operator) *We say that $A : U \to V$ is* injective *if the equation $Ax = y$ has a unique solution for every $y \in \text{Range}(A)$.*

Definition 13.4 (bijective operator) *We say that $A : U \to V$ is* bijective *if the equation $Ax = y$ has exactly one solution for every $y \in V$.*

Definition 13.5 (inverse operator) *We say that A is* invertible *if it is bijective. Then the equation $Ax = y$ has a unique solution for all $y \in V$ and we define $A^{-1}y = x$.*

1. **Exercise:** Show that A^{-1} is a linear operator.
2. **Exercise:** Show that if A^{-1} is invertible, then A^{-1} is also invertible and
$$(A^{-1})^{-1} = A.$$
3. **Exercise:** Show that if A and B are two invertible linear operators, then AB is also invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Proposition 13.6 *A linear operator $A : U \to V$ is invertible iff*

$$\text{Ker}(A) = \{0\} \qquad \textit{and} \qquad \text{Range}(A) = V.$$

Exercise: prove the proposition.

We will use $I_V : V \to V$ to denote the identity operator on V, i.e., $I_V(x) = x$ for all $x \in V$. Moreover, we will skip the subscript V if there is no danger of a mistake. It is easy to see that if $A : U \to V$ is invertible then

$$AA^{-1} = I_V \qquad \text{and} \qquad A^{-1}A = I_U\,.$$

Example: The right shift $T_r : \ell^2 \to \ell^2$ has a trivial kernel and

$$T_l T_r = I.$$

but it is not invertible since $\text{Range}(T_r) \neq \ell^2$. (Indeed, any sequence in the range of T_r has a zero on the first place). Consequently, the equality $AB = I$ along does not imply that $B = A^{-1}$.

Lemma 13.7 *If $A : U \to V$ and $B : V \to U$ are linear operators such that*

$$AB = I_V \qquad \textit{and} \qquad BA = I_U$$

then A and B are both invertible and $B = A^{-1}$.

Proof: The equality $ABy = y$ for all $y \in V$ implies that $\text{Ker}B = \{0\}$ and $\text{Range}A = V$. On the other hand $BAx = x$ for all $x \in U$ implies $\text{Ker}A = \{0\}$ and $\text{Range}B = U$. Therefore both A and B satisfy the definition of invertible operator. $\square$

13.3 Resolvent and spectrum

Let $A : V \to V$ be a linear operator on a vector space V. A complex number λ is an eigenvalue of A if $Ax = \lambda x$ for some $x \neq 0$. This equation is equivalent to $(A - \lambda I)x = 0$. Then we immediately see that $A - \lambda I$ is not invertible since 0 has infinitely many preimages: αx with $\alpha \in \mathbb{C}$.

If V is finite dimensional the reversed statement is also true: if $A - \lambda I$ is not invertible then λ is an eigenvalue of A (recall the Fredholm alternative from the first year Linear Algebra). In the infinite dimensional case this is not necessarily true.

Definition 13.8 (resolvent set and spectrum) The resolvent set *of a linear operator $A : H \to H$ is defined by*

$$R(A) = \{ \lambda \in \mathbb{C} : (A - \lambda I)^{-1} \in B(H,H) \} .$$

The resolvent set consists of regular values. The spectrum *is the complement to the resolvent set in $\mathbb{C}$:*

$$\sigma(A) = \mathbb{C} \setminus R(A) .$$

Note that the definition of the resolvent set assumes existence of the inverse operator $(A - \lambda I)^{-1}$ for $\lambda \in R(A)$. If $\lambda \in \sigma_p(A)$ then $(A - \lambda I)$ is not invertible. Consequently any eigenvalue $\lambda \in \sigma(A)$ and

$$\sigma_p(A) \subseteq \sigma(A) .$$

The spectrum of A can be larger than the point spectrum.

Example: The point spectrum of the right shift operator T_r is empty but since $\operatorname{Range} T_r \neq \ell^2$ it is not invertible and therefore $0 \in \sigma(T_r)$. So $\sigma_p(T_r) \neq \sigma(T_r)$.

Technical lemmas

Lemma 13.9 *If $T \in B(H,H)$ and $\|T\| < 1$, then $(I - T)^{-1} \in B(H,H)$. Moreover*

$$(I - T)^{-1} = I + T + T^2 + T^3 + \dots$$

and

$$\|(I - T)^{-1}\| \leq (1 - \|T\|)^{-1} .$$

Proof: Consider the sequence $V_n = I + T + T^2 + \cdots + T^n$. Since

$$\|T^n x\| \leq \|T\| \, \|T^{n-1} x\|$$

we conclude that $\|T^n\| \leq \|T\|^n$. Consequently for any $m > n$ we have

$$\begin{aligned} \|V_m - V_n\| &= \left\| T^{n+1} + T^{n+2} + \cdots + T^m \right\| \\ &\leq \|T\|^{n+1} + \|T\|^{n+2} + \cdots + \|T\|^m \\ &= \frac{\|T\|^{n+1} - \|T\|^{m+1}}{1 - \|T\|} \leq \frac{\|T\|^{n+1}}{1 - \|T\|} . \end{aligned}$$

Since $\|T\| < 1$, V_n is a Cauchy sequence in the operator norm. The space $B(H,H)$ is complete and there is $V \in B(H,H)$ such that $V_n \to V$. Moreover,

$$\|V\| \leq 1 + \|T\| + \|T\|^2 + \cdots = (1 - \|T\|)^{-1}.$$

Finally, taking the limit as $n \to \infty$ in the equalities

$$\begin{aligned} V_n(I-T) &= V_n - V_nT = I - T^{n+1}, \\ (I-T)V_n &= V_n - TV_n = I - T^{n+1} \end{aligned}$$

and using that $T^{n+1} \to 0$ in the operator norm we get $V(I-T) = (I-T)V = I$. Lemma 13.7 implies $(I-T)^{-1} = V$. □

Lemma 13.10 *Let H be a Hilbert space and $T, T^{-1} \in B(H,H)$. If $U \in B(H,H)$ and $\|U\| < \|T^{-1}\|^{-1}$, then the operator $T+U$ is invertible and*

$$\left\|(T+U)^{-1}\right\| \leq \frac{\|T^{-1}\|}{1 - \|U\|\,\|T^{-1}\|}.$$

Proof: Consider the operator $V = T^{-1}(T+U) = I + T^{-1}U$. Since

$$\|T^{-1}U\| \leq \|T^{-1}\|\,\|U\| < 1,$$

Lemma 13.9 implies that V is invertible and

$$\|V^{-1}\| \leq \left(1 - \|T^{-1}\|\,\|U\|\right)^{-1}.$$

Moreover, since both V and T are invertible

$$I = V^{-1}V = V^{-1}T^{-1}(T+U)$$

implies that $T+U$ is invertible with

$$(T+U)^{-1} = V^{-1}T^{-1}.$$

Finally, $\|(T+U)^{-1}\| \leq \|V^{-1}\|\,\|T^{-1}\|$ implies the desired upper bound for the norm of the inverse operator. □

Properties of the spectrum

Lemma 13.11 *If A is bounded and $\lambda \in \sigma(A)$ then $\bar{\lambda} \in \sigma(A^*)$.*

Proof: If $\lambda \in R(A)$ then $A - \lambda I$ has a bounded inverse:

$$(A - \lambda I)(A - \lambda I)^{-1} = I = (A - \lambda I)^{-1}(A - \lambda I)\,.$$

Taking adjoints we obtain

$$\left((A - \lambda I)^{-1}\right)^* (A^* - \bar{\lambda} I) = I = (A^* - \bar{\lambda} I)\left((A - \lambda I)^{-1}\right)^*.$$

Consequently, $(A^* - \bar{\lambda} I)$ has a bounded inverse $\left((A - \lambda I)^{-1}\right)^*$ (an adjoint of a bounded operator). Therefore $\lambda \in R(A)$ iff $\bar{\lambda} \in R(A^*)$. Since the spectrum is the complement of the resolvent set we also get $\lambda \in \sigma(A)$ iff $\bar{\lambda} \in \sigma(A^*)$. $\square$

Proposition 13.12 *If A is bounded and $\lambda \in \sigma(A)$ then $|\lambda| \le \|A\|_{op}$.*

Proof: Take $\lambda \in \mathbb{C}$ such that $|\lambda| > \|A\|_{\mathrm{op}}$. Since $\|\lambda^{-1}A\|_{\mathrm{op}} < 1$ Lemma 13.9 implies that $I - \lambda^{-1}A$ is invertible and the inverse operator is bounded. Consequently, $A - \lambda I = -\lambda(I - \lambda^{-1}A)$ also has a bounded inverse and so $\lambda \in R(A)$. The proposition follows immediately since $\sigma(A)$ is the complement of $R(A)$. $\square$

Proposition 13.13 *If A is bounded then $R(A)$ is open and $\sigma(A)$ is closed.*

Proof: Let $\lambda \in R(A)$. Then $T = (A - \lambda I)$ has a bounded inverse. Set $U = -\delta I$. Obviously, $\|U\| = |\delta|$. If

$$|\delta| \le \|T^{-1}\|^{-1}$$

Lemma 13.10 implies that $T + U = A - (\lambda + \delta)I$ also has a bounded inverse and so $\lambda + \delta \in R(A)$. Consequently $R(A)$ is open and $\sigma(A) = \mathbb{C} \setminus R(A)$ is closed. $\square$

Example: The spectrum of T_l and of T_r are both equal to the closed unit disk on the complex plane.

Indeed, $\sigma_p(T_l) = \{\,\lambda \in \mathbb{C} : |\lambda| < 1\,\}$. Since $\sigma_p(T_l) \subset \sigma(T_l)$ and $\sigma(T_l)$ is closed, we conclude that $\sigma(T_l)$ includes the closed unit disk. On the other hand, Proposition 13.12 implies that $\sigma(T_l)$ is a subset of the closed disk $|\lambda| \le \|T_l\|_{\mathrm{op}} = 1$. Therefore

$$\sigma(T_l) = \{\,\lambda \in \mathbb{C} : |\lambda| \le 1\,\}\,.$$

Since $T_r = T_l^*$ and $\sigma(T_l)$ is invariant, Lemma 13.11 implies $\sigma(T_r) = \sigma(T_l)$.

13.4 Compact operators

Definition 13.14 *Let X be a normed space and Y be a Banach space. Then a linear operator $A : X \to Y$ is compact if the image of any bounded sequence has a convergent subsequence.*

Obviously a compact operator is bounded. Indeed, otherwise there is a sequence x_n with $\|x_n\| = 1$ such that $\|Ax_n\| > n$ for each n. The sequence Ax_n does not contain a convergent subsequence (it does not even contain a bounded subsequence) and so is not sequentially compact.

Lemma 13.15 *Let X be a normed space and Y be Banach. A linear operator $A : X \to Y$ is compact iff the image of the unit sphere is sequentially compact.*

Example: Any bounded operator with finite-dimensional range is compact. Indeed, in a finite dimensional space any bounded sequence has a convergent subsequence.

Theorem 13.16 *If X is a normed space and Y is a Banach space, then compact linear operators form a closed linear subspace in $B(X,Y)$.*

Proof: If K_1, K_2 are compact operators and $\alpha_1, \alpha_2 \in \mathbb{K}$, then $\alpha_1 K_1 + \alpha_2 K_2$ is also compact. Indeed, take any bounded sequence (x_n) in H. There is a subsequence $x_{n_{1j}}$ such that $K_1 x_{n_{1j}}$ converges. This subsequence is also bounded, so it contains a subsequence $x_{n_{2j}}$ such that $K_2 x_{n_{2j}}$ converges. Obviously $K_1 x_{n_{2j}}$ also converges and therefore $\alpha_1 K_1 x_{n_{2j}} + \alpha_2 K_2 x_{n_{2j}}$ is convergent and consequently $\alpha_1 K_1 + \alpha_2 K_2$ is compact Therefore the compact operators form a linear subspace.

Let us prove that this subspace is closed. Let K_n be a convergent sequence of compact operators: $K_n \to K$ in $B(H,H)$. Take any bounded sequence (x_n) in X. Since K_1 is compact, there is a subsequence $x_{n_{1j}}$ such that $K_1 x_{n_{1j}}$ converges. Since $x_{n_{1j}}$ is bounded and K_2 is compact, there is a subsequence $x_{n_{2j}}$ such that $K_2 x_{n_{2j}}$ converges. Repeat this inductively: for each k there is a subsequence $x_{n_{kj}}$ of the original sequence such that $K_l x_{n_{kj}}$ converges as $j \to \infty$ for all $l \le k$.

Consider the diagonal sequence $y_j = x_{n_{jj}}$. Obviously $(y_j)_{j=k}^{\infty}$ is a subsequence of $(x_{n_{kj}})_{j=1}^{\infty}$. Consequently $K_l y_j$ converges as $j \to \infty$ for every l.

In order to show that K is compact it is sufficient to prove that Ky_j is Cauchy:

$$\begin{aligned} \|Ky_j - Ky_l\| &\le \|Ky_j - K_n y_j\| + \|K_n y_j - K_n y_l\| + \|Ky_l - K_n y_l\| \\ &\le \|K - K_n\| \left(\|y_j\| + \|y_l\|\right) + \|K_n y_j - K_n y_l\| . \end{aligned}$$

Given $\varepsilon > 0$ choose n sufficiently large to ensure that the first term is less than $\frac{\varepsilon}{2}$, then choose N sufficiently large to guarantee that the second term is less than $\frac{\varepsilon}{2}$ for all $j, l > N$. So Ky_j is Cauchy and consequently converges. Therefore K is a compact operator, and the subspace formed by compact operators is closed. □

Proposition 13.17 *The integral operator $A : L^2(a,b) \to L^2(a,b)$ defined by*

$$(Af)(t) = \int_a^b K(t,s)f(s)\,ds \qquad \text{with} \qquad \int_a^b \int_a^b |K(t,s)|^2\,dsdt < \infty$$

is compact.

Proof: Let $\{\,\varphi_k : k \in \mathbb{N}\,\}$ be an orthonormal basis in $L^2(a,b)$. Let $\kappa_{jk} = (A\varphi_j, \varphi_k)$ and $n \in \mathbb{N}$, and define an operator

$$A_n f = \sum_{k=1}^{n} \sum_{j=1}^{n} \kappa_{jk}(f, \varphi_j)\varphi_k\,.$$

Obviously $\dim \operatorname{Range}(A_n) = n$ and consequently A_n is compact. In order to complete the proof we need to show that $A_n \to A$ in the operator norm.[17] □

[17] We will not discuss this proof further.

13.5 Spectral theory for compact self-adjoint operators

Lemma 13.18 *Let H be an infinitely dimensional Hilbert space and $T : H \to H$ a compact self-adjoint operator. Then at least one of $\lambda_\pm = \pm\|T\|_{op}$ is an eigenvalue of T.*

Proof: Assume $T \neq 0$ (otherwise the lemma is trivial). Since

$$\|T\|_{\rm op} = \sup_{\|x\|=1} |(Tx,x)|$$

there is a sequence $x_n \in H$ such that $\|x_n\| = 1$ and $(Tx_n, x_n) \to \pm\|T\|_{\rm op} = \alpha$. Since T is compact, $y_n = Tx_n$ has a convergent subsequence. Relabel this subsequence as x_n and let $y = \lim_{n\to\infty} Tx_n$. Then

$$\|Tx_n - \alpha x_n\|^2 = \|Tx_n\|^2 - 2\alpha(Tx_n,x_n) + \alpha^2 \le 2\alpha^2 - 2\alpha(Tx_n,x_n)\,.$$

The right hand side converges to 0 as $n \to \infty$. Consequently $Tx_n - \alpha x_n \to 0$. On the other hand $Tx_n \to y$ and consequently

$$x_n \to x = \alpha^{-1} y\,.$$

The operator T is continuous and consequently $Tx = \alpha x$. Finally, since $\|x_n\| = 1$ for all n, we have $\|x\| = 1$, and consequently α is an eigenvalue. □

Proposition 13.19 *Let H be an infinitely dimensional Hilbert space and $T : H \to H$ a compact self-adjoint operator. Then $\sigma_p(T)$ is either a finite set or countable sequence tending to zero. Moreover, every non-zero eigenvalue corresponds to a finite dimensional eigenspace.*

Proof: Suppose there is $\varepsilon > 0$ such that T has infinitely many different eigenvalues with $|\lambda_n| > \varepsilon$. Let x_n be corresponding eigenvectors with $\|x_n\| = 1$. Since the operator is self-adjoint, this sequence is orthonormal and for any $n \neq m$

$$\begin{aligned}
\|Tx_n - Tx_m\|^2 &= \|\lambda_n x_n - \lambda_m x_m\|^2 \\
&= (\lambda_n x_n - \lambda_m x_m, \lambda_n x_n - \lambda_m x_m) = |\lambda_n|^2 + |\lambda_m|^2 > 2\varepsilon\,.
\end{aligned}$$

Consequently, (Tx_n) does not have a convergent subsequence. This contradicts to the compactness of T. Consequently, $\sigma_p(T)$ is either finite or a converging to zero sequence.

Now let $\lambda \neq 0$ be an eigenvalue and E_λ the corresponding eigenspace. Let $\tilde{A} : E_\lambda \to E_\lambda$ be the restriction of A onto E_λ. Since $\tilde{A}x = \lambda x$ for any $x \in E_\lambda$, the operator $\tilde{A}$ maps the unit sphere into the sphere of radius λ. Since A is compact, the image of the unit sphere is sequentially compact. Therefore the sphere of radius λ is compact. Since E_λ is a Hilbert (and consequently Banach) space itself, Theorem 3.23 implies that E_λ is finite dimensional. □

Theorem 13.20 (Hilbert-Schmidt theorem) *Let H be a Hilbert space and $T : H \to H$ be a compact self-adjoint operator. Then there is a finite or countable orthonormal sequence (e_n) of eigenvectors of T with corresponding real eigenvalues (λ_n) such that*

$$Tx = \sum_j \lambda_j (x, e_j) e_j \qquad \textit{for all } x \in H.$$

Proof: We construct the sequence e_j inductively. Let $H_1 = H$ and $T_1 = T : H_1 \to H_1$. Lemma 13.18 implies that there is an eigenvector $e_1 \in H_1$ with $\|e_1\| = 1$ and an eigenvalue $\lambda_1 \in \mathbb{R}$ such that $|\lambda_1| = \|T_1\|_{B(H_1,H_1)}$.

Then let $H_2 = \{x \in H_1 : x \perp e_1\}$. If $x \in H_2$ then $Tx \in H_2$. Indeed, since T is self adjoint

$$(Tx, e_1) = (x, Te_1) = \lambda_1 (x, e_1) = 0$$

and $Tx \in H_2$. Therefore the restriction of T onto H_2 is an operator $T_2 : H_2 \to H_2$. Since H_2 is an orthogonal complement, it is closed and so a Hilbert space itself. Lemma 13.18 implies that there is an eigenvector $e_2 \in H_2$ with $\|e_2\| = 1$ and an eigenvalue $\lambda_2 \in \mathbb{R}$ such that $|\lambda_2| = \|T_2\|_{B(H_2,H_2)}$. Then let $H_3 = \{x \in H_2 : x \perp e_2\}$ and repeat the procedure as long as T_n is not zero.

Suppose $T_n = 0$ for some $n \in \mathbb{N}$. Then for any $x \in H$ let

$$y = x - \sum_{j=1}^{n-1} (x, e_j) e_j .$$

Applying T to the equality we get:

$$Ty = Tx - \sum_{j=1}^{n-1} (x, e_j) Te_j = Tx - \sum_{j=1}^{n-1} (x, e_j) \lambda_j e_j .$$

Since $y \perp e_j$ for $j < n$ we have $y \in H_n$ and consequently $Ty = T_n y = 0$. Therefore

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

which is the required formula for T.

Suppose $T_n \neq 0$ for all $n \in \mathbb{N}$. Then for any $x \in H$ and any n consider

$$y_n = x - \sum_{j=1}^{n-1} (x, e_j) e_j .$$

Since $y_n \perp e_j$ for $j < n$ we have $y \in H_n$ and

$$\|x\|^2 = \|y_n\|^2 + \sum_{j=1}^{n-1} |(x, e_j)|^2$$

implies $\|y_n\|^2 \le \|x\|^2$. On the other hand

$$\left\| Tx - \sum_{j=1}^{n-1} (x, e_j)\lambda_j e_j \right\| = \|Ty_n\| \le \|T_n\| \, \|y_n\| \le |\lambda_n| \, \|x\|$$

and since $\lambda_n \to 0$ as $n \to \infty$ we have

$$Tx = \sum_{j=1}^{\infty} (x, e_j)\lambda_j e_j \,. \qquad \square$$

Corollary 13.21 *Let H be an infinite dimensional separable Hilbert space and $T : H \to H$ a compact self-adjoint operator. Then there is an orthonormal basis $E = \{e_j : j \in \mathbb{N}\}$ in H such that $Te_j = \lambda_j e_j$ for all $j \in \mathbb{N}$ and*

$$Tx = \sum_{j=1}^{\infty} \lambda_j (x, e_j) e_j \qquad \textit{for all } x \in H.$$

Exercise: Deduce that operators with finite range are dense among compact self-adjoint operators.

Theorem 13.22 *If H is an infinite dimensional Hilbert space and $T : H \to H$ is a compact self-adjoint operator, then $\sigma(T) = \overline{\sigma_p(T)}$.*

Proposition 13.19 implies that zero is the only possible limit point of $\sigma_p(T)$. Therefore the theorem means that either $\sigma(T) = \sigma_p(T)$ or $\sigma(T) = \sigma_p(T) \cup \{0\}$. In particular, $\sigma(T) = \sigma_p(T)$ if zero is an eigenvalue. Note that if zero is not an eigenvalue, then there is a sequence of eigenvalues which accumulates to zero since H is infinite dimensional.

Proof: According to the Hilbert-Schmidt theorem

$$Tx = \sum_{j=1}^{\infty} \lambda_j (x, e_j) e_j$$

where $\{e_j\}$ is an orthonormal basis in H.[18] Then $x = \sum_{j=1}^{\infty} (x, e_j) e_j$ and for any $\mu \in \mathbb{C}$

$$(T - \mu I)x = \sum_{j=1}^{\infty} (\lambda_j - \mu)(x, e_j) e_j \,.$$

[18]The proof uses a countable basis in H, therefore it assumes that H is separable. The theorem remain valid for a non-separable H but the proof should be slightly modified. The modification is based on the following observation: Proposition 13.19 implies that $(\operatorname{Ker} T)^{\perp}$ has a countable orthonormal basis of eigenvectors $\{e_j\}$. Then for any vector $x \in H$ write $x = P_{\operatorname{Ker} T}(x) + \sum_{j=1}^{\infty} (x, e_j) e_j$ where $P_{\operatorname{Ker} T}$ is the orthogonal projection on the kernel of T. Then follow the arguments of the proof (adding $\mu^{-1} P_{\operatorname{Ker} T}(y)$ to the definition of S).

Let $\mu \in \mathbb{C} \setminus \overline{\sigma_p(T)}$ which is an open subset of $\mathbb{C}$. Consequently there is $\varepsilon > 0$ such that $|\mu - \lambda| > \varepsilon$ for all $\lambda \in \sigma_p(T) \subset \overline{\sigma_p(T)}$. Consider an operator S defined by

$$Sy = \sum_{k=1}^{\infty} \frac{(y, e_k)}{\lambda_k - \mu} e_k .$$

Lemma 7.12 implies that the series converges since $|\lambda_k - \mu| > \varepsilon$ and

$$\|Sy\|^2 = \sum_{k=1}^{\infty} \left| \frac{(y, e_k)}{\lambda_k - \mu} \right|^2 \leq \varepsilon^{-2} \sum_{j=k}^{\infty} |(y, e_k)|^2 = \varepsilon^{-2} \|y\|^2 .$$

In particular we see that S is bounded with $\|S\|_{\text{op}} \leq \varepsilon^{-1}$. Moreover $S = (T - \mu I)^{-1}$. Indeed,

$$(T - \mu I)Sy = \sum_{j=1}^{\infty} (\lambda_j - \mu)(Sy, e_j)e_j = \sum_{j=1}^{\infty} \frac{\lambda_j - \mu}{\lambda_j - \mu} (y, e_j)e_j = y .$$

and

$$S(T - \mu I)x = \sum_{j=1}^{\infty} \frac{((T - \mu I)x, e_j)}{\lambda_j - \mu} e_j = \sum_{j=1}^{\infty} \frac{\lambda_j - \mu}{\lambda_j - \mu} (x, e_j)e_j = x .$$

Then $S = (T - \mu I)^{-1}$ and $\mu \in R(T)$, and so $\sigma(T) \subseteq \overline{\sigma_p(T)}$. On the other hand, $\overline{\sigma_p(T)} \subseteq \sigma(T)$. We conclude $\sigma(T) = \overline{\sigma_p(T)}$. $\square$

14 Sturm-Liouville problems

In this chapter we will study the Sturm-Liouville problem:

$$-\frac{d}{dx}\left(p(x)\frac{du}{dx}\right)+q(x)u=\lambda u \qquad \text{with } u(a)=u(b)=0$$

where p and q are given function. The values of λ for which the problem has a non-trivial solution are called eigenvalues of the Sturm-Liouville problem and the corresponding solutions u are called eigenfunctions. We will prove that the eigenfunctions form an orthonormal basis in $L^2(a,b)$.

We assume that $p\in C^1[a,b]$, $q\in C[a,b]$ and

$$p(x)>0 \qquad \text{and} \qquad q(x)\ge 0 \qquad \text{for } x\in[a,b].$$

For a function $u\in C^2[a,b]$ we define

$$L(u)=-\frac{d}{dx}\left(p(x)\frac{du}{dx}\right)+q(x)u\,.$$

Obviously $L:C^2[a,b]\to C^0[a,b]$ is linear. We will see that $\operatorname{Range}L=C^0[a,b]$ and $\operatorname{Ker}L=\operatorname{Span}\{\,u_1,u_2\,\}$ where u_1,u_2 are two linear independent solutions of the equation $Lu=0$. Therefore the operator is not invertible.

We will restrict L onto the space

$$D(L):=\left\{\,u\in C^2[a,b]:u(a)=u(b)=0\,\right\}$$

and show that $L:D(L)\to C^0[a,b]$ is invertible. We will prove that L^{-1} is a restriction on $\operatorname{Range}L$ of a compact self-adjoint operator $A:L^2(a,b)\to L^2(a,b)$. The Sturm-Liouville theorem states that the eigenfunction of A form an orthonormal basis in $L^2(a,b)$. Moreover, we will see that A and L have the same eigenfunctions.

Lemma 14.1 *If both u_1 and u_2 are solutions of the equation $Lu=0$, then*

$$W_p(u_1,u_2)(x)=p(x)(u_1'(x)u_2(x)-u_1(x)u_2'(x))$$

is constant. Moreover, the solutions u_1 and u_2 are linearly independent iff

$$W_p(u_1,u_2)\neq 0.$$

Proof: The equation $Lu=0$ implies $pu''=-p'u'+qu$. Then differentiate W with respect to x:

$$\begin{aligned} W_p' &= p'(u_1'u_2-u_1u_2')+p(u_1''u_2-u_1u_2'') \\ &= p'(u_1'u_2-u_1u_2')+((-p'u_1'+qu_1)u_2-(-p'u_2'+qu_2)u_1)=0\,. \end{aligned}$$

Therefore W_p is constant.

Finally, since $p \neq 0$, $W_p = 0$ implies $u_1'(x)u_2(x) - u_1(x)u_2'(x) = 0$ for all x. Consequently at all points where u_1 and u_2 do not vanish

$$\frac{u_1'}{u_1} = \frac{u_2'}{u_2} \qquad \Leftrightarrow \qquad \frac{d\ln|u_1|}{dx} = \frac{d\ln|u_2|}{dx},$$

which implies that $u_1 = Cu_2$ for some constant C, i.e., if $W_p = 0$ then u_1 and u_2 are not linearly independent. In the reverse direction the statement is straightforward. □

Lemma 14.2 *The equation $Lu = 0$ has two linearly independent solutions, $u_1, u_2 \in C^2[a,b]$, such that $u_1(a) = u_2(b) = 0$.*

Proof: Let u_1, u_2 be solutions of the Cauchy problems

$$\begin{aligned} Lu_1 &= 0 \qquad u_1(a) = 0,\ u_1'(a) = 1, \\ Lu_2 &= 0 \qquad u_2(b) = 0,\ u_2'(b) = 1. \end{aligned}$$

According to the theory of linear ordinary differential equations u_1 and u_2 exist, belong to $C^2[a,b]$ and are unique.

Moreover, u_1 and u_2 are linearly independent. Indeed, suppose $Lu = 0$ for some $u \in C^2[a,b]$ such that $u(a) = u(b) = 0$. Then

$$\begin{aligned} 0 = (Lu,u) &= \int_a^b \left(-(pu')'u + qu^2\right) dx \\ &= p(x)u'(x)u(x)\Big|_a^b + \int_a^b \left(p(u')^2 + qu^2\right) dx \quad \text{(integration by parts)} \\ &= \int_a^b \left(p(u')^2 + qu^2\right) dx \end{aligned}$$

Since $p > 0$ on $[a,b]$, we conclude that $u' \equiv 0$. Then $u(a) = u(b) = 0$ implies $u(x) = 0$ for all $x \in [a,b]$.

Consequently since $u_2(b) = 0$ and u_2 is not identically zero, $u_2(a) \neq 0$ and so

$$W_p(u_1,u_2) = p(a)(u_1'(a)u_2(a) - u_1(a)u_2'(a)) = p(a)u_1'(a)u_2(a) \neq 0.$$

Therefore u_1, u_2 are linearly independent by Lemma 14.1. □

Proposition 14.3 *If u_1 and u_2 are linearly independent solutions of the equation $Lu = 0$ such that $u_1(a) = u_2(b) = 0$ and*

$$G(x,y) = \frac{1}{W_p(u_1,u_2)} \begin{cases} u_1(x)u_2(y), & a \leq x < y \leq b, \\ u_1(y)u_2(x), & a \leq y \leq x \leq b, \end{cases}$$

then for any $f \in C^0[a,b]$ the function

$$u(x) = \int_a^b G(x,y) f(y)\, dy$$

belongs to $C^2[a,b]$, satisfies the equation $Lu = f$ and the boundary conditions $u(a) = u(b) = 0$.

Proof: The statement is proved by a direct substitution of

$$u(x) = \frac{u_2(x)}{W_p(u_1,u_2)} \int_a^x u_1(y) f(y)\, dy + \frac{u_1(x)}{W_p(u_1,u_2)} \int_x^b u_2(y) f(y)\, dy$$

into the differential equation. Moreover, since $u_1(a) = u_2(b) = 0$, we get $u(a) = u(b) = 0$. □

Lemma 14.4 *The operator $A : L^2(a,b) \to L^2(a,b)$ defined by*

$$(Af)(x) = \int_a^b G(x,y) f(y)\, dy\,.$$

is compact and self-adjoint. Moreover, Range$A$ *is dense in $L^2(a,b)$,* Ker$A = \{0\}$ *and all its eigenfunctions, $Au = \mu u$, belong to $C^2[a,b]$ and satisfy $u(a) = u(b) = 0$.*

Proof: The operator A is compact by Proposition 13.17. Moreover, G is symmetric and so A is self-adjoint. Proposition 14.3 implies the range of A contains all functions from $C^2[a,b]$ such that $u(a) = u(b) = 0$. This set is dense in $L^2(a,b)$.

Now suppose $Au = 0$ for some $u \in L^2[a,b]$. Then for any $v \in L^2$

$$0 = (Au, v) = (u, Av)\,,$$

which implies $u = 0$ because u is orthogonal to a dense set.

Finally, let u be an eigenfunction of A, i.e., $Au = \mu u$. Since $\mu \neq 0$ we can write $u = \mu^{-1} Au = \mu^{-2} A^2 u$. A priori $u \in L^2(a,b)$, then Au is continuous and $A^2 u$ is C^2. Therefore, the eigenvectors of A are smooth. □

Proposition 14.5 *The operator $L : D(L) \to C^0[a,b]$ has a bounded inverse (in the operator norm induced by the L^2 norm on both spaces).*

Proof: Lemma 14.2 implies that Ker$L = \{0\}$. Consequently, L is injective. Proposition 14.3 implies that Range$L = C^0[a,b]$ and the inverse operator is defined by

$$(L^{-1} f)(x) = \int_a^b G(x,y) f(y)\, dy\,.$$

Lemma 14.4 states that this operator is bounded. In other words, L^{-1} coincides with the restriction of A onto Range$L$. □

Theorem 14.6 *The eigenfunctions of the Sturm-Liouville problem form an orthonormal basis in $L^2(a,b)$.*

Proof: Since $A : L^2(a,b) \to L^2(a,b)$ is compact and self adjoint, Theorem 13.21 implies that its eigenvectors form an orthonormal basis in $L^2(a,b)$. If u is an eigenfunction of A, then $Au = \mu u$, $\mu \neq 0$ and $u \in C^2[a,b]$. Consequently $Lu = \lambda u = \mu^{-1} u$, i.e., u is also an eigenvector of L which corresponds to the eigenvalue $\lambda = \mu^{-1}$. □

Example: An application for Fourier series. Consider the Strum-Liouville problem

$$-\frac{d^2u}{dx^2} = \lambda u, \qquad u(0) = u(1) = 0\,.$$

It corresponds to the choice $p = 1$, $q = 0$. Theorem 14.6 implies that the normalised eigenfunctions of this problem form an orthonormal basis in $L^2(0,1)$. In this example the eigenfunctions are easy to find:

$$\left\{ \frac{1}{\sqrt{2}} \sin k\pi x : k \in \mathbb{N} \right\}.$$

Consequently any function $f \in L^2(0,1)$ can be written in the form

$$f(x) = \sum_{k=1}^{\infty} \alpha_k \sin k\pi x$$

where

$$\alpha_k = \frac{1}{2} \int_0^1 f(x) \sin k\pi x \, dx\,.$$

The series converges in the L^2 norm.