Functional Analysis I

James C. Robinson

Introduction

I hope that these notes will be useful. They are, of course, much more wordy than the notes you will have taken in lectures, but the maths itself is usually done in a little more detail and should generally be 'tighter'. You may find that the order in which the material is presented is a little different to the lectures, but this should make things more coherent.

Solutions to the examples sheets will follow separately.

I hope that there are relatively few mistakes, but if you find yourself staring at something thinking that it must be wrong then it most likely is, so do email me at `j.c.robinson@warwick.ac.uk`. I will post a list of errata as and when people find them on my webpage for the course,

`www.maths.warwick.ac.uk/~jcr/FAI`.

The section on Lebesgue integration here is a little less detailed than what was in the lectures. I will post a somewhat more detailed version of the above webpage during the holidays, but this material is not examinable.

These notes will form the basis of the first part of a textbook on functional analysis, so any general comments would also be welcome.

Contents

1

Vector spaces

In all that follows we use $\mathbb{K}$ to denote $\mathbb{R}$ or $\mathbb{C}$, although one can define vector spaces over arbitrary fields.

1.1 Vector spaces and bases

$\mathbb{R}^n$ is the simplest and most natural example of a vector space. We give a formal definition, but it is the closure property inherent in the definitions,

$$f + \lambda g \in V, \qquad f, g \in V, \quad \lambda \in \mathbb{K} = \mathbb{R} \text{ or } \mathbb{C},$$

that one usually has to check.

Definition 1.1 *A vector space V over $\mathbb{K}$ is a set V with operations $+ : V \times V \to V$ and $* : \mathbb{K} \times V \to V$ such that*

- *additive and multiplicative identities exist: there exists a zero element $0 \in V$ such that $x + 0 = x$ for all $x \in V$; and $1 \in \mathbb{K}$ is the identity for scalar multiplication, $1 * x = x$ for all $x \in V$;*
- *there are additive inverses: for every $x \in V$ there exists an element $-x \in V$ such that $x + (-x) = 0$;*
- *addition is commutative and associative, $x+y = y+x$ and $x+(y+z) = (x+y)+z$ for all $x, y, z \in V$;*
- *multiplication is associative*

$$\alpha * (\beta * x) = (\alpha\beta) * x \qquad \textit{for all} \qquad \alpha, \beta \in \mathbb{K}, \; x \in V$$

and distributive

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

for all $\alpha, \beta \in \mathbb{K}$, $x, y \in V$.

The multiplication operator $*$ can usually be understood and so we generally drop this notation.

As remarked above we will only consider $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ here, and will refer to real or complex vector spaces respectively. Generally we will omit the word 'real' or 'complex' unless wishing to make an explicit distinction between real and complex vector spaces.

Examples: $\mathbb{R}^n$ is a real vector space over $\mathbb{R}$; it is not a vector space over $\mathbb{C}$ (since $\mathrm{i} * \mathbf{x} \notin \mathbb{R}^n$ for any $\mathbf{x} \in \mathbb{R}^n$); $\mathbb{C}^n$ is a vector space over both $\mathbb{R}$ and $\mathbb{C}$ (so if we take $\mathbb{K} = \mathbb{R}$ the space $\mathbb{C}^n$ can be thought of, somewhat unnaturally, as a 'real vector space').

Example 1.2 *Define the space* $\ell^2(\mathbb{K})$ *of all* square summable *sequences with elements in* $\mathbb{K}$ *(recall that* $\mathbb{K} = \mathbb{R}$ *or* $\mathbb{C}$*):*

$$\ell^2(\mathbb{K}) = \{\underline{x} = (x_1, x_2, \ldots) : \ x_j \in \mathbb{K}, \ \sum_{j=1}^{\infty} |x_j|^2 < +\infty\}.$$

For $\underline{x}, \underline{y} \in \ell^2(\mathbb{K})$ *set*

$$\underline{x} + \underline{y} = (x_1 + y_1, x_2 + y_2, \ldots),$$

and for $\alpha \in \mathbb{K}$, $\underline{x} \in \ell^2$, *define*

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

With these definitions $\ell^2(\mathbb{K})$ *is a vector space. The only issue is whether* $\underline{x} + \underline{y}$ *is still in* $\ell^2(\mathbb{K})$*; but this follows since*

$$\sum_{j=1}^{n} |x_j + y_j|^2 \le \sum_{j=1}^{n} 2|x_j|^2 + 2|y_j|^2 \le 2\sum_{j=1}^{\infty} |x_j|^2 + 2\sum_{j=1}^{\infty} |y_j|^2 < +\infty.$$

Sometimes we will simply write ℓ^2 for $\ell^2(\mathbb{R})$.

Example 1.3 *The space* $C^0([0,1])$ *of all real-valued continuous functions on the interval* $[0,1]$ *is a vector space with the obvious definitions of addition* *and scalar multiplication, which we give here for the one and only time: for $f, g \in C^0([0,1])$ and $\alpha \in \mathbb{R}$, we denote by $f+g$ the function whose values are given by*

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

and by αf the function whose values are

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

Example 1.4 *Denote by $\tilde{L}^2(0,1)$ the set of all real-valued continuous functions on $(0,1)$ for which*

$$\int_0^1 |f(x)|^2 \,\mathrm{d}x < +\infty.$$

Then $\tilde{L}^2(0,1)$ is a vector space (with the obvious definitions of addition and scalar multiplication).

The only thing to check here is that $f + \lambda g \in \tilde{L}^2(0,1)$ whenever $f, g \in \tilde{L}^2(0,1)$ and $\lambda \in \mathbb{R}$. Clearly $f + \lambda g \in C^0(0,1)$, and we have

$$\begin{aligned} \int_0^1 |(f+\lambda g)(x)|^2 \,\mathrm{d}x &= \int_0^1 |f(x)|^2 + 2|\lambda||f(x)||g(x)| + |\lambda|^2 |g(x)|^2 \,\mathrm{d}x \\ &\le 2\left(\int_0^1 |f(x)|^2 + |\lambda|^2 \int_0^1 |g(x)|^2 \,\mathrm{d}x\right) < +\infty. \end{aligned}$$

Note that if $f \in C^0([0,1])$ then, since it is a continuous function on a closed bounded interval, it is bounded and attains its bounds. It follows that for some $M \ge 0$, $|f(x)| \le M$ for all $x \in [0,1]$, and so

$$\int_0^1 |f(x)|^2 \,\mathrm{d}x \le M < \infty,$$

i.e. $f \in \tilde{L}^2(0,1)$.

But while the function $f(x) = x^{-1/4}$ is not continuous on $[0,1]$, it is continuous on $(0,1)$ and

$$\int_0^1 |x^{-1/4}|^2 \,\mathrm{d}x = \int_0^1 x^{-1/2} \,\mathrm{d}x = \left[2x^{1/2}\right]_0^1 = 2 < \infty,$$

so $f \in \tilde{L}^2(0,1)$. These two examples show that $C^0([0,1])$ is a strict subset of $\tilde{L}^2(0,1)$.

We now discuss spanning sets, linear independence, and bases. Note that the definitions – and the following arguments – also apply to infinite-dimensional spaces. In particular the result of Lemma 1.9 is valid for infinite-dimensional spaces.

Definition 1.5 *The* linear span *of a subset E of a vector space V is the collection of all finite linear combinations of elements of E:*

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

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.

Note that this definition requires v to be expressed as a *finite* linear combination of elements of E. When discussing bases for abstract vector spaces with no additional structure the only option is to take finite linear combinations, since these are defined using only the axioms for a vector space (scalar multiplication and addition of vector space elements). In order to take infinite linear combinations we require some way to discuss convergence, which is not available in a general vector space.

Definition 1.6 *A set E is* linearly independent *if any finite collection of elements of E is linearly independent*

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

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

Definition 1.7 *A* basis *for V is an linearly independent spanning set.*

Expansions in terms of basis elements are unique:

Lemma 1.8 *If E is a basis for V then any element of V can be written uniquely in the form*

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

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

For a proof see Linear Algebra.

If E is a linearly independent set that spans V then it is not possible to find an element of V that can be added to E to obtain a larger linearly independent set (otherwise E would not span V). We now show that this can be reversed.

Lemma 1.9 *If $E \subset V$ is maximal linearly independent set, i.e. a linearly independent set E such that $E \cup \{v\}$ is not linearly independent for any $v \in V \setminus E$. Then E is a basis for V.*

Proof Suppose that E does not span V: in particular take $v \in V$ that cannot be written as any finite linear combination of elements of E. To obtain a contradiction, choose $n \in \mathbb{N}$ and $\{e_j\}_{j=1}^n$ with $e_j \in E$, and suppose that

$$\sum_{j=1}^{n} \alpha_j e_j + \alpha_{n+1} v = 0.$$

Since v cannot be written as a sum of any finite collection of the $\{e_j\}$, we must have $\alpha_{n+1} = 0$, which leaves $\sum_{j=1}^n \alpha_j e_j = 0$. However, $\{e_j\}$ is a finite subset of E and is thus linearly independent by assumption, and so we must have $\alpha_j = 0$ for all $j = 1, \ldots, n+1$. But this says that $E \cup \{v\}$ is linearly independent, a contradiction. So E spans V. $\square$

We recall here the following fundamental theorem:

Theorem 1.10 *Suppose that V has a basis consisting of a finite number of elements. Then every basis of V contains the same number of elements.*

This allows us to make the following definition:

Definition 1.11 *If V has a basis consisting of a finite number of elements then the* dimension *of V is the number of elements in this basis. If V has no finite basis then V is* infinite-dimensional.

Since a basis is a maximal linearly independent set (Lemma 1.9), it follows that a space is infinite-dimensional iff for every $n \in \mathbb{N}$ one can find a set of n linearly independent elements of V.

Example 1.12 *For any $n \in \mathbb{N}$ the n elements in $\ell^2(\mathbb{K})$ given by*

$$(1,0,0,0,\ldots),\ (0,1,0,0,\ldots),\ (0,0,...,0,1,0,\ldots),$$

i.e. elements $\underline{e}^{(j)}$ with $e_j^{(j)} = 1$ and $e_i^{(j)} = 0$ if $i \neq j$, are linearly independent. It follows that $\ell^2(\mathbb{K})$ is an infinite-dimensional vector space.

Example 1.13 *Consider the functions $f_n \in C^0([0,1])$, where f_n is zero for $x \notin I_n = [2^{-n} - 2^{-(n+2)}, 2^{-n} + 2^{-(n+2)}]$ and interpolates linearly between the values*

$$f_n(2^{-n} - 2^{-(n+2)}) = 0 \qquad f_n(2^{-n}) = 1 \qquad f_n(2^{-n} + 2^{-(n+2)}) = 0.$$

The intervals I_n where $f_n \neq 0$ are disjoint, but $f(2^{-n}) = 1$. It follows that for any n the $\{f_j\}_{j=1}^n$ are linearly independent, and so $C^0([0,1])$ is infinite-dimensional.

We end this section with the following powerful-looking theorem:

Theorem 1.14 *Every vector space has a basis.*

However, this theorem is almost immediate from the definition of a finite-dimensional vector space, while the proof for infinite-dimensional spaces relies on Zorn's Lemma and is non-constructive: such an abstract basis (whose existence is assured but which we cannot construct in general) is not very useful. [The theorem is equivalent, using Lemma 1.9, to the statement that every vector space contains a maximal linearly independent set. Zorn's Lemma is a tool for guaranteeing the existence of 'maximal' elements.]

In the case of an infinite-dimensional space the type of basis we have discussed in this section (every $v \in V$ can be expressed as a *finite* linear combination of basis elements) is usually referred to as a *Hamel basis*.

Exercise 1.15 *Show that the space ℓ_f consisting of all sequences that contain only finitely many non-zero terms is a vector space. Show that*

$$\underline{e}_j = (0,\ldots,0,1,0,\ldots)$$

(all zeros except for a single 1 in the jth position) is a Hamel basis for ℓ_f.

This is a very artificial example. No Banach space (a particularly nice kind of vector space, see later) can have a countable Hamel basis.

2

Norms and normed spaces

2.1 Norms and normed spaces

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

(i) $\|x\| \geq 0$ *with equality iff* $x = 0$*;*

(ii) $\|\alpha x\| = |\alpha|\|x\|$*; and*

(iii) $\|x + y\| \leq \|x\| + \|y\|$ *('the triangle inequality').*

A vector space equipped with a norm is called a normed space*.*

Strictly a normed space should be written $(V, \|\cdot\|_V)$ where $\|\cdot\|_V$ is the particular norm on V. However, many normed spaces have standard norms, and so often the norm is not specified. E.g. unless otherwise stated, $\mathbb{R}^n$ is equipped with the standard norm

$$\|\mathbf{x}\| = \left(\sum_{j=1}^{n} |x_j|^2\right)^{1/2}. \tag{2.1}$$

However, others are possible, such as

$$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,n} |x_i| \qquad \text{and} \qquad \|\mathbf{x}\|_1 = \sum_i |x_i|.$$

Exercise 2.2 *Show that $\|\cdot\|$, $\|\cdot\|_\infty$, and $\|\cdot\|_1$ are all norms on $\mathbb{R}^n$.*

Example 2.3 *The standard norm on $\ell^2(\mathbb{K})$ is*

$$\|\underline{x}\| = \left(\sum_{j=1}^{\infty} |x_j|^2\right)^{1/2},$$

where $\underline{x} = (x_1, x_2, x_3, \ldots)$. Note that when $\mathbb{K} = \mathbb{R}$ this is the natural extension of the standard Euclidean norm to a countable collection of real numbers.

One of our main concerns in what follows will be normed spaces that consist of functions. For example, the following are norms on $C^0([0,1])$, the space of all continuous functions on $[0,1]$: the 'sup(remum) norm',

$$\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$$

(convergence in this norm is equivalent to uniform convergence) and the L^1 and L^2 norms

$$\|f\|_{L^1} = \int_0^1 |f(x)|\,dx \qquad \text{and} \qquad \|f\|_{L^2} = \left(\int_0^1 |f(x)|^2\,\mathrm{d}x\right)^{1/2}.$$

Note that of the three candidates here the L^2 norm looks most like the expression (2.1) for the familiar norm in $\mathbb{R}^n$.

Lemma 2.4 $\|\cdot\|_{L^1}$ *is a norm on $C^0([0,1])$.*

Proof The only part that requires much thought is (i), to make sure that $\|f\|_{L^1} = 0$ iff $f = 0$. So suppose that $f \neq 0$. Then $|f(y)| = \delta > 0$ for some $y \in (0,1)$ (if $f(0) \neq 0$ or $f(1) \neq 0$ it follows from continuity that $f(y) \neq 0$ for some $y \in (0,1)$). Since f is continuous, there exists an $\epsilon > 0$ such that for any $x \in (0,1)$ with $|x-y| < \epsilon$ we have

$$|f(x) - f(y)| < \delta/2.$$

If necessary, reduce ϵ so that $[y-\epsilon, y+\epsilon] \in (0,1)$. Then

$$\int_0^1 |f(x)|^2\,\mathrm{d}x \geq \int_{y-\epsilon}^{y+\epsilon} |f(x)|^2\,\mathrm{d}x \geq \int_{y-\epsilon}^{y+\epsilon} \frac{\delta}{2}\,\mathrm{d}x = \epsilon\delta > 0.$$

(ii) and (iii) are clear since

$$\|\alpha f\|_{L^1} = \int |\alpha f(x)|\,\mathrm{d}x = |\alpha| \int |f(x)|\,\mathrm{d}x = |\alpha|\|f\|_{L^1}$$

and

$$\|f+g\|_{L^1} = \int |f(x)+g(x)|\,\mathrm{d}x \le \int |f(x)|+|g(x)|\,\mathrm{d}x \le \|f\|_{L^1} + \|g\|_{L^1}.$$

□

For $\|\cdot\|_{L^2}$ (i) and (ii) are the same as above; we will see (iii) below as a consequence of the Cauchy-Schwarz inequality for inner products.

Definition 2.5 *Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space V are* equivalent *if there exist constants $0 < c_1 \le c_2$ such that*

$$c_1\|x\|_1 \le \|x\|_2 \le c_2\|x\|_1 \qquad \textit{for all} \qquad x \in V.$$

It is clear that the above notion of 'equivalence' is reflexive. It is also transitive:

Lemma 2.6 *Suppose that $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_3$ are all norms on a vector space V, and that $\|\cdot\|_2$ and $\|\cdot\|_3$ are both equivalent to $\|\cdot\|_1$. Then $\|\cdot\|_2$ and $\|\cdot\|_3$ are equivalent.*

Proof There exist constants $0 < \alpha_1 \le \alpha_2$ and $0 < \beta_1 \le \beta_2$ such that

$$\alpha_1\|x\|_2 \le \|x\|_1 \le \alpha_2\|x\|_2 \qquad \text{and} \qquad \beta_1\|x\|_3 \le \|x\|_1 \le \beta_2\|x\|_3,$$

and so

$$\|x\|_2 \ge \alpha_2^{-1}\|x\|_1 \ge \beta_1\alpha_2^{-1}\|x\|_3$$

and

$$\|x\|_2 \le \alpha_1^{-1}\|x\|_1 \le \beta_2\alpha_1^{-1}\|x\|_3,$$

i.e. $\beta_1\alpha_2^{-1}\|x\|_3 \le \|x\|_2 \le \beta_2\alpha_1^{-1}\|x\|_3$ and $\|\cdot\|_2$ and $\|\cdot\|_3$ are equivalent. □

Exercise 2.7 *Show that the norms $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ on $\mathbb{R}^n$ are all equivalent.*

This is a particular case of the general result that all norms on a finite-dimensional vector space are equivalent, which we will prove in the following chapter. As part of this proof, the following proposition – which shows that one can always find a norm on a finite-dimensional vector space – will be useful.

Proposition 2.8 *Let V be an n-dimensional vector space, and $E = \{e_j\}_{j=1}^n$ a basis for V. Define a map $\|\cdot\|_E : V \to [0,\infty)$ by*

$$\left\|\sum_{j=1}^n \alpha_j e_j\right\|_E = \left(\sum_{j=1}^n |\alpha_j|^2\right)^{1/2}$$

(taking the positive square root). Then $\|\cdot\|_E$ is a norm on V.

Proof First, note that any $v \in V$ can be written uniquely as $v = \sum_j \alpha_j e_j$, so the map $v \mapsto \|v\|_E$ is well-defined. We check that $\|\cdot\|_E$ satisfies the three requirements of a norm:

(i) clearly $\|v\|_E \geq 0$, and if $\|v\|_E = 0$ then $v = \sum \alpha_j e_j$ with $\sum |\alpha_j|^2 = 0$; i.e. $\alpha_j = 0$ for $j = 1, \ldots, n$, and so $v = 0$.

(ii) If $v = \sum_j \alpha_j e_j$ then $\lambda v = \sum_j (\lambda\alpha_j)e_j$, and so

$$\|\lambda v\|_E^2 = \sum_j |\lambda\alpha_j|^2 = |\lambda|^2 \sum_j |\alpha_j|^2 = |\lambda|^2\|v\|_E^2.$$

(iii) For the triangle inequality, if $u = \sum_j \alpha_j e_j$ and $v = \sum_j \beta_j e_j$ then, using the Cauchy-Schwarz inequality[1]

$$\begin{aligned}
\|u+v\|_E^2 &= \left\|\sum_j (\alpha_j + \beta_j)e_j\right\|^2 \\
&= \sum_j |\alpha_j + \beta_j|^2 \\
&= \sum_j |\alpha_j|^2 + \alpha_j\overline{\beta_j} + \overline{\alpha_j}\beta_j + |\beta_j|^2 \\
&= \|u\|_E^2 + \sum_j \alpha_j\overline{\beta_j} + \sum_j \overline{\alpha_j}\beta_j + \|v\|_E^2 \\
&\leq \|u\|_E^2 + \left(\sum_j |\alpha_j|^2\right)^{1/2}\left(\sum_j |\beta_j|^2\right)^{1/2} + \|v\|_E^2 \\
&= \|u\|_E^2 + 2\|u\|_E\|v\|_E + \|v\|_E^2 \\
&= (\|u\|_E + \|v\|_E)^2,
\end{aligned}$$

[1] For $a_j, b_j \in \mathbb{C}$,

$$\left|\sum_{j=1}^n a_j b_j\right| \leq \left(\sum_j |a_j|^2\right)^{1/2}\left(\sum_j |b_j|^2\right)^{1/2}$$

We will see a proof of this in Chapter 6.

i.e. $\|u+v\|_E \leq \|u\|_E + \|v\|_E$.

$\square$

2.2 Convergence

In a normed space we can measure the distance between x and y using $\|x-y\|$. So we can define notions of convergence and continuity using this idea of distance:

Definition 2.9 *A sequence $\{x_k\}_{k=1}^{\infty}$ in a normed space X* converges *to a limit $x \in X$ if for any $\epsilon > 0$ there exists an N such that*

$$\|x_k - x\| < \epsilon \qquad \textit{for all} \qquad n \geq N.$$

This definition is sensible in that limits are unique:

Exercise 2.10 *Show that the limit of a convergent sequence is unique.*

The following result shows that if $x_n \to x$ then the norm of x_n converges to the norm of x. This will turn out to be a very useful observation.

Lemma 2.11 *If $x_n \to x$ in $(X, \|\cdot\|)$ then $\|x_n\| \to \|x\|$.*

Proof The triangle inequality gives

$$\|x_n\| \leq \|x\| + \|x_n - x\| \qquad \text{and} \qquad \|x\| \leq \|x_n\| + \|x - x_n\|$$

which implies that

$$\Big| \|x_n\| - \|x\| \Big| \leq \|x_n - x\|.$$

$\square$

Two equivalent norms give rise to the same notion of convergence:

Lemma 2.12 *Suppose that $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on a space X. Then*

$$\|x_n - x\|_1 \to 0 \qquad \textit{iff} \qquad \|x_n - x\|_2 \to 0,$$

i.e. convergence in one norm is equivalent to convergence in the other, with the same limit.

The proof of this lemma is immediate from the definition of the equivalence of norms, since there exist constants $0 < c_1 \le c_2$ such that

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

Using convergence we can also define continuity:

Definition 2.13 *A map $f : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ is* continuous *if*

$$x_n \to x \text{ in } X \quad \Rightarrow \quad f(x_n) \to f(x) \text{ in } Y,$$

i.e. if

$$\|x_n - x\|_X \to 0 \quad \Rightarrow \quad \|f(x_n) - f(x)\|_Y \to 0.$$

Exercise 2.14 *Show that this is equivalent to the ϵ–δ definition of continuity: for each $x \in X$, for every $\epsilon > 0$ there exists a $\delta > 0$ such that*

$$\|y - x\|_X < \epsilon \qquad \Rightarrow \qquad \|f(y) - f(x)\|_Y < \delta.$$

Lemma 2.12 has an immediate implication for continuity:

Corollary 2.15 *Suppose that $\|\cdot\|_{X,1}$ and $\|\cdot\|_{X,2}$ are two equivalent norms on a space X, and $\|\cdot\|_{Y,1}$ and $\|\cdot\|_{Y,2}$ are two equivalent norms on a space Y. Then a function $f : (X, \|\cdot\|_{X,1}) \to (Y, \|\cdot\|_{Y,1})$ is continuous iff it is continuous as a map from $(X, \|\cdot\|_{X,2})$ into $(Y, \|\cdot\|_{Y,2})$.*

We remarked above that all norms on a finite-dimensional space are equivalent, which means that there is essentially only one notion of 'convergence' and of 'continuity'. But in infinite-dimensional spaces there are distinct norms, and the different notions of convergence implied by the norms we have introduced for continuous functions are not equivalent, as we now show.

First we note that convergence of $f_k \in C^0([0,1])$ to f in the supremum norm, i.e.

$$\sup_{x\in[0,1]} |f_k(x) - f(x)| \to 0 \tag{2.2}$$

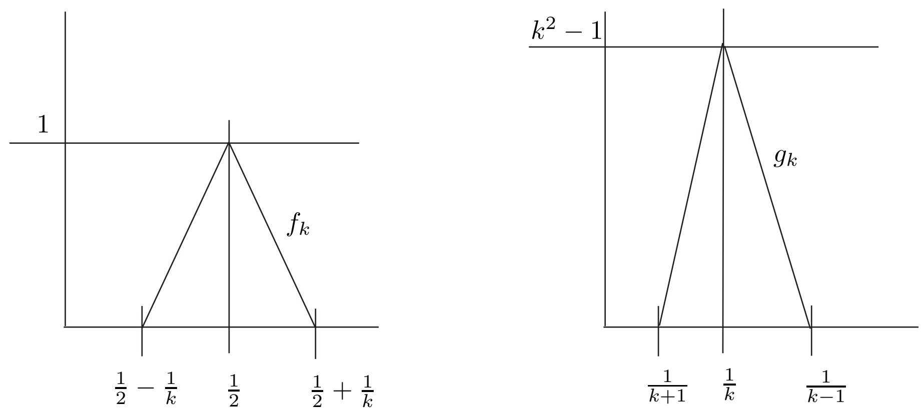

Fig. 2.1. (a) definition of f_k and (b) definition of g_k.

implies that $f_k \to f$ in the L^1 norm, since clearly

$$\int_0^1 |f_k(y) - f(y)| \, \mathrm{d}y \leq \int_0^1 \left(\sup_{x \in [0,1]} |f_k(x) - f(x)| \right) \mathrm{d}y = \sup_{x \in [0,1]} |f_k(x) - f(x)|.$$

This inequality should make very clear the advantage of the shorthand norm notation, since it just says

$$\|f_k - f\|_{L^1} \leq \|f_k - f\|_\infty.$$

It is also clear that if (2.2) holds then $f_k(x) \to f(x)$ for each $x \in [0,1]$, which is 'pointwise convergence'. However, neither pointwise convergence nor L^1 convergence imply uniform convergence:

Example 2.16 *Consider the sequence of functions $\{f_k\}$ as illustrated in Figure 2.2(a). Then $f_k \to 0$ in the L^1 norm, since*

$$\|f_k - 0\|_{L^1} = \|f_k\|_{L^1} = \frac{1}{k}.$$

However, $f_k \not\to 0$ pointwise, since $f_k(\frac{1}{2}) = 1$ for all k.

Example 2.17 *Consider the sequence of functions $\{g_k\}$ as illustrated in Figure 2.2(b). Then $f_k \to 0$ pointwise, but*

$$\|f_k\|_{L^1} = \tfrac{1}{2}(k^2 - 1) \left[\left(\frac{1}{k-1} - \frac{1}{k} \right) + \left(\frac{1}{k} - \frac{1}{k+1} \right) \right] = 1,$$

so $f_k \not\to 0$ in the L^1 norm.

3

Compactness and equivalence of norms

3.1 Compactness

One fundamental property of the real numbers is expressed by the Bolzano-Weierstrass Theorem:

Theorem 3.1 (Bolzano-Weierstrass) *A bounded sequence of real numbers has a convergent subsequence.*

This can easily be generalised to sequences in $\mathbb{R}^n$:

Corollary 3.2 (Bolzano-Weierstrass in $\mathbb{R}^n$) *A bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Proof Let $\{\underline{x}^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)}\}$ be a bounded sequence in $\mathbb{R}^n$. Since $x_1^{(k)}$ is a bounded sequence in $\mathbb{R}$, there is a subsequence $\underline{x}^{(k_{1,j})}$ for which $x_1^{(k_{1,j})}$ converges. Since $\underline{x}^{(k_{1,j})}$ is again a bounded sequence in $\mathbb{R}^n$, $x_2^{(k_{1,j})}$ is a bounded sequence in $\mathbb{R}$. We can therefore find a subsequence $\underline{x}^{(k_{2,j})}$ of $\underline{x}^{(k_{1,j})}$ such that $x_2^{(k_{2,j})}$ converges. Since $\underline{x}^{(k_{2,j})}$ is a subsequence of $\underline{x}^{(k_{1,j})}$, $x_1^{(k_{2,j})}$ still converges. We can continue this process inductively to obtain a subsequence $\underline{x}^{(k_{n,j})}$ such that all the $x_i^{(k_{n,j})}$ for $i = 1, \ldots, n$ converge. □

We now make two definitions:

Definition 3.3 *A subset X of a normed space $(V, \|\cdot\|)$ is* bounded *if there exists an $M > 0$ such that*

$$\|x\| \leq M \qquad \text{for all} \qquad x \in X.$$

Note that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on V then X is bounded wrt $\|\cdot\|_1$ iff it is bounded wrt $\|\cdot\|_2$.

Definition 3.4 *A subset X of a normed space $(V, \|\cdot\|)$ is* closed *if whenever a sequence $\{x_n\}$ with $x_n \in X$ converges to some x, we must have $x \in X$.*

Note that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on V then X is closed wrt $\|\cdot\|_1$ iff it is closed wrt $\|\cdot\|_2$.

Example: any closed interval in $\mathbb{R}$ is closed in this sense. Any product of closed intervals in closed in $\mathbb{R}^n$.

Exercise 3.5 *Show that if $(X, \|\cdot\|)$ is a normed space then the unit ball*

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

and the unit sphere

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

are both closed.

Definition 3.6 *A subset K of a normed space $(V, \|\cdot\|)$ is* compact *if any sequence $\{x_n\}$ with $x_n \in K$ has a convergent subsequence $x_{n_j} \to x^*$ with $x^* \in K$.*

Note that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on V then X is compact wrt $\|\cdot\|_1$ iff it is compact wrt $\|\cdot\|_2$.

Two properties of compact sets are easy to prove:

Theorem 3.7 *A compact set is closed and bounded.*

Proof Let K be a compact set in $(V, \|\cdot\|)$ and $x_n \to x$ with $x_n \in K$. Since K is compact $\{x_n\}$ has a convergent subsequence; its limit must also be x, and from the definition of compactness $x \in K$, and so K is closed.

Suppose that K is not bounded. Then for each $n \in \mathbb{N}$ there exists an

$x_n \in K$ such that $\|x_n\| \geq n$. But $\{x_n\}$ must have a convergent subsequence, and any convergent sequence is bounded, which yields a contradiction. $\square$

It follows from the Bolzano-Weierstrass theorem that any closed bounded set K in $\mathbb{R}^n$ is compact: A sequence in a bounded subset K of $\mathbb{R}^n$ has a convergent subsequence by Corollary 3.2; since K is closed by definition this subsequence converges to an element of K. So K is compact. We have therefore shown:

Theorem 3.8 *A subset of $\mathbb{R}^n$ is compact iff it is closed and bounded.*

We will see later that this characterisation does *not* hold in infinite-dimensional spaces (and this is one way to characterise such spaces).

We now prove two fundamental results about continuous functions on compact sets:

Theorem 3.9 *Suppose that $K \subset (X, \|\cdot\|_X)$ is compact and that f is a continuous map from $(X, \|\cdot\|_X)$ into $(Y, \|\cdot\|_Y)$. Then $f(K)$ is a compact subset of $(Y, \|\cdot\|_Y)$.*

Proof Let $\{y_n\} \in f(K)$. Then $y_n = f(x_n)$ for some $x_n \in K$. Since $\{x_n\} \in K$, and K is compact there is a subsequence of x_n that converges, $x_{n_j} \to x^* \in K$. Since f is continuous it follows that as $j \to \infty$

$$y_{n_j} = f(x_{n_j}) \to f(x^*) = y^* \in f(K),$$

i.e. the subsequence y_{n_j} converges to some $y^* \in f(K)$, and so $f(K)$ is compact. $\square$

From which follows:

Proposition 3.10 *Let K be a compact subset of $(X, \|\cdot\|)$. Then any continuous function $f : K \to \mathbb{R}$ is bounded and attains its bounds, i.e. there exists an $M > 0$ such that $|f(x)| \leq M$ for all $x \in K$, and there exist $\underline{x}, \overline{x} \in K$ such that*

$$f(\underline{x}) = \inf_{x \in K} f(x) \qquad \textit{and} \qquad f(\overline{x}) = \sup_{x \in K} f(x).$$

Proof Since f is continuous and K is compact, $f(K)$ is a compact subset of $\mathbb{R}$, i.e. $f(K)$ is closed and bounded. It follows that

$$\overline{f} = \sup_{y \in f(K)} y \in f(K),$$

and so $\overline{f} = f(\overline{x})$ for some $\overline{x} \in K$. [That $\sup(S) \in S$ for any closed S is clear, since for each n there exists an $s_n \in S$ such that $s_n > \sup(S) - 1/n$. Since $s_n \le \sup(S)$ by definition, $s_n \to \sup(S)$, and it follows from the fact that S is closed that $\sup(S) \in S$.] The argument for $\underline{x}$ is identical. □

This allows one to prove the equivalence of all norms on a finite-dimensional space.

Theorem 3.11 *Let V be a finite-dimensional vector space. Then all norms on V are equivalent.*

Proof Let $E = \{e_j\}_{j=1}^n$ be a basis for V, and let $\|\cdot\|_E$ be the norm on V defined in Proposition 2.8. Let $\|\cdot\|$ be another norm on V. We will show that $\|\cdot\|$ is equivalent to $\|\cdot\|_E$. Since equivalence of norms is an equivalence relation, this will imply that all norms on V are equivalent.

Now, if $u = \sum_j \alpha_j e_j$ then

$$\begin{aligned} \|u\| &= \left\| \sum_j \alpha_j e_j \right\| \\ &\le \sum_j |\alpha_j| \|e_j\| \quad \text{(using the triangle inequality)} \\ &\le \left(\sum_j |\alpha_j|^2 \right)^{1/2} \left(\sum_j \|e_j\|^2 \right)^{1/2} \qquad \text{(using the Cauchy-Schwarz inequality)} \\ &= C_E \|u\|_E, \end{aligned}$$

where $C_E^2 = \sum_j \|e_j\|^2$, i.e. C_E is a constant that does not depend on u.

Now, observe that this estimate $\|u\| \le C_E \|u\|_E$ implies for $u, v \in V$,

$$\|u - v\| \le C_E \|u - v\|_E,$$

and so the map $u \mapsto \|u\|$ is continuous from $(V, \|\cdot\|_E)$ into $\mathbb{R}$.

Now, note that set

$$S_V = \{u \in V : \ \|u\|_E = 1\}$$

is the image of $S_n = \{\underline{\alpha} \in \mathbb{R}^n : \ |\underline{\alpha}| = 1\}$ under the map $\underline{\alpha} \mapsto \sum_{j=1}^n \alpha_j e_j$. Since by definition

$$\left\| \sum_{j=1}^n \alpha_j e_j \right\|_E = |\underline{\alpha}|,$$

this map is continuous. Since S_n is closed and bounded, it is a compact subset of $\mathbb{R}^n$; since S_V is the image of S_n under a continuous map, it is also compact.

Therefore the map $v \mapsto \|v\|$ is bounded on S_V, and attains its bounds. In particular, there exists an $a \geq 0$ such that

$$\|v\| \geq a \qquad \text{for every } v \in V \text{ with } \|v\|_E = 1.$$

Since the bound is attained, there exists a $v \in S_V$ such that $\|v\| = a$. If $a = 0$ then $\|v\| = 0$, i.e. $v = 0$. But since $v \in S_V$ we have $\|v\|_E = 1$, and so v cannot be zero. It follows that $a > 0$. Then for an arbitrary $v \in V$, we have $v/\|v\|_E \in S_V$, and so

$$\left\| \frac{v}{\|v\|_E} \right\| \geq a \quad \Rightarrow \quad \|v\| \geq a\|v\|_E.$$

Combining this with $\|u\| \leq C_E\|u\|_E$ shows that $\|\cdot\|$ and $\|\cdot\|_E$ are equivalent. □

Corollary 3.12 *A subset of a finite-dimensional normed space V is compact iff it is closed and bounded.*

Proof A subset K of $(V, \|\cdot\|)$ is compact iff it is a compact as a subset of $(V, \|\cdot\|_E)$, where $\|\cdot\|_E$ is the norm defined in Proposition 2.8 and used in the proof of the above theorem. The map $f : (V, \|\cdot\|_E) \to \mathbb{R}^n$ defined by

$$f\left(\sum_j \alpha_j e_j\right) = (\alpha_1, \ldots, \alpha_n)$$

is continuous, and its inverse is also continuous. It follows that K is compact iff $f(K)$ is compact. Since $f(K) \subset \mathbb{R}^n$ it is compact iff it is closed and bounded. Since both f and f^{-1} are continuous, it follows that $f(K)$ is closed and bounded iff K is closed and bounded. □

4

Completeness

In the treatment of convergent sequences of real numbers, one natural question is whether there is a way to characterise which sequences converge without knowing their limit. The answer, of course, is yes, and is given by the notion of a Cauchy sequence.

Theorem 4.1 *A sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ converges if and only if it is a Cauchy sequence, i.e. given any $\epsilon > 0$ there exists an N such that*

$$|x_n - x_m| < \epsilon \qquad \textit{for all} \qquad n, m \geq N.$$

Note that the proof makes use of the Bolzano-Weierstrass Theorem, so is in some way entangled with compactness properties of closed bounded subsets of $\mathbb{R}$.

A sequence in a normed space $(X, \|\cdot\|)$ is Cauchy if given any $\epsilon > 0$ there exists an N such that

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

Lemma 4.2 *Any Cauchy sequence is bounded.*

Proof There exists an N such that

$$\|x_n - x_m\| < 1 \qquad \text{for all} \qquad n, m \geq N.$$

It follows that in particular $\|x_n\| \leq \|x_N\| + 1$ for all $n \geq N$, and hence $\|x_n\|$ is bounded. $\square$

Definition 4.3 *A normed space X is* complete *if any Cauchy sequence in X, converges to some $x \in X$. A complete normed space is called a* Banach space.

Theorem 4.1 states that $\mathbb{R}$ with its standard norm is *complete* ('$\mathbb{R}$ is a Banach space'). It follows fairly straightforwardly that the same is true for any finite-dimensional normed space.

Theorem 4.4 *Every finite-dimensional normed space $(V, \|\cdot\|)$ (over $\mathbb{R}$ or $\mathbb{C}$) is complete.*

Proof Choose a basis $E = (e_1, \ldots, e_n)$ of V, and define another norm $\|\cdot\|_E$ on V by

$$\left\| \sum_{j=1}^{n} x_j e_j \right\|_E = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2}.$$

Since all norms on V are equivalent (Theorem 3.11), a sequence $\{x^k\}$ that is Cauchy in $\|\cdot\|$ is Cauchy in $\|\cdot\|_E$.

Writing $x^k = \sum_{j=1}^{n} x_j^k e_j$ it follows that given any $\epsilon > 0$ there exists an N_ϵ such that for $k, l \geq N_\epsilon$

$$\|x^k - x^l\|_E^2 = \sum_{j=1}^{n} |x_j^k - x_j^l|^2 < \epsilon^2. \tag{4.1}$$

In particular $\{x_j^k\}$ is a Cauchy sequence of real numbers for each fixed $j = 1, \ldots, n$. It follows that for each $j = 1, \ldots, n$ we have $x_j^n \to x_j^*$ for some x_j^*. Set $x^* = \sum_{j=1}^{n} x_j^* e_j$.

Letting $l \to \infty$ in (4.1) shows that

$$\|x^k - x^*\|_E^2 = \sum_{j=1}^{n} |x_j^k - x_j^*|^2 \leq \epsilon^2 \qquad \text{for all} \qquad n \geq N_\epsilon,$$

i.e. $x^n \to x^*$ wrt $\|\cdot\|_E$. It follows that $x^n \to x^*$ wrt $\|\cdot\|$, and clearly $x^* \in V$, and so V is complete. □

Note that in particular $\mathbb{R}^n$ is complete.

The completeness of ℓ^2 is a little more delicate, but only in the final steps.

Proposition 4.5 (Completeness of ℓ^2) *The sequence space $\ell^2(\mathbb{K})$ (equipped with its standard norm) is complete.*

Proof Suppose that $\underline{x}^k = (x_1^k, x_2^k, \cdots)$ is a Cauchy sequence in $\ell^2(\mathbb{K})$. Then for every $\epsilon > 0$ there exists an N_ϵ such that

$$\|\underline{x}^n - \underline{x}^m\|^2 = \sum_{j=1}^{\infty} |x_j^n - x_j^m|^2 < \epsilon^2 \qquad \text{for all} \qquad n, m \geq N_\epsilon. \tag{4.2}$$

In particular $\{x_j^k\}_{k=1}^{\infty}$ is a Cauchy sequence in $\mathbb{K}$ for every fixed j. Since $\mathbb{K}$ is complete (recall $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$) it follows that for each $k \in \mathbb{N}$

$$x_j^k \to a_k$$

for some $a_k \in \mathbb{R}$.

Set $\underline{a} = (a_1, a_2, \cdots)$. We want to show that $\underline{a} \in \ell^2$ and that $\|\underline{x}^k - \underline{a}\| \to 0$ as $k \to \infty$. First, since $\{\underline{x}^k\}$ is Cauchy we have from (4.2) that $\|\underline{x}^n - \underline{x}^m\| < \epsilon$ for all $n, m \geq N_\epsilon$, and so in particular for any $N \in \mathbb{N}$

$$\sum_{j=1}^{N} |x_j^n - x_j^m|^2 \leq \sum_{j=1}^{\infty} |x_j^n - x_j^m|^2 \leq \epsilon^2.$$

Letting $m \to \infty$ we obtain

$$\sum_{j=1}^{N} |x_j^n - a_j|^2 \leq \epsilon^2,$$

and since this holds for all N it follows that

$$\sum_{j=1}^{\infty} |x_j^n - a_j|^2 \leq \epsilon^2,$$

and so $\underline{x}^k - \underline{a} \in \ell^2$. But since ℓ^2 is a vector space and $\underline{x}^k \in \ell^2$, this implies that $a \in \ell^2$ and $\|\underline{x}^k - \underline{a}\| \leq \epsilon$. □

Since the norm on ℓ^2 is the natural generalisation of the norm on $\mathbb{R}^n$, and since it is complete, it is tempting to think that ℓ^2 will behave just like $\mathbb{R}^n$. However, it does not have the 'Bolzano-Weierstrass property' (bounded sequences have a convergent subsequence) as we can see easily by considering the sequence $\{\underline{e}_j\}_{j=1}^{\infty}$, where $\underline{e}_j$ consists entirely of zeros apart from a 1 in the jth position. Then clearly $\|\underline{e}_j\| = 1$ for all j; but if $i \neq j$ then

$$\|\underline{e}_i - \underline{e}_j\|^2 = 2,$$

i.e. any two elements of the sequence are always $\sqrt{2}$ away from each other. It follows that no subsequence of the $\{\underline{e}_j\}$ can form a Cauchy sequence, and so there cannot be a convergent subsequence.

This is really the first time we have seen a significant difference between $\mathbb{R}^n$ and the abstract normed vector spaces that we have been considering. The failure of the Bolzano-Weierstrass property is in fact a defining characteristic of infinite-dimensional spaces.

Theorem 4.6 $C^0([0,1])$ *equipped with the sup norm* $\|\cdot\|_\infty$ *is complete.*

Proof Let $\{f_k\}$ be a Cauchy sequence in $C^0([0,1])$: so given any $\epsilon > 0$ there exists an N such that

$$\sup_{x\in[0,1]} |f_n(x) - f_m(x)| < \epsilon \qquad \text{for all} \qquad n, m \geq N. \tag{4.3}$$

In particular $\{f_k(x)\}$ is a Cauchy sequence for each fixed x, so $f_k(x)$ converges for each fixed $x \in [0,1]$: define

$$f(x) = \lim_{k\to\infty} f_k(x).$$

We need to show that in fact $f_k \to f$ uniformly. But this follows since for every $x \in [0,1]$ we have from (4.3)

$$|f_n(x) - f_m(x)| < \epsilon \qquad \text{for all} \qquad n, m \geq N,$$

where N does not depend on x. Letting $m \to \infty$ in this expression we obtain

$$|f_n(x) - f(x)| < \epsilon \qquad \text{for all} \qquad n \geq N,$$

where again N does not depend on x. It follows that

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| < \epsilon \qquad \text{for all} \qquad n \geq N,$$

i.e. f_n converges uniformly to f on $[0,1]$. Completeness of $C^0([0,1])$ then follows from the fact that the uniform limit of a sequence of continuous functions is still continuous. □

For this reason the supremum norm is the 'standard norm' on $C^0([0,1])$; if no norm is mentioned this is the norm that is intended.

Example 4.7 $C^0([0,1])$ *equipped with the* L^2 *norm is* not *complete.*

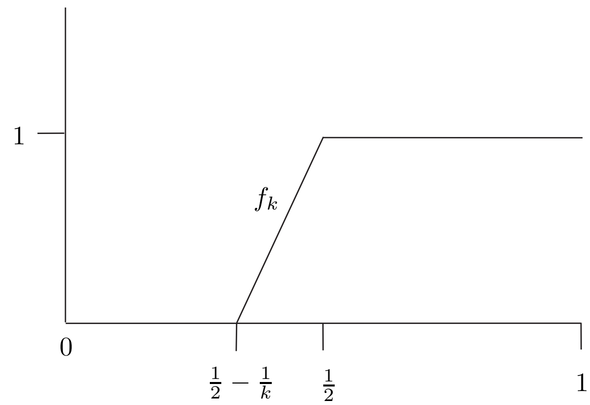

Fig. 4.1. Definition of f_k.

Consider the sequence of functions $\{f_k\}$ defined by

$$f_k(x) = \begin{cases} 0 & 0 \le x \le \frac{1}{2} - \frac{1}{k} \\ k\left[x - \left(\frac{1}{2} - \frac{1}{k}\right)\right] & \frac{1}{2} - \frac{1}{k} < x < \frac{1}{2} \\ 1 & \frac{1}{2} \le x \le 1, \end{cases}$$

see Figure 4.

This sequence is Cauchy in the L^2 norm, since for $n, m \ge N$ we have

$$f_n(x) = f_m(x) \qquad \text{for all} \qquad x < \tfrac{1}{2} - \frac{1}{N},\ x > \tfrac{1}{2},$$

and so

$$\int_0^1 |f_n(x) - f_m(x)|^2 \,\mathrm{d}x = \int_{\frac{1}{2}-\frac{1}{N}}^{\frac{1}{2}} |f_n(x) - f_m(x)|^2 \,\mathrm{d}x \le \frac{1}{N}, \tag{4.4}$$

since $|f_n(x) - f_m(x)| \le 1$ for all $x \in [0, 1]$ and all $n, m \in \mathbb{N}$.

But what is the limit, $f(x)$, as $n \to \infty$? Clearly one would expect

$$f(x) = \begin{cases} 0 & 0 \le x < \frac{1}{2} \\ 1 & \frac{1}{2} < x \le 1, \end{cases}$$

but this is certainly not continuous; nor is it defined at $x = \frac{1}{2}$. In fact whatever one chooses for $x(\frac{1}{2})$, we have

$$\begin{aligned} \|f_k - f\|_{L^2}^2 &= \int_0^1 |f_k(x) - f(x)|^2 \,\mathrm{d}x \\ &= \int_{\frac{1}{2}-\frac{1}{k}}^{\frac{1}{2}} |k(x - \tfrac{1}{2} + \tfrac{1}{k})|^2 \,\mathrm{d}x \\ &= \frac{k^2}{3} \left[x - \tfrac{1}{2} + \tfrac{1}{k}\right]_{x=\frac{1}{2}-\frac{1}{k}}^{\frac{1}{2}} \\ &= \frac{1}{3k} \to 0 \qquad \text{as} \qquad k \to \infty. \end{aligned}$$

So this sequence converges in the L^2 norm but not the sup norm.

Exercise 4.8 *Show that $C^0([0,1])$ is not complete in the L^1 norm.*

4.1 The completion of a normed space

However, *every normed space has a completion*, i.e. a minimal set $\tilde{V}$ such that $\tilde{V} \supset V$ and $(\tilde{V}, \|\cdot\|)$ is a Banach space. Essentially $\tilde{V}$ consists of all limit points of Cauchy sequences in V (and in particular, therefore, contains a copy of V via the constant sequence $v_n = v \in V$).

This implies that for any $v \in \tilde{V}$ there exist $v_n \in V$ such that $v_n \to v$ in the norm $\|\cdot\|$; we say that V is *dense* in $\tilde{V}$.

Definition 4.9 *Let $(V, \|\cdot\|)$ be a normed space. Then X is* dense *in V if given any $v \in V$ there exists a sequence $x_n \in X$ such that*

$$\|x_n - v\| \to 0 \qquad \textit{as} \qquad n \to \infty.$$

If X is dense in V one can usually deduce properties of V by approximating them with elements of X.

Example 4.10 *$\mathbb{R}$ is the completion of $\mathbb{Q}$ in the standard norm on $\mathbb{R}$.*

Exercise 4.11 *Recall that we defined ℓ_f to be the set of all sequences in* *which only a finite numbers of terms are non-zero. Show that ℓ_f is dense in ℓ^2.*

The description of the completion of $(V, \|\cdot\|)$ above is not strictly correct. Clearly it *must* be missing some subtleties, since we are 'adding' to V elements that are not in V and hence, in the setting of a general abstract normed space $(V, \|\cdot\|_V)$, are not defined.

To give a correct description, we first define some terminology. We say that two normed spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are *isometrically isomorphic*, or simply *isometric*, if there exists a linear isomorphism $\varphi : X \to Y$ such that

$$\|x\|_X = \|\varphi(x)\|_Y \qquad \text{for all} \qquad x, y \in X.$$

This guarantees that not only are X and Y isomorphic, but that the norms on X and Y are somehow 'the same'.

A completion of $(V, \|\cdot\|_V)$ is a Banach space $(X, \|\cdot\|_X)$ that contains an isometrically isomorphic image of $(V, \|\cdot\|_V)$ that is dense in X. One can show that there is 'only one' completion in that any two candidates must be isometrically isomorphic:

Theorem 4.12 *Let $(X, \|\cdot\|_X)$ be a normed space. Then there exists a complete normed space $(\mathscr{X}, \|\cdot\|_{\mathscr{X}})$ and a linear map $i : (X, \|\cdot\|_X) \to (\mathscr{X}, \|\cdot\|_{\mathscr{X}})$ that is an isometry between X and its image, such that $i(X)$ is a dense subspace of $\mathscr{X}$. Furthermore $\mathscr{X}$ is unique up to isometry; if $(\mathscr{Y}, \|\cdot\|_{\mathscr{Y}})$ is a complete normed space and $j : (X, \|\cdot\|_X) \to (\mathscr{Y}, \|\cdot\|_{\mathscr{Y}})$ is an isometry between X and its image, such that $j(X)$ is a dense subspace of $\mathscr{Y}$, then $\mathscr{X}$ and $\mathscr{Y}$ are isometric.*

Proof We consider Cauchy sequences in X, writing

$$\underline{x} = (x_1, x_2, \ldots) \qquad x_j \in X$$

for a sequence in X. We say that two Cauchy sequences $\underline{x}$ and $\underline{y}$ are equivalent, $\underline{x} \sim \underline{y}$, if

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

We let $\mathscr{X}$ be the space of equivalence classes of Cauchy sequences in X (i.e. $\mathscr{X} = X/\sim$). It is clear that $\mathscr{X}$ is a vector space, since the sum of two Cauchy sequences in X is again a Cauchy sequence in X. We define a candidate for our norm on $\mathscr{X}$: if $\eta \in \mathscr{X}$ then

$$\|\eta\|_{\mathscr{X}} = \lim_{n\to\infty} \|x_n\|_X, \tag{4.5}$$

for any $\underline{x} \in \eta$ (recall that η is an equivalence class of Cauchy sequences).

Note that (i) if $\underline{y}$ is a Cauchy sequence in X, then $\{\|y_n\|\}$ forms a Cauchy sequence in $\mathbb{R}$, so for a particular choice of $\underline{y} \in \eta$ the right-hand side of (4.5) exists, and (ii) if $\underline{x}, \underline{y} \in \eta$ then

$$\begin{aligned}\left|\lim_{n\to\infty}\|x_n\| - \lim_{n\to\infty}\|y_n\|\right| &= \left|\lim_{n\to\infty}\|x_n\| - \|y_n\|\right| \\ &= \lim_{n\to\infty}\Big|\|x_n\| - \|y_n\|\Big| \\ &\le \lim_{n\to\infty}\|x_n - y_n\| = 0\end{aligned}$$

since $\underline{x} \sim \underline{y}$. So the map in (4.5) is well-defined, and it is easy to check that it satisfies the three requirements of a norm.

Now we define a map $i : X \to \mathscr{X}$, by setting

$$i(x) = [(x, x, x, x, x, x, \ldots)].$$

Clearly i is linear, and an isometry between X and its image. We want to show that $i(X)$ is a dense subset of $\mathscr{X}$.

For any given $\eta \in \mathscr{X}$, choose some $\underline{x} \in \eta$. Since $\underline{x}$ is Cauchy, for any given $\epsilon > 0$ there exists an N such that

$$\|x_n - x_m\|_X < \epsilon \qquad \text{for all} \quad n, m \ge N.$$

In particular, $\|x_n - x_N\|_X < \epsilon$ for all $n \ge N$, and so

$$\|\eta - i(x_N)\|_{\mathscr{X}} = \lim_{n\to\infty} \|x_n - x_N\|_X < \epsilon,$$

which shows that $i(X)$ is dense in $\mathscr{X}$.

Finally, we have to show that $\mathscr{X}$ is complete, i.e. that any Cauchy sequence in $\mathscr{X}$ converges to another element of $\mathscr{X}$. A Cauchy sequence in $\mathscr{X}$ is a Cauchy sequence of equivalence classes of Cauchy sequences in X! Take such a Cauchy sequence, $\{\eta^{(k)}\}_{k=1}^\infty$. For each k, find $x_k \in X$ such that

$$\|i(x_k) - \eta^{(k)}\|_{\mathscr{X}} < 1/k,$$

using the density of $i(X)$ in $\mathscr{X}$. Now let

$$\underline{x} = (x_1, x_2, x_3, \ldots).$$

We will show (i) that $\underline{x}$ is a Cauchy sequence, and so $[\underline{x}] \in \mathscr{X}$, and (ii) that $\eta^{(k)}$ converges to $[\underline{x}]$. This will show that $\mathscr{X}$ is complete.

(i) To show that $\underline{x}$ is Cauchy, observe that

$$\begin{aligned}\|x_n - x_m\|_X &= \|i(x_n) - i(x_m)\|_{\mathscr{X}} \\ &= \|i(x_n) - \eta^{(n)} + \eta^{(n)} - \eta^{(m)} + \eta^{(m)} - i(x_m)\|_{\mathscr{X}} \\ &\le \|i(x_n) - \eta^{(n)}\|_{\mathscr{X}} + \|\eta^{(n)} - \eta^{(m)}\|_{\mathscr{X}} + \|\eta^{(m)} - i(x_m)\|_{\mathscr{X}} \\ &\le \frac{1}{n} + \|\eta^{(n)} - \eta^{(m)}\|_{\mathscr{X}} + \frac{1}{m}.\end{aligned}$$

So now given $\epsilon > 0$, choose N such that $\|\eta^{(n)} - \eta^{(m)}\|_{\mathscr{X}} < \epsilon/3$ for $n, m \ge N$. If $N' = \max(N, 3/\epsilon)$, it follows that

$$\|x_n - x_m\|_X < \epsilon \quad \text{for all} \quad n, m \ge N',$$

i.e. $\underline{x}$ is Cauchy. So $[\underline{x}] \in \mathscr{X}$.

(ii) To show that $\eta^{(k)} \to [\underline{x}]$, simply observe that

$$\|[\underline{x}] - \eta^{(k)}\|_{\mathscr{X}} \le \|[\underline{x}] - i(x_k)\|_{\mathscr{X}} + \|i(x_k) - \eta^{(k)}\|_{\mathscr{X}}.$$

Given $\epsilon > 0$, choose N large enough that $\|x_n - x_m\|_X < \epsilon/2$ for all $n, m \ge N$, and then set $N' = \max(N, 2/\epsilon)$. It follows that for $k \ge N'$,

$$\|[\underline{x}] - i(x_k)\|_{\mathscr{X}} = \lim_{n\to\infty} \|x_n - x_k\| < \epsilon/2$$

and $\|i(x_k) - \eta^{(k)}\|_{\mathscr{X}} < \epsilon/2$, i.e.

$$\|[\underline{x}] - \eta^{(k)}\|_{\mathscr{X}} < \epsilon,$$

and so $\eta^{(k)} \to [\underline{x}]$.

We will not prove the uniqueness of $\mathscr{X}$ here.

□

The space $\mathscr{X}$ in the above theorem is a very abstract one, and we are fortunate that in most situations there is a more concrete description of the completion of 'interesting' normed spaces.

Definition 4.13 *The space $L^2(0,1)$ is the completion of $C^0([0,1])$ with respect to the L^2 norm.*

Note that with this definition it is immediate that $L^2(0,1)$ is complete, and that $C^0([0,1])$ is dense in $L^2(0,1)$.

What is this space $L^2(0,1)$? There are a number of possible answers:

• Heuristically, $L^2(0,1)$ consists of all functions that can arise as the limit (with respect to the L^2 norm) of sequences $f_n \in C^0([0,1])$.

However, how do we characterise these limits? Certainly $L^2(0,1)$ is larger than $C^0([0,1])$ (and larger than $C^0(0,1)$). We saw above that it contains functions that are not continuous, and even functions whose values at individual points (e.g. $x = \frac{1}{2}$) are not defined.

• Formally, $L^2(0,1)$ is isometrically isomorphic to the equivalence class of sequences in $C^0([0,1])$ that are Cauchy in the L^2 norm, where $\{f_k\} \sim \{g_k\}$ if

$$\int_0^1 |f_k(x) - g_k(x)|^2 \, \mathrm{d}x \to 0 \qquad \text{as} \qquad k \to \infty.$$

This is hardly helpful.

• The space $L^2(0,1)$ consists of all real-valued functions f such that

$$\int_0^1 |f(x)|^2 \, \mathrm{d}x$$

is finite, where the integral is understood in the sense of Lebesgue integration. We say that $f = g$ in $L^2(0,1)$ (the functions are essentially 'the same' if

$$\int_0^1 |f(x) - g(x)|^2 \, \mathrm{d}x = 0;$$

equivalently, if $f = g$ almost everywhere.

This is the most intrinsic definition, and some ways the most 'useful'. But note that given this definition it is certainly not obvious that $L^2(0,1)$ is complete, nor that $C^0([0,1])$ is dense in $L^2(0,1)$. We will assume these properties in what follows, but at the risk of over-emphasis: if we use Definition 4.13 to define L^2 these properties come for free. If we use the 'useful' definition above there is actually some work to do to check these (which would be part of a proper development of the Lebesgue integral and corresponding 'Lebesgue spaces').

Although we cannot discuss the theory of Lebesgue integration in detail here, we can give a quick overview of its fundamental features and give a rigorous definition of the notion of 'almost everywhere'. Essentially the Lebesgue integral extends more elementary definitions of the integral in a mathematically consistent way.

5

Lebesgue integration

We follow the presentation in Priestley (1997), and start the construction of the Lebesgue integral by defining the integral of simple functions for which there can be no argument as to the correct definition.

We define the measure (or length) $|I|$ of an interval $I = [a, b]$ to be

$$|I| = b - a.$$

We will say that a set $A \subset \mathbb{R}$ has "measure zero" if, given any $\epsilon > 0$, one can find a (possibly countably infinite) set of intervals $[a_j, b_j]$ that cover A but whose total length is less than ϵ:

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

Exercise 5.1 *Show that if A_j has measure zero for all $j = 1, \ldots$ then*

$$\bigcup_{j=1}^{\infty} A_j$$

also has measure zero. [Hint: $\sum_{j=n+1}^{\infty} 2^{-j} = 2^{-n}$.]

A property is said to hold for 'almost every $x \in [a, b]$' (or 'almost everywhere in $[a, b]$') if the set of points at which it does not hold has measure zero.

Exercise 5.2 *Show that if each property P_j, $j = 1, 2, \ldots$, holds almost* *everywhere in an interval I then all the P_j hold simultaneously at almost every point in I.*

The class $L^{\text{step}}(\mathbb{R})$ of *step functions* on $\mathbb{R}$ consists of all those functions $s(x)$ which are piecewise constant on a finite number of intervals, i.e.

$$s(x) = \sum_{j=1}^{n} c_j \chi[I_j](x), \tag{5.1}$$

where $c_j \in \mathbb{R}$, each I_j is an interval, and $\chi[A]$ denotes the characteristic function of the set A,

$$\chi[A](x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

We define the integral of $s(x)$ by

$$\int s = \sum_{j=1}^{n} c_j |I_j|. \tag{5.2}$$

It is tedious but fairly elementary to check that this integral is well-defined on $L^{\text{step}}(\mathbb{R})$, so that if $s(x)$ is given by two possible expressions (5.1) then the integrals in (5.2) agree.

It is also relatively simple to check that this integral satisfies the following three fundamental properties:

(L) Linearity: if $\phi, \psi \in L^{\text{step}}(\mathbb{R})$ and $\lambda \in \mathbb{R}$ then $\phi + \lambda\psi \in L^{\text{step}}(\mathbb{R})$ and

$$\int (\phi + \lambda\psi) = \int \phi + \lambda \int \psi.$$

(P) Positivity: If $\phi \in L^{\text{step}}(\mathbb{R})$ with $\phi \geq 0$ then $\int \phi \geq 0$. (Note that combining this with the linearity implies that if $\phi, \psi \in L^{\text{step}}(\mathbb{R})$ and $\phi \geq \psi$ then $\int \phi \geq \int \psi$.)

(M) Modulus property: If $\phi \in L^{\text{step}}(\mathbb{R})$ then $|\phi| \in L^{\text{step}}(\mathbb{R})$ and $|\int \phi| \leq \int |\phi|$.

(T) Translation invariance: Take $\phi \in L^{\text{step}}(\mathbb{R})$. For $t \in \mathbb{R}$ define $\phi_d(x) = \phi(d + x)$. Then $\phi_d \in L^{\text{step}}(\mathbb{R})$ and $\int \phi_d = \int \phi$.

Now, if $s_n(x)$ is a monotonically increasing sequence of functions in $L^{\text{step}}(\mathbb{R})$ ($s_{n+1}(x) \geq s_n(x)$ for each $x \in \mathbb{R}$), then it follows from property (P) that the sequence

$$\int s_n \tag{5.3}$$

is also monotonically increasing. Provided that the integrals in (5.3) are uniformly bounded in n,

$$\lim_{n\to\infty} \int s_n$$

exists.

One can show that each monotonic sequence $s_n(x)$ with (5.3) uniformly bounded tends pointwise to a function $f(x)$ *almost everywhere*, i.e. except on a set of measure zero. We denote the set of all functions which can be arrived at in this way by $L_{\text{inc}}(\mathbb{R})$, and for such functions we can define

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

Again, we have to check that this definition does not depend on exactly which sequence $\{s_n\}$ we have chosen.

Finally, we define the space of integrable functions on $\mathbb{R}$, written $L^1(\mathbb{R})$, to be all functions of the form $f(x) = f_1(x) - f_2(x)$ with f_1 and f_2 in $L_{\text{inc}}(\mathbb{R})$, and set

$$\int f = \int f_1 - \int f_2.$$

It follows from this definition (although it is not immediately obvious) that any two functions that agree almost everywhere have the same integral, i.e. if $f = g$ almost everywhere then

$$\int f = \int g.$$

Properties (L), (P), (M), and (T) all hold for this definition of the integral on $L^1(\mathbb{R})$, where now properties are required to hold only almost everywhere.

There are three fundamental theorems for the Lebesgue integral. The first is the Monotone Convergence Theorem, which looks like the construction of the Lebesgue integral, but with a monotone sequence of step functions replaced by a monotone sequence of integrable functions.

Theorem 5.3 (Monotone Convergence Theorem) *Suppose that $f_n \in L^1(\mathbb{R})$, $f_n(x) \le f_{n+1}(x)$ almost everywhere, and $\int f_n \le K$ for some K independent of n. Then there exists a $g \in L^1(\mathbb{R})$ such that $f_n \to g$ almost* *everywhere, and*

$$\int g = \lim \int f_n.$$

Theorem 5.4 (Dominated Convergence Theorem) *Suppose that $f_n \in L^1(\mathbb{R})$ and that $f_n \to f$ almost everywhere. If there exists a function $g \in L^1(\mathbb{R})$ such that $|f_n(x)| \leq g(x)$ for almost every x, and for every n, then $f \in L^1(\mathbb{R})$ and*

$$\int f = \lim \int f_n.$$

To define an integral of a function of two variables one would naturally proceed by analogy with the construction above: take 'step functions' that are constant on rectangles, construct an integral on $L^{\text{inc}}(\mathbb{R}^2)$ by taking limits of monotonic sequences, and the construct $L^1(\mathbb{R}^2)$ as the limits of differences. But this does not related 'double integrals' to single integrals. This is achieved by the Fubini and Tonelli theorems. We give a less-than-rigorous formulation:

Theorem 5.5 (Fubini-Tonelli) *If $f : \mathbb{R}^2 \to \mathbb{R}$ is such that either*

$$\int \left(\int |f(x,y)| \, \mathrm{d}x \right) \mathrm{d}y < \infty \qquad \textit{or} \qquad \int \left(\int |f(x,y)| \, \mathrm{d}y \right) \mathrm{d}x < \infty$$

then $f \in L^1(\mathbb{R}^2)$ and

$$\int \left(\int f(x,y) \, \mathrm{d}x \right) \mathrm{d}y = \int \left(\int f(x,y) \, \mathrm{d}y \right) \mathrm{d}x.$$

(The less-than-rigorous nature is that the conditions require integrability properties of $|f(x,y)|$: first that for almost every $y \in \mathbb{R}$, $|f(\cdot,y)| \in L^1(\mathbb{R})$, and then that the resulting function $g(y) = \int |f(x,y)| \, dy$ is again in $L^1(\mathbb{R})$.)

5.1 The Lebesgue space L^2 and Hilbert spaces

Having defined the integral in this way, we denote by $L^2(0,1)$ the set of all functions defined on $(0,1)$ such that

$$\int_0^1 |f(x)|^2 \, \mathrm{d}x$$

is finite, where $\int$ is the Lebesgue integral. The standard norm on $L^2(0,1)$ is that derived from the inner product

$$(f,g) = \int f(x)\overline{g(x)}\,\mathrm{d}x,$$

i.e.

$$\|f\|_{L^2} = \left(\int |f(x)|^2\,\mathrm{d}x\right)^{1/2}.$$

We showed earlier that this is indeed a norm on $C^0([0,1])$; in that case the main thing we had to check was that $\|f\| = 0$ implies that $f = 0$.

This does not hold for $L^2(0,1)$, unless we 'identify' functions that agree almost everywhere, since for any two such functions we will have

$$\|f - g\|_{L^2} = 0,$$

since $|f(x) - g(x)|^2$ will be zero almost everywhere.

In particular, this implies, strictly, that no element of $L^2(0,1)$ has a well-defined value at a particular point. But one often has a well-defined 'representative element' in mind, e.g. one can view any $f \in C^0([0,1])$ as an element of $L^2(0,1)$, and it would be perverse to insist that f does not have well-defined values at any point, when this continuous representative is in some ways the most 'natural' choice.

6

Inner product spaces

If $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ are two elements of $\mathbb{R}^n$ then we define their dot product as

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n. \tag{6.1}$$

This is one concrete example of an inner product on a vector space:

Definition 6.1 *An inner product $(\cdot, \cdot)$ 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 $\alpha \in \mathbb{K}$,*

(i) $(x, x) \geq 0$ *with equality iff* $x = 0$,
(ii) $(x + y, z) = (x, z) + (y, z)$,
(iii) $(\alpha x, y) = \alpha(x, y)$, *and*
(iv) $(x, y) = \overline{(y, x)}$.

Note that

- in a real vector space the complex conjugate in (iv) is unnecessary;
- in the complex case the restriction that $(y, x) = \overline{(x, y)}$ implies in particular that $(x, x) = \overline{(x, x)}$, i.e. that (x, x) is real, and so the requirement that $(x, x) \geq 0$ makes sense; and
- (iii) and (iv) imply that the inner product is conjugate linear in its second element, i.e. $(x, \alpha y) = \bar{\alpha}(x, y)$.

A vector space equipped with an inner product is known as an inner product space.

Example 6.2 *In the space* $\ell^2(\mathbb{K})$ *of square summable sequences, for* $\underline{x} = (x_1, x_2, \ldots)$ *and* $\underline{y} = (y_1, y_2, \ldots)$ *one can define an inner product*

$$(\underline{x}, \underline{y}) = \sum_{j=1}^{\infty} x_j \bar{y}_j.$$

This is well-defined since $\sum_j |x_j \bar{y}_j| \le \frac{1}{2}(\sum_j |x_j|^2 + |y_j|^2)$.

Example 6.3 *The expression*

$$(f, g) = \int_a^b f(x) g(x)\,\mathrm{d}x$$

defines an inner product on the space $L^2(a, b)$.

6.1 Inner products and norms

Given an inner product we can define $\|v\|$ by setting

$$\|v\|^2 = (v, v). \tag{6.2}$$

We will soon shows that $\|\cdot\|$ defines a norm; we say that it is the *norm* induced by the inner product $(\cdot, \cdot)$.

6.2 The Cauchy-Schwarz inequality

Lemma 6.4 (Cauchy-Schwarz inequality) *Any inner product satisfies the inequality*

$$|(x, y)| \le \|x\|\|y\| \qquad \textit{for all} \qquad x, y \in V, \tag{6.3}$$

where $\|\cdot\|$ *is defined in (6.2).*

Proof If $x = 0$ or $y = 0$ then (6.3) is clear; so suppose that $x \neq 0$ and $y \neq 0$. For any $\lambda \in \mathbb{K}$ we have

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

Setting $\lambda = (x,y)/\|y\|^2$ we obtain

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

which implies (6.3). □

The Cauchy-Schwarz inequality allows us to show easily that the map $x \mapsto \|x\|$ is a norm on V. Property (i) is clear, since $\|x\| \geq 0$ and if $\|x\|^2 = (x,x) = 0$ then $x = 0$. Property (ii) is also clear, since

$$\|\alpha x\|^2 = (\alpha x, \alpha x) = \alpha\overline{\alpha}(x,x) = |\alpha|^2\|x\|^2.$$

Property (iii), the triangle inequality, follows from the Cauchy-Schwarz inequality (6.3), since

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

i.e. $\|x+y\| \leq \|x\| + \|y\|$.

As an example of the Cauchy-Schwarz inequality, consider the standard inner product on $\mathbb{R}^n$. As we would expect, the norm derived from this inner product is just

$$\|x\| = \left(\sum_{j=1}^n |x_j|^2\right)^{1/2}.$$

The Cauchy-Schwarz inequality says that

$$|(x,y)|^2 = \left|\sum_{j=1}^n x_j y_j\right|^2 \leq \left(\sum_{j=1}^n |x_j|^2\right)\left(\sum_{j=1}^n |y_j|^2\right), \tag{6.4}$$

or just $|x \cdot y| \leq |x||y|$.

Exercise 6.5 *The norm on the sequence space ℓ^2 derived from the inner product $(\underline{x}, \underline{y}) = \sum x_j \bar{y}_j$ is*

$$\|\underline{x}\|_{\ell^2} = \left(\sum_{j=1}^{\infty} |x_j|^2 \right)^{1/2}.$$

Obtain the Cauchy-Schwarz inequality for ℓ^2 using (6.4) and a limiting argument rather than Lemma 6.4.

The Cauchy-Schwarz inequality in $L^2(a,b)$ gives the very useful:

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

for $f, g \in L^2(a,b)$. (This shows in particular that if $f, g \in L^2(a,b)$ then $fg \in L^1(a,b)$.)

6.3 The relationship between inner products and their norms

Norms derived from inner products have one key property in addition to (i)–(iii) of Definition 2.1:

Lemma 6.6 (Parallelogram law) *Let V be an inner product space with induced norm $\|\cdot\|$. Then*

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2) \qquad \text{for all} \qquad x, y \in V. \tag{6.5}$$

Proof Simply expand the inner products:

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

$\square$

Exercise 6.7 *Show that there is no inner product on $C^0([0,1])$ which induces the sup or L^1 norms,*

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

Given a norm that is derived from an inner product, one can reconstruct the inner product as follows:

Lemma 6.8 (Polarisation identity) *Let V be an inner product space with induced norm $\|\cdot\|$. Then if V is real*

$$4(x,y) = \|x+y\|^2 - \|x-y\|^2, \tag{6.6}$$

while if V is complex

$$4(x,y) = \|x+y\|^2 - \|x-y\|^2 + \mathrm{i}\|x+\mathrm{i}y\|^2 - \mathrm{i}\|x-\mathrm{i}y\|^2. \tag{6.7}$$

Proof Once again, rewrite the right-hand sides as inner products, multiply out, and simplify. □

If V is an real/complex inner product space and $\|\cdot\|$ is a norm on V that satisfies the parallelogram law then (6.6) or (6.7) defines an inner product on V. In other words, the parallelogram law characterises those norms that can be derived from inner products. (This argument is non-trivial.)

Lemma 6.9 *If V is an inner product space with inner product $(\cdot,\cdot)$ and derived norm $\|\cdot\|$, then $x_n \to x$ and $y_n \to y$ implies that*

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

Proof Since x_n and y_n converge, $\|x_n\|$ and $\|y_n\|$ are bounded (the proof is a simple exercise). Then

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

implies that $(x_n, y_n) \to (x, y)$. □

This lemma is extremely useful: that we can swap limits and inner products means that if

$$\sum_{j=1}^{n} x_j$$

converges (so that $\sum_{j=1}^{n} x_j \to x = \sum_{j=1}^{\infty} x_j$) then

$$\left(\sum_{j=1}^{\infty} x_j, y\right) = \sum_{j=1}^{\infty} (x_j, y),$$

i.e. we can swap inner products and sums.

Definition 6.10 *A* Hilbert space *is a complete inner product space.*

Examples: $\mathbb{R}^n$ with inner product and norm

$$\Big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\Big) = \sum_{j=1}^{n} x_j y_j \qquad \|(x_1, \ldots, x_n)\| = \left(\sum_{j=1}^{n} |x_j|^2\right)^{1/2} ;$$

$\mathbb{C}^n$ with inner product and norm

$$\Big((w_1, \ldots, w_n), (z_1, \ldots, z_n)\Big) = \sum_{j=1}^{n} w_j \bar{z}_j \qquad \|(w_1, \ldots, w_n)\| = \left(\sum_{j=1}^{n} |w_j|^2\right)^{1/2} ;$$

$\ell^2(\mathbb{K})$ with inner product and norm

$$(\underline{x}, \underline{y}) = \sum_{j=1}^{\infty} x_j \bar{y}_j \qquad \|\underline{x}\| = \left(\sum_{j=1}^{\infty} |x_j|^2\right)^{1/2}$$

(the complex conjugate is redundant if $\mathbb{K} = \mathbb{R}$); and $L^2(I)$ with inner product and norm

$$(f, g) = \int_I f(x)\overline{g(x)}\,\mathrm{d}x \qquad \|f\|_{L^2} = \left(\int_I |f(x)|^2\,\mathrm{d}x\right)^{1/2} .$$

From now on we will assume unless explicitly stated that all the above spaces are equipped by their standard inner product (and corresponding norm).

7

Orthonormal bases in Hilbert spaces

From now on we will denote by H an arbitrary Hilbert space, with inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$; we take $\mathbb{K}=\mathbb{C}$, since the case $\mathbb{K}=\mathbb{R}$ is simplified only by removing the complex conjugates.

Our aim in this chapter is to discuss orthonormal bases for Hilbert spaces. In contrast to the *Hamel basis* we considered earlier, we are now going to allow infinite linear combinations of basis elements (called a *Schauder basis*).

7.1 Orthonormal sets

Definition 7.1 *Two elements x and y of an inner product space are said to be orthogonal if $(x,y)=0$. (We sometimes write $x\perp y$.)*

Definition 7.2 *A set E is orthonormal if $\|e\|=1$ for all $e\in E$ and $(e_1,e_2)=0$ for any $e_1,e_2\in E$ with $e_1\neq e_2$.*

Any orthonormal set must be linearly independent.

Clearly if $(x,y)=0$ then

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

(Pythagoras). Sums of orthogonal vectors are therefore very useful in calculations, since all the cross terms in their norm vanish:

Lemma 7.3 *Let $e_1,\ldots,e_n$ be an orthonormal set in an inner product space*

V. Then for any $\alpha_j \in \mathbb{K}$.

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

Proof Use Pythagoras repeatedly. □

Example 7.4 *The set $\{\underline{e}_j\}_{j=1}^{\infty}$, where*

$$\underline{e}_j = (0, 0, \ldots, 1, \ldots, 0, \ldots)$$

(with the 1 in the jth position), is an orthonormal set in ℓ^2.

Example 7.5 *Consider the space $L^2(-\pi, \pi)$ and the set*

$$E = \frac{1}{\sqrt{2\pi}},\ \frac{1}{\sqrt{\pi}}\cos t,\ \frac{1}{\sqrt{\pi}}\sin t,\ \frac{1}{\sqrt{\pi}}\sin 2t,\ \frac{1}{\sqrt{\pi}}\cos 2t,\ \ldots.$$

Then E is orthonormal, since

$$\int_{-\pi}^{\pi} \cos^2 nt \, \mathrm{d}t = \int_{-\pi}^{\pi} \sin^2 nt \, \mathrm{d}t = \pi;$$

for any n, m

$$\int_{-\pi}^{\pi} \cos nt \, \mathrm{d}t = \int_{-\pi}^{\pi} \sin nt \, \mathrm{d}t = \int_{-\pi}^{\pi} \sin nt \cos mt \, \mathrm{d}t = 0;$$

and for any $n \neq m$

$$\int_{-\pi}^{\pi} \cos nt \, \cos mt \, \mathrm{d}t = \int_{-\pi}^{\pi} \sin nt \sin mt \, \mathrm{d}t = \int_{-\pi}^{\pi} \sin nt \cos mt \, \mathrm{d}t = 0.$$

The following lemma, proved using the Gram-Schmidt orthonormalisation process (which we will revisit later), guarantees the existence of an orthonormal basis in any *finite-dimensional* inner product space.

Lemma 7.6 *Let $(\cdot, \cdot)$ be any inner product on a vector space V of dimension n. Then there exists an orthonormal basis $\{e_j\}_{j=1}^{n}$ of V.*

It follows that in some sense the dot product (6.1) is the canonical inner product on a finite-dimensional space. Indeed, with respect to any the orthonormal basis $\{e_j\}$ the inner product $(\cdot,\cdot)$ has the form (6.1), i.e.

$$\left(\sum_{j=1}^{n} x_j e_j, \sum_{k=1}^{n} y_k e_k\right) = \sum_{i,j=1}^{n} x_j \bar{y}_k (e_j, e_k) = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n.$$

7.2 Convergence and orthonormality in Hilbert spaces

In an infinite-dimensional Hilbert space we cannot hope to find a finite basis, since then the space would by definition be finite-dimensional. The best that we can hope for is to find a countable basis $\{e_j\}_{j=1}^{\infty}$, in terms of which to expand any $x \in H$ as potentially infinite series,

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

We make the obvious definition of what this equality means.

Definition 7.7 *Let $(X, \|\cdot\|_X)$ be a normed space. Then*

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

iff the partial sums converge to x in the norm of X, i.e.

$$\left\| \left(\sum_{j=1}^{n} \alpha_n \mathrm{e}_n\right) - x \right\|_X \to 0 \qquad \textit{as} \qquad n \to \infty.$$

We now formalise our notion of a basis for a Hilbert space:

Definition 7.8 *A set $\{e_j\}_{j=1}^{\infty}$ is a* basis *for H if every x can be written uniquely in the form*

$$x = \sum_{j=1}^{\infty} \alpha_j e_j \tag{7.1}$$

for some $\alpha_j \in \mathbb{K}$. If in addition $\{e_j\}_{j=1}^{\infty}$ is an orthonormal set then we refer to it as an orthonormal basis*.*

(Note that if $\{e_j\}$ is a basis in the sense of Definition 7.8, i.e. the expansion in terms of the e_j is unique then the e_j are linearly independent, since if

$$0 = \sum_{j=1}^{n} \alpha_j e_j$$

there is a unique expansion for zero and so we must have $\alpha_j = 0$ for all $j = 1, \ldots, n$.)

Rather than discuss general bases, we concentrate on orthonormal bases. Neglecting for the moment the question of convergence, and of conditions to guarantee that $\{e_j\}$ really is a basis, suppose that the equality (7.1) holds for some $x \in H$. To find the coefficients α_j, simply take the inner product with some e_k to give

$$(x, e_k) = \left(\sum_{j=1}^{\infty} \alpha_j e_j, e_k \right) = \sum_{j=1}^{\infty} \alpha_j (e_j, e_k) = \alpha_k,$$

and so we would expect $\alpha_k = (x, e_k)$, and so for an *orthonormal* basis we would expect to obtain the expansion

$$x = \sum_{j=1}^{\infty} (x, e_j) e_j.$$

Assuming that the Pythagoras result of Lemma 7.3 holds for infinite sums, we would expect that

$$\sum_{j=1}^{\infty} |(x, e_j)|^2 = \|x\|^2.$$

In some ways this says that 'the projections onto the e_j capture all of x'. Presumably if $\{e_j\}$ do not form an orthonormal basis we should be able to find an x such that

$$\sum_{j=1}^{\infty} |(x, e_j)|^2 < \|x\|^2.$$

We *have not* proved any of this yet, since we are assuming that (7.1) holds and taking no care with swapping the inner product and the infinite sum; but it motivates the following lemma, whose result is known as Bessel's inequality.

Lemma 7.9 (Bessel's inequality) *Let V be an inner product space and $\{e_n\}_{n=1}^{\infty}$ an orthonormal sequence. Then for any $x \in V$ we have*

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

and in particular the left-hand side converges.

Proof Let us denote by x_k the partial sum

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

Clearly

$$\|x_k\|^2 = \sum_{j=1}^{k} |(x, e_j)|^2$$

and so we have

$$\begin{aligned}
\|x - x_k\|^2 &= (x - x_k, x - x_k) \\
&= \|x\|^2 - (x_k, x) - (x, x_k) + \|x_k\|^2 \\
&= \|x\|^2 - \sum_{j=1}^{k} (x, e_j)(e_j, x) - \sum_{j=1}^{k} \overline{(x, e_j)}(x, e_j) + \|x_k\|^2 \\
&= \|x\|^2 - \|x_k\|^2.
\end{aligned}$$

It follows that

$$\sum_{j=1}^{k} |(x, e_j)|^2 = \|x_k\|^2 \leq \|x\|^2 - \|x - x_k\|^2 \leq \|x\|^2.$$

□

We now use Bessel's inequality to give a simple criterion for the convergence of a sum $\sum_{j=1}^{\infty} \alpha_j e_j$ when the $\{e_j\}$ are orthonormal.

Lemma 7.10 *Let H be a Hilbert space and $\{e_n\}$ an orthonormal sequence in H. The series $\sum_{n=1}^{\infty} \alpha_n e_n$ converges iff*

$$\sum_{n=1}^{\infty} |\alpha_n|^2 < +\infty$$

and then

$$\left\|\sum_{n=1}^{\infty}\alpha_n e_n\right\|^2=\sum_{n=1}^{\infty}|\alpha_n|^2. \tag{7.2}$$

We could rephrase this as $\sum_{n=1}^{\infty}\alpha_n e_n$ converges iff $\underline{\alpha}=(\alpha_1,\alpha_2,\dots)\in\ell^2$.

Proof Suppose that $\sum_{j=1}^{n}\alpha_j e_j$ converges to x as $n\to\infty$; then

$$\left\|\sum_{j=1}^{n}\alpha_j \mathrm{e}_j\right\|^2=\sum_{j=1}^{n}|\alpha_j|^2$$

converges to $\|x\|^2$ as $n\to\infty$ (see Lemma 2.11).

Conversely, if $\sum_{j=1}^{\infty}|\alpha_j|^2<+\infty$ then $\{\sum_{j=1}^{n}|\alpha_j|^2\}$ is a Cauchy sequence. Setting $x_n=\sum_{j=1}^{n}\alpha_j e_j$ we have, taking wlog $m>n$,

$$\|x_n-x_m\|^2=\left\|\sum_{j=n+1}^{m}\alpha_j e_j\right\|^2=\sum_{j=n+1}^{m}|\alpha_j|^2,$$

and so $\{x_n\}$ is a Cauchy sequence and therefore converges to some $x\in H$, since H is complete. The equality in (7.2) follows as above. □

By combining this lemma with Bessel's inequality we obtain:

Corollary 7.11 *Let H be a Hilbert space and $\{e_n\}_{n=1}^{\infty}$ an orthonormal sequence in H. Then for any x the sequence*

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

converges.

7.3 Orthonormal bases in Hilbert spaces

We now show that $\{e_n\}$ forms a basis for H iff

$$\|x\|^2=\sum_{j=1}^{\infty}|(x,e_j)|^2\qquad\text{for all}\qquad x\in H.$$

Proposition 7.12 *Let $E = \{e_j\}_{j=1}^{\infty}$ be an orthonormal set in a Hilbert space H. Then the following are equivalent to the statement that E is an orthonormal basis for H:*

(a) $x = \sum_{n=1}^{\infty}(x, e_n)e_n$ for all $x \in H$;
(b) $\|x\|^2 = \sum_{n=1}^{\infty}|(x, e_n)|^2$ for all $x \in H$; and
(c) $(x, e_n) = 0$ for all n implies that $x = 0$.

Proof If E is an orthonormal basis for H then we can write

$$x = \sum_{j=1}^{\infty} \alpha_j e_j, \quad \text{i.e.} \quad x = \lim_{j\to\infty} \sum_{j=1}^{n} \alpha_j e_j.$$

Clearly if $k \leq n$ we have

$$(\sum_{j=1}^{n} \alpha_j e_j, e_k) = \alpha_k,$$

and using the properties of the inner product of limits we obtain $\alpha_k = (x, e_k)$ and hence (a) holds. The same argument shows that if we assume (a) then this expansion is unique and so E is a basis.

(a) $\Rightarrow$ (b) is immediate from (2.11).

(b) $\Rightarrow$ (c) is immediate since $\|x\| = 0$ implies that $x = 0$.

(c) $\Rightarrow$ (a) Take $x \in H$ and let

$$y = x - \sum_{j=1}^{\infty}(x, e_j)e_j.$$

For each $m \in \mathbb{N}$ we have, using Lemma 6.9 (continuity of the inner product),

$$\begin{aligned} (y, e_m) &= (x, e_m) - \lim_{n\to\infty}\left(\sum_{j=1}^{n}(x, e_j)e_j, e_m\right) \\ &= 0 \end{aligned}$$

since eventually $n \geq m$. It follows from (c) that $y = 0$, i.e. that

$$x = \sum_{j=1}^{\infty}(x, e_j)e_j$$

as required. □

Example 7.13 *The set $\{\underline{e}_j\}_{j=1}^{\infty}$, where*

$$\underline{e}_j = (0, 0, \ldots, 1, \ldots, 0, \ldots)$$

(with the 1 *in the* j*th position), is an orthonormal basis for* ℓ^2*, since it clear that if* $(\underline{x}, \underline{e}_j) = x_j = 0$ *for all* j *then* $\underline{x} = \underline{0}$*.*

Example 7.14 *The sine and cosine functions given in example 7.5 are an o-n basis for* $L^2(-\pi, \pi)$*.*

8

Closest points and approximation

8.1 Closest points in convex subsets

We start with a general result about closest points.

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

Lemma 8.2 *Let A be a non-empty closed convex subset of a Hilbert space H and let $x \in H$. Then there exists a unique $\hat{a} \in A$ such that*

$$\|x - \hat{a}\| = \inf\{\|x - a\| : \ a \in A\}.$$

Proof Set $\delta = \inf\{\|x-a\| : \ a \in A\}$ and find a sequence $a_n \in A$ such that

$$\|x - a_n\|^2 \leq \delta^2 + \frac{1}{n}. \tag{8.1}$$

We will show that $\{a_n\}$ is a Cauchy sequence. To this end, we use the parallelogram law:

$$\|(x-a_n)+(x-a_m)\|^2 + \|(x-a_n)-(x-a_m)\|^2 = 2[\|x-a_n\|^2 + \|x-a_m\|^2].$$

Which gives

$$\|2x - (a_n + a_m)\|^2 + \|a_n - a_m\|^2 < 4\delta^2 + \frac{2}{m} + \frac{2}{n}$$

or

$$\|a_n - a_m\|^2 \leq 4\delta^2 + \frac{2}{m} + \frac{2}{n} - 4\|x - \tfrac{1}{2}(a_n + a_m)\|^2.$$

Since A is convex, $a_n + a_m \in A$, and so $\|x - \frac{1}{2}(a_n + a_m)\|^2 \geq \delta^2$, which gives

$$\|a_n - a_m\|^2 \leq \frac{2}{m} + \frac{2}{n}.$$

It follows that $\{a_n\}$ is Cauchy, and so $a_n \to \hat{a}$. Since A is closed, $\hat{a} \in A$.

To show that $\hat{a}$ is unique, suppose that $\|u - a^*\| = \delta$ with $a^* \neq \hat{a}$. Then $\|u - \frac{1}{2}(a^* + \hat{a})\| \geq \delta$ since A is convex, and so, using the parallelogram law again,

$$\|a^* - \hat{a}\|^2 \leq 4\gamma^2 - 4\gamma^2 = 0,$$

i.e. $a^* = \hat{a}$ and $\hat{a}$ is unique. □

8.2 Linear subspaces and orthogonal complements

In an infinite-dimensional space, linear subspaces need not be closed. For example, the space $\ell_f(\mathbb{R})$ of all real sequences with only a finite number of non-zero terms is a linear subspace of $\ell^2(\mathbb{R})$, but is not closed (consider the sequence $\underline{x}^{(n)} = (1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, 0, \ldots))$.

If X is a subset of H then the orthogonal complement of X in H is

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

Lemma 8.3 *If X is a subset of H then $X^\perp$ is a closed linear subspace of H.*

Proof It is clear that $X^\perp$ is a linear subspace of H: if $u, v \in X^\perp$ and $\alpha \in K$ then

$$(u + v, x) = (u, x) + (v, x) = 0 \qquad \text{and} \qquad (\alpha u, x) = \alpha(u, x) = 0$$

for every $x \in X$. To show that $X^\perp$ is closed, suppose that $u_n \in X^\perp$ and $u_n \to u$ then

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

and so $X^\perp$ is closed. □

Note that Proposition 7.12 shows that E is a basis for H iff $E^\perp = \{0\}$ (since this is just a rephrasing of (c): $(u, e_j) = 0$ for all j implies that $u = 0$).

Note also that if Span(E) denotes the linear span of E, i.e.

$$\mathrm{Span}(E) = \{u \in H : \; u = \sum_{j=1}^{n} \alpha_j e_j \;\; n \in \mathbb{N}, \alpha_j \in \mathbb{K}, \; e_j \in E\}$$

then $E^\perp = (\mathrm{Span}(E))^\perp$.

Lemma 8.4 *Any infinite-dimensional Hilbert H space contains a countably infinite orthonormal sequence.*

Proof Suppose that H contains an orthonormal set $E_k = \{e_j\}_{j=1}^k$. Then E_k does not form a basis for H, since H is infinite-dimensional. It follows that $E_k^\perp \neq \{0\}$. Therefore there exists a non-zero $x \in H$ such that $x \in E_k^\perp$. Setting $e_{k+1} = x/\|x\|$ yields an e_{k+1} such that

$$\|e_{k+1}\| = 1 \qquad \text{and} \qquad e_{k+1} \in E_k^\perp, \text{ i.e. } (e_{k+1}, e_j) = 0 \;\forall\; j = 1, \ldots, k.$$

So $E_{k+1} = \{e_j\}_{j=1}^{k+1}$ is an orthonormal set. The result follows by induction, start with $e_1 = x/\|x\|$ for any non-zero $x \in H$. □

Theorem 8.5 *A Hilbert space is finite-dimensional iff its unit ball is compact.*

Proof The unit ball is closed and bounded. If H is finite-dimensional this is equivalent to compactness by Corollary 3.12. If H is infinite-dimensional then it contains a countable orthonormal set $\{e_j\}_{j=1}^\infty$, and for $i \neq j$

$$\|e_i - e_j\|^2 = 2.$$

The $\{e_j\}$ form a sequence in the unit ball that can have no convergent subsequence. □

Proposition 8.6 *If U is a closed linear subspace of a Hilbert space H then any $x \in H$ can be written uniquely as*

$$x = u + v \qquad \textit{with} \qquad u \in U, \quad v \in U^\perp,$$

i.e. $H = U \oplus U^\perp$. The map $P_U : H \to U$ defined by

$$P_U x = u$$

is called the orthogonal projection of x onto U, *and satisfies*

$$P_U^2 x = P_U x \qquad \textit{and} \qquad \|P_U x\| \leq \|x\| \qquad \textit{for all} \qquad x \in H.$$

Proof If U is a closed linear subspace then U is closed and convex, so the above result shows that given $x \in H$ there is a unique closest point $u \in U$. It is now simple to show that $x - u \in U^{\perp}$ and then such a decomposition is unique.

Indeed, consider $v = x - u$; the claim is that $v \in U^{\perp}$, i.e. that

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

Consider $\|x - (u - ty)\| = \|v + ty\|$; then

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

We know from the construction of u that $\|v + ty\|$ is minimal when $t = 0$. If t is real then this implies that $\mathrm{d}\Delta/\mathrm{d}t(0) = 2\,\mathrm{Re}\{(y, v)\} = 0$. If $t = \mathrm{i}s$, with s real, then $\mathrm{d}\Delta(\mathrm{i}s)/\mathrm{d}s = -2\,\mathrm{Im}\{(y, v)\} = 0$. So $(y, v) = 0$ for any $y \in U$, i.e. $v \in U^{\perp}$ is claimed.

Finally, the uniqueness follows easily: if $x = u_1 + v_1 = u_2 + v_2$, then $u_1 - u_2 = v_2 - v_1$, and so

$$|v_1 - v_2|^2 = (v_1 - v_2, v_1 - v_2) = (v_1 - v_2, u_2 - u_1) = 0,$$

since $u_1 - u_2 \in U$ and $v_2 - v_1 \in U^{\perp}$.

If $P_U x$ denotes the closest point to x in U then clearly $P_U^2 = P_U$, and it follows from the definition of u that

$$\|x\|^2 = \|u\|^2 + \|x - u\|^2,$$

thus ensuring that

$$\|P_U x\| \leq \|x\|,$$

i.e. the projection can only decrease the norm. □

We will find an expression for P_U in Theorem 8.9

8.3 Closed linear span

Recall that the linear span of a subset E of H is the collection of all finite linear combinations of elements of E,

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

It is easy to show that this is a linear subspace of H, but it need not be closed.

The *closed linear span* of E is the intersection of all closed subsets of H that contain E. More usefully, it is the closure of Span(E), i.e. the set of elements of H that can be approximated arbitrarily closely by finite linear combinations of elements of E:

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

In general it is *not* true that clin(E) is equal to

$$\{u \in H : \ u = \sum_{j=1}^{\infty} \alpha_j e_j, \ \alpha_j \in \mathbb{K}, \ e_j \in E\}.$$

Indeed, if $\{w_j\}$ is an orthonormal basis for H, consider

$$\begin{aligned} e_1 &= w_1 \\ e_2 &= w_1 + \frac{1}{2} w_2 \\ e_3 &= w_1 + \frac{1}{2} w_2 + \frac{1}{3} w_3 \\ &\vdots \\ e_n &= \sum_{j=1}^{n} \frac{1}{j}\, w_j. \end{aligned}$$

Then

$$e_n \to x = \sum_{j=1}^{\infty} \frac{1}{j}\, w_j,$$

since

$$\sum \frac{1}{j}^2 = \frac{\pi^2}{6} < \infty.$$

However, one cannot write

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

since this taking the inner product of both sides with w_k implies that

$$\frac{1}{k} = (x, w_k) = \sum_{j=1}^{\infty} \alpha_j (e_j, w_k) = \frac{1}{k} \sum_{j=k}^{\infty} \alpha_j,$$

i.e.

$$1 = \sum_{j=1}^{\infty} \alpha_j, \qquad \frac{1}{2} = \frac{1}{2} \sum_{j=2}^{\infty} \alpha_j \qquad , \cdots , \qquad \frac{1}{k} = \frac{1}{k} \sum_{j=k}^{\infty} \alpha_j,$$

which implies that $\alpha_j = 0$ for all j, clearly impossible.

However, if E is orthonormal then:

Lemma 8.7 *Suppose that $E = \{e_j\}_{j=1}^{\infty}$ is an orthonormal set such that $H = \text{clin}(E)$. Then E is an orthonormal basis for H.*

Proof Take a $y \in H$ such that $(y, e_j) = 0$ for all $j \in \mathbb{N}$. Suppose that $x_n \in \operatorname{span} E$ such that $x_n \to y$. Then

$$(y, y) = \lim_{n \to \infty} (y, x_n) = 0,$$

i.e. $\|y\|^2 = 0$ and so $y = 0$. It follows from part (c) of Proposition 7.12 that E is a basis for H. □

As a corollary:

Corollary 8.8 *Suppose that $E = \{e_j\}_{j=1}^{\infty}$ is an orthonormal set. Then*

$$\text{clin}(E) = \{u \in H : \ u = \sum_{j=1}^{\infty} \alpha_j e_j, \ \alpha_j \in \mathbb{K}, \ e_j \in E\}. \tag{8.2}$$

Proof Let $\mathscr{H} = \text{clin}(E)$. We show that $\mathscr{H}$ is a Hilbert space; it then follows from the above lemma that E is a basis for $\mathscr{H}$, i.e. that (8.2) holds.

It is clear that $\mathscr{H}$ is a vector space; we equip $\mathscr{H}$ with the inner product of H. We only need show that $\mathscr{H}$ is complete. If $\{u_n\}$ is a Cauchy sequence in $\mathscr{H}$, with

$$u^{(n)} = \sum_{j=1}^{\infty} \alpha_j^{(n)} e_j$$

then the argument of Proposition 4.5 shows that $\alpha_j^{(n)} \to \alpha_j^*$ for each j, and that $u^{(n)} \to u^* = \sum_j \alpha_j^* e_j$, i.e. $\mathscr{H}$ is complete. □

8.4 Best approximations

We now investigate the best approximation of elements of H using the closed linear span of an orthonormal set E. Of course, if E is a basis then there is no approximation involved.

Theorem 8.9 *Let E be an orthonormal set $E = \{e_j\}_{j\in\mathscr{J}}$, where $\mathscr{J} = (1, 2, \dots, n)$ or $\mathbb{N}$. Then for any $u \in H$, the closest point to x in* $\mathrm{clin}(E)$ *is given by*

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

In particular the orthogonal projection of X onto $\mathrm{clin}(E)$ *is given by*

$$P_{\mathrm{clin}(E)}x = \sum_{j\in\mathscr{J}} (x, e_j)e_j.$$

Proof Consider $x - \sum_j \alpha_j e_j$. Then

$$\begin{aligned}
\left\| x - \sum_j \alpha_j e_j \right\|^2 &= \|x\|^2 - \sum_j (x, \alpha_j e_j) - \sum_j (\alpha_j e_j, x) + \sum_j |\alpha_j|^2 \\
&= \|x\|^2 - \sum_j \bar{\alpha}_j (x, e_j) - \sum_j \alpha_j \overline{(x, e_j)} + \sum_j |\alpha_j|^2 \\
&= \|x\|^2 - \sum_j |(x, e_j)|^2 \\
&\qquad + \sum_j \left[|(x, e_j)|^2 - \bar{\alpha}_j (x, e_j) - \alpha_j \overline{(x, e_j)} + |\alpha_j|^2 \right] \\
&= \|x\|^2 - \sum_j |(x, e_j)|^2 + \sum_j |(x, e_j) - \alpha_j|^2,
\end{aligned}$$

and so the minimum occurs when $\alpha_j = (x, e_j)$ for all j. $\square$

Example 8.10 *The best approximation of an element* $\underline{x} \in \ell^2$ *in terms of* $\{\underline{e}_j\}_{j=1}^n$ *(elements of the standard basis) is simply*

$$\sum_{j=1}^n (\underline{x}, \underline{e}_j)\underline{e}_j = (x_1, x_2, \ldots, x_n, 0, 0, \ldots).$$

Example 8.11 *If* $E = \{e_j\}_{j=1}^\infty$ *is an orthonormal basis in* H *then the best approximation of an element of* H *in terms of* $\{e_j\}_{j=1}^n$ *is just given by the partial sum*

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

Now suppose that E is a finite or countable set that is not orthonormal. We can still find the best approximation to any $u \in H$ that lies in $\mathrm{clin}(E)$ by using the Gram-Schmidt orthonormalisation process:

Proposition 8.12 (Gram-Schmidt orthonormalisation) *Given a set* $E = \{e_j\}_{j \in \mathscr{J}}$ *with* $\mathscr{J} = \mathbb{N}$ *or* $\mathscr{J} = (1, \ldots, n)$ *there exists an orthonormal set* $\tilde{E} = \{\tilde{e}_j\}_{j \in \mathscr{J}}$ *such that*

$$\mathrm{Span}(e_1, \ldots, e_k) = \mathrm{Span}(\tilde{e}_1, \ldots, \tilde{e}_k)$$

for every $k \in \mathbb{N}$.

Proof First omit all elements of $\{e_n\}$ which can be written as a linear combination of the preceding ones.

Now suppose that we already have an orthonormal set $(\tilde{e}_1, \ldots, \tilde{e}_n)$ whose span is the same as $(e_1, \ldots, e_n)$. Then we can define $\tilde{e}_{n+1}$ by setting

$$e'_{n+1} = e_{n+1} - \sum_{i=1}^n (e_{n+1}, \tilde{e}_i)\tilde{e}_i \qquad \text{and} \qquad \tilde{e}_{n+1} = \frac{e'_{n+1}}{\|e'_{n+1}\|}.$$

The span of $(\tilde{e}_1, \ldots, \tilde{e}_{n+1})$ is clearly the same as the span of $(\tilde{e}_1, \ldots, \tilde{e}_n, e_{n+1})$, which is the same as the span of $(e_1, \ldots, e_n, e_{n+1})$ using the induction hypothesis. Clearly $\|\tilde{e}_{n+1}\| = 1$ and for $m \leq n$ we have

$$(\tilde{e}_{n+1}, \tilde{e}_m) = \frac{1}{\|e'_{n+1}\|}\Big((e_{n+1}, \tilde{e}_m) - \sum_{i=1}^n (e_{n+1}, \tilde{e}_i)(\tilde{e}_i, \tilde{e}_m)\Big) = 0$$

since $(\tilde{e}_1, \ldots, \tilde{e}_n)$ are orthonormal. Setting $\tilde{e}_1 = e_1/\|e_1\|$ starts the induction. $\square$

Example 8.13 *Consider approximation of functions in $L^2(-1,1)$ with polynomials of degree up to n. We can start with the set $\{1, x, x^2, \ldots, x^n\}$, and then use the Gram-Schmidt process to construct polynomials that are orthonormal w.r.t. the $L^2(-1,1)$ inner product.*

We begin with $e_1 = 1/\sqrt{2}$, and then consider

$$e_2' = x - \left(x, \frac{1}{\sqrt{2}}\right) \frac{1}{\sqrt{2}} = x - \frac{1}{2} \int_{-1}^{1} t \, \mathrm{d}t = x$$

so

$$\|e_2'\|^2 = \int_{-1}^{1} t^2 \, \mathrm{d}t = \frac{2}{3}; \qquad e_2 = \sqrt{\frac{3}{2}}\, x.$$

Then

$$\begin{aligned} e_3' &= x^2 - \left(x^2, \sqrt{\frac{3}{2}} x\right) \sqrt{\frac{3}{2}} x - \left(x^2, \frac{1}{\sqrt{2}}\right) \frac{1}{\sqrt{2}} \\ &= x^2 - \frac{3x}{2} \int_{-1}^{1} \frac{3t^3}{2} \, \mathrm{d}t - \frac{1}{2} \int_{-1}^{1} t^2 \, \mathrm{d}t \\ &= x^2 - \frac{1}{3}, \end{aligned}$$

so

$$\|e_3'\|^2 = \int_{-1}^{1} \left(t^2 - \frac{1}{3}\right)^2 \mathrm{d}t = \left[\frac{t^5}{5} - \frac{2t^3}{9} + \frac{t}{9}\right]_{-1}^{1} = \frac{8}{45}$$

which gives

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

Exercise 8.14 *Show that $e_4 = \sqrt{\frac{7}{8}}\,(5x^3 - 3x)$, and check that this is orthogonal to e_1, e_2, and e_3.*

Using these orthonormal functions we can find the best approximation of any function $f \in L^2(-1,1)$ by a degree three polynomial:

$$\left(\frac{7}{8} \int_{-1}^{1} f(t)(5t^3 - 3t) \, \mathrm{d}t\right) (5x^3 - 3x) + \left(\frac{5}{8} \int_{-1}^{1} f(t)(3t^2 - 1) \, \mathrm{d}t\right) (3x^2 - 1)$$
$$+ \left(\frac{3}{2} \int_{-1}^{1} f(t) t \, \mathrm{d}t\right) x + \frac{1}{2} \int_{-1}^{1} f(t) \, \mathrm{d}t.$$

Example 8.15 *The best approximation of* $f(x) = |x|$ *by a third degree polynomial is*

$$f_3(x) = \left(\frac{5}{4}\int_0^1 3t^3 - t\,\mathrm{d}t\right)(3x^2-1) + \int_0^1 t\,\mathrm{d}t = \frac{5}{16}(3x^2-1) + \frac{1}{2} = \frac{15x^2+3}{16}.$$

We have (after some tedious integration)

$$\|f - f_3\|^2 = \frac{2}{16^2}\int_0^1 (15x^2 - 16x + 3)^2\,\mathrm{d}x = \frac{3}{16}$$

Of course, the meaning of the 'best approximation' is that this choice minimises the L^2 norm of the difference. It is not the 'best approximation' in terms of the supremum norm: at $x = 0$ the value of $f_3(x) = 3/16$, while

$$\sup_{x\in[0,1]} |x - (x^2 + \frac{1}{8})| = \frac{1}{8}.$$

Exercise 8.16 *Find the best approximation (w.r.t.* $L^2(-1,1)$ *norm) of* $\sin x$ *by a third degree polynomial.*

Exercise 8.17 *Find the first four polynomials that are orthogonal on* $L^2(0,1)$ *with respect to the usual* L^2 *inner product.*

By using different 'weights' in our definition of the inner product we can find sets of polynomials that are orthonormal in different senses. For a function $w > 0$ on (a,b) define the $L^2_w(a,b)$ inner product by

$$(f,g)_w = \int_a^b w(t)f(t)g(t)\,\mathrm{d}t.$$

It is simple to check that this is an inner product on the space of all functions for which

$$\int_a^b w(t)|f(t)|^2\,\mathrm{d}t < +\infty.$$

Exercise 8.18 *The Tchebyshev polynomials are obtained by taking* $w(x) = (1-x^2)^{-1/2}$ *on the interval* $(-1,1)$*. Find the first three Tchebyshev polynomials.*

9

Separable Hilbert spaces and ℓ^2

We start with a definition.

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

This is an approximation property: one can find a countable set $\{x_n\}_{n=1}^{\infty}$ such that given any $u \in H$ and $\epsilon > 0$, there exists an x_j such that

$$\|x_j - u\| < \epsilon.$$

Example 9.2 $\mathbb{R}$ *is separable, since* $\mathbb{Q}$ *is a countable dense subset. So is* $\mathbb{R}^n$*, since* $\mathbb{Q}^n$ *is countable and dense.* $\mathbb{C}$ *is separable since complex numbers of the form* $q_1 + \mathrm{i}q_2$ *with* $q_1, q_2 \in \mathbb{Q}$ *is countable and dense.*

Example 9.3 ℓ^2 *is separable, since sequences of the form*

$$\underline{x} = (x_1, \ldots, x_n, 0, 0, 0, \ldots)$$

with $x_1, \ldots, x_n \in \mathbb{Q}$ *are dense.*

We now show that $C^0([0,1])$ is separable, by proving the Weierstrass approximation theorem: every continuous function can be approximated arbitrarily closely (in the supremum norm) by a polynomial.

Theorem 9.4 *Let $f(x)$ be a real-valued continuous function on $[0,1]$. Then the sequence of polynomials*

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

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

Proof Start with the identity

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

Differentiate with respect to x and multiply by x to give

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

differentiate twice with respect to x and multiply by x^2 to give

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

It follows that if we write

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

we have

$$\sum_{p=0}^{n}r_p(x)=1,\quad \sum_{p=0}^{n}pr_p(x)=nx,\quad \text{and}\quad \sum_{p=0}^{n}p(p-1)r_p(x)=n(n-1)x^2.$$

Therefore

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

Since f is continuous on the closed bounded interval it is bounded, $|f(x)|\le M$ for some $M>0$. It also follows that f is uniformly continuous on $[0,1]$, so for any $\epsilon>0$ there exists a $\delta>0$ such that

$$|x-y|<\delta \qquad \Rightarrow |f(x)-f(y)|<\epsilon.$$

Since $\sum_{p=0}^n r_p(x) = 1$ we have

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

For the first term on the right-hand side we have

$$\left|\sum_{|(p/n)-x|\leq\delta} (f(x) - f(p/n)) r_p(x)\right| \leq \epsilon \sum r_p(x) = \epsilon,$$

and for the second term on the right-hand side

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

which tends to zero as $n \to \infty$.

□

One could also state this as: the set of polynomials is dense in $C^0([0,1])$ equipped with the supremum norm.

Proposition 9.5 $C^0([0,1])$ *is separable.*

Proof Given any $f \in C^0([0,1])$, it can be approximated to within ϵ by some polynomial, i.e.

$$\left\| f - \left(\sum_{j=1}^N a_n x^n\right)\right\|_\infty < \epsilon.$$

While the set of all polynomials in not countable, the set of all polynomials with rational coefficients is. Since

$$\left\| \left(\sum_{j=1}^{N} a_n x^n \right) - \left(\sum_{j=1}^{N} b_n x^n \right) \right\|_\infty \leq \sum_{j=1}^{N} |a_n - b_n|,$$

one can choose $b_n \in \mathbb{Q}$ such that $|a_n - b_n| < \epsilon/2N$, and then

$$\left\| f - \left(\sum_{j=1}^{N} b_n x^n \right) \right\|_\infty < \epsilon.$$

□

If we now use the fact that $C^0([0,1])$ is dense in $L^2(0,1)$, it follows that:

Proposition 9.6 $L^2(0,1)$ *is separable.*

Proof Take $f \in L^2(0,1)$. Given $\epsilon > 0$ there exists a $g \in C^0([0,1])$ such that $\|f-g\|_{L^2} < \epsilon/2$. We know from above that there exists a polynomial h with rational coefficients such that

$$\|g - h\|_\infty < \epsilon/2.$$

Since

$$\|g - h\|_{L^2}^2 = \int_0^1 |g(x) - h(x)|^2 \, \mathrm{d}x \leq \int_0^1 \|g - h\|_\infty^2 \, \mathrm{d}x = \|g - h\|_\infty^2,$$

it follows that

$$\|f - h\|_{L^2} \leq \|f - g\|_{L^2} + \|g - h\|_\infty < \epsilon.$$

□

The property of separability seems very strong, but it is a simple consequence of the existence of a countable orthonormal basis, as we now show.

Proposition 9.7 *An infinite-dimensional Hilbert space is separable iff it has a countable orthonormal basis.*

Note that this shows immediately that the unit ball in a separable Hilbert space is not compact.

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

If H is separable, let $E = \{x_n\}$ be a countable dense subset. In particular, the closed linear span of E is the whole of H. The Gram-Schmidt process now provides a countable orthonormal set whose closed linear span is all of H, i.e. a countable orthonormal basis. $\square$

Note that there are Hilbert spaces that are not separable. For example, if Γ is uncountable then the space $\ell^2(\Gamma)$ consisting of all functions $f : \Gamma \to \mathbb{R}$ such that

$$\sum_{\gamma\in\Gamma} |f(\gamma)|^2 < \infty$$

is a Hilbert space but is not separable.

By using these basis elements we can construct an isomorphism between any separable Hilbert space and ℓ^2, so that in some sense ℓ^2 is "the only" separable infinite-dimensional Hilbert space:

Theorem 9.8 *Any infinite-dimensional separable Hilbert space H is isometric to $\ell^2(\mathbb{K})$.*

Proof Since H is separable it has a countable orthonormal basis $\{e_j\}$. Define $\varphi : H \to \ell^2$ by the map

$$u \mapsto \Big((u, e_1), (u, e_2), \ldots, (u, e_n), \ldots\Big);$$

clearly the inverse map is given by

$$\underline{\alpha} \mapsto \sum_{j=1}^{\infty} \alpha_j e_j \qquad \text{where} \qquad \underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots).$$

By the result of Lemma 7.10, $u \in H \Rightarrow \varphi(u) \in \ell^2$ and $\underline{\alpha} \in \ell^2 \Rightarrow \varphi^{-1}(\underline{\alpha}) \in H$, while the characterisation of a basis in Proposition 7.12 shows that $\|u\|_H = \|\varphi(u)\|_{\ell^2}$. $\square$

10

Linear maps between Banach spaces

We now consider linear maps between Banach spaces.

10.1 Bounded linear maps

Definition 10.1 *If U and V are vector spaces over $\mathbb{K}$ then an operator A from U into V is* linear *if*

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

The collection $L(U, V)$ of all linear operators from U into V is a vector space.

Definition 10.2 *A linear operator A from a normed space $(X, \|\cdot\|_X)$ into another normed space $(Y, \|\cdot\|_Y)$ is* bounded *if there exists a constant M such that*

$$\|Ax\|_Y \le M\|x\|_X \qquad \text{for all} \qquad x \in X. \tag{10.1}$$

Linear operators on infinite-dimensional spaces need not be bounded.

Lemma 10.3 *A linear operator $T : X \to Y$ is continuous iff it is bounded.*

Proof Suppose that T is bounded; then for some $M > 0$

$$\|Tx_n - Tx\|_Y = \|T(x_n - x)\|_Y \le M\|x_n - x\|_X,$$

and so T is continuous. Now suppose that T is continuous; then in particular it is continuous at zero, and so, taking $\epsilon = 1$ in the definition of continuity, there exists a $\delta > 0$ such that

$$\|Tx\| \leq 1 \qquad \text{for all} \qquad \|x\| \leq \delta.$$

It follows that

$$\|Tz\| = \left\| T\left(\frac{\|z\|}{\delta}\frac{\delta z}{\|z\|}\right)\right\| = \frac{\|z\|}{\delta}\left\| T\left(\frac{\delta z}{\|z\|}\right)\right\| \leq \frac{1}{\delta}\|z\|,$$

and so T is bounded. □

The space of all bounded linear operators from X into Y is denoted by $B(X, Y)$.

Definition 10.4 *The* operator norm *of an operator A (from X into Y) is the smallest value of M such that (10.1) holds,*

$$\|A\|_{B(X,Y)} = \inf\{M : \text{ (10.1) holds}\}. \tag{10.2}$$

Note that – and this is the key point – it follows that

$$\|Ax\|_Y \leq \|A\|_{B(X,Y)}\|x\|_X \qquad \text{for all} \qquad x \in X.$$

(Since for each $x \in X$, $\|Ax\|_Y \leq M\|x\|_X$ for every $M > \|A\|_{B(X,Y)}$).

Lemma 10.5 *The following is an equivalent definition of the operator norm:*

$$\|A\|_{B(X,Y)} = \sup_{\|x\|_X = 1} \|Ax\|_Y. \tag{10.3}$$

Proof Let us denote by $\|A\|_1$ the value defined in (10.2), and by $\|A\|_2$ the value defined in (10.3). Then given $x \neq 0$ we have

$$\left\| A\frac{x}{\|x\|_X}\right\|_Y \leq \|A\|_2 \qquad \text{i.e.} \quad \|Ax\|_Y \leq \|A\|_2\|x\|_X,$$

and so $\|A\|_1 \leq \|A\|_2$. It is also clear that if $\|x\|_X = 1$ then

$$\|Ax\|_Y \leq \|A\|_1\|x\|_X = \|A\|_1,$$

and so $\|A\|_2 \leq \|A\|_1$. It follows that $\|A\|_1 = \|A\|_2$. □

Exercise 10.6 *Show that also*

$$\|A\|_{B(X,Y)} = \sup_{\|x\|_X \le 1} \|Ax\|_Y$$

and

$$\|A\|_{B(X,Y)} = \sup_{x \neq 0} \frac{\|Ax\|_Y}{\|x\|_X}. \tag{10.4}$$

When there is no room for confusion we will omit the $B(X,Y)$ subscript on the norm, sometimes adding the subscript "op" (for "operator") to make things clearer ($\|\cdot\|_{\mathrm{op}}$).

If $T : X \to Y$ then in order to find $\|T\|_{\mathrm{op}}$ one can try the following: first show that

$$\|Tx\|_Y \le M\|x\|_X \tag{10.5}$$

for some $M > 0$, i.e. show that T is bounded. It is then clear that $\|T\|_{\mathrm{op}} \le M$ (since $\|T\|_{\mathrm{op}}$ is the infimum of all M such that (10.5) holds). Then, in order to show that in fact $\|T\|_{\mathrm{op}} = M$, find an example of a particular $z \in X$ such that

$$\|Tz\|_Y = M\|z\|_X.$$

This shows from the definition in (10.4) that $\|T\|_{\mathrm{op}} \ge M$ and hence that in fact $\|T\|_{\mathrm{op}} = M$.

Example 10.7 *Consider the right and left shift operators on ℓ^2, σ_{r} and σ_{l} given by*

$$\sigma_{\mathrm{r}}(\underline{x}) = (0, x_1, x_2, \ldots) \qquad \textit{and} \qquad \sigma_{\mathrm{l}}(\underline{x}) = (x_2, x_3, x_4, \ldots).$$

Both operators are clearly linear. We have

$$\|\sigma_{\mathrm{r}}(\underline{x})\|_{\ell^2}^2 = \sum_{i=1}^{\infty} |x_i|^2 = \|\underline{x}\|_{\ell^2}^2,$$

so that $\|\sigma_{\mathrm{r}}\|_{\mathrm{op}} = 1$, and

$$\|\sigma_{\mathrm{l}}(\underline{x})\|_{\ell^2}^2 = \sum_{i=2}^{\infty} |x_i|^2 \le \|\underline{x}\|_{\ell^2}^2,$$

so that $\|\sigma_{\mathrm{l}}\|_{\mathrm{op}} \le 1$. However, if we choose an $\underline{x}$ with

$$\underline{x} = (0, x_2, x_3, \ldots)$$

then we have

$$\|\sigma_1(\underline{x})\|_{\ell^2}^2 = \sum_{j=2}^{\infty} |x_j|^2 = \|\underline{x}\|_{\ell^2}^2,$$

and so we must have $\|\sigma_1\|_{\mathrm{op}} = 1$.

In slightly more involved examples one might have to be a little more crafty; for example, given the bound $\|Tx\|_Y \leq M\|x\|_X$ find a sequence $z_n \in X$ such that

$$\frac{\|Tz_n\|_Y}{\|z_n\|_X} \to M$$

as $n \to \infty$, which again shows using (10.4) that $\|T\|_{\mathrm{op}} \geq M$ and hence that $\|T\|_{\mathrm{op}} = M$.

Example 10.8 *Consider the space* $L^2(a,b)$ *with* $-\infty < a < b < +\infty$ *and the multiplication operator* T *from* $L^2(a,b)$ *into itself given by*

$$Tx(t) = f(t)x(t) \qquad t \in [a,b]$$

where $f \in C^0([a,b])$. *Then clearly* T *is linear and*

$$\begin{aligned} \|Tx\|^2 &= \int_a^b |f(t)x(t)|^2 \,\mathrm{d}t \\ &= \int_a^b |f(t)|^2 |x(t)|^2 \,\mathrm{d}t \\ &\leq \max_{a\leq t\leq b} |f(t)|^2 \int_a^b |x(t)|^2 \,\mathrm{d}t, \end{aligned}$$

and so

$$\|Tx\|_{L^2} \leq \|f\|_\infty \|x\|_{L^2},$$

i.e. $\|T\|_{\mathrm{op}} \leq \|f\|_\infty$.

Now let s *be a point at which* $|f|$ *is maximum. Assume for simplicity that* $s \in (a,b)$, *and for each* $\epsilon > 0$ *consider*

$$x_\epsilon(t) = \begin{cases} 1 & |t-s| < \epsilon \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\frac{\|Tx_\epsilon\|^2}{\|x_\epsilon\|^2} = \frac{1}{2\epsilon} \int_{s-\epsilon}^{s+\epsilon} |f(t)|^2 \,\mathrm{d}t \to |f(s)|^2 \qquad \text{as} \qquad \epsilon \to 0$$

since f is continuous. Therefore in fact

$$\|T\|_{\rm op} = \|f\|_\infty.$$

If $s = a$ then we can replace $|t-s| < \epsilon$ in the definition of x_ϵ by $a \le t < a+\epsilon$, and if $s = b$ we replace it by $b-\epsilon < t \le b$; the rest of the argument is identical.

We now give a very important particular example.

Example 10.9 *Consider the map from $L^2(a,b)$ into itself given by the integral*

$$(Tx)(t) = \int_a^b K(t,s)x(s)\,\mathrm{d}s \qquad \textit{for all} \qquad t \in [a,b]$$

where

$$\int_a^b \int_a^b |K(t,s)|^2\,\mathrm{d}s\,\mathrm{d}t < +\infty.$$

Then T is clearly linear, and

$$\begin{aligned} \|Tx\|^2 &= \int_a^b \left| \int_a^b K(t,s)x(s)\,\mathrm{d}s \right|^2 \mathrm{d}t \\ &\le \int_a^b \left[\int_a^b |K(t,s)|^2\,\mathrm{d}s \int_a^b |x(s)|^2\,\mathrm{d}s \right] \mathrm{d}t \quad \textit{by Cauchy-Schwarz} \\ &= \left(\int_a^b \int_a^b |K(t,s)|^2\,\mathrm{d}s\,\mathrm{d}t \right) \|x\|^2, \end{aligned}$$

and so

$$\|T\|_{\rm op}^2 \le \int_a^b \int_a^b |K(t,s)|^2\,\mathrm{d}s\,\mathrm{d}t.$$

Note that this upper bound on the operator norm can be strict. Indeed, we have already shown that $e_1 = 1/\sqrt{2}$ and $e_2(x) = \sqrt{3/2}x$ are orthonormal functions in $L^2(-1,1)$; consider

$$K(t,s) = 1 + 6ts = 2e_1(t)e_1(s) + 4e_2(t)e_2(s).$$

Then

$$\begin{aligned} \int_{-1}^1 \int_{-1}^1 |K(t,s)|^2\,\mathrm{d}t\,\mathrm{d}s &= \int_{-1}^1 \int_{-1}^1 \left(1 + 12ts + 36t^2s^2\right)\,\mathrm{d}t\,\mathrm{d}s \\ &= \int_{-1}^1 2 + 24s^2\,\mathrm{d}s = 4 + 16 = 20, \end{aligned}$$

while

$$Tx(t) = \int_a^b K(t,s)x(s)\,\mathrm{d}s = 2e_1(t)(x,e_1) + 4e_1(t)(x,e_2).$$

Since e_1 and e_2 are orthonormal,

$$\|Tx\|_{L^2}^2 = 4|(x,e_1)|^2 + 16|(x,e_2)|^2 \le 16(|(x,e_1)|^2 + |(x,e_2)|^2) \le 16\|x\|_{L^2}^2,$$

and so $\|T\|_{\mathrm{op}} \le 4$. In fact, since $Te_2 = 4e_2$, we have $\|T\|_{\mathrm{op}} = 4 < \sqrt{20}$.

The space $B(X,Y)$ is a Banach space whenever Y is a Banach space. Remarkably this does not depend on whether the space X is complete or not.

Theorem 10.10 *Let X be a normed space and Y a Banach space. Then $B(X,Y)$ is a Banach space.*

Proof Let $\{A_n\}$ be a Cauchy sequence in $B(X,Y)$. We need to show that $A_n \to A$ for some $A \in \mathcal{L}(X,Y)$. Since $\{A_n\}$ is Cauchy, given $\epsilon > 0$ there exists an N such that

$$\|A_n - A_m\|_{\mathrm{op}} \le \epsilon \qquad \text{for all} \quad n,m \ge N. \tag{10.6}$$

We now show that for every fixed $x \in X$ the sequence $\{A_n x\}$ is Cauchy in Y. This follows since

$$\|A_n x - A_m x\|_Y = \|(A_n - A_m)x\|_Y \le \|A_n - A_m\|_{\mathrm{op}}\,\|x\|_X, \tag{10.7}$$

and $\{A_n\}$ is Cauchy in $B(X,Y)$. Since Y is complete, it follows that

$$A_n x \to y,$$

where y depends on x. We can therefore define a mapping $A : X \to Y$ by $Ax = y$. We still need to show, however, that $A \in B(X,Y)$ and that $A_n \to A$ in the operator norm.

First, A is linear since

$$A(x + \lambda y) = \lim_{n\to\infty} A_n(x+\lambda y) = \lim_{n\to\infty} A_n x + \lambda \lim_{n\to\infty} A_n y = Ax + \lambda Ay.$$

To show that A is bounded take $n,m \ge N$ (from (10.6)) in (10.7), and let $m \to \infty$. Since $A_m x \to Ax$ this shows that

$$\|A_n x - Ax\|_Y \le \epsilon \|x\|_X. \tag{10.8}$$

Since (10.8) holds for every x it follows that

$$\|A_n - A\|_{\mathrm{op}} \le \epsilon, \tag{10.9}$$

and so $A_n - A \in B(X, Y)$. Since $B(X, Y)$ is a vector space and $A_n \in B(X, Y)$, it follows that $A \in B(X, Y)$, and (10.9) shows that $A_n \to A$ in $B(X, Y)$. $\square$

10.2 Kernel and range

Given a linear operator, we define its kernel

$$\operatorname{Ker} T = \{x \in X : \; Tx = 0\}$$

and its range

$$\operatorname{Range} T = \{y \in Y : \; y = Tx \text{ for some } x \in X\}.$$

Corollary 10.11 *If $T \in B(X, Y)$ then $\operatorname{Ker} T$ is a closed linear subspace of X.*

Proof It is easy to show that $\operatorname{Ker}(T)$ is a linear subspace, since if $x, y \in \operatorname{Ker}(T)$ then

$$T(\alpha x + \beta y) = \alpha Tx + \beta Ty = 0.$$

Furthermore if $x_n \to x$ and $Tx_n = 0$ then since T is continuous $Tx = \lim_{n\to\infty} Tx_n = 0$, so $\operatorname{Ker}(T)$ is closed. $\square$

Note that the range is not necessarily closed. Indeed, consider the map from ℓ^2 into itself given by

$$T\underline{x} = (x_1, \frac{x_2}{2}, \frac{x_3}{3}, \frac{x_4}{4}, \ldots).$$

Then clearly $\|T\|_{\mathrm{op}} \le 1$, so T is bounded. Now consider

$$\underline{y}^{(n)} = T(\underbrace{1, 1, \ldots, 1}_{n \text{ times}}, 0 \ldots) = (1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, 0, \ldots).$$

Then clearly $\underline{y}^{(n)} \to \underline{y}$ with $y_j = j^{-1}$, and since

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6} < \infty$$

it follows that $\underline{y} \in \ell^2$. However, there is no $\underline{x} \in \ell^2$ such that $T(\underline{x}) = \underline{y}$: the only candidate is $\underline{x} = (1, 1, 1, \ldots)$, but this is not in ℓ^2 since its ℓ^2 norm is not finite.

11

The Riesz representation theorem and the adjoint operator

If U is a normed space over $\mathbb{K}$ then a linear map from U into $\mathbb{K}$ is called a *linear functional* on U.

Since $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ is complete then by Theorem 10.10 the collection of all linear functionals on U, $B(U, \mathbb{K})$, is a Banach space. This space is termed the *dual space* of U, and is denoted by U^*.

For any $f \in U^*$,

$$\|f\|_{U^*} = \sup_{\|u\|=1} |f(u)|.$$

Example 11.1 *Take $U = C^0([a,b])$, and consider δ_x defined for $x \in [a,b]$ by*

$$\delta_x(f) = f(x) \qquad \text{for all} \qquad f \in U.$$

Then

$$|\delta_x(f)| = |f(x)| \leq \|f\|_\infty,$$

so that $\delta_x \in U^*$ with $\|\delta_x\| \leq 1$. Choosing a function $f \in C^0([a,b])$ such that $|f(x)| = \|f\|_\infty$ shows that in fact $\|\delta_x\| = 1$.

(Note: this shows that – at least for this particular choice of U – knowledge of $T(f)$ for all $T \in U^*$ determines $f \in U$. This result is in fact true in general.)

Example 11.2 *Let U be the real vector space $L^2(a,b)$, and take $\phi \in C^0([a,b])$.*

Consider

$$f(u) = \int_a^b \phi(t)u(t)\,\mathrm{d}t.$$

Then

$$\begin{aligned} |f(u)| &= \left|\int_a^b \phi(t)u(t)\,\mathrm{d}t\right| \\ &= |(\phi, u)_{L^2}| \\ &\le \|\phi\|_{L^2}\|u\|_{L^2} \quad \text{using the Cauchy-Schwarz inequality,} \end{aligned}$$

and so $f \in U^*$ *with*

$$\|f\| \le \|\phi\|_{L^2}.$$

If we choose $u = \frac{\phi}{\|\phi\|_{L^2}}$ *then* $\|u\|_{L^2} = 1$ *and*

$$|f(u)| = \int_a^b \frac{|\phi(t)|^2}{\|\phi\|_{L^2}}\,\mathrm{d}t = \|\phi\|_{L^2}$$

and so $\|f\| = \|\phi\|$.

Exercise 11.3 *Let* U *be* $C^0([a,b])$ *and for some* $\phi \in U$ *consider* f_ϕ *defined as*

$$f_\phi(u) = \int_a^b \phi(t)u(t)\,\mathrm{d}t \qquad \text{for all} \qquad u \in U.$$

Show that $f_\phi \in U^*$ *with* $\|f_\phi\| \le \int_a^b |\phi(t)|\,\mathrm{d}t$. *Show that this is in fact an equality by choosing an appropriate sequence of functions* $u_n \in U$ *for which* $|f_\phi(u_n)|/\|u_n\|_\infty \to \int_a^b |\phi(t)|\,\mathrm{d}t$.

Example 11.4 *Let* U *be a Hilbert space. Given any* $y \in H$, *define*

$$l_y(x) = (x, y). \tag{11.1}$$

Then l_y *is clearly linear, and*

$$|l_y(x)| = |(x,y)| \le \|x\|\|y\|$$

using the Cauchy-Schwarz inequality. It follows that $l_y \in H^*$ *with* $\|l_y\| \le \|y\|$. *Choosing* $x = y$ *in (11.1) shows that*

$$|l_y(y)| = (y,y) = \|y\|^2$$

and hence $\|l_y\| = \|y\|$.

The Riesz Representation Theorem shows that this example can be 're-versed', i.e. every linear functional on H corresponds to some inner product:

Theorem 11.5 (Riesz Representation Theorem) *For every bounded linear functional f on a Hilbert space H there exists a unique element $y \in H$ such that*

$$f(x) = (x, y) \qquad \text{for all} \qquad x \in H \tag{11.2}$$

and $\|y\|_H = \|f\|_{H^*}$.

Proof Let $K = \operatorname{Ker} f$, which is a closed linear subspace of H.

First we show that $K^\perp$ is a one-dimensional linear subspace of H. Indeed, given $u, v \in K^\perp$ we have

$$l(l(u)v - l(v)u) = 0. \tag{11.3}$$

Since $u, v \in K^\perp$ it follows that $l(u)v - l(v)u \in K^\perp$, while (11.3) shows that $l(u)v - l(v)u \in K$. It follows[1] that

$$l(u)v - l(v)u = 0,$$

and so u and v are proportional.

Therefore we can choose $z \in K$ such that $\|z\| = 1$, and decompose any $x \in H$ as

$$x = (x, z)z + w \qquad \text{with} \qquad w \in K.$$

Therefore

$$l(x) = (x, z)l(z) = (x, \overline{l(z)}z).$$

Setting $y = \overline{l(z)}z$ we obtain (11.2).

To show that this choice of y is unique, suppose that

$$(x, y) = (x, \hat{y}) \qquad \text{for all} \qquad x \in H.$$

Then $(x, y - \hat{y}) = 0$ for all $x \in H$, i.e. $y - \hat{y} \in H^\perp = \{0\}$.

Finally, the calculation in (11.4) shows the equality of the norms of y and f. $\square$

We now use this to define the *adjoint* of an operator.

[1] We always have $U \cap U^\perp = \{0\}$: if $x \in U$ and $x \in U^\perp$ then $\|x\|^2 = (x, x) = 0$.

Theorem 11.6 *Let H and K be Hilbert spaces and $T \in B(H, K)$. Then there exists a unique operator $T^* \in B(K, H)$, the* adjoint of T, *such that*

$$(Tx, y)_K = (x, T^*y)_H$$

for all $x \in H$, $y \in K$. In particular, $\|T^*\|_{B(K,H)} \leq \|T\|_{B(H,K)}$.

Proof Let $y \in K$ and consider the function $f : H \to \mathbb{K}$ defined by $f(x) = (Tx, y)_K$. Then clearly f is linear and

$$\begin{aligned} |f(x)| &= |(Tx, y)_K| \\ &\leq \|Tx\|_K \|y\|_K \\ &\leq \|T\| \|x\|_H \|y\|_K. \end{aligned}$$

It follows that $f \in H^*$, and so by the Riesz representation theorem there exists a unique $z \in H$ such that

$$(Tx, y)_K = (x, z)_H \qquad \text{for all} \qquad x \in H.$$

Define $T^*y = z$. Then by definition

$$(Tx, y)_K = (x, T^*y) \qquad \text{for all} \qquad x \in H, \ y \in K.$$

However, it remains to show that $T^* \in B(K, H)$. First, T^* is linear since

$$\begin{aligned} (x, T^*(\alpha y_1 + \beta y_2))_H &= (Tx, \alpha y_1 + \beta y_2)_K \\ &= \bar{\alpha}(Tx, y_1)_K + \bar{\beta}(Tx, y_2)_K \\ &= \bar{\alpha}(x, T^*y_1)_H + \bar{\beta}(x, T^*y_2)_H \\ &= (x, \alpha T^*y_1 + \beta T^*y_2)_H, \end{aligned}$$

i.e.

$$T^*(\alpha y_1 + \beta y_2) = \alpha T^*y_1 + \beta T^*y_2.$$

To show that T^* is bounded, we have

$$\begin{aligned} \|T^*y\|^2 &= (T^*y, T^*y)_H \\ &= (TT^*y, y)_K \\ &\leq \|TT^*y\|_K \|y\|_K \\ &\leq \|T\| \|T^*y\|_H \|y\|_K. \end{aligned}$$

If $\|T^*y\|_H = 0$ then clearly $\|T^*y\|_H \leq \|T\| \|y\|_K$, otherwise we can divide both side by $\|T^*y\|_H$ to obtain the same conclusion (that T^* is bounded from K into H). So $\|T^*\| \leq \|T\|$.

Finally suppose that $(x, T^*y)_H = (x, \hat{T}y)_H$ for all $x \in H$, $y \in K$. Then for each $y \in H$, $(x, (T^* - \hat{T})y)_H = 0$ for all $x \in H$: this shows that $(T^* - \hat{T})y = 0$ for all $y \in H$, i.e. that $T = T^*$. $\square$

Example 11.7 *Let $H = K = \mathbb{C}^n$ with its standard inner product. Then given a matrix $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ we have*

$$\begin{aligned} (Ax, y) &= \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} x_j \right) \bar{y}_i \\ &= \sum_{j=1}^n x_j \sum_{i=1}^n \overline{(\bar{a}_{ij} y_i)} \\ &= (x, A^* y), \end{aligned}$$

where A^* is the Hermitian conjugate of A, i.e. $A^* = \overline{A}^T$.

Example 11.8 *Consider $H = K = L^2(0,1)$ and the integral operator*

$$(Tx)(t) = \int_0^1 K(t,s) x(s) \, \mathrm{d}s.$$

Then for $x, y \in L^2(0,1)$ we have

$$\begin{aligned} (Tx, y)_H &= \int_0^1 \int_0^1 K(t,s) x(s) \, \mathrm{d}s y(t) \, \mathrm{d}t \\ &= \int_0^1 \int_0^1 K(t,s) x(s) y(t) \, \mathrm{d}s \, \mathrm{d}t \\ &= \int_0^1 x(s) \left(\int_0^1 K(t,s) y(t) \, \mathrm{d}t \right) \, \mathrm{d}s \\ &= (x, T^* y)_H, \end{aligned}$$

where

$$T^* y(s) = \int_0^1 K(t,s) y(t) \; dt.$$

Exercise 11.9 *Show that the adjoint of the integral operator $T : L^2(0,1) \to L^2(0,1)$ defined as*

$$(Tx)(t) = \int_0^t K(t,s) x(s) \, \mathrm{d}s$$

is given by

$$(T^* y)(t) = \int_t^1 K(s,t) y(s) \, \mathrm{d}s.$$

Example 11.10 *Let $H = K = \ell^2$ and consider $\sigma_{\mathrm{r}}\underline{x} = (0, x_1, x_2, \dots)$ then*

$$\begin{aligned}(\sigma_{\mathrm{r}}\underline{x}, \underline{y}) &= x_1y_2 + x_2y_3 + x_3y_4 + \dots \\ &= (\underline{x}, \sigma_{\mathrm{r}}^*\underline{y}),\end{aligned}$$

where $\sigma_{\mathrm{r}}^* = \sigma_{\mathrm{l}}\underline{y} = (y_2, y_3, y_4, \dots)$. Similarly $\sigma_{\mathrm{l}}^* = \sigma_{\mathrm{r}}$.

The following lemma gives some elementary properties of the adjoint:

Lemma 11.11 *Let H, K, and J be Hilbert spaces, $R, S \in B(H, K)$ and $T \in B(K, J)$, then*

(a) $(\alpha R + \beta S)^* = \bar{\alpha}R^* + \bar{\beta}S^*$ and
(b) $(TR)^* = R^*T^*$.

Proof

(a) Exercise
(b) Clearly

$$\begin{aligned}(x, (TR)^*)_H &= (TRx, y)_J \\ &= (Rx, T^*y)_K = (x, R^*T^*y)_H.\end{aligned}$$

□

Less trivially we have the following:

Theorem 11.12 *Let H and K be Hilbert spaces and $T \in B(H, K)$. Then*

(a) $(T^*)^* = T$,
(b) $\|T^*\| = \|T\|$, and
(c) $\|T^*T\| = \|T\|^2$.

Proof

(a) Since $T^* \in B(K, H)$, $(T^*)^* \in B(H, K)$. For all $x \in K$, $y \in H$ we have

$$\begin{aligned}(x, (T^*)^*y)_K &= (T^*x, y)_H \\ &= \overline{(y, T^*x)_H} \\ &= \overline{(Ty, x)_K} \\ &= (x, Ty)_K,\end{aligned}$$

i.e. $(T^*)^*y = Ty$ for all $y \in H$, i.e. $(T^*)^* = T$.

(b) We have already shown in the proof of Theorem 11.6 that $\|T^*\| \leq \|T\|$. Applying this inequality to T^* rather than to T we have $\|(T^*)^*\| = \|T\| \leq \|T^*\|$, and so $\|T^*\| = \|T\|$.

(c) Since $\|T\| = \|T^*\|$ we have

$$\|T^*T\| \leq \|T^*\|\|T\| = \|T\|^2.$$

But also we have

$$\begin{aligned} \|Tx\|^2 &= (Tx, Tx) = (x, T^*Tx) \\ &\leq \|x\|\|T^*Tx\| \leq \|T^*T\|\|x\|^2, \end{aligned}$$

i.e. $\|T\|^2 \leq \|T^*T\|$.

□

11.1 Linear operators from H into H

Definition 11.13 *If H is a Hilbert space and $T \in B(H, H)$ then T is* normal *if*

$$TT^* = T^*T$$

and self-adjoint *if $T = T^*$.*

Note that if T is normal then TT^* is self-adjoint, since $(TT^*)^* = (T^*)^*T^* = TT^*$ using (b) of Lemma 11.11 ($(TR)^* = R^*T^*$) and (a) of Theorem 11.12 ($(T^*)^* = T$).

Equivalently $T \in B(H, H)$ is self-adjoint iff

$$(x, Ty) = (Tx, y) \qquad \text{for all} \qquad x, y \in H.$$

Example 11.14 *Let $H = \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. Then A is self-adjoint if $A = A^T$, i.e. if A is symmetric.*

Example 11.15 *Let $H = \mathbb{C}^n$ and $A \in \mathbb{C}^{n\times n}$. Then A is self-adjoint if $A = \overline{A}^T$, i.e. if A is Hermitian.*

Example 11.16 *Consider the right-shift operator on ℓ^2, for which $\sigma_{\mathrm{r}}^* = \sigma_{\mathrm{l}}$. Then σ_{r} is not normal, since*

$$\sigma_{\mathrm{r}}\sigma_{\mathrm{l}}\underline{x} = \sigma_{\mathrm{r}}(x_2, x_3, \ldots) = (0, x_2, x_3, \ldots)$$

and

$$\sigma_{\mathrm{l}}\sigma_{\mathrm{r}}\underline{x} = \sigma_{\mathrm{l}}(0, x_1, x_2, \ldots) = (x_1, x_2, x_3, \ldots).$$

Example 11.17 *The integral operator $T : L^2 \to L^2$ given by*

$$Tf(t) = \int_0^1 K(t,s)f(s)\,\mathrm{d}s$$

is self-adjoint if $K(t,s) = K(s,t)$, i.e. if K is symmetric.

It would be nice, of course, to have yet another expression for $\|T\|$, and we can do this when T is self-adjoint.

Theorem 11.18 *Let H be a Hilbert space and $T \in B(H,H)$ a self-adjoint operator. Then*

(a) (Tx,x) is real for all $x \in H$ and
(b) $\|T\| = \sup\{|(Tx,x)| : \ x \in H, \ \|x\| = 1\}$.

Proof For (a) we have

$$(Tx,x) = (x,Tx) = \overline{(Tx,x)},$$

and so (Tx,x) is real. Now let $M = \sup\{|(Tx,x) : \ x \in H, \ \|x\| = 1\}$. Clearly

$$|(Tx,x)| \leq \|Tx\|\|x\| \leq \|T\|\|x\|^2 = \|T\|$$

when $\|x\| = 1$, and so $M \leq \|T\|$.

For any $u, v \in H$ we have

$$\begin{aligned} 4(Tu,v) &= (T(u+v), u+v) - (T(u-v), u-v) \\ &\leq M(\|u+v\|^2 + \|u-v\|^2) \\ &\leq 2M(\|u\|^2 + \|v\|^2) \end{aligned}$$

using the parallelogram law. If $Tu \neq 0$ choose

$$v = \frac{\|u\|}{\|Tu\|}Tu$$

to obtain, since $\|v\| = \|u\|$, that

$$4\|u\|\|Tu\| \leq 4M\|u\|^2,$$

i.e. $\|Tu\| \leq M\|u\|$. This also holds if $Tu = 0$. It follows that $\|T\| \leq M$ and therefore that $\|T\| = M$. □

12

Spectral Theory I: General theory

12.1 Spectrum and point spectrum

Let H be a Hilbert space and $T \in B(H, H)$, then the *point spectrum* of T consists of the set of all eigenvalues,

$$\sigma_p(T) = \{\lambda \in \mathbb{C} : \ Tx = \lambda x \text{ for some non-zero } x \in H\}.$$

Clearly $|\lambda| \leq \|T\|_{\mathrm{op}}$ for any $\lambda \in \sigma_p(T)$, since if $Tx = \lambda x$ then

$$|\lambda|\|x\| = |\lambda x| = \|Tx\| \leq \|T\|_{\mathrm{op}}\|x\|.$$

Example 12.1 *Consider the right shift operator σ_{r} on ℓ^2. This operator has no eigenvalues, since $\sigma_r \underline{x} = \lambda \underline{x}$ implies that*

$$(0, x_1, x_2, \ldots) = \lambda(x_1, x_2, x_3, \ldots)$$

and so

$$\lambda x_1 = 0, \quad \lambda x_2 = x_1, \quad \lambda x_3 = x_2, \ldots.$$

If $\lambda \neq 0$ then this implies that $x_1 = 0$, and then $x_2 = x_3 = x_4 = \ldots = 0$, and so λ is not an eigenvalue. If $\lambda = 0$ then we also obtain $\underline{x} = 0$, and so there are no eigenvalues, i.e. $\sigma_{\mathrm{p}}(\sigma_{\mathrm{r}}) = \emptyset$.

Example 12.2 *Consider the left shift operator σ_{l} on ℓ^2; $\lambda \in \mathbb{C}$ is an eigenvalue if $\sigma_{\mathrm{l}} \underline{x} = \lambda \underline{x}$, i.e. if*

$$(x_2, x_3, x_4 \ldots) = \lambda(x_1, x_2, x_3, \ldots),$$

i.e. if

$$x_2 = \lambda x_1, \quad x_3 = \lambda x_2, \quad x_4 = \lambda x_3.$$

Given $\lambda \neq 0$ *this gives a candidate 'eigenfunction'*

$$\underline{x} = (1, \lambda, \lambda^2, \lambda^3, \ldots),$$

which is an element of ℓ^2 *provided that*

$$\sum_{n=1}^{\infty} |\lambda|^{2n} = \frac{1}{1 - |\lambda|^2} < \infty$$

which is the case for any λ *with* $|\lambda| < 1$. *It follows that*

$$\{\lambda \in \mathbb{C} : \ |\lambda| < 1\} \subseteq \sigma_{\mathrm{p}}(\sigma_{\mathrm{l}}).$$

If A is a linear operator on a finite-dimensional space V then $\lambda \in \mathbb{C}$ is an eigenvalue of A if

$$Ax = \lambda x \text{ for some non-zero } x \in V.$$

In this case λ is an eigenvalue if and only if $A - \lambda I$ is not invertible (recall that you can find the eigenvalues of an $n \times n$ matrix by solving $\det(A - \lambda I) = 0$).

However, this is no longer true in infinite-dimensional spaces. Before we define the spectrum, we briefly discuss the definition of the inverse of a linear operator.

As with the theory of matrices, the concept of the inverse of a general linear operator is extremely useful. We say that A is *invertible* if the equation

$$Ax = y$$

has a unique solution for every $y \in \mathrm{range}(A)$ (i.e. if A is injective). In this case we define the *inverse of* A, A^{-1}, by $A^{-1}y = x$. It is easy to check that

$$AA^{-1}u = u \quad \text{for all } u \in \mathrm{range}(A) \quad \text{and} \quad A^{-1}Au = u \quad \text{for all } u \in H.$$

(In some contexts 'invertible' will mean more, e.g. that A^{-1} is bounded; but we will retain this definition in these notes. [In contrast to the messy distinction I made in lectures.])

Exercise 12.3 *Show that if* A *is linear and* A^{-1} *exists then it is linear too.*

The invertibility of A is equivalent to the triviality of its kernel.

Lemma 12.4 *A is invertible iff* $\mathrm{Ker}(A) = \{0\}$.

Proof Suppose that A is invertible. Then the equation $Ax = y$ has a unique solution for any $y \in \mathrm{Range}(A)$. However, if $\mathrm{Ker}(A)$ contain some non-zero element z then $A(x+z) = y$ also, so $\mathrm{Ker}(A)$ must be $\{0\}$. Conversely, if A is not invertible then for some $y \in \mathrm{Range}(A)$ there are two distinct solutions, x_1 and x_2, of $Ax = y$, and so $A(x_1 - x_2) = 0$, giving a non-zero element of $\mathrm{Ker}(A)$. □

We can now make the following definition:

Definition 12.5 *The* resolvent set of T, $R(T)$, *is*

$$R(T) = \{\lambda \in \mathbb{C} : \ (T - \lambda I)^{-1} \in B(H,H)\}.$$

The spectrum $\sigma(T)$ *of* $T \in B(H,H)$ *is the complement of* $R(T)$,

$$\sigma(T) = \mathbb{C} \setminus R(T),$$

i.e. the spectrum of T is the set of all complex λ for which $T - \lambda I$ does not have a bounded inverse defined on all of H.

Clearly $\sigma_{\mathrm{p}}(T) \subseteq \sigma(T)$, since if there is a non-zero z with $Tz = \lambda z$ then $\mathrm{Ker}(T - \lambda I) \neq \{0\}$ and so $(T - \lambda I)$ is not invertible (using Lemma 12.4). But the spectrum can be much larger than the point spectrum; a nice example will be a consequence of the fact that $\sigma(T^*) = \overline{\sigma(T)}$.

Lemma 12.6 $\sigma(T^*) = \{\bar{\lambda} : \ \lambda \in \sigma(T)\}$.

Proof If $\lambda \notin \sigma(T)$ then $T - \lambda I$ has a bounded inverse,

$$(T - \lambda I)(T - \lambda I)^{-1} = I = (T - \lambda I)^{-1}(T - \lambda I).$$

Taking adjoints we obtain

$$[(T - \lambda I)^{-1}]^*(T^* - \bar{\lambda} I) = I = (T^* - \bar{\lambda} I)[(T - \lambda I)^{-1}]^*,$$

and so $T^* - \bar{\lambda} I$ has a bounded inverse, i.e. $\bar{\lambda} \notin \sigma(T^*)$. Starting instead with T^* we deduce that $\lambda \notin \sigma(T^*) \Rightarrow \bar{\lambda} \notin \sigma(T)$, which completes the proof. □

Example 12.7 *Let σ_{r} be the right-shift operator on ℓ^2. We saw above that*

σ_r has no eigenvalues, but that for $\sigma_r^* = \sigma_l$ the interior of the unit disc is contained in the point spectrum. It follows that

$$\{\lambda \in \mathbb{C} : \ |\lambda| < 1\} \subseteq \sigma(\sigma_{\rm r})$$

even though $\sigma_{\rm p}(\sigma_{\rm r}) = \emptyset$.

We have already seen that any eigenvalue λ of T must satisfy $|\lambda| \leq \|T\|_{\rm op}$. We now show that this also holds for any $\lambda \in \sigma(T)$; the argument is more subtle, and based on considering how to solve the linear equation $(I-T)x = y$.

Theorem 12.8 *Suppose that $T \in B(H,H)$ with $\|T\| < 1$. Then $(I-T)^{-1} \in B(H,H)$ and*

$$(I-T)^{-1} = I + T + T^2 + \cdots$$

with

$$\|(I-T)\|^{-1} \leq (1 - \|T\|)^{-1}.$$

Proof Since

$$\|T^n x\| \leq \|T\| \|T^{n-1}x\|$$

it follows that $\|T^n\| \leq \|T\|^n$. Therefore if we consider

$$V_n = I + T + \cdots + T^n$$

we have (for $n > m$)

$$\begin{aligned} \|V_n - V_m\| &= \|T^{m+1} + \cdots + T^{n-1} + T^n\| \\ &\leq \|T^{m+1}\| + \cdots + \|T^{n-1}\| + \|T^n\| \\ &\leq \|T\|^{m+1} + \cdots + \|T\|^{n-1} + \|T\|^n \\ &\leq \|T\|^{m+1} \frac{1}{1 - \|T\|}. \end{aligned}$$

It follows that $\{V_n\}$ is Cauchy in the operator norm, and so converges to some $V \in B(H,H)$ with

$$\|V\| \leq 1 + \|T\| + \|T\|^2 + \cdots = [1 - \|T\|]^{-1}.$$

Clearly

$$V(I-T) = (I+T+T^2+\cdots)(I-T) = (I+T+T^2+\cdots) - (T+T^2+T^3) = I$$

and similarly $(I-T)V = I$. $\square$

As promised:

Corollary 12.9 *The spectrum of* T, $\sigma(T) \subseteq \{\lambda \in \mathbb{C} : |\lambda| \leq \|T\|_{\text{op}}\}$.

Proof We have $T - \lambda I = \lambda(\frac{1}{\lambda}T - I)$. So if $I - \frac{1}{\lambda}T$ is invertible, $\lambda \notin \sigma(T)$. But for $|\lambda| > \|T\|_{\text{op}}$ we have $\|\frac{1}{\lambda}T\|_{\text{op}} < 1$, and the above theorem shows that $T - \lambda I$ is invertible and the result follows. □

We now show that the spectrum must also be closed, by showing that it complement (the resolvent set) is open. To this end, we prove the following theorem, which shows that the set of bounded linear operators with bounded inverses defined on all of H is open, i.e. that this property is stable under perturbation.

Theorem 12.10 *Let* H *be a Hilbert space and* $T \in B(H,H)$ *such that* $T^{-1} \in B(H,H)$. *Then for any* $U \in B(H,H)$ *with*

$$\|U\| < \|T^{-1}\|^{-1}$$

the operator $T + U$ *is invertible with*

$$\|(T+U)^{-1}\| \leq \frac{\|T^{-1}\|}{1 - \|U\|\|T^{-1}\|}. \tag{12.1}$$

Proof Let $P = T^{-1}(T+U) = I + T^{-1}U$. Then since by assumption $\|T^{-1}\|\|U\| < 1$ it follows from Theorem 12.8 that P is invertible with

$$\|P\|^{-1} \leq \frac{1}{1 - \|T^{-1}\|\|U\|}.$$

Using the definition of P we have

$$T^{-1}(T+U)P^{-1} = P^{-1}T^{-1}(T+U) = I;$$

from the first of these identities we have

$$(T+U)P^{-1}T^{-1} = I$$

and so

$$(T+U)^{-1} = P^{-1}T^{-1}$$

and (12.1) follows. □

Corollary 12.11 *If* $T \in B(H,H)$ *then the spectrum of* T *is closed.*

Proof We show that the resolvent set $R(T)$, the complement of $\sigma(T)$, is open. Indeed, if $\lambda \in R(T)$ then $T - \lambda I$ is invertible. Theorem 12.10 show that $(T - \lambda I) - \delta I$ is invertible for all

$$|\delta| < \|(T - \lambda I)^{-1}\|^{-1},$$

i.e. $R(T)$ is open. □

Lemma 12.12 *The spectrum of σ_l and of σ_r are both equal to the unit disc in the complex plane:*

$$\sigma(\sigma_\cdot) = \{\lambda \in \mathbb{C} : \ |\lambda| \le 1\}.$$

Proof We showed earlier that for the shift operators σ_{r} and σ_l on ℓ^2,

$$\sigma(\sigma_{\mathrm{r}}) = \sigma(\sigma_{\mathrm{l}}) \supseteq \{\lambda \in \mathbb{C} : \ |\lambda| < 1\}.$$

We have already shown that $\|\sigma_\cdot\|_{\mathrm{op}} = 1$, so we know that at most

$$\sigma(\sigma_\cdot) \supseteq \{\lambda \in \mathbb{C} : \ |\lambda| \le 1\},$$

but since it follows from Corollary 12.11 that

$$\sigma(\sigma_\cdot) \subseteq \{\lambda \in \mathbb{C} : \ |\lambda| \le 1\}.$$

It follows that in fact

$$\sigma(\sigma_\cdot) = \{\lambda \in \mathbb{C} : \ |\lambda| \le 1\}.$$

□

Note that the final assignment shows that if T is self-adjoint then $\sigma(T) \subseteq \mathbb{R}$.

13

Spectral theory II: compact self-adjoint operators

We now consider eigenvalues of compact self-adjoint linear operators on a Hilbert space. It is convenient to restrict attention to Hilbert spaces over $\mathbb{C}$, but this is no restriction, since we can always consider the 'complexification' of a Hilbert space over $\mathbb{R}$.

13.1 Complexification and real eigenvalues

Exercise 13.1 *Let H be a Hilbert space over $\mathbb{R}$, and define its complexification $H_{\mathbb{C}}$ as the vector space*

$$H_{\mathbb{C}} = \{x + \mathrm{i}y : \; x, y \in H\},$$

equipped with operations $+$ and $*$ defined via

$$(x + \mathrm{i}y) + (w + \mathrm{i}z) = (x + w) + \mathrm{i}(y + z), \qquad x, y, w, z \in V$$

and

$$(a + \mathrm{i}b) * (x + \mathrm{i}y) = (ax - by) + \mathrm{i}(bx + ay) \qquad a, b \in \mathbb{R}, \; x, y \in V.$$

Show that equipped with the inner product

$$(x + \mathrm{i}y, w + \mathrm{i}z)_{H_{\mathbb{C}}} = (x, w) + \mathrm{i}(y, w) - \mathrm{i}(x, z) + (y, z)$$

$H_{\mathbb{C}}$ is a Hilbert space.

Just as we can complexify a Hilbert space H to give $H_{\mathbb{C}}$, we can complexify a linear operator T that acts on H to a linear operator $T_{\mathbb{C}}$ that acts on $H_{\mathbb{C}}$:

Exercise 13.2 *Let H be a real Hilbert space and $H_{\mathbb{C}}$ its complexification. Given $T \in B(H,H)$, extend T to a linear operator $\tilde{T} : H_{\mathbb{C}} \to H_{\mathbb{C}}$ via the definition*

$$\tilde{T}(x + \mathrm{i}y) = Tx + \mathrm{i}Ty \qquad x, y \in H.$$

Show $\tilde{T} \in B(H_{\mathbb{C}}, H_{\mathbb{C}})$, that any eigenvalue of T is an eigenvalue of $\tilde{T}$, and that any real eigenvalue of $\tilde{T}$ is an eigenvalue of T.

Lemma 13.3 *Let H be a real Hilbert space and $T \in B(H,H)$ a self-adjoint operator. Then $\tilde{T}$ as defined above is a self-adjoint operator on $H_{\mathbb{C}}$.*

Proof For $\xi, \eta \in H_{\mathbb{C}}$, $\xi = x + \mathrm{i}y$, $\eta = u + \mathrm{i}v$, $x, y, u, v \in H$,

$$\begin{aligned}
(\tilde{T}\xi, \eta) &= (T(x + \mathrm{i}y), u + \mathrm{i}v)_{H_{\mathbb{C}}} \\
&= (Tx + \mathrm{i}Ty, u + \mathrm{i}v) \\
&= (Tx, u) - \mathrm{i}(Tx, v) + \mathrm{i}(Ty, u) + (Ty, v) \\
&= (x, Tu) - \mathrm{i}(x, Tv) + \mathrm{i}(y, Tu) + (y, Tv) \\
&= (x + \mathrm{i}y, Tu + \mathrm{i}Tv) \\
&= (\xi, \tilde{T}\eta).
\end{aligned}$$

□

If T is self-adjoint then all its eigenvalues are real:

Theorem 13.4 *Let T be a self-adjoint operator on a Hilbert space H. Then all the eigenvalues of T are real and the eigenvectors corresponding to distinct eigenvalues are orthogonal.*

Proof Suppose that $Tx = \lambda x$ with $x \neq 0$. Then

$$\lambda \|x\|^2 = (\lambda x, x) = (Tx, x) = (x, T^* x) = (x, Tx) = (x, \lambda x) = \bar{\lambda} \|x\|^2,$$

i.e. $\lambda = \bar{\lambda}$.

Now if λ and μ are distinct eigenvalues with $Tx = \lambda x$ and $Ty = \mu y$ then

$$0 = (Tx, y) - (x, Ty) = (\lambda x, y) - (x, \mu y) = (\lambda - \mu)(x, y),$$

and so $(x, y) = 0$. □

Corollary 13.5 *If T is a self-adjoint operator on a real Hilbert space H, and $\tilde{T}$ is its complexification defined above, $\sigma_p(T) = \sigma_p(\tilde{T})$.*

13.2 Compact operators

We will develop our spectral theory for operators that are self-adjoint and 'compact' according to the following definition:

Definition 13.6 *Let X and Y be normed spaces. Then a linear operator $T : X \to Y$ is* compact *if for any bounded sequence $\{x_n\} \in X$, the sequence $\{Tx_n\} \in Y$ has a convergent subsequence (whose limit lies in Y).*

Note that a compact operator must be bounded, since otherwise there exists a sequence in X with $\|x_n\| = 1$ but $\|Tx_n\| \to \infty$, and clearly $\{Tx_n\}$ cannot have a convergent subsequence.

Example 13.7 *Take $T \in B(X, Y)$ with finite-dimensional range. Then T is compact, since any bounded sequence in a finite-dimensional space has a convergent subsequence.*

Theorem 13.8 *Suppose that X is a normed space and Y is a Banach space. If $\{K_n\}$ is a sequence of compact (linear) operators in $B(X, Y)$ converging to some $K \in B(X, Y)$ in the operator norm, i.e.*

$$\sup_{\|x\|_X=1} \|K_n x - Kx\|_Y \to 0$$

as $n \to \infty$, then K is compact.

Proof Let $\{x_n\}$ be a bounded sequence in X. Then since K_1 is compact, $K_1(x_n)$ has a convergent subsequence, $K_1(x_{n_{1j}})$. Since $x_{n_{1j}}$ is bounded, $K_2(x_{n_{1j}})$ has a convergent subsequence, $K_2(x_{n_{2j}})$. Repeat this process to get a family of nested subsequences, $x_{n_{kj}}$, with $K_l(x_{n_{kj}})$ convergent for all $l \le k$.

Now consider the diagonal sequence $y_j = x_{n_{jj}}$. Since $\{y_j\}$ is a subsequence of $\{x_i^n\}$ for $j \ge n$, it follows that $K_n(y_j)$ converges (as $j \to \infty$) for every n.

We now show that $K(y_j)$ is Cauchy, and hence convergent, to complete the proof. Choose $\epsilon > 0$, and use the triangle inequality to write

$$\begin{aligned} &\|K(y_i) - K(y_j)\|_Y \\ &\le \|K(y_i) - K_n(y_i)\|_Y + \|K_n(y_i) - K_n(y_j)\|_Y + \|K_n(y_j) - K(y_j)\|_Y. \end{aligned}$$

Since $\{y_j\}$ is bounded and $K_n \to K$ in the operator norm, pick n large enough that

$$\|K(y_j) - K_n(y_j)\|_Y \le \epsilon/3$$

for all y_j in the sequence. For such an n, the sequence $K_n(y_j)$ is Cauchy, and so there exists an N such that for $i, j \ge N$ we can guarantee

$$\|K_n(y_i) - K_n(y_j)\|_Y \le \epsilon/3.$$

So now

$$\|K(y_i) - K(y_j)\|_Y \le \epsilon \qquad \text{for all} \qquad i, j \ge N,$$

and $\{K(y_n)\}$ is a Cauchy sequence. □

We now use this theorem to show the following:

Proposition 13.9 *The integral operator* $T : L^2(a,b) \to L^2(a,b)$ *given by*

$$[Tu](x) = \int_a^b K(x,y)u(y)\,\mathrm{d}y$$

with

$$\int_a^b \int_a^b |K(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y < \infty$$

is compact.

Proof Let $\{\phi_j\}$ be an orthonormal basis for $L^2(a,b)$. It follows that $\{\phi_i(x)\phi_j(y)\}$ is an orthonormal basis for $L^2((a,b)\times(a,b))$. If we write $K(x,y)$ in terms of this basis we have

$$K(x,y) = \sum_{j,k=1}^{\infty} k_{ij}\phi_i(x)\phi_j(y),$$

where the coefficients k_{ij} are given by

$$k_{ij} = \int_a^b \int_a^b K(x,y)\phi_i(x)\phi_j(y)\,\mathrm{d}x\,\mathrm{d}y,$$

and the sum converges in $L^2((a,b)\times(a,b))$. Since $\{\phi_i(x)\phi_j(y)\}$ is a basis we have

$$\|K\|^2_{L^2((a,b)\times(a,b))} = \int_a^b \int_a^b |K(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y = \sum_{i,j=1}^{\infty} |k_{ij}|^2. \tag{13.1}$$

We now approximate T by operators derived from finite truncations of the expansion of $K(x,y)$. We set

$$K_n(x,y) = \sum_{i,j=1}^{n} k_{ij}\phi_i(x)\phi_j(y),$$

and

$$[T_n u](x) = \int_a^b K_n(x,y)u(y)\,\mathrm{d}y.$$

If $u \in L^2(\Omega)$ is given by $u = \sum_{l=1}^{\infty} c_l\phi_l$, then

$$\begin{aligned}(T_n u)(x) &= \sum_{i,j=1}^{n}\sum_{l=1}^{\infty}\int_a^b k_{ij}\phi_i(x)\phi_j(y)c_l\phi_l(y)\,\mathrm{d}y \\ &= \sum_{i,j=1}^{n} k_{ij}c_j\phi_i(x).\end{aligned}$$

Since $T_n u$ is therefore a linear combination of $\{\phi_i\}_{i=1}^n$, the range of T_n has dimension n. It follows that $_n$ is compact for each n; if we can show that $T_n \to T$ in the operator norm then we can use theorem 13.8 to show that T is compact.

This is straightforward, since

$$\begin{aligned}\|(T-T_n)u\|^2 &= \int_a^b\int_a^b |K(x,y)u(y) - K_n(x,y)u(y)|^2\,\mathrm{d}x\,\mathrm{d}y \\ &\le \left(\int_a^b\int_a^b |K(x,y) - K_n(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y\right)\int_a^b |u(y)|^2\ \ dy,\end{aligned}$$

i.e.

$$\begin{aligned}\|T-T_n\|^2 &\le \int_a^b\int_a^b |K(x,y) - K_n(x,y)|^2\,\mathrm{d}x\,\mathrm{d}y \\ &\le \int_a^b\int_a^b \left|\sum_{i,j=n+1}^{\infty} k_{ij}\phi_i(x)\phi_j(y)\right|^2\,\mathrm{d}x\,\mathrm{d}y \\ &= \sum_{i,j=n+1}^{\infty}|k_{ij}|^2,\end{aligned}$$

using the expansion of K and K_n. Convergence of T_n to T follows since the sum in (13.1) is finite. $\square$

We now show that any compact self-adjoint operator has at least one eigenvalue. (Recall that σ_{r} is not even normal, so this is no contradiction.)

Theorem 13.10 *Let H be a Hilbert space and $T \in B(H, H)$ a compact self-adjoint operator. Then at least one of $\pm\|T\|_{\mathrm{op}}$ is an eigenvalue of T.*

Proof We assume that $T \neq 0$, otherwise the result is trivial. From Theorem 11.18,

$$\|T\|_{\mathrm{op}} = \sup_{\|x\|=1} |(Tx, x)|.$$

Thus there exists a sequence x_n, of unit vectors, such that

$$(Tx_n, x_n) \to \pm\|T\|_{\mathrm{op}} = \alpha.$$

Since T is compact there is a subsequence x_{n_j} such that Tx_{n_j} is convergent to some y. Relabel x_{n_j} as x_n again.

Now consider

$$\begin{aligned} \|Tx_n - \alpha x_n\|^2 &= \|Tx_n\|^2 + \alpha^2 - 2\alpha(Tx_n, x_n) \\ &\leq 2\alpha^2 - 2\alpha(Tx_n, x_n); \end{aligned}$$

by the choice of x_n, the right-hand side tends to zero as $n \to \infty$. It follows, since $Tx_n \to y$, that

$$\alpha x_n \to y,$$

and since $\alpha \neq 0$ is fixed we must have $x_n \to x$ for some $x \in H$. Therefore $Tx_n \to Tx = \alpha x$. It follows that

$$Tx = \alpha x$$

and clearly $x \neq 0$, since $\|y\| = |\alpha|\|x\| = \|T\|_{\mathrm{op}} \neq 0$. □

Note that since any eigenvalue must satisfy $|\lambda| \leq \|T\|_{\mathrm{op}}$, since if $Tx = \lambda x$ we have

$$\lambda\|x\|^2 = (\lambda x, x) = (Tx, x) \leq \|Tx\|\|x\| \leq \|T\|_{\mathrm{op}}\|x\|^2,$$

it follows that $\|T\|_{\mathrm{op}} = \sup\{\lambda : \ \lambda \in \sigma_p(T)\}$.

Proposition 13.11 *Let T be a compact self-adjoint operator on a Hilbert space H. Then $\sigma_{\mathrm{p}}(T)$ is either finite or consists of a countable sequence tending to zero. Furthermore every distinct non-zero eigenvalue corresponds to a finite number of linearly independent eigenvectors.*

Proof Suppose that T has infinitely many eigenvalues that do not form a sequence tending to zero. Then for some $\epsilon > 0$ there exists a sequence of distinct eigenvalues with $|\lambda_n| > \epsilon$. Let x_n be a corresponding sequence of eigenvectors with $\|x_n\| = 1$; then

$$\|Tx_n - Tx_m\|^2 = (Tx_n - Tx_m, Tx_n - Tx_m) = |\lambda_n|^2 + |\lambda_m|^2 \geq 2\epsilon^2$$

since $(x_n, x_m) = 0$. It follows that $\{Tx_n\}$ can have no convergent subsequence, which contradicts the compactness of T.

Now suppose that for some eigenvalue λ there exists an infinite number of linearly independent eigenvectors $\{e_n\}_{n=1}^{\infty}$. Using the Gram-Schmidt process we can find a countably infinite orthonormal set of eigenvectors, since any linear combination of the $\{e_j\}$ is still an eigenvector:

$$T(\sum_{j=1}^{n} \alpha_j e_j) = \sum_{j=1}^{n} \alpha_j T e_j = \lambda(\sum_{j=1}^{n} \alpha_j e_j).$$

Now, we have

$$\|Te_n - Te_m\| = \|\lambda e_n - \lambda e_m\| = |\lambda| \|e_n - e_m\| = \sqrt{2}|\lambda|.$$

It follows that $\{Te_n\}$ can have no convergent subsequence, again contradicting the compactness of T. (Note that this second part does not in fact use the fact that T is self-adjoint.) $\square$

Lemma 13.12 *Let $T \in B(H, H)$ and let S be a closed linear subspace of H such that $TS \subseteq S$. Then $T^* S^\perp \subseteq S^\perp$.*

Proof Let $x \in S^\perp$ and $y \in S$. Then $Ty \in S$ and so $(Ty, x) = (y, T^*x) = 0$ for all $y \in S$, i.e. $T^*x \in S^\perp$. $\square$

Since we will apply this lemma when T is self-adjoint, for such a T we have

$$TS \subseteq S \qquad \Rightarrow \qquad TS^\perp \subseteq S^\perp$$

for any closed linear subspace S of H.

Theorem 13.13 *(Hilbert-Schmidt Theorem). Let H be a Hilbert space and $T \in B(H, H)$ be a compact self-adjoint operator. Then there exists a finite* *or countably infinite orthonormal sequence $\{w_n\}$ of eigenvectors of T with corresponding non-zero real eigenvalues $\{\lambda_n\}$ such that for all $x \in H$*

$$Tx = \sum_j \lambda_j (x, w_j) w_j. \tag{13.2}$$

Proof By Theorem 13.10 there exists a w_1 such that $\|w_1\| = 1$ and $Tw_1 = \pm\|T\|w_1$. Consider the subspace of H perpendicular to w_1,

$$H_2 = w_1^{\perp}.$$

Then since T is self-adjoint, Lemma 13.12 shows that T leaves H_2 invariant. If we consider $T_2 = T|_{H_2}$ then we have $T_2 \in B(H_2, H_2)$ with T_2 compact; this operator is still self-adjoint, since for all $x, y \in H_2$

$$(x, T_2 y) = (x, Ty) = (T^* x, y) = (Tx, y) = (T_2 x, y).$$

Now apply Theorem 13.10 to the operator T_2 on the Hilbert space H_2 find an eigenvalue $\lambda_2 = \pm\|T_2\|$ and an eigenvector $w_2 \in H_2$ with $\|w_2\| = 1$. Continue this process as long as $T_n \neq 0$.

If $T_n = 0$ for some n then for any $x \in H$ we have

$$y := x - \sum_{j=1}^{n-1} (x, w_j) w_j \in H_n.$$

Then

$$0 = T_n y = Ty = Tx - \sum_{j=1}^{n-1} (x, w_j) T w_j = Tx - \sum_{j=1}^{n-1} \lambda_j (x, w_j) w_j$$

which is (13.2).

If T_n is never zero then consider

$$y_n := x - \sum_{j=1}^{n-1} (x, w_j) w_j \in H_n.$$

Then we have

$$\|x\|^2 = \|y_n\|^2 + \sum_{j=1}^{n-1} |(x, w_j)|^2,$$

and so $\|y_n\| \leq \|x\|$. It follows that

$$\left\| Tx - \sum_{j=1}^{n-1} \lambda_j (x, w_j) w_j \right\| = \|Ty_n\| \leq \|T_n\| \|y_n\| = |\lambda_n| \|x\|,$$

and since $|\lambda_n| \to 0$ as $n \to \infty$ we have (13.2). $\square$

There is a partial converse to this theorem on the last examples sheet: if H is a Hilbert space, $\{e_j\}$ is an orthonormal set in H, and $T \in B(H,H)$ is such that

$$Tu = \sum_{j=1}^{\infty} \lambda_j (u, e_j) e_j \qquad \text{with} \qquad \lambda_j \in \mathbb{R} \quad \text{and} \quad \lambda_j \to 0 \quad \text{as} \quad j \to \infty$$

then T is compact and self-adjoint.

The orthonormal sequence constructed in this theorem is only a basis for the range of T; however, we do have:

Corollary 13.14 *Let H be an infinite-dimensional separable Hilbert space and $T \in B(H,H)$ a compact self-adjoint operator. Then there exists an orthonormal basis of H consisting of eigenvectors $\{w_j\}$ of T, and for any $x \in H$*

$$Tx = \sum_{j=1}^{\infty} \lambda_j (x, w_j) w_j.$$

where $Tw_j = \lambda_j w_j$.

Proof From Theorem 13.13 we have a finite or countable sequence $\{w_n\}$ of eigenvectors of T such that

$$Tx = \sum_{j=1}^{\infty} \lambda_j (x, w_j) w_j. \tag{13.3}$$

Now let $\{e_j\}$ be an orthonormal basis for $\operatorname{Ker} T$ (this exists since H, and so $\operatorname{Ker}(T)$, is separable); each e_j is an eigenvector of T with eigenvalue zero, and since $Te_j = 0$ but $Tw_j = \lambda_j w_j$ with $\lambda_j \neq 0$, we know that $(w_j, e_k) = 0$ for all j, k. So $\{w_j\} \cup \{e_k\}$ is a countable orthonormal set in H.

Now, (13.3) implies that

$$T\left[x - \sum_{j=1}^{\infty} (x, w_j) w_j\right] = 0,$$

i.e. that $x - \sum_{j=1}^{\infty} (x, w_j) w_j \in \operatorname{Ker} T$, and therefore

$$x - \sum_{j=1}^{\infty} (x, w_j) w_j = \sum_{k=1}^{\infty} \alpha_k e_k,$$

since $\{e_k\}$ is a basis for $\operatorname{Ker} T$. It follows that $\{w_j\} \cup \{e_k\}$ is a countably infinite orthonormal basis for H. $\square$

Exercise 13.15 *Show that if T is invertible and satisfies the conditions of Theorem 13.13 then there is an orthonormal basis of H consisting of eigenvectors corresponding to non-zero eigenvalues of T.*

In general it is a hard problem to find the eigenvalues and eigenvectors of an arbitrary operator T; and even for the integral operator that we have considered repeatedly. However, there are a class of operators for which this is very straightforward:

Theorem 13.16 *Suppose that $\{e_j(x)\}$ is an orthonormal set in $L^2(a,b)$ (real-valued), and that*

$$K(x,y) = \sum_{j=1}^{\infty} \lambda_j e_j(x) e_j(y) \qquad \textit{with} \qquad \sum_{j=1}^{\infty} |\lambda_j|^2 < \infty.$$

Then the $\{e_j(x)\}$ are eigenvectors of

$$(Tu)(x) = \int_a^b K(x,y) u(y) \, \mathrm{d}y$$

with corresponding eigenvalues λ_j. There are no other eigenvectors corresponding to non-zero eigenvalues.

Proof We have

$$\begin{aligned} Tu &= \int_a^b K(x,y) u(y) \, \mathrm{d}y \\ &= \int_a^b \sum_{j=1}^{\infty} \lambda_j e_j(x) e_j(y) u(y) \, \mathrm{d}y \\ &= \sum_{j=1}^{\infty} \lambda_j e_j (e_j, u), \end{aligned}$$

and so $Te_k = \lambda_k e_k$.

To show that these are the only eigenvalues and eigenvectors, if $u \in L^2(a,b)$ with $u = w + \sum_{k=1}^{\infty} (u, e_k) e_k$ and $Tu = \lambda u$ then, since $w \perp e_j$

for all j,

$$\begin{aligned} Tu &= Tw + \sum_{k=1}^{\infty}(u,e_k)Te_k \\ &= \sum_{j,k=1}^{\infty}\int_a^b \lambda_j e_j(x)e_j(y)(u,e_k)e_k(y)\,\mathrm{d}y \\ &= \sum_{j=1}^{\infty}\lambda_j(u,e_j)e_j(x) \end{aligned}$$

and

$$\lambda u = \sum_{j=1}^{\infty}\lambda(u,e_j)e_j(x).$$

Taking the inner product with each e_k yields

$$\lambda(u,e_k) = \lambda_k(u,e_k),$$

so either $(u,e_k)=0$ or $\lambda=\lambda_k$. □

Example 13.17 *Let $(a,b)=(-\pi,\pi)$ and consider $K(t,s)=\cos(t-s)$. Then K is clearly symmetric, so the corresponding T is self-adjoint. We have*

$$\cos(t-s) = \cos t\cos s - \sin t\sin s;$$

recall (Example 7.5) that $\cos t$ and $\sin t$ are orthogonal in $L^2(-\pi,\pi)$. We have

$$\begin{aligned} T(\sin t) &= \int_{-\pi}^{\pi} K(t,s)\sin s\,\mathrm{d}s \\ &= \int_{-\pi}^{\pi} \cos t\cos s\sin s - \sin t\sin^2 s\,\mathrm{d}s \\ &= -2\pi\sin t \end{aligned}$$

and

$$\begin{aligned} T(\cos t) &= \int_{-\pi}^{\pi} K(t,s)\cos s\,\mathrm{d}s \\ &= \int_{-\pi}^{\pi} \cos t\cos^2 s - \sin t\sin s\cos s\,\mathrm{d}s \\ &= 2\pi\cos t. \end{aligned}$$

Example 13.18 *Let* $(a, b) = (-1, 1)$ *and let*

$$K(t, s) = 1 - 3(t - s)^2 + 9t^2 s^2.$$

Then in fact

$$K(t, s) = 4\left(\sqrt{\frac{3}{2}}t\right)\left(\sqrt{\frac{3}{2}}s\right) + \frac{8}{5}\left(\sqrt{\frac{5}{8}}(3t^2 - 1)\right)\left(\sqrt{\frac{5}{8}}(3s^2 - 1)\right),$$

and so, since $\{\sqrt{\frac{3}{2}}t, \sqrt{\frac{5}{8}}(3t^2 - 1)\}$ *are orthonormal (they are some of the Legendre polynomials from Chapter 6) the integral operator associated with* $K(t, s)$ *has*

$$T(t) = 4t \qquad \text{and} \qquad T(3t^2 - 1) = \frac{8}{5}(3t^2 - 1).$$

There are no other eigenvalues/vectors.

We end this chapter with a corollary of Corollary 13.14 (!) that shows that the eigenvalues are essentially all of the spectrum of a compact self-adjoint operator.

Theorem 13.19 *Let* T *be a compact self-adjoint operator on a separable Hilbert space. Then* $\sigma(T) = \overline{\sigma_p(T)}$.

Note that this means that either $\sigma(T) = \sigma_p(T)$ or $\sigma(T) = \sigma_p(T) \cup \{0\}$, since $\sigma_p(T)$ has no limit points except perhaps zero. So $\sigma(T) = \sigma_p(T)$ unless there are an infinite number of eigenvalues but zero itself is not an eigenvalue.

Proof By the corollary of the HS Theorem, we have

$$Tx = \sum_j \lambda_j (x, e_j) e_j$$

for some orthonormal basis $\{e_j\}$ of H.

Now take $\mu \notin \overline{\sigma_p(T)}$. For such μ, it follows that there exists a $\delta > 0$ such that

$$\sup_j |\mu - \lambda_j| \geq \delta > 0$$

(otherwise $\mu \in \overline{\sigma_p(T)}$). We use this to show that $T - \mu I$ is invertible with bounded inverse.

Now,

$$(T - \mu I)x = y \qquad \Leftrightarrow \qquad \sum_j (\lambda_j - \mu)(x, e_j)e_j = \sum_j (y, e_j)e_j.$$

Taking the inner product of both sides with e_k, we have

$$(T - \mu I)x = y \qquad \Leftrightarrow \qquad (\lambda_k - \mu)(x, e_k) = (y, e_k), \ \ k = 1, 2, \ldots.$$

So we must have

$$(x, e_k) = \frac{(y, e_k)}{\lambda_k - \mu}.$$

Since $\sum |(y, e_k)|^2 < \infty$ and $|\lambda_k - \mu| \geq \delta$, it follows that

$$x = \sum_k (x, e_k)e_k$$

converges, and that $\|x\| \leq \delta^{-1}\|y\|$. So $(T - \mu I)^{-1}$ exists and is bounded. $\square$

14

Sturm-Liouville problems

We consider the Sturm-Liouville problem

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}u}{\mathrm{d}x}\right)+q(x)u=\lambda u \qquad \text{with} \qquad u(a)=u(b)=0. \tag{14.1}$$

As a shorthand, we write $L[u]$ for the left-hand side of (14.1), i.e.

$$L[u]=-(p(x)u')'+q(x)u$$

We will assume that $p(x)>0$ on $[a,b]$ and that $q(x)\geq 0$ on $[a,b]$.

It was one of the major concerns of applied mathematics throughout the nineteenth century to show that the solutions $\{u_n(x)\}$ of (14.1) form a complete basis for some appropriate space of functions (generalising the use of Fourier series as a basis for L^2). We can do this easily using the theory developed in the last section.

However, first we have to turn the differential equation (14.1) into an integral equation. We do this as follows:

Lemma 14.1 *Let $u_1(x)$ and $u_2(x)$ be two linearly independent non-zero solutions of*

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}u}{\mathrm{d}x}\right)+q(x)u=0.$$

Then

$$W_p(u_1,u_2)(x):=p(x)[u_1'(x)u_2(x)-u_2'(x)u_1(x)]$$

is a constant.

Proof First we show that W_p is constant. Differentiate W_p with respect to x, then use the fact that $L[u_1] = L[u_2] = 0$ to substitute for $pu'' = qu - p'u'$ to give:

$$\begin{aligned} W_p' &= p'u_1'u_2 + pu_1''u_2' + pu_1'u_2' - p'u_1u_2' - pu_1'u_2' - pu_1u_2'' \\ &= p'(u_1'u_2 - u_2'u_1) + p(u_1''u_2 - u_2''u_1) \\ &= p'(u_1'u_2 - u_2'u_1) + u_2(qu_1 - p'u_1') - u_1(qu_2 - p'u_2') \\ &= 0. \end{aligned}$$

Now, if $W_p \equiv 0$ then, since $p \neq 0$, we have

$$u_1'u_2 - u_2'u_1 = 0 \qquad \Rightarrow \qquad \frac{u_1'}{u_1} = \frac{u_2'}{u_2},$$

which can be integrated to give $\ln u_1 = \ln u_2 + c$, i.e. $u_1 = \mathrm{e}^c u_2$, which implies that u_1 and u_2 are proportional, contradicting the fact that they are linearly independent. □

Theorem 14.2 *Suppose that $u_1(x)$ and $u_2(x)$ are two linearly independent non-zero solutions of*

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}y}{\mathrm{d}x}\right) + q(x)y = 0,$$

with $u_1(a) = 0$ and $u_2(b) = 0$. Set $C = W_p(u_1, u_2)^{-1}$ and define

$$G(x,y) = \begin{cases} Cu_1(x)u_2(y) & a \le x < y \\ Cu_2(x)u_1(y) & y \le x \le b; \end{cases} \tag{14.2}$$

then the solution of $L[u] = f$ is given by

$$u(x) = \int_a^b G(x,y)f(y)\,\mathrm{d}y. \tag{14.3}$$

Proof Writing (14.3) out in full we have

$$u(x) = Cu_2(x)\int_a^x u_1(y)f(y)\,\mathrm{d}y + Cu_1(x)\int_x^b u_2(y)f(y)\,\mathrm{d}y.$$

Now,

$$\begin{aligned} u'(x) &= Cu_2(x)u_1(x)f(x) + Cu_2'(x)\int_a^x u_1(y)f(y)\,\mathrm{d}y - Cu_1(x)u_2(x)f(x) \\ &\quad + Cu_1'(x)\int_x^b u_2(y)f(y)\,\mathrm{d}y \\ &= Cu_2'(x)\int_a^x u_1(y)f(y)\,\mathrm{d}y + Cu_1'(x)\int_x^b u_2(y)f(y)\,\mathrm{d}y, \end{aligned}$$

and then since $CW_p(u_1, u_2) = 1$,

$$\begin{aligned} u''(x) &= Cu_2'(x)u_1(x)f(x) + Cu_2''(x)\int_a^x u_1(y)f(y)\,\mathrm{d}y - Cu_1'(x)u_2(x)f(x) \\ &\quad + Cu_1''(x)\int_x^b u_2(y)f(y)\,\mathrm{d}y \\ &= -\frac{f(x)}{p(x)} + Cu_2''(x)\int_a^x u_1(y)f(y)\,\mathrm{d}y + Cu_1''(x)\int_x^b u_2(y)f(y)\,\mathrm{d}y. \end{aligned}$$

Now we have $L[u] = -pu'' - p'u' + qu$, and since L is linear with $L[u_1] = L[u_2] = 0$ it follows that

$$L[u] = f(x)$$

as claimed. $\square$

We can now define a linear operator on $L^2(a, b)$ by the right-hand side of (14.3):

$$Tf(x) = \int_a^b G(x, y)f(y)\,\mathrm{d}y.$$

Since G is symmetric (see (14.2)), it follows that T is a self-adjoint (see Example 11.17), and we have proved that such a T is compact in Proposition 13.9.

If $L[u] = \lambda u$, then $u = T(\lambda u)$. Since T is linear,

$$L[u] = \lambda u \qquad \Leftrightarrow \qquad u = \lambda T u.$$

If we can show that $\lambda, 1/\lambda \neq 0$ then the eigenvectors (or 'eigenfunctions') of the ODE boundary value problem $L[u] = \lambda u$ (which is just (14.1)) will be exactly those of the operator T (for which $Tu = \frac{1}{\lambda}u$). Since the eigenvectors of T form an orthonormal basis for $L^2(a, b)$ (we will see that $\mathrm{Ker}(T) = \{0\}$), the same is true of the eigenfunctions of the Sturm-Liouville problem.

Theorem 14.3 *The eigenfunctions of the problem (14.1) form a complete orthonormal basis for $L^2(a,b)$.*

Proof We show first that $\lambda = 0$ is not an eigenvalue of (14.1), i.e. there is no non-zero u for which $L[u] = 0$. Indeed, if $L[u] =$ then we have

$$\begin{aligned}
0 = (L[u], u) &= \int_a^b -(pu')'u + q|u|^2 \, \mathrm{d}x \\
&= \int_a^b p|u'|^2 + q|u|^2 \, \mathrm{d}x - \left[_a^b p(x)u'(x)u(x)\right] \\
&= \int_a^b p|u'|^2 + q|u|^2 \, \mathrm{d}x.
\end{aligned}$$

Since $p > 0$ on $[a,b]$, it follows that $u' = 0$ on $[a,b]$, and so u must be constant on $[a,b]$. Since $u(a) = 0$, it follows that $u \equiv 0$.

We now use show that $\mathrm{Ker}(T) = 0$. Indeed, Tf is the solution of $L[u] = f$, i.e. $f = L[Tf]$. So if $Tf = 0$, it follows that $f = 0$.

So ϕ is an eigenfunction of the SL problem iff it is an eigenvector for T:

$$L[\phi] = \lambda\phi \qquad \Leftrightarrow \qquad T\phi = \frac{1}{\lambda}\phi.$$

Since $G(x,y)$ is symmetric and bounded, it follows from Examples 10.9 and 11.8 that T is a bounded self-adjoint operator; Proposition 13.9 shows that T is also compact. It follows from Theorem 13.13 that T has a set of orthonormal eigenfunctions $\{\phi_j\}$ with

$$T\phi_j = \mu_j \phi_j,$$

and since $\mathrm{Ker}(T) = \{0\}$ the argument of Corollary **??** shows that those form an orthonormal basis for $L^2(a,b)$.

Comparing this with our original problem we obtain an infinite set of eigenfunctions $\{\phi_j\}$ with corresponding eigenvalues $\lambda_j = \mu_j^{-1}$. Note that now $\lambda_j \to \infty$ as $j \to \infty$. As above, the eigenfunctions $\{\phi_j\}$ form an orthonormal basis of $L^2(a,b)$. □

This has immediate applications for Fourier series. Indeed, if we consider

$$-\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} = \lambda u \qquad u(0) = 0, \quad u(1) = 0,$$

which is (14.1) with $p = 1$, $q = 0$, it follows that the eigenfunctions of this equation will form a basis for $L^2(0,1)$. These are easily found by elementary methods, and are

$$\{\sin k\pi x\}_{k=1}^{\infty}.$$

It follows that, appropriately normalised, these functions form an orthonormal basis for $L^2(a,b)$, i.e. that any $f \in L^2(a,b)$ can be expanded in the form

$$f(x) = \sum_{k=1}^{\infty} \alpha_k \sin k\pi x.$$

Thus begins the theory of Fourier series...

Exercise 14.4 *Show that the solution of* $-\mathrm{d}^2u/\mathrm{d}x^2 = f$ *is given by*

$$u(x) = \int_0^1 G(x,y)f(y)\,\mathrm{d}y,$$

where

$$G(x,y) = \begin{cases} x(1-y) & 0 \le x < y \\ y(1-x) & y \le x \le 1. \end{cases}$$