Functional Analysis - MT4515

Course Notes

Contents

Introductory Remarks about Functional Analysis

• In functional analysis, functions are considered as points or vectors in some big space. We study analysis (convergence, continuity, etc) in terms of these points rather like points or vectors in $\mathbb{R}$ or $\mathbb{R}^n$.

• Such function spaces are (generally infinite dimensional) normed spaces, in particular Banach spaces and Hilbert spaces, and we study the structure of such spaces in their own right.

• Transformations such as $f \mapsto \frac{\mathrm{d}f}{\mathrm{d}x}$ and $f \mapsto \int_0^t f(u)\,\mathrm{d}u$ are $l$inear mappings on such spaces, so we develop linear algebra in this context. It is more involved than in the finite dimensional setting since questions of convergence arise. Notions like eigenvalues and diagonalisation of matrices generalise (almost beyond recognition) to $s$pectral theory.

• The flavour of the work is of analysis and calculus, but ideas from linear algebra are important.

• Functional analysis has many applications across mathematics, e.g. to differential equations, fluid dynamics, quantum theory, statistics. The language of functional analysis is standard in many areas.

NOTATION Note that when we write $(x_n) \in X$, we mean a sequence of points x_n in X. The $:=$ sign is means that a quantity is being defined by the equation. The $\lightning$ symbol indicates a contradiction. We use the usual 'iff' to stand for 'if and only if'.

Some books

Rynne, B.P. & Youngson, M.A., *Linear Functional Analysis*, Springer, 2nd Ed 2007.

Young, N., *An Introduction to Hilbert Space*, Cambridge University Press, 1988.

Griffel, D.H., *Applied Functional Analysis*, Dover, 2nd Ed 2002.

Bollobás, B, *Linear Analysis: An Introductory Course*, Cambridge University Press, 1999.

Grateful thanks are due to Scott Thomson who typeset a large part of these notes and also to Donald Robertson for his help with them.

CHAPTER 1

Normed Spaces

Useful Inequalities

We establish some useful inequalities which we put in boxes for easy reference. The first inequality is the *triangle inequality* on $\mathbb{R}$:

$$\boxed{\forall x, y \in \mathbb{R} \qquad |x+y| \leqslant |x| + |y|. \qquad \text{(Triangle Inequality)}}$$

Cauchy's Inequality or the *Cauchy-Schwarz Inequality* is frequently used:

$$\boxed{\sum_{i=1}^{N} |a_i b_i| \leqslant \left(\sum_{i=1}^{N} |a_i|^2\right)^{1/2} \left(\sum_{i=1}^{N} |b_i|^2\right)^{1/2}. \qquad \text{(Cauchy's Inequality)}}$$

Proof Note that $0 \leqslant \sum_i (ta_i + b_i)^2$ for real t. Multiplying out

$$t^2 \sum_i a_i^2 + 2t \sum_i a_i b_i + \sum_i b_i^2 \geqslant 0.$$

This can be viewed as a quadratic in t. The discriminant condition requires that '$B^2 - 4AC \leqslant 0$' and by substituting the appropriate coefficients into this condition we get Cauchy's Inequality. □

As a corollary to this, we have the *Distance inequality*:

$$\boxed{\left(\sum_{i=1}^{N} (a_i + b_i)^2\right)^{1/2} \leqslant \left(\sum_{i=1}^{N} a_i^2\right)^{1/2} + \left(\sum_{i=1}^{N} b_i^2\right)^{1/2}. \qquad \text{(Distance Inequality)}}$$

Proof

$$\begin{aligned}
\sum_i (a_i + b_i)^2 &= \sum_i a_i^2 + 2 \sum_i a_i b_i + \sum_i b_i^2 \\
\text{(by Cauchy's Inequality)} \qquad &\leqslant \left[\left(\sum_i a_i^2\right)^{1/2}\right]^2 + 2\left(\sum_i a_i^2\right)^{1/2} \left(\sum_i b_i^2\right)^{1/2} + \left[\left(\sum_i b_i^2\right)^{1/2}\right]^2 \\
&= \left[\left(\sum_i a_i^2\right)^{1/2} + \left(\sum_i b_i^2\right)^{1/2}\right]^2.
\end{aligned}$$

□

Hölder's Inequality is a generalisation of Cauchy's inequality. Suppose $p, q > 1$ are conjugate indices, that is $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $a_i, b_i \in \mathbb{R}$,

$$\boxed{\sum_{i=1}^{N} |a_i b_i| \leqslant \left(\sum_{i=1}^{N} |a_i|^p\right)^{1/p} \left(\sum_{i=1}^{N} |b_i|^q\right)^{1/q}. \qquad \text{(Hölder's Inequality)}}$$

Cauchy's inequality is the $p = q = 2$ case of this.

Minkowski's inequality is a generalisation of the distance inequality. Let $p \geqslant 1$. Then

$$\boxed{\left(\sum_{i=1}^{N}|a_i+b_i|^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{N}|a_i|^p\right)^{1/p} + \left(\sum_{i=1}^{N}|b_i|^p\right)^{1/p}. \qquad \text{(Minkowski's Inequality)}}$$

Analogous inequalities, with similar derivations, hold for infinite sums and integrals. For example, Cauchy's inequality can appear in the forms

$$\boxed{\sum_{i=1}^{\infty}|a_ib_i| \leqslant \left(\sum_{i=1}^{\infty}|a_i|^2\right)^{1/2}\left(\sum_{i=1}^{\infty}|b_i|^2\right)^{1/2} \quad \text{and} \quad \int|fg| \leqslant \left(\int|f|^2\right)^{1/2}\left(\int|g|^2\right)^{1/2}.}$$

Moreover, such inequalities are also valid for complex numbers, with $\|\ \|$ taken to mean the modulus of a complex number.

The following are useful when working with specific functions:

$$|\cos x - \cos y| \leqslant |x-y|, \tag{1.1}$$

$$|\sin x - \sin y| \leqslant |x-y|, \tag{1.2}$$

which follow easily from the mean value theorem.

Also useful when working with integrals are:

$$\left|\int f(x)\,\mathrm{d}x\right| \leqslant \int\big|f(x)\big|\,\mathrm{d}x$$

and

$$\left|\int (f(x)+g(x))\,\mathrm{d}x\right| \leqslant \int\big|f(x)\big|\,\mathrm{d}x + \int\big|g(x)\big|\,\mathrm{d}x.$$

Normed Spaces

Throughout it is assumed that all vector spaces, or linear spaces as they are sometimes called, are over $\mathbb{R}$ or $\mathbb{C}$. Most results will hold for either $\mathbb{R}$ or $\mathbb{C}$ unless otherwise stated.

♦ DEFINITION 1.1 (NORM, NORMED SPACE). *Let X be a vector space. A function $\|\cdot\| : X \to \mathbb{R}$ is called a* norm *if the following hold:*

(1) $\|x\| \geqslant 0$ *for every* $x \in X$*, and* $\|x\| = 0$ *when, and only when,* $x = 0$ (Positivity)*;*
(2) $\|\lambda x\| = |\lambda|\,\|x\|$ *for every* $x \in X$ *and every* λ *in the field of scalars* (Scalar property)*;*
(3) $\|x+y\| \leqslant \|x\| + \|y\|$ *for evey* $x, y \in X$ (Triangle inequality).

We call the pair $(X, \|\cdot\|)$ *a* normed space.

We immediately make the following observations:

- By repeatedly applying the triangle inequality we get that

$$\left\|\sum_{i=1}^{N} x_i\right\| \leqslant \sum_{i=1}^{N}\|x_i\| \quad (x_i \in X).$$

- We have the *reverse triangle inequality* :

$$\big|\|x\| - \|y\|\big| \leqslant \|x-y\| \qquad \forall x, y \in X.$$

Proof of the reverse triangle inequality Let $x, y \in X$. Then

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

so

$$\|x\| - \|y\| \leqslant \|x-y\|.$$

Interchanging the roles of x and y completes the proof. □

It is extremely useful to think of $\|x\|$ as the length of the vector x and $\|x-y\|$ as the distance between the two points x and y. In this way a normed space can be though of as a metric space in a natural way.

♦ LEMMA 1.2. *Define* $d\colon X \times X \to \mathbb{R}$ *by*

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

then d *is* a metric *on* X *i.e., for every* $x, y, z \in X$ *it satisfies*

(1) $d(x,y) \geqslant 0 \quad \forall x, y \in X$, *and* $d(x,y) = 0$ *iff* $x = y$,
(2) $d(x,y) = d(y,x) \quad \forall x, y \in X$,
(3) *and* $d(x,y) \leqslant d(x,z) + d(z,y) \quad \forall x, y, z \in X$.

Proof This is easy to verify from the defining properties of the norm, with (3) following from the norm triangle inequality. □

Examples of Normed Spaces

Standard examples of normed spaces include the following:

(1) The space $(\mathbb{R}, |\ \cdot\ |)$.
(2) The space $\mathbb{R}^N$, where $\mathbb{R}^N = \{x = (x_1, x_2, \ldots, x_N) : x_i \in \mathbb{R}\}$ can be normed. In this case we have a number of possible norms:
   (a) $\|x\|_\infty = \max_{1 \leqslant i \leqslant N} |x_i|$ (*maximum norm*);
   (b) $\|x\|_1 = \sum_{i=1}^N |x_i|$ (1-*norm*);
   (c) $\|x\|_2 = \left(\sum_{i=1}^N |x_i|^2\right)^{1/2}$ (2-*norm* or *Euclidean norm*);
   (d) $\|x\|_p = \left(\sum_{i=1}^N |x_i|^p\right)^{1/p}$ (for $p \geqslant 1$ only) (p-*norm*).

   [By way of an example we check that (2c) is actually a norm: it is clearly non-negative and zero if and only if $x = 0$, the scalar property holds from the definition and the triangle inequality is just the distance inequality.]
(3) The set $C[0,1]$ of real-valued continuous functions on $[0,1]$ forms a linear space under the obvious operations

$$(f+g)(t) = f(t) + g(t) \qquad \text{and} \qquad (\lambda f)(t) = \lambda f(t).$$

   Possible norms on $C[0,1]$ include:
   (a) $\|f\|_\infty := \sup_{t \in [0,1]} |f(t)|$ is known as the *uniform norm* or *supremum norm*.
   To verify the triangle inequality:

$$\begin{aligned}\|f+g\|_\infty &= \sup_{t\in[0,1]} |f(t)+g(t)| \leqslant \sup_{t\in[0,1]} (|f(t)| + |g(t)|)\\ &\leqslant \sup_{t\in[0,1]} |f(t)| + \sup_{u\in[0,1]} |g(u)| = \|f\|_\infty + \|g\|_\infty.\end{aligned}$$

   (b) $\|f\|_1 := \int_0^1 |f| \,\mathrm{d}t$ is known as the 1-*norm*. Note that a function could have a large $\|\cdot\|_\infty$-norm but a small $\|\cdot\|_1$-norm.
   (c) $\|f\|_p := \left(\int_0^1 |f|^p \,\mathrm{d}t\right)^{1/p}$ is the p-*norm*.
(4) In order to establish the L^p spaces we first say that two functions f, g on $[0,1]$ are *e*quivalent if they are equal 'almost everywhere', that is

$$\int_X |f - g| \,\mathrm{d}x = 0.$$

   Then the space $L^p[0,1]$ consists of all equivalence classes of functions on $[0,1]$ such that $\int_0^1 |f|^p \,\mathrm{d}x$ is finite. We define the p-norm (for $p \geqslant 1$) on this space of functions by

$$\|f\|_p := \left(\int_0^1 |f|^p \,\mathrm{d}x\right)^{1/p}.$$

(It is necessary to regard such functions as equivalent to ensure that $\|f\|_p = 0$ implies f=0. The interested reader can find more detail on the Lebesgue L^p spaces in the literature on measure theory.)

(5) The space $C^1[0,1]$ of continuously differentiable real-valued functions on $[0,1]$ becomes a normed space taking

$$\|f\| = \sup_{t\in[0,1]} |f(t)| + \sup_{t\in[0,1]} |f'(t)|.$$

(6) The set of infinite sequences by $x \equiv (x_1, x_2, \ldots)$ is a vector space under the natural componentwise operations $(x_1, x_2, \ldots) + (y_1, y_2, \ldots) = (x_1 + y_1, x_2 + y_2 \ldots)$ and $\lambda(x_1, x_2, \ldots) = (\lambda x_1, \lambda x_2, \ldots)$. Certain subspaces are natural normed spaces,

(a) l^∞ is the space of bounded sequences with norm given by

$$\|x\|_\infty := \sup_{1\leqslant i<\infty} |x_i|.$$

(b) l^p $(p \geqslant 1)$ is the space of sequences such that $\sum_{i=1}^{\infty} |x_i|^p < \infty$ which is a normed space under the p-norm

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

Convergence and Continuity

Norms allows us to define notions such as convergence by requiring that the distance between sequence elements and the limit tends to 0 (The reader unfamiliar with the formal definitions of convergence and continuity may consult the extensive literature on real and abstract analysis.)

♦ Definition 1.3 (Convergence). *Let $(X, \|\cdot\|)$ be a normed space, let (x_n) be a sequence in X and let $x \in X$. We say that (x_n)* converges to a limit x *iff*

$$\lim_{n\to\infty} \|x_n - x\| = 0$$

and we write $x_n \to x$.

In other words, $x_n \to x$ iff for every $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $\|x_n - x\| < \varepsilon$ whenever $n \geqslant n_0$.

♦ Lemma 1.4. *Let $x_n \to x$ and $y_n \to y$ in $(X, \|\cdot\|)$. Let $\lambda \in \mathbb{R}$. Then*

(1) $(x_n + y_n) \to (x + y)$;

(2) $\lambda x_n \to \lambda x$.

Proof For (1), we have

$$\|(x_n + y_n) - (x + y)\| \leqslant \|x_n - x\| + \|y_n - y\| \to 0.$$

For (2), $\|\lambda x_n - \lambda x\| = |\lambda|\, \|x_n - x\| \to 0$. □

♦ Example 1.5 (Convergence in different norms). *Sometimes different norms will give different convergent sequences. For, if*

$$f_n(t) := \begin{cases} 1 - nt & t \in [0, \frac{1}{n}] \\ 0 & t \in [\frac{1}{n}, 1] \end{cases},$$

then $\|f_n - 0\|_1 = 1/2n \to 0$, but $\|f_n - 0\|_\infty = 1 \nrightarrow 0$.

For finite-dimensional spaces, all norms give the same convergent sequences, as will be seen in Chapter 3.

♦ Definition 1.6 (Continuity). *Let* $(X, \| \|_X)$ *and* $(Y, \| \|_Y)$ *be normed spaces and let* $f : X \to Y$*. We say that* f *is continuous at* $x \in X$ *if for all* (x_n) *with* $x_n \to x$ *in* $\| \|_X$ *we have* $f(x_n) \to f(x)$ *in* $\| \|_Y$*. The function* f *is said to be* continuous on X *if* f *is continuous at every* $x \in X$*.*

There are equivalent 'ε-δ' definitions of continuity but here this sequence characterisation is adequate for our purposes.

Open, Closed and Dense Sets

We need a few basic ideas from topology that generalise the notions of open and closed intervals on the real line to subsets of a normed space $(X, \| \cdot \|)$.

♦ Definition 1.7 (Open Set). *A set* $S \subseteq X$ *is called* open *if for every* $x \in X$*, there exists* $\delta > 0$ *such that*

$$\{y \in X : \|y - x\| < \delta\} \subseteq S.$$

Intuitively, an open set is one that does not contain its boundary points.

♦ Example 1.8 (Open Ball). *The* unit open ball

$$B_O := \{x \in X : \|x\| < 1\}$$

is an open set.

Proof Fix $x \in B_O$, so that $\|x\| < 1$. Let $\delta = 1 - \|x\|$. If y is such that $\|y - x\| < \delta$, then

$$\|y\| = \|(y - x) + x\| < \delta + (1 - \delta) = 1.$$

So $y \in B_O$. □

♦ Definition 1.9 (Limit Point, Closed Set). *Let* $S \subseteq X$*. Then* $x \in X$ *is called* a limit point of S *if there exists a sequence* $(x_n) \in S$ *such that* $x_n \to x$*. A set is* closed *if it contains all its limit points.*

♦ Example 1.10 (Closed Ball). *The set*

$$B := \{x \in X : \|x\| \leqslant 1\} \text{ is closed.}$$

Proof Suppose x is a limit point of B; that is, there exists $(x_n) \in B$ with $x_n \to x$. By the reverse triangle inequality, we get

$$0 \leqslant \big| \|x_n\| - \|x\| \big| \leqslant \|x_n - x\| \to 0.$$

Hence $\|x_n\| \to \|x\|$ in $\mathbb{R}$, so $\|x\| \leqslant 1$ and $x \in B$. □

Note that 'open' and 'closed' are not mutually exclusive properties. A set could be neither, or both. In particular the sets $\emptyset$ and X are both open and closed.

♦ Proposition 1.11. *Let* S *be a subset of the normed space* $(X, \| \cdot \|)$*. Then* S *is closed iff the complement* $S^c := X \setminus S$ *is open.*

Proof (Only if): Let S be closed. Suppose for a contradiction that $X \setminus S$ is not open. From the definition of openness, there exists $x \in X \setminus S$ such that for every $\delta > 0$

$$\{y : \|y - x\| < \delta\} \cap S \neq \emptyset.$$

Choose a sequence $(x_n) \in S$ such that

$$\|x_n - x\| < \frac{1}{n} \qquad \text{(taking } \delta = 1/n\text{)}.$$

Then $x_n \to x$ but $x \notin S$. Hence S is not closed.

(If): Assume that $X \setminus S$ is open. Let $(x_n) \in S$ be such that $x_n \to x$. If $x \notin S$, i.e., $x \in X \setminus S$, there exists $\delta > 0$ such that

$$\{y : \|x - y\| < \delta\} \subseteq X \setminus S$$

since $X \setminus S$ is open. In particular, since $(x_n) \in S$, $\|x - x_n\| \geqslant \delta$, a contradiction. We conclude that $x \in S$, so that S is closed. □

Example 1.12. *We showed that the unit closed ball B is closed. Hence*

$$X \setminus B = \{x : \|x\| > 1\} \text{ is open.}$$

Definition 1.13 (Closure). *If $S \subseteq X$, then the* closure $\overline{S}$ *of the set S is the set of all limit points of S.*

Proposition 1.14 (Properties of the Closure). *The following hold:*

(1) *The closure $\overline{S}$ of a set S is closed.*
(2) *$\overline{S}$ is the smallest closed set containing S. (That is, if T is closed and $T \supseteq S$, then $T \supseteq \overline{S}$.)*

Proof (1): To show $\overline{S}$ is closed, let x be a limit point of $\overline{S}$, so that there exists $x_n \in \overline{S}$ with $x_n \to x$. Since $x_n \in \overline{S}$, we may find $y_n \in S$ such that $\|x_n - y_n\| < 1/n$. Then

$$0 \leqslant \|y_n - x\| = \|(y_n - x_n) + (x_n - x)\| \leqslant \frac{1}{n} + \|x_n - x\| \to 0.$$

So x is a limit point of S, i.e., $x \in \overline{S}$.

(2): Let $x \in \overline{S}$. Then we can find a sequence $(x_n) \in S$ with $x_n \to x$. But $x_n \in T$ so the limit x is in T since T is closed. Thus $\overline{S} \subseteq T$. □

Definition 1.15 (Dense Set). *Let $S \subseteq T \subseteq (X, \|\cdot\|)$. We say that S is* dense *in T iff $T \subseteq \overline{S}$. Equivalently if either of the following statements hold:*

- *Every $x \in T$ is a limit point of S: there exists $(x_n) \in S$ such that $x_n \to x$.*
- *For every $x \in T$ and every $\varepsilon > 0$, there exists $y \in S$ such that $\|x - y\| < \varepsilon$.*

We think of a dense set S as one for which we can approximate points in $T \supseteq S$ by points lying in S.

Example 1.16. *$\mathbb{Q}$ is dense in $\mathbb{R}$ with the $|\cdot|$-norm.*

We will derive the following more substantial density result in Chapter 4.

Theorem 1.17 (Weierstrass Approximation Theorem). *The polynomials over $[0,1]$ are dense in $\big(C[0,1], \|\cdot\|_\infty\big)$. That is, given any continuous function $f\colon [0,1] \to \mathbb{R}$ and any $\varepsilon > 0$, there exists a polynomial $p(t)$ such that*

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

Proof This result is a corollary of Theorem 4.3. □

The following result illustrates how the Weierstrass Approximation Theorem may be used to transfer a property that is easy to show for polynomials to all continuous functions.

Lemma 1.18 (Riemann-Lebesgue Lemma). *Let $f \in C[0,1]$. Then the Fourier cosine coefficients*

$$\int_0^1 f(t)\cos(2n\pi t)\,\mathrm{d}t \to 0 \quad \text{as} \quad n \to \infty.$$

Proof Let $\varepsilon > 0$. By the Weierstrass Approximation Theorem, there exists a polynomial $p(x)$ such that

$$\|f - p\|_\infty < \frac{\varepsilon}{2}.$$

Using obvious inequalities and integration by parts,

$$\begin{aligned}
I := \left|\int_0^1 f(t)\cos(2n\pi t)\,\mathrm{d}t\right| &= \left|\int_0^1 \big(f(t)-p(t)\big)\cos(2n\pi t)\,\mathrm{d}t + \int_0^1 p(t)\cos(2n\pi t)\,\mathrm{d}t\right| \\
&\leqslant \left|\int_0^1 \big(f(t)-p(t)\big)\cos(2n\pi t)\,\mathrm{d}t\right| + \left|\int_0^1 p(t)\cos(2n\pi t)\,\mathrm{d}t\right| \\
&\leqslant \int_0^1 \big|f(t)-p(t)\big|\,\big|\cos(2n\pi t)\big|\,\mathrm{d}t + \Bigg|\underbrace{\left[p(t)\frac{\sin(2n\pi t)}{2n\pi}\right]_0^1}_{=0} \\
&\qquad - \frac{1}{2n\pi}\int_0^1 \sin(2n\pi t)p'(t)\,\mathrm{d}t\Bigg| \\
&\leqslant \|f-p\|_\infty + \frac{1}{2n\pi}\left|\int_0^1 p'(t)\sin(2n\pi t)\,\mathrm{d}t\right| \\
&\leqslant \|f-p\|_\infty + \frac{1}{2n\pi}\int_0^1 |p'(t)|\,\mathrm{d}t.
\end{aligned}$$

If we choose n large enough, we can make the second term less than $\varepsilon/2$ and thus obtain $I < \varepsilon$. $\square$

CHAPTER 2

Completeness and the Contraction Mapping Theorem

Basic Notions

Completeness is a central concept in functional analysis. The following definition, which is familiar in real analysis, extends easily to normed spaces.

♦ Definition 2.1 (Cauchy Sequence). *A sequence (x_n) in a normed space $\big(X, \|\cdot\|\big)$ is called a* Cauchy sequence *(or just* Cauchy*) if for all $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that*

$$\|x_n - x_m\| < \varepsilon \qquad \text{for every } n, m \geqslant n_o.$$

♦ Example 2.2. *Let $f_n(t) = t^n$ in $\big(C[0,1], \|\cdot\|_1\big)$.Then*

$$\|f_n - f_m\|_1 = \int_0^1 |t^n - t^m| \,\mathrm{d}t \leqslant \int_0^1 |t^n| + |t^m| \,\mathrm{d}t = \frac{1}{n+1} + \frac{1}{m+1}.$$

Then if $\varepsilon > 0$, simply set $n, m > 2/\varepsilon$ to obtain $\|f_n - f_m\|_1 < \varepsilon$. Thus, (f_n) is Cauchy.

♦ Proposition 2.3. *In a normed space, every convergent sequence is Cauchy.*

Proof Suppose $x_n \to x$ and let $\varepsilon > 0$. There must exist $n_0 \in \mathbb{N}$ such that

$$\|x_n - x\| < \frac{\varepsilon}{2}$$

for $n > n_0$. Then

$$\|x_n - x_m\| \leqslant \|x_n - x\| + \|x - x_m\| < \varepsilon \qquad \text{for every } n, m > n_0.$$

□

The converse of this proposition is in general *not* true. The Cauchy sequences may be thought of sequences that are 'trying to converge' and a space is complete if there is somewhere for them to converge to.

♦ Definition 2.4 (Complete Normed Space, Banach Space). *Let $\big(X, \|\cdot\|\big)$ be a normed space. If every Cauchy sequence in X converges to a limit in X, we call $\big(X, \|\cdot\|\big)$* a complete normed space *or a* Banach space. *Furthermore, if S is a subset of some normed space X and every Cauchy sequence in S converges to a limit in S, we call S* a complete subset *of X.*

Many important normed spaces are complete and we will give some of these below. It is useful to have the completeness property in a normed space since we can assert that limits of sequences exist without explicitly having to find the limit.

One way of demonstrating the completeness of sets such as $\mathbb{R}$ is via the following result:

♦ Lemma 2.5. *Every bounded sequence of real numbers has a convergent subsequence. (That is, if there exists $M \in \mathbb{R}$ such that $|x_n| \leqslant M$ for every n, then there exists $x \in \mathbb{R}$ and a subsequence $(x_{n_i})_{i\in\mathbb{N}}$ such that $x_{n_i} \to x$.)*

Proof Define

$$x := \inf\big\{\alpha : (-\infty, \alpha) \text{ contains infinitely many terms of } (x_n)\big\}. \tag{2.1}$$

Clearly, $-M \leqslant x \leqslant M$. Given $\varepsilon > 0$, there are infinitely many terms of the sequence in $(x-\varepsilon, x+\varepsilon)$, since there are only finitely many in $(-\infty, x-\varepsilon)$ but infinitely many in $(-\infty, x+\varepsilon)$. Now simply let $(x_{n_i})_{i\in\mathbb{N}}$ be such that

$$x_{n_i} \in (x - \tfrac{1}{i}, x + \tfrac{1}{i}) \text{ and such that } n_1 < n_2 < \cdots .$$

Hence $x_{n_i} \to x$. □

Note that in order to prove this result we have had to assume in (2.1) an axiomatic property of real numbers, namely that every bounded set of real numbers has a least upper bound or infimum.

Important Complete Normed Spaces

We give some important examples of complete normed spaces.

♦ PROPOSITION 2.6 (GENERAL PRINCIPLE OF CONVERGENCE). *The space* $(\mathbb{R}, |\cdot|)$ *is complete.*

Proof Let (x_n) be Cauchy. Then there is an n_0 such that $|x_n - x_m| < 1$ for every $n, m \geqslant n_0$. If $n \geqslant n_0$ then, by the triangle inequality,

$$|x_n| \leqslant |x_n - x_{n_0}| + |x_{n_0}| \leqslant 1 + |x_{n_0}|.$$

So $|x_n| \leqslant \max\{|x_1|, |x_2|, \dots, |x_n|, 1 + |x_{n_0}|\}$ for all n, so (x_n) is bounded. By Lemma 2.5 we can find a convergent subsequence $(x_{n_i})_{i\in\mathbb{N}}$ with limit x. Given $\varepsilon > 0$ we may find n_0 such that for $n, m \geqslant n_0$, we have $|x_n - x_m| < \frac{\varepsilon}{2}$. We may then find i such that $n_i \geqslant n_0$ and $|x_{n_i} - x| < \frac{\varepsilon}{2}$, giving

$$|x_n - x| \leqslant |x_n - x_{n_i}| + |x_{n_i} - x| < \varepsilon.$$

□

Hence, a sequence of real numbers converges if and only if it is Cauchy, by this proposition and Proposition 2.3. Note that we proved completeness of $\mathbb{R}$ from Lemma 2.5, which in turn relied on a bounded nonempty set having an infimum. It turns out that this property of subsets of $\mathbb{R}$ is equivalent to the completeness of $\mathbb{R}$. A serious textbook on real analysis will have more to say on this rather deep matter.

Completeness of other spaces often follows from completeness of $\mathbb{R}$.

♦ EXAMPLE 2.7. *The spaces* $(\mathbb{R}^N, \|\cdot\|_\infty)$, $(\mathbb{R}^N, \|\cdot\|_1)$ *and* $(\mathbb{R}^N, \|\cdot\|_p)$ *(with* $p \geqslant 1$*) are all complete.*

Sample proof for $(\mathbb{R}^N, \|\cdot\|_1)$. Let $(x_n) = ((x_{n,1}, \dots, x_{n,N}))$ be Cauchy in $(\mathbb{R}^N, \|\cdot\|_1)$. For each i,

$$|x_{n,i} - x_{m,i}| \leqslant \sum_{i=1}^{N} |x_{n,i} - x_{m,i}| = \|x_n - x_m\|_1.$$

Thus $(x_{n,i})_i$ is a Cauchy sequence for each i and $x_{n,i} \to x_i$ for some $x_i \in \mathbb{R}$ by completeness of $(\mathbb{R}, |\cdot|)$. Then

$$\|x_n - x\|_1 = \sum_{i=1}^{N} |x_{n,i} - x_i| \to 0,$$

so $x_n \to x$ where $x = (x_1, \dots, x_N)$. □

The next result is important for later work since many of the later results rely on this space being complete.

♦ PROPOSITION 2.8. *The space* $(C[0,1], \|\cdot\|_\infty)$ *is a Banach space.*

Proof Suppose $(f_n)_n$ is a Cauchy sequence in $(C[0,1], \|\cdot\|_\infty)$. Then for each $t \in [0,1]$ we have,

$$|f_n(t) - f_m(t)| \leqslant \sup_{u\in[0,1]} |f_n(u) - f_m(u)| = \|f_n - f_m\|_\infty.$$

Hence, given $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that, whenever $n, m \geqslant n_0$, we have

$$|f_n(t) - f_m(t)| < \varepsilon \qquad \text{for every } t \in [0,1]. \tag{2.2}$$

So for each t, the sequence $\big(f_n(t)\big)_n$ is Cauchy and by completeness of $\mathbb{R}$ converges to a limit $f(t)$, say. We need to show that $f(t) \in C[0,1]$. Letting $m \to \infty$ in (2.2) it follows that if $n \geqslant n_0$ then $|f_n(t) - f(t)| \leqslant \varepsilon$ for every $t \in [0,1]$. Thus $f_n \to f$ uniformly on $[0,1]$, that is in $\|\cdot\|_\infty$. A standard result of real analysis states that the uniform limit of a sequence of continuous functions is continuous, so since the f_n are continuous f is continuous. □

Note that the spaces $\big(C[0,1], \|\cdot\|_1\big)$ and $\big(C[0,1], \|\cdot\|_p\big)$ are not Banach spaces.

We list a few more complete spaces for reference:

♦ EXAMPLE 2.9. *The spaces* $\big(L^1[0,1], \|\cdot\|_1\big)$, $\big(L^2[0,1], \|\cdot\|_2\big)$ *and* $\big(L^p[0,1], \|\cdot\|_p\big)$ *are all Banach spaces.*

♦ EXAMPLE 2.10. *The sequence spaces* $\big(l^1[0,1], \|\cdot\|_1\big)$, $\big(l^2[0,1], \|\cdot\|_2\big)$ *and* $\big(l^p[0,1], \|\cdot\|_p\big)$ *are Banach spaces.*

Completeness of subsets is inherited from the space provided that the subset is closed:

♦ PROPOSITION 2.11. *Let* $\big(X, \|\cdot\|\big)$ *be a complete normed space, and* Y *be a closed subset of* X. *Then* Y *is complete.*

Proof Let (x_n) be a Cauchy sequence in Y. Then (x_n) is a Cauchy sequence in X, and thus $x_n \to x \in X$. Since Y is closed, $x \in Y$; hence Y is complete. □

The Contraction Mapping Theorem

♦ NOTATION 2.12. *If* $T\colon X \to X$ *we write* $T^n\colon X \to X$ *to denote the* n*th iterate of* T, *given by*

$$T^n(x) = \underbrace{T(T(T(\cdots(T}_{n \text{ times}}x)\cdots)))$$

(i.e., $T^{n+1}x = T(T^n x)$ *where* $T^1(x) = T(x)$*).*

♦ DEFINITION 2.13 (CONTRACTION MAPPING). *Let* $\big(X, \|\cdot\|\big)$ *be a normed space. Let* $T\colon X \to X$. *We call* T a contraction mapping *if there exists* $k \in (0,1)$ *such that*

$$\|Tx - Ty\| \leqslant k\|x - y\| \qquad \text{for every } x, y \in X.$$

Thus contractions reduce distances by a factor k or less. The following theorem (which in fact holds for any complete metric space) is sometimes called Banach's fixed point theorem.

♦ THEOREM 2.14 (CONTRACTION MAPPING THEOREM). *Let* $T\colon X \to X$ *be a contraction mapping on a Banach space* $\big(X, \|\cdot\|\big)$. *Then* T *has a unique fixed point, that is there exists a unique* $x \in X$ *such that*

$$Tx = x.$$

Moreover, for every $y \in X$, *the sequence* $T^n y \to x$ *as* $n \to \infty$. *This theoem remains true if* X *is taken to be a complete subset of a normed space.*

Proof For every $n \in \mathbb{N}$ and every $y \in X$, we have $\|T^n y - T^{n-1}y\| = \|T(T^{n-1})y - T(T^{n-2})y\| \leqslant k\|T^{n-1}y - T^{n-2}y\|$, so by induction we get $\|T^n y - T^{n-1}y\| \leqslant k^{n-1}\|Ty - y\|$. Then if $n > m$

$$\begin{aligned}
\|T^n y - T^m y\| &= \|(T^n y - T^{n-1}y) + (T^{n-1}y - T^{n-2}y) + \cdots + (T^{m+1}y - T^m y)\| \\
&\leqslant \|T^n y - T^{n-1}y\| + \|T^{n-1}y - T^{n-2}y\| + \cdots + \|T^{m+1}y - T^m y\| \\
&\leqslant [k^{n-1} + k^{n-2} + \cdots + k^m]\|Ty - y\| \\
&\leqslant \frac{k^m}{1-k}\|Ty - y\| \qquad (2.3)
\end{aligned}$$

where the last step comes from summing the geometric series. If $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that

$$\frac{k^m}{1-k}\|Ty - y\| < \varepsilon \qquad \text{if } m \geqslant n_0$$

so by (2.3) the sequence $(T^n y)_{n\in\mathbb{N}}$ is Cauchy. The completeness of X implies that $T^n y \to x$ for some $x \in X$. To show that x is a fixed point of T, note that

$$\|Tx - x\| = \|(Tx - T^n y) + (T^n y - x)\| \leqslant k\|x - T^{n-1}y\| + \|T^n y - x\| \to 0.$$

So $\|Tx - x\| = 0$, that is $Tx = x$.

To show uniqueness, suppose that x and x' are both fixed points. Then

$$0 \leqslant \|x - x'\| = \|Tx - Tx'\| \leqslant k\|x - x'\|;$$

since $0 \leqslant k < 1$ this implies $\|x - x'\| = 0$ so $x = x'$.

Finally, for any $z \in X$, the sequence $T^n z$ converges to some fixed point of T, as in the first part of the proof; since this point is unique $T^n z \to x$ for all $z \in X$. □

Note that for all $z \in X$, we have $\|T^n z - x\| \leqslant k^n \|z - x\|$. If k is small then we will have rapid convergence.

The contraction mapping theorem is a very powerful tool with applications in a wide range of settings.

♦ Example 2.15 (Roots of Equations). *Investigate the real roots of the equation* $x^3 + 3x - 1 = 0$.

Rearranging gives

$$x = \frac{1}{x^2+3}.$$

The mapping $T\colon \mathbb{R} \to \mathbb{R}$ such that $x \mapsto 1/(x^2+3)$ is a contraction on the complete space $(\mathbb{R}, |\cdot|)$, since

$$\begin{aligned}|Tx - Ty| &= \left|\frac{1}{x^2+3} - \frac{1}{y^2+3}\right| = \left|\frac{(y-x)(y+x)}{(x^2+3)(y^2+3)}\right| \\ &\leqslant |x-y|\left(\frac{|x|}{x^2+3}\frac{1}{y^2+3} + \frac{|y|}{y^2+3}\frac{1}{x^2+3}\right) \\ &\leqslant \tfrac{2}{3}|x-y|.\end{aligned}$$

(We could have also used the mean value theorem thus

$$|f(x) - f(y)| \leqslant |x-y| \max_c |f'(c)|$$

where c ranges over the interval in question.) By the CMT the equation has a unique real solution. Moreover, for any $y \in \mathbb{R}$, the sequence $(T^n y)_n$ converges to the solution, so taking $y = 0.5$ we get the sequence of approximations

$$0.5, \quad 0.308, \quad 0.323, \quad 0.3221, \quad 0.32219\ldots.$$

We next turn to the question of whether a given differential equation has a solution and if so whether it is unique. This is not always clear, for example, the equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2|y|^{1/2} \qquad \text{with } y(0) = 0$$

has $y = 0$ and $y = x^2$ both as solutions, in fact, it has infinitely many solutions. However, the CMT may used to show that that a wide variety of differential equations have unique solutions.

♦ Example 2.16 (Differential Equations). *The differential equation*

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \tfrac{1}{4}\cos\bigl(f(t)\bigr) + t^3 \qquad \textit{with } f(0) = 0.$$

has a unique solution for $0 \leqslant t \leqslant 1$.

This differential equation is equivalent to the integral equation

$$f(t) = \tfrac{1}{4}\int_0^t \cos\bigl(f(u)\bigr)\,\mathrm{d}u + \tfrac{1}{4}t^4,$$

which includes the initial conditions. Define the mapping T on $(C[0,1], \|\cdot\|_\infty)$ by

$$(Tf)(t) = \tfrac{1}{4}\int_0^t \cos\big(f(u)\big)\,\mathrm{d}u + \tfrac{1}{4}t^4.$$

Then,

$$\begin{aligned}
\big|(Tf)(t) - (Tg)(t)\big| &\leqslant \tfrac{1}{4}\int_0^t \big|\cos(f(u)) - \cos(g(u))\big|\,\mathrm{d}u \\
&\leqslant \tfrac{1}{4}\int_0^t \big|f(u) - g(u)\big|\,\mathrm{d}u \\
&\leqslant \tfrac{1}{4}t\|f-g\|_\infty \leqslant \tfrac{1}{4}\|f-g\|_\infty \text{ since } t \in [0,1].
\end{aligned}$$

In particular, $\|Tf - Tg\|_\infty \leqslant \frac{1}{4}\|f-g\|_\infty$. Hence T is a contraction on the complete normed space $(C[0,1], \|\cdot\|_\infty)$. This means that there is a unique 'fixed point' f of T that is the solution to the integral equation and thus to the differential equation. We could even use iterative numerical methods to find a sequence of approximations $T^n(g)$ to this f for any initial $g \in C([0,1])$.

CHAPTER 3

Finite Dimensional Spaces

In this short chapter we will show that all norms on a given finite dimensional space are equivalent in that they define the same convergent sequences, etc. We first prove a lemma which is a generalisation of Lemma 2.5 to $\mathbb{R}^N$.

♦ LEMMA 3.1. *Every bounded sequence in* $\left(\mathbb{R}^N, \|\cdot\|_\infty\right)$ *has a convergent subsequence.*

PROOF. Recall (Lemma 2.5) that every bounded sequence in $\mathbb{R}$ has a convergent subsequence. So let $(x_n)_n$ be a bounded sequence in $\mathbb{R}^N$. Then there is a real M such that $\|x_n\|_\infty \leqslant M$ for every $n \in \mathbb{N}$. We write members of the sequence in coordinate form $x_n = (x_{n,1}, x_{n,2}, \ldots, x_{n,N})$. Apply the established result for $\mathbb{R}$ to each coordinate:

Since $(x_{n,1})$ is bounded it has a convergent subsequence $(x_{n_{k(1)},1})_{k(1)} \to x_1$; since $(x_{n_{k(1)},2})$ is bounded it has a convergent subsequence $(x_{n_{k(2)},2})_{k(2)} \to x_2$, and so on. At each stage we are considering only the points left in the sequence $(x_n)_n$ after deleting to obtain a convergent subsequence in the previous coordinate. We are left with all coordinates converging, that is $(x_{n_{k(N)},i})_{k(N)} \to x_i$ for all $i = 1, \ldots, N$. □

♦ DEFINITION 3.2 (EQUIVALENT NORMS). *Two norms* $\|\cdot\|_A$ *and* $\|\cdot\|_B$ *on the same space* X *are called* equivalent *if there are* $a, b > 0$ *such that*

$$a\|x\|_A \leqslant \|x\|_B \leqslant b\|x\|_A \qquad \text{for every } x \in X.$$

We note that if a sequence converges in one of the norms, it converges in the other, since $a\|x_n - x\|_A \leqslant \|x_n - x\|_B \leqslant b\|x_n - x\|_A$, so $x_n \to x$ in $\|\cdot\|_A$ iff $x_n \to x$ in $\|\cdot\|_B$. Similarly, if a sequence is Cauchy in one norm it is Cauchy in the other, etc. It can be shown that the equivalence of norms is an equivalence relation.

In the case of finite-dimensional spaces all norms on a space are equivalent:

♦ THEOREM 3.3. *Let* X *be a vector space with finite dimension* N. *Then all norms on* X *are equivalent.*

PROOF. Let $\{e_1, e_2, \ldots, e_N\}$ be a basis for X, and define

$$\left\|\sum_{i=1}^N x_i e_i\right\|_\infty = \max_{1\leqslant i\leqslant N} |x_i|.$$

We will show that any other given norm $\|\cdot\|$ is equivalent to this norm.

First,

$$\left\|\sum_i x_i e_i\right\| \leqslant \sum_i |x_i|\,\|e_i\| \leqslant \max_i |x_i| \sum_i \|e_i\| = \left\|\sum_i x_i e_i\right\|_\infty \underbrace{\sum_i \|e_i\|}_{b}.$$

Thus $\|\sum_i x_i e_i\| \leqslant b\|\sum_i x_i e_i\|_\infty$.

Secondly, to show that $\|x\|_\infty \leqslant c\|x\|$ we argue by contradiction. Suppose that there is no c with this property: then for every $n \in \mathbb{N}$, there is an $x_n \in X$ such that

$$\|x_n\|_\infty \geqslant n\|x_n\|.$$

Put $y_n = x_n/\|x_n\|_\infty$, so that $\|y_n\|_\infty = 1$ and $\|y_n\|_\infty \geqslant n\|y_n\|$, so $\|y_n\| \leqslant 1/n$.

Since $(y_n)_n$ is bounded, by Lemma 3.1 it has a convergent subsequence $(y_{n_i})_i$ with $y_{n_i} \to y$ in the $\|\cdot\|_\infty$-norm. Then

$$\big| \|y_{n_i}\|_\infty - \|y\|_\infty \big| \leqslant \|y_{n_i} - y\|_\infty \to 0$$

which implies $\|y\|_\infty = 1$. However, using the triangle inequality,

$$\|y\| \leqslant \|y - y_{n_i}\| + \|y_{n_i}\| \leqslant b\|y - y_{n_i}\|_\infty + \frac{1}{n_i} \to 0.$$

Thus $\|y\| = 0$, so $y = 0$ which contradicts that $\|y\|_\infty = 1$. □

♦ COROLLARY 3.4. *In a finite dimensional space X, Cauchy sequences, convergent sequences, limit points, closed sets etc. are independent of the norm chosen. In particular, every finite dimensional normed space $(X, \|\cdot\|)$ is complete since, as we showed in Chapter 2, $(X, \|\cdot\|_1)$ is complete.*

♦ EXAMPLE 3.5. *Norms on $P_N = \{a_0 + a_1 t + a_2 t^2 + \ldots + a_N t^N : a_i \in \mathbb{R}\}$, the space of polynomials of degree at most N on $[0, 1]$.*

The space P_N has dimension $N + 1$. We define $\|\cdot\|$ on P_N by

$$\|a_0 + a_1 t + a_2 t^2 + \ldots + a_N t^N\| = \max_i |a_i|.$$

which can be checked is a norm. Another possible norm is

$$\|a_0 + a_1 t + a_2 t^2 + \ldots + a_N t^N\|_\infty = \sup_{t\in[0,1]} |p(t)|.$$

These norms are equivalent by Theorem 3.3. Hence a sequence $(p_n)_n$ of polynomials converges in one norm iff it converges in the other. In other words, a sequence of polynomials (p_n) in P_N converges to p uniformly, that is in $\|\cdot\|_\infty$, if and only each sequence of individual coefficients converges.

CHAPTER 4

The Space $\big(C[K], \|\cdot\|_\infty\big)$

♦ Notation 4.1. *We write $(C[K], \|\cdot\|_\infty)$ to denote the vector space of continuous functions on K, where K is a compact topological or metric space, equipped with the norm*

$$\|f\|_\infty = \sup_{t\in K} |f(t)|. \tag{4.1}$$

For our purposes K will normally be the closed interval $[0,1]$, or $[a,b]$, or the interval $[0,2\pi]$ with 0 and 2π identified (thus forming a circle), or the square $[0,1]\times[0,1]$. We will develop the theory for $(C[0,1], \|\cdot\|_\infty)$ but identical arguments hold in the other cases. We will state the Stone-Weierstrass theorem for real-valued continuous functions on $[0,1]$. We work with the supremum norm (4.1) where the space K is $[0,1]$.

♦ Definition 4.2 (Algebra). *A vector subspace A of $C[0,1]$ is called an* algebra *if $fg \in A$ whenever $f \in A$ and $g \in A$. (By fg is meant the function defined by $(fg)(t) = f(t)g(t)$ for every $t \in C[0,1]$.)*

For example $C[0,1]$ is itself an algebra and the polynomials form a subalgebra. As usual the closure $\overline{A}$ of A consists of the limit points of A; if A is a subalgebra of $C[0,1]$ then so is $\overline{A}$.

The Stone-Weierstrass Theorem gives simple conditions for an algebra to be dense in $C[0,1]$, in other words tells us when every $f \in C[0,1]$ can be approximated arbitraily closely by members of A.

♦ Theorem 4.3 (Stone-Weierstrass Theorem). *Let A be a subalgebra of $(C[0,1], \|\cdot\|_\infty)$ such that*

(1) *A contains the constant functions,*
(2) *A 'separates the points of $[0,1]$', i.e., for all $t \neq u \in [0,1]$ we may find $f \in A$ such that $f(t) \neq f(u)$.*

Then A is dense in $(C[0,1], \|\cdot\|_\infty)$, that is $\overline{A} = C[0,1]$.

The proof uses several lemmas.

♦ Lemma 4.4. *Given $\varepsilon > 0$ there exists a polynomial p on $[-1,1]$ such that $\big|p(t) - |t|\big| < \varepsilon$ for all $-1 \leqslant t \leqslant 1$. (Thus $|t|$ is approximable by polynomials in $(C[-1,1], \|\cdot\|_\infty)$.)*

Proof We use an iterative trick. Define a sequence of polynomials by $p_1(t) = 0$, and

$$p_{n+1}(t) = \tfrac{1}{2}t^2 + p_n(t) - \tfrac{1}{2}p_n(t)^2 \qquad (n = 1, 2, \ldots). \tag{4.2}$$

Clearly $p_n(t)$ is a polynomial for all n. We claim that $p_n(t) \to |t|$ in $(C[-1,1], \|\cdot\|_\infty)$.

Note that if $0 \leqslant a \leqslant 1$ then $0 \leqslant a - \frac{1}{2}a^2 \leqslant \frac{1}{2}$ by a simple calculus check. Hence from (4.2), if $0 \leqslant p_n(t) \leqslant 1$ for all $t \in [-1,1]$, then $0 \leqslant p_{n+1}(t) \leqslant 1$ for all $t \in [-1,1]$. Thus by induction, $0 \leqslant p_n(t) \leqslant 1$ for all $t \in [-1,1]$ and all n. Moreover

$$\begin{aligned} |p_{n+1}(t) - |t|| &= \big|\tfrac{1}{2}t^2 + p_n(t) - \tfrac{1}{2}p_n(t)^2 - |t|\big| \\ &= \big|p_n(t) - |t|\big|\,\big|1 - \tfrac{1}{2}(p_n(t) + |t|)\big| \\ &\leqslant \big|p_n(t) - |t|\big|\big(1 - \tfrac{1}{2}|t|\big) \end{aligned}$$

for $t \in [-1,1]$. Iterating this,

$$\big|p_n(t) - |t|\big| \leqslant \big|p_1(t) - |t|\big|\,\big(1 - \tfrac{1}{2}|t|\big)^{n-1} = |t|\,\big(1 - \tfrac{1}{2}|t|\big)^{n-1}. \tag{4.3}$$

So given $0 < \varepsilon < 1$, choose N large enough to get $\big(1 - \frac{1}{2}|\varepsilon|\big)^{n-1} < \varepsilon$ for all $n \geqslant N$. It follows that if $|t| < \varepsilon$, then

$$\big|p_n(t) - |t|\big| \leqslant |t| < \varepsilon, \quad \text{by (4.3)},$$

and if $\varepsilon \leqslant |t| \leqslant 1$, then

$$\big|p_n(t) - |t|\big| \leqslant \big(1 - \tfrac{1}{2}|t|\big)^{n-1} < \big(1 - \tfrac{1}{2}|\varepsilon|\big)^{n-1} < \varepsilon; \quad \text{again by (4.3).}$$

Hence $\big|p_n(t) - |t|\big| < \varepsilon$ for all $t \in [-1, 1]$ and $n \geqslant N$, i.e., $\|p_n - |t|\,\|_\infty < \varepsilon$. □

♦ LEMMA 4.5. *Let A be a closed subalgebra of $C[0,1]$. If $f \in A$ then $|f| \in A$.*

Proof We may assume that $\|f\|_\infty \leqslant 1$ on multiplying f by a scalar. Given $\varepsilon > 0$ choose a polynomial p as in Lemma 4.4 with $\big|p(t) - |t|\big| < \varepsilon$ for $-1 \leqslant t \leqslant 1$. Then $\big|p(f(u)) - |f(u)|\big| < \varepsilon$ for all $u \in [0,1]$, i.e., $\big\|p(f) - |f|\big\|_\infty < \varepsilon$. But $p(f)$ is just the sum of powers of f so is in the algebra A since $f \in A$. As ε is arbitrary, $|f| \in \overline{A} = A$ since A is closed. □

♦ LEMMA 4.6. *Let A be a closed subalgebra of $C[0,1]$. If $f, g \in A$ then $\max\{f,g\} \in A$ and $\min\{f,g\} \in A$.*

Proof We have

$$\max\{f,g\} = \tfrac{1}{2}\big((f+g) + |f-g|\big) \in A,$$
$$\min\{f,g\} = \tfrac{1}{2}\big((f+g) - |f-g|\big) \in A;$$

by using Lemma 4.5 and that A is an algebra. □

Proof of the Stone-Weierstrass Theorem: To prove the Theorem we assume that A is a closed subalgebra and show that $A = C[0,1]$. (If A is an algebra satisfying (1) and (2) then so is $\overline{A}$.)

Note that (1) and (2) in the statement of the Theorem imply that given $t \neq u \in [0,1]$ there always exists $h \in A$ such that $h(t)$ and $h(u)$ take any prescribed values. (Take $h(s) = af(s) + b$ for suitable constants a, b, with f as in (2).)

Let $f \in C[0,1]$ and $\varepsilon > 0$; we will find $g \in A$ such that $\|f - g\|_\infty < \varepsilon$. For $t \neq u \in [0,1]$ take $f_{t,u} \in A$ such that $f_{t,u}(s) = f(s)$ when $s = t$ and $s = u$. Then

$$O_{t,u} = \{s \in [0,1] : f_{t,u}(s) < f(s) + \varepsilon\}$$

is an open subset of $[0,1]$ containing t and u. Thus for each t we have $\bigcup_{u \in [0,1]} O_{t,u} = [0,1]$. A standard result (the Heine-Borel theorem or the compactness of $[0,1]$) states that any cover of $[0,1]$ by a collection of open sets has a finite subcover. Thus there is a *f*inite collection of points $\{u_1, \dots, u_k\}$ such that

$$\bigcup_{i=1}^{k} O_{t,u_i} = [0,1].$$

Define $f_t = \min_i f_{t,u_i}$, which is in A, by Lemma 4.6. Then

$$f_t(t) = f(t) \quad \text{and } f_t(s) < f(s) + \varepsilon \text{ for all } s \in [0,1], \tag{4.4}$$

since if $s \in [0,1]$ then $s \in O_{t,u_i}$ for some i.

In the same way, $O_t = \{s \in [0,1] : f_t(s) > f(s) - \varepsilon\}$ is an open subset of $[0,1]$ containing t. Then $\bigcup_{t \in [0,1]} O_t = [0,1]$ so by the standard result there is a *f*inite collection of points $\{t_1, \dots, t_m\}$ such that $\bigcup_{i=1}^{m} O_{t_i} = [0,1]$. Define $g = \max_i f_{t_i}$ which is in A, by Lemma 4.6. Then if $s \in [0,1]$ we have

$$g(s) \leqslant \max_i f_{t_i}(s) < f(s) + \varepsilon \qquad \text{by (4.4)},$$

but also $s \in O_{t_i}$ for some i, so

$$g(s) \geqslant f_{t_i}(s) > f(s) - \varepsilon.$$

Thus $\|g - f\|_\infty < \varepsilon$. It follows that A is dense in $C[0,1]$, so since A is closed, $A = C[0,1]$. □

♦ COROLLARY 4.7 (TO STONE-WEIERSTRASS THEOREM). *The following follow immediately from the Stone-Weierstrass Theorem:*

(1) *Weierstrass's Theorem: The polynomials on* $[0,1]$ *are dense in* $(C[0,1], \|\cdot\|_\infty)$.
(2) *The infinitely differentiable functions on* $[0,1]$ *are dense in* $(C[0,1], \|\cdot\|_\infty)$.
(3) *The two variable polynomials* $p(x,y)$ *on* $[0,1]\times[0,1]$ *are dense in* $(C([0,1]\times[0,1]), \|\cdot\|_\infty)$.
(4) *The trigonometric polynomials on* $[0,2\pi]$, *i.e., functions of the form*

$$\sum_{j=1}^{n} a_j \sin(jt) + \sum_{j=0}^{m} b_j \cos(jt),$$

are dense in $(C[0,2\pi], \|\cdot\|_\infty)$ *(with* 0 *and* 2π *identified).*

Proof One only needs to show that these sets satisfy the assumptions of the theorem which is straightforward. To show the last set is an algebra, note that

$$\sin(rt)\cos(st) = \tfrac{1}{2}\big(\sin((r+s)t) + \sin((r-s)t)\big).$$

□

These approximation theorems often allow us to obtain extend results which are easy to prove for polynomials, say, to general continuous functions.

♦ EXAMPLE 4.8. *Functions determined by moments.*

As above, the polynomials are dense in $(C[0,1], \|\cdot\|_\infty)$. Suppose $f \in C[0,1]$ is such that its 'moments' are all zero, i.e.

$$\int_0^1 f(t)t^k \,\mathrm{d}t = 0 \qquad \text{for every integer } k \geqslant 0. \tag{4.5}$$

We claim that $f(t) = 0$ for all $t \in [0,1]$. To see this, first let $\varepsilon > 0$, and choose a polynomial p such that $\|f-p\|_\infty \leqslant \varepsilon$. Note that

$$f^2 = f(f-p) + fp.$$

We then have

$$\int_0^1 f^2(t)\,\mathrm{d}t = \int_0^1 f(t)\big(f(t)-p(t)\big)\,\mathrm{d}t + \underbrace{\int_0^1 f(t)p(t)\,\mathrm{d}t}_{=0 \text{ by } (4.5)}$$

$$\text{so} \qquad \left|\int_0^1 f^2(t)\,\mathrm{d}t\right| \leqslant \int_0^1 |f(t)|\big|f(t)-p(t)\big|\,\mathrm{d}t \leqslant \varepsilon \int_0^1 |f(t)|\,\mathrm{d}t.$$

Thus $\int_0^1 f^2(t)\,\mathrm{d}t = 0$ so $f = 0$.

We may conclude that a function $f \in C[0,1]$ is uniquely determined by its moments, that is by the values given for every integer $k \geqslant 0$ by

$$M_k(f) := \int_0^1 f(t)t^k\,\mathrm{d}t.$$

If for every k we have $M_k(g) = M_k(h)$, then

$$\int_0^1 \big(g(t)-h(t)\big)t^k\,\mathrm{d}t = 0,$$

so from the above, $g = h$, that is we have uniqueness.

CHAPTER 5

Operators on Normed Spaces

In this chapter we generalise the notions of linear mappings, eigenvalues and eigenvectors to more general spaces. This eventually leads us to Spectral Theory. We begin with some basic terminology:

♦ Definition 5.1 (Linear, Continuous and Bounded Operators). *Let $\big(X, \|\cdot\|_X\big)$ and $\big(Y, \|\cdot\|_Y\big)$ be normed spaces where X, Y are vector spaces over either $\mathbb{R}$ or $\mathbb{C}$. We call a mapping $T\colon X \to Y$ an* operator*. We say that T is* linear *if*

$$T(\lambda x + \mu x') = \lambda T(x) + \mu T(x')$$

for every $x, x' \in X$ and all scalars λ, μ. We call T continuous *if given a sequence (x_n) in X we have*

$$x_n \to x \Rightarrow T(x_n) \to T(x).$$

T is called bounded *iff there exists $c > 0$ such that*

$$\|Tx\|_Y \leqslant c\|x\|_X \qquad \text{for every } x \in X.$$

Note that we often omit the subscript on the norm when it is clear what the domain of the norm is. We also write Tx and $T(x)$ interchangeably when the meaning is clear. In cases where T is linear and has a matrix representation this allows us to write Tx where T is the matrix; but this means more or less the same as $T(x)$, for our purposes at least.

♦ Example 5.2. *Let $A\colon \mathbb{R}^N \to \mathbb{R}^N$ be given by $x \mapsto Ax$ where A is an $N \times N$ matrix. A is clearly linear. If the matrix A has entries (a_{ij}) then $(Ax)_i = \sum_{k=1}^N a_{ik}x_k$ (where the subscripts refer to the ith and kth coordinates). We have that if $(x_{n,1}, \dots, x_{n,N}) \to (x_1, \dots, x_N)$ in the $\|\cdot\|_\infty$-norm, then $((Ax_n)_1, \dots, (Ax_n)_N) \to ((Ax)_1, \dots, (Ax_N))$ in $\|\cdot\|_\infty$, so A is continuous. We also have that*

$$\|Ax\|_\infty \leqslant \underbrace{\left(\max_i \sum_{j=1}^N |a_{ij}|\right)}_{c} \|x\|_\infty,$$

and the constant c is only dependent on A, so A is bounded.

♦ Example 5.3. *The following are linear, continuous and bounded on $(C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$:*

(1) *$T\colon (Tf)(t) = (t^2 + 2)f(t)$;*

(2) *$T\colon (Tf)(t) = \displaystyle\int_0^t u^2 f(u)\,\mathrm{d}u$;*

(3) *$T\colon (Tf)(t) = \displaystyle\int_0^1 g(t,u)f(u)\,\mathrm{d}u$ where $g(t,u)$ is continuous.*

Proof of (2) Linearity of T follows from linearity of the integral. To see that T is continuous, let $f_n \to f$ in $\|\cdot\|_\infty$. Then for every $n$we have

$$\big|(Tf_n)(t) - (Tf)(t)\big| \leqslant \int_0^t u^2 |f_n(u) - f(u)|\,\mathrm{d}u \leqslant \int_0^t u^2 \|f_n - f\|_\infty\,\mathrm{d}u = \frac{1}{3}\|f_n - f\|_\infty.$$

Then $\|Tf_n - Tf\|_\infty = \sup_{t\in[0,1]}\big|(Tf_n)(t) - (Tf)(t)\big| \to 0$, establishing continuity.

For boundedness, we have

$$\big|(Tf)(t)\big| \leqslant \int_0^t u^2 \|f\|_\infty\,\mathrm{d}u \leqslant \tfrac{1}{3}\|f\|_\infty$$

for all $t \in [0,1]$, so in particular $\|Tf\|_\infty \leqslant \frac{1}{3}\|f\|_\infty$. □

♦ EXAMPLE 5.4. *Let $D\colon C^1[0,1] \to C[0,1]$ with $D\colon f \mapsto \dfrac{\mathrm{d}f}{\mathrm{d}t}$, where $C^1[0,1]$ denotes the space of functions on $[0,1]$ with continuous derivative and the norm in both spaces is the $\|\cdot\|_\infty$-norm. By the usual linearity of derivatives, D is linear. It is not continuous or bounded, for if $f_n(t) := n^{-1/2}\sin(nt)$, then $f_n \to f := 0$ in $\|\cdot\|_\infty$. But $Df_n = n^{1/2}\cos(nt) \not\to 0 = Df$. Clearly D is not bounded.*

The following property, that a linear operator is bounded if and only if it is continuous is fundamental to all that follows.

♦ PROPOSITION 5.5. *Let $T\colon \big(X, \|\cdot\|_X\big) \to \big(Y, \|\cdot\|_Y\big)$ be linear. Then T is continuous if and only if it is bounded.*

Proof (⇒): Assume T is continuous but not bounded. Then choose $0 \neq x_n \in X$ such that

$$\|Tx_n\|_Y \geqslant n^2\|x_n\|_X.$$

By replacing x_n by $x_n/\|x_n\|_X$ we may assume $\|x_n\|_X = 1$. Considering the sequence x_n/n, we have

$$\Big\|\frac{x_n}{n}\Big\|_X = \frac{1}{n} \to 0. \tag{5.1}$$

We also have

$$\Big\|T\left(\frac{x_n}{n}\right)\Big\|_Y = \frac{1}{n}\|T(x_n)\|_Y \geqslant n\|x_n\|_X = n \to \infty. \tag{5.2}$$

The equations (5.1) and (5.2) imply that T is *not* continuous. ↯

(⇐): Let T be bounded and let $x_n \to x$ in $\|\cdot\|_X$. Then, we have by linearity and boundedness,

$$\|Tx_n - Tx\|_Y = \|T(x_n - x)\|_Y \leqslant c\|x_n - x\|_X \to 0.$$

This implies T is continuous. □

♦ EXAMPLE 5.6. *Let $T\colon (C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$ be given by*

$$(Tf)(t) = \int_0^1 (t^2 + u^2) f(u)\,\mathrm{d}u.$$

Then T is linear and bounded, for it is easy to show that $\|Tf\|_\infty \leqslant \frac{4}{3}\|f\|_\infty$. Continuity of T follows from the last lemma.

The Space of Operators

♦ DEFINITION 5.7 (SPACE OF OPERATORS, INDUCED NORM). *We write $\mathcal{B}(X,Y)$ to denote the* space of bounded linear operators *from X to Y. If $Y = X$ we simply write $\mathcal{B}(X)$ for the* space of bounded operator $T\colon X \to X$ on X. *If $T \in \mathcal{B}(X,Y)$, define $\|T\|$ to be the smallest (i.e., infimum) number c such that $\|Tx\| \leqslant c\|x\|$ for all $x \in X$. Thus*

$$\|T\| := \sup_{x\neq 0}\frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\substack{x\in X\\ \|x\|=1}}\frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\substack{x\in X\\ \|x\|=1}}\|Tx\|.$$

In particular we have for every $x \in X$ that $\|Tx\| \leqslant \|T\|\,\|x\|$. We call this norm of T the induced norm *on T.*

The definition does not make it clear that $\mathcal{B}(X,Y)$ is actually a normed linear space. We will go on to show that, in fact, it is.

♦ EXAMPLE 5.8. *We find the induced norm for T where $T\colon (C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$ is defined by*

$$(Tf)(t) = \int_0^t uf(u)\,\mathrm{d}u.$$

First, $\big|(Tf)(t)\big| \leqslant \displaystyle\int_0^t |u|\,\|f\|_\infty\,\mathrm{d}u \leqslant \tfrac{1}{2}\|f\|_\infty$. So in particular, for all $f \in C[0,1])$,

$$\|(Tf)(t)\|_\infty \leqslant \tfrac{1}{2}\|f\|_\infty$$

i.e., $\|T\| \leqslant \frac{1}{2}$ .Furthermore if $f(t) = 1$ for every $t \in [0,1]$, then we have $\|Tf\|_\infty = \frac{1}{2}$. This means that $\frac{1}{2}$ is the smallest number such that $\|Tf\|_\infty \leqslant \frac{1}{2}\|f\|_\infty$ for arbitrary f so we conclude that $\|T\| = \frac{1}{2}$.

♦ DEFINITION 5.9. *Let $T, U \in \mathcal{B}(X,Y)$ and let λ be a scalar. Then we define $T+U$ and λT by*

$$(T+U)(x) = Tx + Ux \qquad \text{and} \qquad (\lambda T)(x) = \lambda T(x). \tag{5.3}$$

Also, if $V \in \mathcal{B}(Y,Z)$ then we write VT for the composition $V \circ T\colon X \to Z$ defined by

$$(VT)(x) = V(T(x)) \qquad \text{for every } x \in X.$$

We will show that our definitions are 'good' ones in the results that follow.

♦ THEOREM 5.10. *The set $\mathcal{B}(X,Y)$ forms a linear space under the operations* (5.3) *and this is a normed space with the induced norm $\|\cdot\|$. Furthermore if Y is a Banach space, then so is $\big(\mathcal{B}(X,Y), \|\cdot\|\big)$.*

Proof To verify $\mathcal{B}(X,Y)$ is a linear space, we must do a few routine checks. So suppose $T, U \in \mathcal{B}(X,Y)$ and that λ is a scalar. It is a standard result that $T+U$ and λT are linear mappings. We must check that these are both bounded. So,

$$\|(T+U)(x)\| \leqslant \|Tx\| + \|Ux\| \leqslant \|T\|\,\|x\| + \|U\|\,\|x\| = \big(\|T\| + \|U\|\big)\|x\|, \tag{5.4}$$

and

$$\|(\lambda T)(x)\| = \|\lambda(Tx)\| = |\lambda|\,\|Tx\| \leqslant |\lambda|\,\|T\|\,\|x\|. \tag{5.5}$$

Thus, $T+U \in \mathcal{B}(X,Y)$ and $\lambda T \in \mathcal{B}(X,Y)$. We must also check that $\mathcal{B}(X,Y)$ is normed. Clearly $\|T\| \geqslant 0$ and $\|0\| = 0$. If $\|T\| = 0$, then $\|Tx\| \leqslant 0\|x\|$ for every $x \in X$, so $T = 0$. The triangle inequality follows from (5.4). The scalar property follows from (5.5) noting that

$$\|\lambda T\| = \sup_{\|x\|=1} \|\lambda Tx\| = \sup_{\|x\|=1} |\lambda|\|T(x)\| = |\lambda|\|T\|.$$

To check the rest of the linear space axioms is routine.

Now suppose Y is complete and let T_n be a Cauchy sequence in $\mathcal{B}(X,Y)$. So if $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $\|T_n - T_m\| < \varepsilon$ whenever $n, m \geqslant n_0$. If $x \in X$, we have

$$\|T_n x - T_m x\| = \|(T_n - T_m)x\| \leqslant \|T_n - T_m\|\,\|x\| < \varepsilon\|x\| \qquad \text{whenever } n, m \geqslant n_0. \tag{5.6}$$

Then, since Y is complete, the sequence $(T_n x)_n$ is Cauchy in Y and converges to a limit which we call $Tx \in Y$. It is easy to check that T is a linear operator. Also, if we let $m \to \infty$ in (5.6) we get that if $n \geqslant n_0$, then $\|T_n x - Tx\| < \varepsilon\|x\|$. Thus

$$\|Tx\| \leqslant \|Tx - T_n x\| + \|T_n x\| \leqslant \big(\varepsilon + \|T_n\|\big)\|x\|.$$

Hence, T is bounded. Moreover, if $n \geqslant n_0$ then $\|(T - T_n)x\| \leqslant \varepsilon\|x\|$; that is

$$\|T - T_n\| \leqslant \varepsilon \qquad \text{whenever } n \geqslant n_0,$$

so $T_n \to T$ in $\|\cdot\|$. □

♦ LEMMA 5.11. *Let $T \in \mathcal{B}(X,Y)$ and let $U \in \mathcal{B}(Y,Z)$. Then*

$$UT \in \mathcal{B}(X,Z) \qquad \text{and} \qquad \|TU\| \leqslant \|T\|\,\|U\|.$$

Proof It is easy to check that UT is linear. If $x \in X$, then

$$\|(UT)(x)\| = \|U(Tx)\| \leqslant \|U\| \, \|Tx\| \leqslant \|U\| \, \|T\| \, \|x\|.$$

□

♦ Definition 5.12 (Identity Operator, Invertible Operator). *We define the* identity operator *in* $\mathcal{B}(X)$ *by*

$$I(x) = x \text{ for every } x \in X.$$

We say that $T \in \mathcal{B}(X)$ *is* invertible *with inverse* T^{-1} *iff there exists* $T^{-1} \in \mathcal{B}(X)$ *such that* $TT^{-1} = T^{-1}T = I$.

Note that the condition $T^{-1} \in \mathcal{B}(X)$ requires T^{-1} to be bounded as well as being the set-theoretic inverse of T. Since $\|Ix\| = \|x\|$ for all x it is clear that $\|I\| = 1$.

We give an example of an operator and its inverse:

♦ Example 5.13. *Let* $T\colon (C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$ *be such that* $(Tf)(t) = (t^2+2)f(t)$. *If we define the operator* T^{-1} *by*

$$(T^{-1}g)(t) = \frac{g(t)}{t^2+2},$$

then we see that $T^{-1}\colon (C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$ *and that* $T^{-1}T = I$. *That is,* T^{-1} *is the inverse of* T. *But we must also check that it is bounded. It is, for* $|(T^{-1}g)(t)| \leqslant \frac{1}{2}|g(t)|$ *and thus* $\|T^{-1}g\|_\infty \leqslant \frac{1}{2}\|g\|_\infty$.

♦ Example 5.14. *Let* $T\colon (C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$ *be given by* $(Tf)(t) = (3t^2-1)f(t)$. *Then* T *is not invertible (in our sense), since it has no bounded inverse.*

♦ Proposition 5.15. *Let* $T, U \in \mathcal{B}(X)$. *Then* TU *is invertible iff* T *and* U *are both invertible. The inverse of* TU, *if it exists, is* $U^{-1}T^{-1}$.

Proof ($\Leftarrow$): If both T and U are invertible, then the product $U^{-1}T^{-1}$ must be bounded. It is the inverse, since $U^{-1}T^{-1}TU = I$.

($\Rightarrow$): If TU is invertible, and V is its inverse, then $(TU)V = I = T(UV)$. Thus $UV = T^{-1}$. (Note that right inverses are always equal to left inverses.) Similarly for U^{-1}. □

♦ Lemma 5.16. *Let* $T \in \mathcal{B}(X)$. *Then* T *is invertible iff both the following hold:*

(1) T *is a surjection.*
(2) *There exists* $c > 0$ *such that* $\|Tx\| \geqslant c\|x\|$ *for every* $x \in X$.

Proof ($\Rightarrow$): If T is invertible, then T must be surjective. As T^{-1} is bounded we must have for some $c' > 0$ that $\|T^{-1}y\| \leqslant c'\|y\|$ for every $y \in X$. So, taking $y = Tx$ we get $\|x\| \leqslant c'\|Tx\|$; i.e., $\|Tx\| \geqslant c\|x\|$ where $c = 1/c'$.

($\Leftarrow$): If $\|Tx\| \geqslant c\|x\|$ for every $x \in X$, then

$$\|Tx - Tx'\| \geqslant c\|x - x'\|.$$

This means that if $x \neq x'$, then $Tx \neq Tx'$, so T is injective. T is then bijective since (by hypothesis) it is a surjection. We must now show that the set-theoretic inverse T^{-1} is linear and bounded. For linearity, choose $y, y' \in X$ and let x, x' be the unique elements of X such that

$$Tx = y \quad \text{and} \quad Tx' = y'.$$

Then, since T is linear,

$$y + y' = T(x + x') \qquad \text{giving} \qquad T^{-1}(y + y') = T^{-1}y + T^{-1}y'. \tag{5.7}$$

For a scalar λ,

$$\lambda y = T(\lambda x) \qquad \text{so} \qquad T^{-1}(\lambda y) = \lambda x = \lambda T^{-1}(y). \tag{5.8}$$

The equations (5.7) and (5.8) imply that T^{-1} is linear. Finally, T^{-1} is bounded, for

$$\|T^{-1}y\| = \|x\| \leqslant \tfrac{1}{c}\|Tx\| = \tfrac{1}{c}\|y\|.$$

□

♦ Proposition 5.17. *Let $(X, \|\cdot\|)$ be a Banach space, and let $T \in \mathcal{B}(X)$ with $\|T\| < 1$. Then $I - T$ is invertible and*

$$(I-T)^{-1} = I + T + T^2 + T^3 + \cdots \quad .$$

This series converges in $\mathcal{B}(X)$.

Proof We first prove that the series does actually converge in $\mathcal{B}(X)$. First, we have, if $n > m$:

$$\begin{aligned}\left\|\sum_{k=0}^{n} T^k - \sum_{k=0}^{m} T^k\right\| &= \left\|\sum_{k=m+1}^{n} T^k\right\| \leqslant \sum_{k=m+1}^{n} \|T^k\| \\ &\leqslant \sum_{k=m+1}^{n} \|T\|^k \leqslant \sum_{k=m+1}^{\infty} \|T\|^k \\ &= \frac{\|T\|^{m+1}}{1-\|T\|},\end{aligned}$$

summing a geometric series. Then, given $\varepsilon > 0$, we can choose n_0 such that

$$\frac{\|T\|^{n_0+1}}{1-\|T\|} < \varepsilon.$$

This means $\left(\sum\limits_{k=0}^{n} T^k\right)_n$ is Cauchy in $(\mathcal{B}(X), \|\cdot\|)$. Since $\mathcal{B}(X)$ is complete by Theorem 5.10, this sequence of partial sums converges. Now we set

$$U = \lim_{n\to\infty} \sum_{k=0}^{n} T^k$$

and check that this is the inverse of $I - T$. So,

$$\begin{aligned}U(I-T) &= \big(U - (I + T + T^2 + \cdots + T^n) + (I + T + T^2 + \cdots + T^n)\big)\big(I-T\big) \\ &= \big(U - (I + T + T^2 + \cdots + T^n)\big)\big(I-T\big) + I - T^{n+1}\end{aligned}$$

and hence

$$\begin{aligned}\|U(I-T) - I\| &\leqslant \big\|\big(U - (I + T + T^2 + \cdots + T^n)\big)\big(I-T\big)\big\| + \|T^{n+1}\| \\ &\leqslant \big\|\big(U - (I + T + T^2 + \cdots + T^n)\big)\big\|\big\|\big(I-T\big)\big\| + \|T\|^{n+1} \to 0.\end{aligned}$$

This means $\|U(I-T) - I\| = 0$; that is, $U(I-T) = I$. The proof that U is also the right inverse of $I - T$ is similar. □

The Spectrum of an Operator

From this point onwards we will take $(X, \|\cdot\|)$ to be a complex Banach space and consider operators in $\mathcal{B}(X)$.

♦ Definition 5.18 (Eigenvalue, Eigenfunction, Spectrum). *A number $\lambda \in \mathbb{C}$ is called an* eigenvalue *of $T \in \mathcal{B}(X)$ with corresponding* eigenvector *(or* eigenfunction*) $x \in X$ iff*

$$(T - \lambda I)x = 0 \quad \text{and} \quad x \neq 0.$$

For a given λ, the set of all eigenvectors corresponding to it is called the eigenspace *of λ (and is easily checked to be a subspace of X). The set of all eigenvalues of T is called the* point spectrum *of T. The* spectrum *$\sigma(T)$ of T is the set*

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ is not invertible}\}.$$

We note that if λ is an eigenvalue of T, then $T - \lambda I$ annihilates its eigenspace (of nonzero vectors), so cannot be injective. Thus $T - \lambda I$ is not invertible and λ is in the spectrum of T. It is not always the case that the spectrum is contained in the set of eigenvalues, though in the finite-dimensional case it is. We have the following three possibilities for $\lambda \in \sigma(T)$:

(1) $T - \lambda I$ is not injective (λ is an eigenvalue);
(2) $T - \lambda I$ is injective but not surjective;
(3) $T - \lambda I$ is a bijection but $(T - \lambda I)^{-1}$ is not bounded.

Note that if $\dim(X) < \infty$, then (1) and (2) are equivalent and correspond exactly to $\lambda \in \sigma(T)$. In a Banach space, (3) is in fact impossible, though this is quite hard to prove. For finite dimensional spaces, the theory of eigenvalues and the spectrum does not depend on the norms and is covered in the basic linear algebra course. Things are rather more subtle for infinite dimensional spaces where the rôle of the norm is crucial.

♦ EXAMPLE 5.19. *Let* $T\colon (C[0,1], \|\cdot\|_\infty) \to (C[0,1], \|\cdot\|_\infty)$ *be given by* $(Tf)(t) = (t^2 - 3)f(t)$. *Then we have that* $(T - \lambda I)f(t) = (t^2 - 3 - \lambda)f(t)$. *If* $(t_0^2 - 3 - \lambda) = 0$ *for some* $t_0 \in [0,1]$, *then* $T - \lambda I$ *is not surjective since* $(T - \lambda I)f(t_0) = 0$ *for every* $f \in C[0,1]$. *Now if* $\lambda \in [-3,-2]$, *then such a* t_0 *does exist, so* $[-3,-2] \subset \sigma(T)$. *In fact,* $[-3,-2] = \sigma(T)$, *since for* $\lambda \notin [-3,-2]$,

$$(T - \lambda I)^{-1} g(t) = \frac{g(t)}{t^2 - 3 - \lambda},$$

and this is bounded.

♦ EXAMPLE 5.20. *Let* $T\colon \big(l^\infty, \|\cdot\|_\infty\big) \to \big(l^\infty, \|\cdot\|_\infty\big)$ *be given by*

$$T\colon (x_1, x_2, \dots,) \mapsto \left(x_1, \frac{x_2}{2}, \frac{x_3}{2^2}, \dots, \frac{x_i}{2^{i-1}}, \dots\right).$$

Then the set $\Lambda = \left\{1, \frac{1}{2}, \frac{1}{2^2}, \dots\right\}$ *is a set of eigenvalues since*

$$T(0,\dots,0,\underset{\substack{\uparrow \\ n\text{th place}}}{1},0,\dots) = \frac{1}{2^{n-1}}(0,\dots,0,1,0,\dots).$$

If $\lambda \notin \Lambda$ *and* $\lambda \neq 0$, *then* $(T - \lambda I)(x_1, x_2, \dots) = \big((1-\lambda)x_1, (\frac{1}{2} - \lambda)x_2, \dots\big)$, *so*

$$(T - \lambda I)^{-1}(y_1, y_2, \dots) = \left(\frac{y_1}{1-\lambda}, \frac{y_2}{\frac{1}{2} - \lambda}, \dots\right).$$

This operator will be bounded if $\sup_{n=0,1,2,\dots}\left|\frac{1}{(1/2^n) - \lambda}\right| < \infty$, *which is the case if* $\lambda \notin \Lambda$ *and* $\lambda \neq 0$. *Thus the spectrum is* $\sigma(T) = \Lambda \cup \{0\}$.

♦ THEOREM 5.21. *Let* $T \in \mathcal{B}(X)$ *where* $\big(X, \|\cdot\|\big)$ *is a Banach space. Then*

(1) *$\sigma(T)$ is bounded; in fact if $\lambda \in \sigma(T)$ then $|\lambda| \leqslant \|T\|$;*
(2) *$\sigma(T)$ is a closed subset of $\mathbb{C}$;*
(3) *$\sigma(T) \neq \emptyset$.*

Proof (1): Let $|\lambda| > \|T\|$. We show that this would imply $\lambda \notin \sigma(T)$. Note that $(T - \lambda I) = -\lambda(I - \frac{1}{\lambda}T)$. Now,

$$\left\|\frac{T}{\lambda}\right\| = \frac{1}{|\lambda|}\|T\| < 1 \qquad \text{since } |\lambda| > \|T\|.$$

Hence, by Proposition 5.17 we have that

$$(T - \lambda I)^{-1} = -\frac{1}{\lambda}\left(I - \frac{T}{\lambda}\right)^{-1} \qquad \text{(since this inverse exists).}$$

So $T - \lambda I$ is invertible and thus $\lambda \notin \sigma(T)$.

(2): We show that the complement $\mathbb{C} \setminus \sigma(T)$ is open. Let $\lambda_0 \notin \sigma(T)$. We show that if λ is close enough to λ_0, then $\lambda \notin \sigma(T)$. Note that $T - \lambda_0 I$ is invertible. Also, identically,

$$(T - \lambda I) = (T - \lambda_0 I)\big(I - (T - \lambda_0 I)^{-1}(\lambda - \lambda_0)\big). \tag{5.9}$$

Again by Proposition 5.17 we have that

$$\text{if } |\lambda - \lambda_0| < \frac{1}{\|(T-\lambda_0 I)^{-1}\|}, \text{ then (5.9) is invertible.}$$

Then, by definition $\mathbb{C} \setminus \sigma(T)$ is open, so $\sigma(T)$ is closed.

(3): (Sketch Proof) Replacing λ by z in (5.9) and taking inverses, we get

$$(T-zI)^{-1} = \underbrace{\Big(I - (T-z_0 I)^{-1}(z-z_0)\Big)^{-1}}_{\Omega}(T-z_0 I)^{-1}.$$

If $z_0 \notin \sigma(T)$ then for any z close to z_0 we have a convergent power series expression (by Proposition 5.17) for the bracket Ω, thus

$$(T-zI)^{-1} = \Big(I + (T-z_0 I)^{-1}(z-z_0) + (T-z_0 I)^{-2}(z-z_0)^2 + \cdots\Big)(T-z_0 I)^{-1}.$$

Hence if $z_0 \notin \sigma(T)$ then $(T-zI)^{-1}$ has a convergent power series in z about z_0: this means that $(T-zI)^{-1}$ is analytic for $z \in \mathbb{C}$ near z_0. If we assume, for a contradiction, that $\sigma(T) = \emptyset$, then $(T-zI)^{-1}$ is analytic in $\mathbb{C}$. But by (1) we have that

$$(T-zI)^{-1} = -\frac{1}{z}\left(I - \frac{T}{z}\right)^{-1},$$

for large z, so the function $\|(T-zI)^{-1}\|$ is bounded. By an analogue of Liouville's theorem in complex analysis (which states that a function that is analytic and bounded on $\mathbb{C}$ is constant), we conclude that $(T-zI)^{-1}$ is constant, which it clearly is not. $\lightning$ Thus $\sigma(T) \neq \emptyset$. □

♦ Theorem 5.22 (Spectral Mapping Theorem). *Let $p(t)$ be a polynomial and let $T \in \mathcal{B}(X)$. Then*

$$\sigma\big(p(T)\big) = p\big(\sigma(T)\big).$$

Proof. Let $p(z) = c(z-z_1)(z-z_2)\cdots(z-z_n)$ where $c \neq 0$. Then

$$\begin{aligned} 0 \notin \sigma(p(T)) &\Leftrightarrow (p(T))^{-1} \text{ exists} \\ &\Leftrightarrow \big(c(T-z_1 I)(T-z_2 I)\cdots(T-z_n I)\big)^{-1} \text{ exists} \\ &\Leftrightarrow (T-z_i I)^{-1} \text{ exists for each } i = 1,\ldots,n \\ &\Leftrightarrow z_1, z_2, \ldots, z_n \notin \sigma(T) \\ &\Leftrightarrow 0 \notin p(\sigma(T)). \end{aligned} \tag{5.10}$$

To complete the proof,

$$\begin{aligned} \lambda \notin \sigma(p(T)) &\Leftrightarrow (p(T)-\lambda I)^{-1} \text{ exists i.e., } \big((p(T)-\lambda I) - 0I\big)^{-1} \text{ exists} \\ &\Leftrightarrow 0 \notin \sigma(p(T)-\lambda I) \\ &\Leftrightarrow 0 \notin p(\sigma(T)) - \lambda \quad \text{by (5.10)} \\ &\Leftrightarrow \lambda \notin p(\sigma(T)). \end{aligned}$$

□

CHAPTER 6

Inner Products and Hilbert Space

Basic Ideas

◆ Definition 6.1. *An* inner product space *is a vector space X over $\mathbb{R}$ or $\mathbb{C}$, together with real or complex inner product $\langle x, y\rangle$ defined for all $y, x \in X$, such that*

(1) $\langle \lambda x + \mu y, w\rangle = \lambda\langle x, w\rangle + \mu\langle y, w\rangle$ *for every $x, y, w \in X$ and all scalars μ, λ;*
(2) $\langle x, y\rangle = \overline{\langle y, x\rangle}$ *(where the bar denotes complex conjugation);*
(3) $\langle x, x\rangle$ *is real,* $\langle x, x\rangle \geqslant 0$, *and* $\langle x, x\rangle = 0$ *iff* $x = 0$.

We write $\big(X, \langle\,\cdot\,,\,\cdot\,\rangle\big)$ to denote the space and its inner product.

Note that (1) and (2) above imply that

$$\langle w, \lambda x + \mu y\rangle = \bar{\lambda}\langle w, x\rangle + \bar{\mu}\langle w, y\rangle,$$

and

$$\langle x + y, w + z\rangle = \langle x, z\rangle + \langle x, w\rangle + \langle y, z\rangle + \langle y, w\rangle.$$

We can easily turn X into a normed space by setting

$$\|x\| = \sqrt{\langle x, x\rangle}. \tag{6.1}$$

To prove that this is a norm, we need the following well-known and important result:

◆ Lemma 6.2 (Schwarz' Inequality). *Let $x, y \in X$ where X is an inner product space with normgiven by* (6.1). *Then*

$$\boxed{|\langle x, y\rangle| \leqslant \|x\|\,\|y\|. \qquad \text{(Schwarz' Inequality)}}$$

Proof Assume $y \neq 0$. Then for all scalars λ,

$$0 \leqslant \langle x - \lambda y, x - \lambda y\rangle = \langle x, x\rangle - \lambda\langle y, x\rangle - \bar{\lambda}\langle x, y\rangle + \lambda\bar{\lambda}\langle y, y\rangle.$$

In particular we can choose $\lambda = \langle x, y\rangle/\langle y, y\rangle$, which gives

$$0 \leqslant \langle x, x\rangle - \frac{\langle x, y\rangle\overline{\langle x, y\rangle}}{\langle y, y\rangle} - \frac{\overline{\langle x, y\rangle}\langle x, y\rangle}{\langle y, y\rangle} + \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle},$$

so

$$0 \leqslant \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2},$$

and rearranging gives the Schwarz inequality. □

◆ Definition 6.3 (Hilbert Space). *let $\big(X, \langle\,\cdot\,,\,\cdot\,\rangle\big)$ be an inner product space with norm on X is given by* (6.1). *Then we call $\big(X, \langle\,\cdot\,,\,\cdot\,\rangle\big)$* a Hilbert space *if X is complete when equipped with this norm.*

Examples of Hilbert Spaces

♦ EXAMPLE 6.4. *The simplest example of a Hilbert space is* $\mathbb{R}^N$ *with the usual scalar product*

$$\langle (x_1, x_2, \ldots, x_N), (y_1, y_2, \ldots, y_N) \rangle = \sum_{i=1}^{N} x_i y_i.$$

The norm is the $\|\cdot\|_2$*-norm.*

♦ EXAMPLE 6.5. *The space* l^2 *over either* $\mathbb{R}$ *(or alternatively* $\mathbb{C}$*)*

$$l^2 = \left\{ (x_1, x_2, \ldots) : x_i \in \mathbb{R}\,(\textit{or}\,\mathbb{C}) \textit{ and } \sum_{i=1}^{\infty} |x_i|^2 < \infty \right\}$$

is a complete inner product space with the inner product

$$\langle (x_1, x_2, \ldots), (y_1, y_2, \ldots) \rangle = \sum_{i=1}^{\infty} x_i \overline{y_i}.$$

♦ EXAMPLE 6.6. *The space* $C[0,1]$ *of continuous functions becomes an inner product space with*

$$\langle f, g \rangle = \int_0^1 f(t)\overline{g(t)}\,\mathrm{d}t \tag{6.2}$$

and the norm is the $\|\cdot\|_2$*-norm. This space is* not *complete.*

However the *completion* of this space is illustrated in the next example:

♦ EXAMPLE 6.7. *The space* $L^2[0,1]$ *of square-integrable functions is a Hilbert space with inner product*

$$\langle f, g \rangle = \int_0^1 f(t)\overline{g(t)}\,\mathrm{d}t \tag{6.3}$$

which gives the $\|\cdot\|_2$*-norm.*

We can extend this last inner product to include 'weighted' inner products.

♦ EXAMPLE 6.8. *Let* w *be a real positive continuous 'weight function' . Let*

$$H = \left\{ f \in L^2[0,1] : \int_0^1 w(u)|f(u)|^2\,\mathrm{d}u < \infty \right\}$$

and set

$$\langle f, g \rangle = \int_0^1 w(u) f(u)\overline{g(u)}\,\mathrm{d}u.$$

Then H *is a Hilbert space under this inner product.*

Orthonormal Systems

♦ LEMMA 6.9 (PARALLELOGRAM LAW). *For every* x, y *in an inner product space,*

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

Proof

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x,x\rangle + \langle y,y\rangle + \langle x,y\rangle + \langle y,x\rangle \\ &\quad + \langle x,x\rangle + \langle y,y\rangle - \langle x,y\rangle - \langle y,x\rangle \\ &= 2\|x\|^2 + 2\|y\|^2.\end{aligned}$$

□

♦ DEFINITION 6.10 (ORTHOGONAL, ORTHONORMAL). *Let* $\big(H, \langle\,\cdot\,,\,\cdot\,\rangle\big)$ *be a Hilbert space. We say that* $x, y \in H$ *are* orthogonal *or* perpendicular *if* $\langle x, y\rangle = 0$. *A set of elements* $\{e_i\}_i$ *in* H *is called* orthonormal *if* $\langle e_i, e_i\rangle = \|e_i\|^2 = 1$ *for all* i *and* $\langle e_i, e_j\rangle = 0$ *if* $i \neq j$.

♦ DEFINITION 6.11 (SPAN, COMPLETE ORTHONORMAL SYSTEM). *The* span *of a set* $\{x_i\}$ *of vectors in* H, *written* span $\{x_i\}$, *is the set of all finite linear combinations of the* x_i, *that is* $\{\sum_i^n a_i e_i : a_i \in \mathbb{R}\}$. *If* $\{e_i\}$ *is an orthonormal set and* span $\{e_i\}$ *is dense in* H, *then we call* $\{e_i\}$ *a* complete orthonormal system.

(Note that 'complete' in this context is not the same as the completeness property of a normed space.)

♦ EXAMPLE 6.12 (TWO COMPLETE ORTHONORMAL SYSTEMS). *The following orthonormal systems underly the theory of Fourier series:*

(1) *The set* $\{1\} \cup \left\{\sqrt{2}\cos(2\pi jt)\right\}_{j=1}^{\infty} \cup \left\{\sqrt{2}\sin(2\pi jt)\right\}_{j=1}^{\infty}$ *is a complete orthonormal system in* $L^2[0,1]$.
(2) *The set* $\left\{e^{2\pi int}\right\}_{n=-\infty}^{\infty}$ *is a complex complete orthonormal system.*

Here is a useful check for a vector to be zero depending on orthogonality.

♦ LEMMA 6.13. *Let* S *be a dense subset of a Hilbert space* H. *Let* $x \in H$ *and suppose that* $\langle x, s\rangle = 0$ *for all* $s \in S$. *Then* $x = 0$. *In particular, if* $\{e_i\}$ *is an complete orthonormal set and* $\langle x, e_i\rangle = 0$ *for all* i *then* $x = 0$.

Proof If S is dense in H, then we can find $(x_n)_n \in S$ such that $x_n \to x$. Then,

$$\begin{aligned} 0 \leqslant |\langle x, x\rangle| &= |\langle x - x_n, x\rangle + \langle x_n, x\rangle| \leqslant |\langle x - x_n, x\rangle + 0| \qquad (\text{since } x_n \in S) \\ (\text{by Schwarz}) \qquad &\leqslant \|x - x_n\|\,\|x\| \to 0. \end{aligned}$$

So $\langle x, x\rangle = 0$ and hence $x = 0$. □

♦ THEOREM 6.14 (PARSEVAL'S EQUALITY). *Let* $\{e_i\}$ *be a (countable) complete orthonormal system in the Hilbert space* H. *Then if* $x \in H$, *we have*

$$x = \sum_{i=1}^{\infty} \langle x, e_i\rangle e_i,$$

with this series convergent in H. *Also,*

$$\|x\|^2 = \sum_{i=1}^{\infty} \big|\langle x, e_i\rangle\big|^2. \qquad (\text{Parseval's Equality})$$

Proof Let $x \in H$. Write

$$x_n = \sum_{i=1}^{n} \langle x, e_i\rangle e_i.$$

Then

$$\langle x - x_n, e_i\rangle = \langle x, e_i\rangle - \langle x_n, e_i\rangle = 0 \qquad \text{where } i = 1, 2, \ldots, n.$$

This gives $\langle x - x_n, x_n\rangle = 0$ since x_n is a linear combination of $\{e_1, \ldots, e_n\}$. Then

$$\begin{aligned} \|x\|^2 &= \langle x, x\rangle \\ &= \langle (x - x_n) + x_n, (x - x_n) + x_n\rangle \\ &= \|x - x_n\|^2 + \|x_n\|^2 + \underbrace{\langle x - x_n, x_n\rangle}_{=0} + \underbrace{\langle x_n, x - x_n\rangle}_{=0}; \end{aligned}$$

hence $\|x\|^2 \geqslant \|x_n\|^2$. Thus for each n we have

$$\sum_{i=1}^{n} \big|\langle x, e_i\rangle\big|^2 = \|x_n\|^2 \leqslant \|x\|^2,$$

which means that $\sum_{i=1}^{n} |\langle x, e_i\rangle|^2$ is convergent.

Moreover, suppose $\varepsilon > 0$, then if n and m are large enough,

$$\|x_n - x_m\|^2 = \left\| \sum_{i=m+1}^{n} \langle x, e_i\rangle e_i \right\|^2 = \sum_{i=m+1}^{n} \big|\langle x, e_i\rangle\big|^2 < \varepsilon,$$

since $\sum_{i=1}^{\infty} |\langle x, e_i\rangle|^2$ converges. Since H is complete, there must be $y \in H$ such that $x_n \to y$. Hence, for each i,

$$\langle x - y, e_i\rangle = \underbrace{\langle x - x_n, e_i\rangle}_{=0 \text{ if } n \geqslant i} + \underbrace{\langle x_n - y, e_i\rangle}_{\to 0 \text{ by Schwarz}} \to 0.$$

Thus $\langle x - y, e_i\rangle = 0$ for every i. Since $\{e_i\}$ is complete, the previous lemma gives $x = y$, and

$$\|x\|^2 = \lim_{n\to\infty} \|x_n\|^2 = \sum_{i=1}^{\infty} \big|\langle x, e_i\rangle\big|^2.$$

□

If we drop the requirement in Theorem 6.14 that the set $\{e_i\}$ is complete (i.e., that its span is dense in H), we get *Bessel's Inequality*:

$$\sum_{i=1}^{\infty} \big|\langle x, e_i\rangle\big|^2 \leqslant \|x\|^2. \qquad \text{(Bessel's Inequality)}$$

♦ Example 6.15 (Fourier Series). *Recall that*

$$\operatorname{span}\left\{e^{2\pi i n t}\right\}_{n=-\infty}^{\infty} \text{ is a complete o.n. system in } L^2[0,1].$$

Thus if $f \in L^2[0,1]$ *we get*

$$f(t) = \sum_{n=-\infty}^{\infty} \left(\int_0^1 f(t) e^{-2\pi i n t}\, \mathrm{d}t\right) e^{2\pi i n t},$$

$$\text{and} \quad \int_0^1 |f(t)|^2\, \mathrm{d}t = \sum_{n=-\infty}^{\infty} \left| \int_0^1 f(t) e^{-2\pi i n t}\, \mathrm{d}t \right|^2.$$

The following theorem is of theoretical significance and its proof provides a computational method for finding orthonormal bases.

♦ Theorem 6.16 (Gram-Schmidt Orthonormalisation). *Let* $\{x_1, x_2, \ldots\}$ *be a linearly independent set in* H *(i.e. if* $\sum_1^N \lambda_i x_i = 0$ *then* $\lambda_i = 0$ *for each* i*). Then we may find an orthonormal sequence* $\{e_1, e_2, \ldots\} \subset H$ *such that for every* $n = 1, 2, \ldots,$ *we have*

$$\operatorname{span}\{e_1, e_2, \ldots\} = \operatorname{span}\{x_1, x_2, \ldots\}.$$

Proof Let

$$e_1 = \frac{x_1}{\|x_1\|}.$$

Then define the remaining terms in the sequence by the formula

$$e_{n+1} = \frac{x_{n+1} - \sum_{i=1}^{n} \langle x_{n+1}, e_i\rangle e_i}{\|x_{n+1} - \sum_{i=1}^{n} \langle x_{n+1}, e_i\rangle e_i\|}.$$

Clearly $\|e_i\| = 1$ for each i, and $\langle e_n, e_i\rangle = 0$ for $i \neq n$. □

♦ DEFINITION 6.17 (BASIS). *An orthonormal set* $\{e_i\}$ *is called a* basis *for* H *if for every* $x \in H$ *we may find scalars* $a_1, a_2, \ldots$ *such that*

$$x = \sum_{i=1}^{\infty} a_i e_i,$$

and the sum converges (i.e., $\sum_i |a_i|^2 < \infty$*).*

♦ COROLLARY 6.18 (TO GRAM-SCHMIDT THEOREM). *If* H *is a Hilbert space and there exists a sequence* $(x_n)_n$ *in* H *such that* $\operatorname{span}\{x_1, x_2, \ldots\}$ *is dense in* H*, then* H *has an orthonormal basis.*

Proof The Gram-Schmidt process gives an orthonormal set $\{e_i\}$ with $\operatorname{span}\{e_i\} = \operatorname{span}\{x_i\}$ which must be dense in H. By Theorem 6.14 the set $\{e_i\}$ is a basis. □

CHAPTER 7

The Spectral Theorem

The aim of this chapter is to prove the spectral theorem for compact self-adjoint operators on an infinite dimensional complex Hilbert space H. This should be thought if as being analogous to diagonalising symmetric matrices by finding a suitable basis.

♦ DEFINITION 7.1 (SELF-ADJOINT OPERATOR). *An operator* $T \in \mathcal{B}(H)$ *is* self-adjoint *or* Hermitian *iff*

$$\langle Tx, y\rangle = \langle x, Ty\rangle \qquad \text{for all } x, y \in H.$$

(Self-adjoint operators correspond to symmetric or Hermitian matrices.) Note that if T is self-adjoint then $\langle Tx, x\rangle$ is real, since $\langle Tx, x\rangle = \langle x, Tx\rangle = \overline{\langle Tx, x\rangle}$.

♦ EXAMPLE 7.2. *The following operators are easily seen to be self-adjoint.*

(1) $T\colon l^2 \to l^2$ *given by* $T(x_1, x_2, x_3, \ldots) = (x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \ldots)$.
(2) $T\colon L^2[0,1] \to L^2[0,1]$ *given by* $(Tf)(t) = f(t)(t^3 + 3)$.
(3) $T\colon L^2[0,1] \to L^2[0,1]$ *given by* $(Tf)(t) = \int_0^1 f(u)(t^2 + u^2)\,\mathrm{d}u$.

♦ LEMMA 7.3. *If* T *is self-adjoint then*

(1) *the eigenvalues of* T *are all real,*
(2) *eigenvectors corresponding to different eigenvalues are orthogonal.*

Proof (1): Suppose that $Tx = \lambda x$. Then

$$\lambda\langle x, x\rangle = \langle \lambda x, x\rangle = \langle Tx, x\rangle = \langle x, Tx\rangle = \langle x, \lambda x\rangle = \bar{\lambda}\langle x, x\rangle.$$

If $x \neq 0$ then $\lambda = \bar{\lambda}$, and hence λ is real.

(2): Suppose that $Tx = \lambda x$ and $Ty = \mu y$. Then

$$\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle Tx, y\rangle = \langle x, Ty\rangle = \langle x, \mu y\rangle = \bar{\mu}\langle x, y\rangle = \mu\langle x, y\rangle,$$

since μ is real. Thus if $\lambda \neq \mu$ then $\langle x, y\rangle = 0$. □

It is useful to have an alternative form for the norm of a self-adjoint operator:

♦ LEMMA 7.4. *Let* T *be self-adjoint. Then* $\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|$.

Proof Write $m = \sup_{\|x\|=1} |\langle Tx, x\rangle|$. First note that if $\|x\| = 1$ then

$$|\langle Tx, x\rangle| \leqslant \|Tx\|\,\|x\| \leqslant \|T\|\,\|x\|\,\|x\| = \|T\|$$

using Schwarz' inequality; so $m \leqslant \|T\|$.

For the opposite inequality, let x be such that $\|x\| = 1$ and $Tx \neq 0$, and write $y = Tx/\|Tx\|$ so that $\|y\| = 1$. Then $\langle Tx, y\rangle = \langle Tx, Tx\rangle/\|Tx\| = \|Tx\|$. By expanding the inner product and noting that $m = \sup_{\|z\|\neq 0} |\langle Tz, z\rangle|/\|z\|^2$, we have

$$\begin{aligned}\|Tx\| = \langle Tx, y\rangle &= \tfrac{1}{2}\big(\langle Tx, y\rangle + \langle y, Tx\rangle\big) = \tfrac{1}{2}\big(\langle Tx, y\rangle + \langle Ty, x\rangle\big)\\ &= \tfrac{1}{4}\big(\langle T(x+y), (x+y)\rangle - \langle T(x-y), (x-y)\rangle\big)\\ &\leqslant \tfrac{1}{4}m\left(\|x+y\|^2 + \|x-y\|^2\right)\\ &= \tfrac{1}{4}m \times 2(\|x\|^2 + \|y\|^2) = m,\end{aligned}$$

using the parallelogram law. Thus $\|T\| \leqslant m$ as required. □

We now turn our attention to compact operators, which may be thought of as operators with images that are 'nearly finite-dimensional'.

♦ Definition 7.5 (Compact Operator). *An operator $T \in \mathcal{B}(H)$ is* compact *if for every bounded sequence (x_n) the sequence (Tx_n) has a convergent subsequence.*

In a finite dimensional space every bounded sequence has a convergent subsequence so any operator on a finite dimensional space is necessarily compact. Checking that operators on infinite dimensional spaces are compact can be more awkward. For the Hilbert space L^2 the following proposition is helpful.

♦ Proposition 7.6. *Let $T\colon L^2[0,1] \to L^2[0,1]$ be such that the following 'equicontinuity condition' holds: for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|Tf(x) - Tf(y)| < \varepsilon$ whenever $|x - y| < \delta$ for all $f \in L^2[0,1]$ such that $\|f\|_2 \leqslant 1$. Then T is compact.*

Proof This follows from the Arzelà-Ascoli theorem which gives a condition for a family of continuous functions to have a convergent subsequence. The proof of this is technical but not particularly difficult, see, for example, the book by Griffel. □

♦ Example 7.7. *The following operators are compact:*

(1) $T\colon l^2 \to l^2$ *given by* $T(x_1, x_2, x_3, \ldots) = (x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \ldots)$.
(2) $T\colon L^2[0,1] \to L^2[0,1]$ *given by* $(Tf)(t) = \int_0^1 f(u)(t^2 + u^2)\,\mathrm{d}u$.
(3) $T\colon L^2[0,1] \to L^2[0,1]$ *given by* $(Tf)(t) = \int_0^t f(u)(u^2 + 1)\,\mathrm{d}u$.

Proposition 7.6 may be sued to verify that (2) and (3) are compact.

♦ Lemma 7.8. *If T is compact and self-adjoint and $r > 0$, then there are at most finitely many eigenvalues (counted according to multiplicity of independent eigenvectors) such that $|\lambda| \geqslant r$.*

Proof Suppose to the contrary. Then we may find an infinite sequence of eigenvectors (e_i) and corresponding eigenvalues λ_i all with $|\lambda_i| \geqslant r$. For each distinct eigenvalue we may select a basis of the corresponding eigenspace consisting of orthonormal eignevectors (using the Gram-Schmidt process), so together with Lemma 7.3 (2) this means we can assume that the infinite sequence (e_i) is orthonormal. Then for $i \neq j$ we have

$$\langle Te_i, Te_j\rangle = \langle \lambda_i e_i, \lambda_j e_j\rangle = \lambda_i \bar{\lambda}_j \langle e_i, e_j\rangle = 0,$$

so expanding,

$$\begin{aligned}\|Te_i - Te_j\|^2 = \langle Te_i - Te_j, Te_i - Te_j\rangle &= \|Te_i\|^2 - \langle Te_i, Te_j\rangle - \langle Te_j, Te_i\rangle + \|Te_j\|^2 \\ &= |\lambda_i|^2 - 0 - 0 + |\lambda_j|^2 \geqslant 2r^2.\end{aligned}$$

Thus $(Te_i)_i$ cannot have a convergent subsequence (since no subsequence can be a Cauchy sequence), which contradicts the compactness of T. □

The key property that we need in order to prove the spectral theorem is that the operator has at least one non-zero eignevalue. We show this (and more) in the following lemma.

♦ Lemma 7.9. *Let T be a compact self-adjoint operator on H. Then either $\|T\|$ or $-\|T\|$ is an eigenvalue of T.*

Proof Assume that $\|T\| > 0$ (otherwise the result is trivial). Writing $m = \sup_{\|x\|=1} |\langle Tx, x\rangle| = \|T\|$ by Lemma 7.4, we show that T has an eigenvalue λ with $|\lambda| = m$. Noting that $\langle Tx, x\rangle$ is real, suppose that $\sup_{\|x\|=1} \langle Tx, x\rangle = m$. (A symmetrical argument applies if $\inf_{\|x\|=1} \langle Tx, x\rangle = -m$.) Choose a sequence (x_n) with $\|x_n\| = 1$ such that $\langle Tx_n, x_n\rangle \to m$. Then expanding

$$\begin{aligned}\|Tx_n - mx_n\|^2 &= \langle Tx_n - mx_n, Tx_n - mx_n\rangle \\ &= \|Tx_n\|^2 - 2m\langle Tx_n, x_n\rangle + m^2\|x_n\|^2 \\ &\leqslant \|T\|^2\|x_n\|^2 - 2m\langle Tx_n, x_n\rangle + m^2\|x_n\|^2 \\ &\leqslant 2m^2 - 2m\langle Tx_n, x_n\rangle \to 0 \qquad \text{as } n \to \infty.\end{aligned}$$

Since T is compact, we may assume, by taking a subsequence if necessary, that $Tx_n \to y$ for some $y \neq 0$. Then

$$mx_n = Tx_n + (mx_n - Tx_n) \to y + 0.$$

Thus $x_n \to y/m$ and by continuity $Tx_n \to (Ty)/m$ so $(Ty)/m = y$ giving that m is an eigenvalue with eigenvector y. □

We put together the above properties to get the spectral theorem.

♦ Theorem 7.10 (Spectral theorem for compact self-adjoint operators). *Let T be a compact self-adjoint operator on H. Then there is an orthonormal sequence (e_n) of eigenvectors in H, with $Te_n = \lambda_n e_n$, say, with λ_n real, $\lambda_n \neq 0$ and $\lambda_n \to 0$, such that for all $x \in H$ we may write*

$$x = \sum_{n=1}^{\infty} a_n e_n + y,$$

where $a_n = \langle x, e_n\rangle$, with $\sum_{n=1}^{\infty} |a_n|^2 < \infty$, and $y \in H$ satisfying $\langle y, e_n\rangle = 0$ for all n and $Ty = 0$. In particular

$$Tx = \sum_{n=1}^{\infty} a_n \lambda_n e_n.$$

Proof For each eigenvalue $\lambda \neq 0$ let H_λ be the eigenspace $H_\lambda = \{x : Tx = \lambda x\}$. By Lemma 7.8, each H_λ has finite dimension so has a finite orthonormal basis (by the Gram-Schmidt process). Let $\{e_n\}$ be the aggregate of the basis vectors obtained in this way from all $\lambda \neq 0$; by Lemma 7.3 (2) the e_n are all orthonormal (though they need not form a basis of H).

Suppose $Te_n = \lambda_n e_n$ for each n. As T is self-adjoint the λ_n are real (Lemma 7.3 (1)) and since T is compact $\lambda_n \to 0$ (Lemma 7.8). Given $x \in H$, we may write

$$x = \sum_{n=1}^{\infty} a_n e_n + y \quad \text{where } a_n = \langle x, e_n\rangle$$

for some $y \in H$; since the e_n are orthonormal $\sum_{n=1}^{\infty} |a_n|^2 < \infty$ and the series converges by Bessel's inequality. Taking inner products with e_n we see that $\langle y, e_n\rangle = 0$ for all n.

To show that $Ty = 0$, let $H_0 = \{y : \langle y, e_n\rangle = 0 \text{ for all } n\}$. If $H_0 = \{0\}$ there is nothing left to show. Otherwise, H_0 is closed so may itself be regarded as a Hilbert space. Moreover, if $y \in H_0$ then

$$\langle Ty, e_n\rangle = \langle y, Te_n\rangle = \langle y, \lambda_n e_n\rangle = \lambda_n\langle y, e_n\rangle = 0,$$

so $Ty \in H_0$. Thus T maps H_0 into itself, so we may regard T as a compact self-adjoint operator on H_0. By Lemma 7.9 there is an eigenvalue λ with $|\lambda| = \sup_{0 \neq x \in H_0} \frac{\|Tx\|}{\|x\|}$ and a (nonzero) eigenvector $w \in H_0$ such that $Tw = \lambda w$. But $w \in H_0$ is orthogonal to all eigenvectors with non-zero eigenvalues, so $\lambda = 0$, and hence $Ty = 0$ for all $y \in H_0$.

We conclude that

$$Tx = \sum_{n=1}^{\infty} a_n \lambda_n e_n.$$

To check this formally, note that

$$\left\| Tx - \sum_{n=1}^{N} a_n \lambda_n e_n \right\| = \left\| T\Big(x - \sum_{n=1}^{N} a_n e_n\Big) \right\| = \left\| T\Big(y + \sum_{n=N+1}^{\infty} a_n e_n\Big) \right\|$$

$$= \left\| T\Big(\sum_{n=N+1}^{\infty} a_n e_n\Big)\right\| \leqslant \|T\| \left\|\sum_{n=N+1}^{\infty} a_n e_n\right\| = \|T\| \sum_{n=N+1}^{\infty} |a_n|^2 \to 0. \quad \square$$

♦ Corollary 7.11. *Let T be a compact self-adjoint operator on H. Then H has an orthonormal basis of eigenvectors of T.*

Proof Take $\{e_n\}$ as in the spectral theorem together with an orthonormal basis of H_0. □

If T is compact self-adjoint and $x = \sum_{n=1}^{\infty} a_n e_n + y$, then applying the spectral theorem repeatedly gives that

$$T^k x = \sum_{n=1}^{\infty} \lambda_n^k a_n e_n.$$

In particular, if λ_1 is the unique largest eigenvalue (in absolute value), then

$$\frac{T^k x}{\lambda_1^k} \to a_1 e_1 \qquad \text{and} \qquad \frac{T^k x}{\|T^k x\|} \to e_1,$$

giving an iterative method of finding the eigenvector corresponding to the largest eigenvalue λ_1.

If T is compact self-adjoint and $\lambda \neq 0$ is not an eigenvalue, then we may solve $(T - \lambda I)x = w$ given $w = \sum_{n=1}^{\infty} a_n e_n + y$, by noting that

$$(T - \lambda I)\Big(\sum_{n=1}^{\infty} \frac{a_n e_n}{\lambda_n - \lambda} - \frac{y}{\lambda}\Big) = \sum_{n=1}^{\infty} a_n e_n + y.$$

This solution is in H since

$$\left\|\sum_{n=1}^{\infty} \frac{a_n e_n}{\lambda_n - \lambda}\right\| = \sum_{n=1}^{\infty} \frac{|a_n|^2}{|\lambda_n - \lambda|^2} \leqslant c \sum_{n=1}^{\infty} |a_n|^2$$

where $c = \max_n 1/|\lambda_n - \lambda|^2 < \infty$.

Sturm-Liouville Theory

Sturm-Liouville theory applies the spectral theorem to the eigenvalue structure of linear differential equations.

We illustrate this with the eigenvalue problem for the homogeneous differential equation on the interval $[0, 1]$ with boundary conditions:

$$y'' + \mu y = 0 \qquad\qquad y(0) = y(1) = 0 \tag{7.1}$$

where μ is a number, and its inhomogeneous counterpart:

$$y'' + f = 0 \qquad\qquad y(0) = y(1) = 0 \tag{7.2}$$

for a given function f. We proceed using several steps.

(**A**): Define a continuous function g by

$$g(s, t) = \begin{cases} s(1 - t) & 0 \leqslant s \leqslant t \\ t(1 - s) & t \leqslant s \leqslant 1 \end{cases} \qquad \text{where } s, t \in [0, 1]. \tag{7.3}$$

Then g is called the *Green's function* for the problem (7.2). We check that integrating $f(s) \in C[0, 1]$ against $g(s, t)$ solves (7.2). Let

$$\begin{aligned} y(t) &= \int_0^1 g(s, t) f(s)\,\mathrm{d}s \\ &= \int_0^t s(1 - t) f(s)\,\mathrm{d}s + \int_t^1 t(1 - s) f(s)\,\mathrm{d}s. \end{aligned}$$

Clearly $y(0) = y(1) = 0$. Differentiating:

$$\begin{aligned} y'(t) &= t(1-t)f(t) - \int_0^t sf(s)\,\mathrm{d}s - t(1-t)f(t) + \int_t^1 (1-s)f(s)\,\mathrm{d}s \\ &= -\int_0^t sf(s)\,\mathrm{d}s + \int_t^1 (1-s)f(s)\,\mathrm{d}s, \\ y''(t) &= -tf(t) - (1-t)f(t) \\ &= -f(t). \end{aligned}$$

(**B**): With g as in (7.3) we define the *Green's function operator* G by

$$(Gf)(t) = \int_0^1 g(s,t)f(s)\,\mathrm{d}s. \tag{7.4}$$

We also have the differential operator L, with

$$(Ly)(t) = -y''(t).$$

We need to consider the spaces on which these operators act. As well as using $C[0,1]$ and $L^2[0,1]$, we introduce $C^2_B[0,1]$ for the space of functions f on $[0,1]$ with continuous second derivative which satisfy the boundary conditions $f(0) = f(1) = 0$.

It follows from (**A**) that $G\colon C[0,1] \to C^2_B[0,1]$ and also that $LG\colon C[0,1] \to C[0,1]$ is the identity on $C[0,1]$ (since $LG(f) = -(Gf)'' = f$ for $f \in C[0,1]$).

Moreover $L\colon C^2_B[0,1] \to C[0,1]$ is injective (since if $0 = Ly = -y''$ then $y(t) = at + b$, so if $y(0) = y(1) = 0$ then $y(t) = 0$), and surjective (since LG, and thus L, maps onto the whole of $C[0,1]$).

Thus $L\colon C^2_B[0,1] \to C[0,1]$ is invertible with inverse G, that is $LG = I_{C[0,1]}$ and $GL = I_{C^2_B[0,1]}$, and we have the picture:

$$C[0,1] \underset{L}{\overset{G}{\rightleftarrows}} C^2_B[0,1].$$

Note that G is a much easier operator to work with than L since G is bounded and L is not.

(**C**): We may also regard G as a bounded operator on the Hilbert space $L^2[0,1]$. We claim that

$$G\colon L^2[0,1] \longrightarrow C[0,1] \subset L^2[0,1].$$

To see this note that if $f \in L^2[0,1]$ then by Cauchy's inequality

$$\left|Gf(t)\right| = \left|\int_0^1 g(s,t)f(s)\,\mathrm{d}s\right| \leqslant \left(\int_0^1 g(s,t)^2\,\mathrm{d}s\right)^{1/2}\left(\int_0^1 f(s)^2\,\mathrm{d}s\right)^{1/2} \tag{7.5}$$

and

$$\begin{aligned} \left|Gf(t) - Gf(u)\right| &= \left|\int_0^1 \big(g(s,t) - g(s,u)\big)f(s)\,\mathrm{d}s\right| \\ &\leqslant \left(\int_0^1 \left|g(s,t) - g(s,u)\right|^2\mathrm{d}s\right)^{1/2}\left(\int_0^1 f(s)^2\,\mathrm{d}s\right)^{1/2}, \end{aligned} \tag{7.6}$$

so $Gf(t)$ is bounded and continuous. From (7.5)

$$\|Gf\|_2 \leqslant \|Gf\|_\infty \leqslant \left(\int_0^1 g(s,t)^2\,\mathrm{d}s\right)^{1/2}\|f\|_2.$$

♦ THEOREM 7.12. *The map G, regarded as a bounded operator on $L^2[0,1]$, is compact and self-adjoint.*

Proof (Compact): If $\|f\|_2 \leqslant 1$ and $\varepsilon > 0$, then by (7.6),

$$\left|Gf(t) - Gf(u)\right| \leqslant \left(\int_0^1 \left|g(s,t) - g(s,u)\right|^2\mathrm{d}s\right)^{1/2} \leqslant \sup_{s\in[0,1]}\left|g(s,t) - g(s,u)\right| < \varepsilon$$

provided $|t-u|$ is small enough, since g is continuous. By the equicontinuity criterion for a compact operator (Proposition 7.6 above), G is compact.

(Self-adjoint): Since $g(s,t)$ is real, we have for $f, h \in L^2[0,1]$ that

$$\langle Gf, h\rangle = \int_0^1 Gf(t)\,\overline{h(t)}\,\mathrm{d}t = \int_0^1 \left(\int_0^1 g(s,t)f(s)\,\mathrm{d}s\right)\overline{h(t)}\,\mathrm{d}t = \int_0^1\int_0^1 f(s)\overline{g(s,t)h(t)}\,\mathrm{d}s\,\mathrm{d}t$$
$$= \int_0^1 f(s)\overline{\left(\int_0^1 g(s,t)h(t)\,\mathrm{d}t\right)}\,\mathrm{d}s = \langle f, Gh\rangle,$$

so G is self-adjoint. □

(**D**): We now apply the spectral theorem to the compact self-adjoint operator G on the Hilbert space $L^2[0,1]$.

♦ THEOREM 7.13. *We have the following:*

(1) *The number* 0 *is not an eigenvalue of* G.
(2) *The space* $L^2[0,1]$ *has an orthonormal basis* $\{f_n\}$ *of eigenfunctions of* G, *with* $Gf_n = \lambda_n f_n$, *where* λ_n *are real and* $|\lambda_n| \to 0$ *as* $n \to \infty$. *Moreover* $f_n \in C^2_B[0,1]$ *for all* n.

Proof (1): Suppose that $Gf = 0$ for some $f \in L^2[0,1]$. If $h \in C[0,1]$, then $0 = \langle Gf, h\rangle = \langle f, Gh\rangle$. But by (**B**), $\{Gh : h \in C[0,1]\} = C^2_B[0,1]$, which is a space that is $\|\cdot\|_2$-dense in $L^2[0,1]$. (This may be established by showing that $C^2_B[0,1]$ is dense in $C[0,1]$ which in turn is dense in $L^2[0,1]$ in the $\|\cdot\|_2$-norm.) Thus $f = 0$, so 0 is not an eigenvalue of G.

(2): Applying the spectral theorem to G on $L^2[0,1]$ the first part of (2) follows immediately, since by (1), $\{y : Ty = 0\} = \{0\}$ (in the notation used in the spectral theorem), so that the eigenvectors with non-zero eigenvalues form a basis.

If $f_n \in L^2[0,1]$ is an eigenfunction of G then

$$\lambda_n f_n = Gf_n \in C[0,1], \text{ since } G\colon L^2[0,1] \to C[0,1],$$

so $f_n \in C[0,1]$ since $\lambda_n \neq 0$, and

$$\lambda_n f_n = Gf_n \in C^2_B[0,1], \text{ since } G\colon C[0,1] \to C^2_B[0,1],$$

so $f_n \in C^2_B[0,1]$. □

(**E**): We now transfer these results to the differential operator $L\colon C^2_B[0,1] \to C[0,1]$.

We call μ an *eigenvalue* of L with *eigenfunction* f if $Lf = \mu f$ where $0 \neq f \in C^2_B[0,1]$. Note in particular that the eigenfunctions of L satisfy the boundary conditions.

♦ LEMMA 7.14. *The number* λ *is an eigenvalue of* $G\colon L^2[0,1] \to L^2[0,1]$ *with eigenfunction* f *if and only if* $\mu = 1/\lambda$ *is an eigenvalue of* $L\colon C^2_B[0,1] \to C[0,1]$ *with eigenfunction* f.

Proof (⇒): If $Gf = \lambda f$ with $0 \neq f \in L^2[0,1]$, then $f \in C^2_B[0,1]$ and $\lambda \neq 0$ by (**D**). Thus $f = LGf = L(\lambda f) = \lambda Lf$, so λ^{-1} is an eigenvalue of L.

(⇐): If $Lf = \mu f$ with $0 \neq f \in C^2_B[0,1]$, then $\mu \neq 0$ (since L is injective; see (**B**)) and $f = GLf = G(\mu f) = \mu Gf$, so μ^{-1} is an eigenvalue of G. □

(**F**): The Sturm-Liouville Theorem collects together these conclusions in the context of differential equations.

♦ THEOREM 7.15 (STURM-LIOUVILLE THEOREM). *The eigenvalue problem* (7.1), *which may be written as*

$$Ly = \mu y \qquad (y(0) = y(1) = 0),$$

has countably many eigenvalues $\mu_n \neq 0$, *all real, with* $|\mu_n| \to \infty$. *For each eigenvalue there is a unique eigenfunction (to within a scalar multiple)* $f_n \in C^2_B[0,1]$, *which may be taken to be a real-valued function, with* $Lf_n = \mu_n f_n$. *These eigenfunctions form an orthonormal basis of* $L^2[0,1]$.

Proof Most of this is immediate from (**D**) translated from G to L using the lemma in (**E**).

To see that we may assume that the eigenfunctions are real, note that if $\mu f = Lf = -f''$ then, since μ is real and L maps real functions to real functions, we may take the real and imaginary parts $\operatorname{Re} f$ and $\operatorname{Im} f$ as two separate *real* eigenfunctions, with $\mu\,(\operatorname{Re} f) = L(\operatorname{Re} f) = -(\operatorname{Re} f)''$ and $\mu\,(\operatorname{Im} f) = L(\operatorname{Im} f) = -(\operatorname{Im} f)''$.

Finally, to see that the eigenfunction for each eigenvalue μ is essentially unique, note that the *initial value* problem $\mu f = Lf = -f''$ with $f(0) = 0$ and $f'(0) = a$ has a unique solution. (One way to see this is to observe that the solution must satisfy $f(t) = \mu \int_0^t f(s)(s-t)\,\mathrm{d}s + at$ and show uniqueness by a contraction mapping method.) Thus any non-zero solution of the boundary value problem $\mu f = Lf = -f''$ with $f(0) = f(1) = 0$ is uniquely determined by the value of $f'(0)$ with all solutions scalar multiples of each other. □

For our original problem (7.1), elementary calculus gives the eigenvalues as $\lambda_n^{-1} = \mu_n = n^2\pi^2$ with orthonormal eigenfunctions $f_n(t) = \sqrt{2}\sin(\pi n t)$, and these form a basis for $L^2[0,1]$ by the Sturm-Liouville Theorem. Thus, going back to the spectral theorem, we may express every $f \in L^2[0,1]$ as

$$f(t) = \sum_{n=1}^{\infty} 2\left(\int_0^1 f(s)\sin(\pi n s)\,\mathrm{d}s\right)\sin(\pi n t)$$

where this 'Fourier series' converges in the $\|\cdot\|_2$-norm.

(**G**): The importance of Sturm-Liouville theory is that all the above applies to *every* second order linear differential equation boundary value problem. The reader may have noticed that after paragraph (**A**) we worked entirely with the operators G and L (until we specialised again at the end). If (7.1) is replaced by

$$a(t)y''(t) + b(t)y'(t) + c(t)y + \mu y = 0 \qquad \text{with } y(0) = y(1) = 0 \tag{7.7}$$

(subject to natural requirements on $a(t), b(t), c(t)$ such as continuity and differentiability), then we define the differential operator

$$(Ly)(t) = -\Big(a(t)y''(t) + b(t)y'(t) + c(t)y\Big).$$

We may solve the initial value problem (by turning it into an integral equation) to find real valued functions u and v such that

$$Lu = 0 \text{ with } u(0) = 0, u'(0) = 1 \quad \text{and} \quad Lv = 0 \text{ with } v(1) = 0, v'(1) = 1.$$

We then define the Green's function $g(s,t)$ by

$$g(s,t) = \begin{cases} cu(s)v(t) & 0 \leqslant s \leqslant t \\ cv(s)u(t) & t \leqslant s \leqslant 1 \end{cases} \qquad \text{where } s,t \in [0,1],$$

and where c is a suitable normalising constant. The Green's function operator G is then defined just as in (7.4). Then L and G operate on exactly the same spaces as before and satisfy the same relationships $LG = I_{C[0,1]}$ and $GL = I_{C_B^2[0,1]}$. The theory goes through in just the same way, and the statement of the Sturm-Liouville theorem is exactly the same with (7.1) replaced by (7.7). We also note that this theory can work for different sets of boundary conditions, e.g., $y'(0) = y'(1) = 0$.