LINEAR ANALYSIS COURSE NOTES

JIM WRIGHT

* Star system – subsections, exercises, theorems, propositions and lemmas tagged with a star * are less important and can be skipped on a first reading of these course notes. The details of steps in a construction or the details of a proof which is starred can be safely skipped. However you should know what the construction is accomplishing and you should know the statements of *unstarred* theorems, propositions, lemmas whose proofs are starred.

The Introduction below is mainly for motivation and attempts to set the scene for the course. The Course Notes start properly with the following section, *Normed linear spaces and Banach spaces.*

1. Introduction

Linear analysis is a beautiful mix of analysis and algebra. We will be doing analysis, using ideas and techniques you have seen in Honours Analysis with roots from Fundamentals of Pure Maths, in the setting of vector spaces whose theory you were first exposed to in first year (Linear Algebra) but then developed further in Honours Algebra. We will often call vector spaces *linear spaces* so for this course,

$$\text{vector space} \ = \ \text{linear space}.$$

In order to carry out analysis in the setting of a vector or linear space X, we will introduce a *size* $\|\cdot\|$ to vectors called a **norm** (so that to every $x \in X$, a nonnegative number $\|x\| \geq 0$ is assigned, giving a size to x) whose definition we will come to momentarily. A norm will in turn give us a metric d on X; more precisely, $d(x,y) := \|x-y\|$. This endows X with a metric space structure which allows us to use concepts from Honours Analysis such as convergence and continuity. We will also be able to discuss and use the notions of open, closed and compact subsets of X. A vector space with a norm will be called a **normed linear space**. Normed linear spaces and the various concepts associated to them will be the central object of study in this course.

As with many mathematical topics, we will in time impose more structure on normed linear spaces (the historical development of the subject is guided by key examples and important applications) and this will lead us to **Banach spaces** and then to **Hilbert spaces**, the latter of which provides the perfect, abstract, *euclidean-like* setting to carry out geometric investigations which are almost identical to the classical geometry one encounters in Euclidean spaces. Many of you will be reminded of analogous situations in other subject areas. In group theory, adding structure provides a passage from groups $\to$ simple groups $\to$ finite simple groups.

Given a class of objects, it is natural to introduce and study the *natural* maps between them; maps which preserve the underlying structure. For example,

$$\begin{array}{ccc} \text{Groups} & \longleftrightarrow & \text{group homomorphisms} \\ \text{Metric spaces} & \longleftrightarrow & \text{continuous maps} \\ \text{Vector spaces} & \longleftrightarrow & \text{linear maps} \end{array}$$

$$\boxed{\text{Normed linear spaces} \quad \longleftrightarrow \quad \text{continuous, linear maps}}$$

One purpose of these natural maps is to set up equivalences between objects; more precisely, we say spaces X and Y are *isomorphic* if there is a bijective natural map $\phi : X \to Y$, mapping X onto Y. To be more accurate, we should also require that the inverse map $\phi^{-1} : Y \to X$ be in the same category. This sometimes follows automatically which is the case for group homomorphims but it is not always the case. For example, it is not always true that the inverse of a bijective continuous map is continuous.[1] Interestingly, when we move to the category of Banach spaces, it is an amazing fact that the inverse of a bijective continuous, linear map between Banach spaces is continuous! This will be proved in the Functional Analysis course.

A certain amount of the development of pure mathematics is driven by the quest to address the following structural issues for some basic class of objects:

1. Up to the equivalence described above, classify the objects under study. In our case, one would want to classify normed linear spaces up to continuous, linear bijections (similar statements hold for Banach and Hilbert spaces).

2. Characterise (or at least come to some good understanding of) all natural maps between two given objects in a class. In our case, given two normed linear spaces (or Banach or Hilbert spaces), one would like to try to characterise all linear, continuous maps between them.

We emphasise that these are only two issues, albeit important ones. There are many, many other issues/questions/problems which drive the development of an area forward. Some come from applications arising in neighbouring areas, some come from intrinsic geometric/analytic questions, etc... the list is an endless, vast sea.

1.1. **General Example.** Consider the situation for vector spaces with no added structure. Here the natural maps between vector spaces are linear maps and we say that vector spaces X and Y are isomorphic if there is a bijective linear map from X to Y (here is another situation when the inverse map is automatically linear and

[1] Consider the map $f : [0, 2\pi) \to S^1$ defined by $f(\theta) = (\cos(\theta), \sin(\theta))$ where S^1 is the unit circle in $\mathbb{R}^2$. This is a continuous bijection but f^{-1} is not continuous at $(1, 0)$. Or one can reason as follows: if f^{-1} is continuous, then $[0, 2\pi) = f^{-1}(S^1)$ would be compact since S^1 is compact.

hence in the same class). There is a wonderful invariant called the *dimension*. If X is a vector space, we denote by $\dim(X)$ the dimension of X.

Theorem 1.2. *Two vector spaces X and Y are isomorphic if and only if* $\dim(X) = \dim(Y)$. *Furthermore,*

$$\big\{\text{all linear maps between X and Y}\big\} \text{ “ = ” } \big\{\text{all } \dim(\mathrm{Y}) \times \dim(\mathrm{X}) \text{ matrices}\big\}.$$

You have seen this in your Honours Algebra course in the finite dimensional case but it is not too difficult to extend it to the general case. So this theorem answers both structural questions above. Although nice, it is probably one of the reasons why very few people work only in the category of vector spaces.

1.3. **Concrete examples.** Here we will discuss two classical examples and highlight certain notions and properties.

① $\mathbb{R}^n = \{x = (x_1, \ldots, x_n) : x_j \in \mathbb{R}\}$, n-tuples of real numbers. This is a standard example that you have seen in many different contexts, but most relevant for us here is in the context of metric spaces that you studied in Honours Analysis.

For $x, y \in \mathbb{R}^n$, the *euclidean distance/metric* d is defined as

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

which arises from a norm; namely, $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$. With this distance, we can talk about convergence of sequences, continuity of functions, open, closed and compact subsets of $\mathbb{R}^n$. This was done to a large extent in your Honours Analysis course.

As we know, the linear maps between $\mathbb{R}^n$ and $\mathbb{R}^k$ can be characterised by the collection of $k \times n$ matrices. Furthermore, it turns out, that all linear maps between euclidean spaces are automatically continuous with respect to the distance d above. Try this now as an exercise or wait until we come back to this matter later in the course. So once again, the two structural issues above are settled in the case of euclidean spaces $\mathbb{R}^n$ with the euclidean norm $\|\cdot\|$ given above.

Nevertheless we could examine $\mathbb{R}^n$ with another *norm*, say $\|x\|_\infty = \max_j |x_j|$ which induces the metric $d_\infty(x,y) = \|x - y\|_\infty = \max_j |x_j - y_j|$. This changes the geometry of $\mathbb{R}^n$; for example, the basic balls

$$B_r(x) = \{y \in \mathbb{R}^n : \|x - y\| < r\} \quad \longleftrightarrow \quad B_r^\infty = \{x \in \mathbb{R}^n : \|x - y\|_\infty < r\}$$

are very different (when $n = 2$, the former are discs and the latter are squares). Hence questions/problems of a geometric nature may depend on which norm on $\mathbb{R}^n$ we consider. For example, questions in the area of finite-dimensional spectral geometry depend on the choice of norm.

BUT a central fact in the finite-dimensional part of our theory is the following:

$$\boxed{\text{All norms on a finite dimensional vector space are equivalent!}}$$

This means if X is a finite dimensional vector space and $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms defined on X, then there is a constant C such that $\|x\|_1 \le C\|x\|_2$ and $\|x\|_2 \le C\|x\|_1$ for all $x \in X$. Importantly, the constant C needs to be chosen *not* to depend on x. Let us consider the norms $\|x\| = \sqrt{x_1^2 + \cdot + x_n^2}$ and $\|x\|_\infty = \max(|x_1|, \ldots, |x_n|)$ defined on $\mathbb{R}^n$. Clearly, we have $|x_j| \le \sqrt{x_1^2 + \cdots + x_n^2}$ for every $1 \le j \le n$ and so the inequality $\|x\|_\infty \le \|x\|$ holds for all $x \in \mathbb{R}^n$ with a constant equal to 1. On the other hand, $x_1^2 + \cdots + x_n^2 \le n \max(x_1^2, \ldots, x_n^2)$ clearly holds and so $\|x\| \le \sqrt{n}\|x\|_\infty$. Hence if we choose $C = \sqrt{n}$, we see that the two norms $\|\cdot\|$ and $\|\cdot\|_\infty$ are equivalent.

Two equivalent norms on a linear space X may induce different *geometry* on X, but nonetheless they induce the same *topology*! That is, they give rise to the same collection of open, closed and compact subsets of X. Hence the class of convergent sequences and also the class of continuous functions are the same in both cases. In particular we have a complete resolution of the two structural issues 1. and 2. stated above for the class of finite dimensional normed linear spaces. That is, any two finite dimensional normed linear spaces X and Y are equivalent if and only if they have the same dimension. Furthermore, the family of continuous, linear maps between X and Y is precisely the family of $\dim(Y) \times \dim(X)$ matrices.

The situation for infinite dimensional normed linear spaces is spectacularly different.

The most interesting examples for this course are infinite dimensional ones!

② $C[0,1]$ the space of real-valued, continuous functions on the closed interval $[0,1]$.

This will be one of our standard *infinite dimensional* examples which we will refer to time and time again. The linear space $C[0,1]$ comes with many different norms but the most natural one (we will see why later) is the *sup* or *max* norm, $\|f\| = \sup_{0 \le x \le 1} |f(x)|$. From your Fundamentals of Pure Maths course, we know that the supremum is attained for continuous functions and so we can replace the *sup* with a *max* without changing anything.

The *sup* norm gives rise to the uniform metric $d_\infty(f,g) = \sup_{x\in[0,1]} |f(x) - g(x)|$ discussed extensively in your Honours Analysis course. Therefore it is more convenient to rename the *sup* norm as the *uniform norm* and write $\|\cdot\|_{L^\infty}$ instead of $\|\cdot\|$. Recall that a sequence of continuous functions $\{f_k\}$ converges in this uniform metric if and only if it *converges uniformly* and furthermore, the uniform limit must necessarily be continuous. These are some of the gems from your Honours Analysis course.

We could also consider the 1-metric $d_1(f,g) = \int_0^1 |f(x) - g(x)| dx$ on $C[0,1]$ which also arises from a norm,

$$\|f\|_{L^1} \;:=\; \int_0^1 |f(x)|\, dx.$$

So here we have two different normed linear spaces but in each case, the underlying vector space is the same, namely, $C[0,1]$. Although we have the straightforward comparison $\|f\|_{L^1} \le \|f\|_{L^\infty}$ for every $f \in C[0,1]$, it is not difficult to see that there does not exist a constant C such that $\|f\|_{L^\infty} \le C\|f\|_{L^1}$ for all $f \in C[0,1]$.[2] Therefore these two norms on $C[0,1]$ are not equivalent and hence the class of convergent sequences, open sets, closed sets, etc... are different for these two different normed linear spaces.

2. Normed linear spaces and Banach spaces

In this section, we begin the course notes properly and define some of the notions from the Introduction precisely. In order for us to define a norm on a vector space X, it is essential that the underlying scalar field $\mathbb{F}$ carries an *absolute value*. Of course in the case of the real field $\mathbb{R}$, we are all familiar with the absolute value $|a|$ of a real number $a \in \mathbb{R}$. This absolute value on the real field $\mathbb{R}$ extends to the complex field $\mathbb{C}$; if $w = u + iv$, then the absolute value $|w|$ of w (sometimes referred to as the *modulus* of w) is simply given as $|w| = \sqrt{u^2 + v^2}$. In both cases, the absolute value has the following distinguishing properties: (i) $|a| = 0$ if and only if $a = 0$; (ii) $|ab| = |a||b|$; and (iii) $|a + b| \le |a| + |b|$. It goes beyond the scope of this course to consider other scalar fields which carry a nontrivial absolute value and for this reason, we restrict our attention from now on to vector spaces with real or complex scalars.

Only vector spaces X over $\mathbb{R}$ or over $\mathbb{C}$ will be considered in this course!

For most things we will be discussing, it will not matter much whether we are using real or complex scalars. Hence we will use $\mathbb{F}$ to denote either $\mathbb{R}$ or $\mathbb{C}$ when talking about scalars. If and when we are talking about a vector space X over the complex field $\mathbb{C}$, it is sometimes useful to realise that X can also be thought of as a vector space over the real field $\mathbb{R}$, simply by restricting the scalars to be real (the axioms of a vector space still hold when we make such a restriction).[3] This will be particularly useful when we talk about *convex* sets.

Definition 2.1. Let X be a vector space over $\mathbb{F}$. A *norm* on X is an assignment of a *nonnegative* number $\|x\|$ to every $x \in X$ which possesses the following properties:

- $\|x\| = 0$ if and only if $x = 0$;
- $\|\alpha x\| = |\alpha|\|x\|$ for all $x \in X$ and $\alpha \in \mathbb{F}$;
- $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in X$.

The three bullet points above are called the *axioms of a norm.* A vector space X over $\mathbb{F}$ with norm $\|\cdot\|$ is called a *normed linear space* and we usually denote this by the pair $(X, \|\cdot\|)$.

The first two axioms are natural ones; it makes sense to assign the size 0 only to the zero vector and the second axiom says that the size of a vector scales linearly

[2] As an exercise, construct a sequence $f_k \in C[0,1]$ of nonnegative functions such that $\int_0^1 f_k(x)dx = 1$ while $\|f_k\|_{L^\infty} = k$.

[3] It turns out that we can reverse this process, starting with a vector space X over the reals, we can *complexify* X, turning it into a vector space over the complex numbers. But we will not need to do this in this course.

under scalar multiplication which is also natural. It is the third axiom, the so-called *triangle inequality*, which does not seem completely natural but nevertheless, it is the key property which makes the entire theory work.

As we will see, given a candidate for a norm on a vector space, the first two axioms tend to be readily verified but the triangle inequality often requires some work and can be difficult to verify sometimes. We will devote an entire section to establishing some classical inequalities which will help us verify the triangle inequality for a number of important examples.

A natural question immediately arises. Does every vector space over $\mathbb{F}$ support some norm? See Exercise 18 below.

2.2. **Examples.**

Before we move on, we should list some examples of normed linear spaces.

A. Let $X = \mathbb{F}^n$ and for any $p \geq 1$, set

$$\|x\|_p \; = \; \bigl(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\bigr)^{1/p} \tag{p_1}$$

for $x = (x_1, x_2, \ldots, x_n) \in \mathbb{F}^n$. Although the first two properties of a norm are easily verified for any $p > 0$, the triangle inequality only holds when we restrict to $p \geq 1$ (see Exercise 6 below). We will verify the triangle inequality in the next section. Hence we have a continuum of norms (one for every $p \geq 1$) on $\mathbb{R}^n$ or $\mathbb{C}^n$, making $(\mathbb{F}^n, \|\cdot\|_p)$ a normed linear space.

In the Introduction, we discussed the 2-norm $\|\cdot\|_2$ but we also discussed the norm $\|x\|_\infty = \max_j |x_j|$ which is not listed above. This is one of the rare norms where the triangle inequality is easy to verify; indeed, for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, we have $|x_j + y_j| \leq |x_j| + |y_j| \leq \|x\|_\infty + \|y\|_\infty$ for every j. Hence $\|x + y\|_\infty = \max_j |x_j + y_j| \leq \|x\|_\infty + \|y\|_\infty$. Interestingly, the norm $\|\cdot\|_\infty$ is the limit of the p-norms as $p \to \infty$. See Exercise 3 below.

Analysis in the metric space setting $(\mathbb{F}^n, d_p)$ where $d_p(x,y) = \|x - y\|_p$ was thoroughly investigated in your Honours Analysis course. In particular, let us recall the beautiful result called the **Heine-Borel Theorem**: on the metric space $(\mathbb{F}^n, d_p)$, a set $K \subset \mathbb{F}^n$ is compact if and only if K is closed and bounded. It is important to get one's hands on compact sets and understand which sets are compact. The Heine-Borel theorem gives a simple characterisation of compact sets in $\mathbb{R}^n$ or $\mathbb{C}^n$. We will see that the story remains the same on any finite dimensional normed linear space but the story dramatically changes when we turn to infinite dimensions.

B. Let X be a finite dimensional vector space over $\mathbb{F}$ with dimension n, say. Then every basis for X has precisely n elements. Let us fix a basis $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ so that every $x \in X$ has a unique representation $x = x_1 e_1 + \cdots + x_n e_n$ where $x_j \in \mathbb{F}$. Hence $x \leftrightarrow (x_1, \ldots, x_n)$ gives a linear isomorphism between X and $\mathbb{F}^n$. For every basis $\mathcal{B}$ and for every $p \geq 1$, we can define a norm

$$\|x\|_{p,\mathcal{B}} \; := \; (|x_1|^p + \cdots + |x_n|^p)^{1/p} \tag{p_2}$$

on X. If we denote by $T : X \to \mathbb{F}^n$ the linear isomorphism given above, we also have $\|Tx\|_p = \|x\|_{p,\mathcal{B}}$ for every $x \in X$ which shows that from the perspective of a normed linear space structure, the two spaces $(X, \|\cdot\|_{p,\mathcal{B}})$ and $(\mathbb{F}^n, \|\cdot\|_p)$ are indistinguishable – they are the same. In particular, we can use T to carry results over to $(X, \|\cdot\|_{p,\mathcal{B}})$ from $(\mathbb{F}^n, \|\cdot\|_p)$. For instance, the Heine-Borel theorem is valid in this setting: a set K is compact in $(X, \|\cdot\|_{p,\mathcal{B}})$ if and only if it is closed and bounded. We will give a more detailed discussion later in this section, stating the Heine-Borel theorem in this setting as Corollary 2.26.

We should remark that there are many more norms on X than just those arising from a basis $\mathcal{B}$ and an exponent $p \geq 1$; that is, many more norms than $\|\cdot\|_{p,\mathcal{B}}$. The next example illustrates this.

C. Let $\mathcal{P}_d$ denote the space of polynomials with coefficients in $\mathbb{F}$ of degree at most d. So $f \in \mathcal{P}_d$ if and only if $f(x) = a_d x^d + \cdots + a_1 x + a_0$ where each $a_j \in \mathbb{F}$. As we know, $\mathcal{P}_d$ is a vector space over $\mathbb{F}$ with dimension $d+1$ and so Example **B** shows that $\mathcal{P}_d$ is indistinguishable from $\mathbb{F}^{d+1}$ and for every $p \geq 1$ (including $p = \infty$), we can define the norm $\|\cdot\|_p$ with respect to the canonical basis $\mathcal{B} = \{1, x, x^2, \ldots, x^d\}$.

However we can take advantage of the fact that polynomials can also be viewed as functions on $\mathbb{R}$, the real line, to define yet another continuum of norms, again one for each $p \geq 1$; namely

$$\|f\|_{L^p} \ := \ \Big(\int_0^1 |f(x)|^p \, dx \Big)^{1/p}. \tag{Lp}$$

Also $\|f\|_{L^\infty} = \sup_{x \in [0,1]} |f(x)|$. It is tempting to use the notation $\|\cdot\|_p$ once again in this context but this will certainly confuse the matter as the underlying space is the same in both cases.

D. Now let us move from $\mathcal{P}_d$ to $\mathbb{F}[X]$, the ring of *all* polynomials of any finite degree with coefficients in $\mathbb{F}$. Of course $\mathbb{F}[X]$ is also a vector space over $\mathbb{F}$ but now it is infinite dimensional. Both norms defined in (p_2) and (Lp) above lift to $\mathbb{F}[X]$ and define norms on $\mathbb{F}[X]$. Check this!

What is the analogue of $\mathbb{F}[X]$ back in the sequence world $\mathbb{F}^n$? It is the space

$$\mathbb{F}_0^\infty \ := \ \big\{ a = (a_1, a_2, a_3, \ldots) : \text{each } a_j \in \mathbb{F} \text{ and only finitely many } a_j \text{ are nonzero} \big\}$$

of all finite sequences. The norms defined in (p_1) lift to give us norms on $\mathbb{F}_0^\infty$. With respect to any of these p-norms, the map $T : \mathbb{F}_0^\infty \to \mathbb{F}[X]$ defined by $T(a_1, a_2, \ldots) = \sum_{j \geq 0} a_{j-1} x^j$ is a linear isomorphism which perserves the respective norms as before. Hence the spaces $\mathbb{F}_0^\infty$ and $\mathbb{F}[X]$ are indistinguishable.

E. Let us begin with the vector space

$$\mathbb{F}^\infty \ := \ \{ x = (x_1, x_2, x_3, \ldots) : x_j \in \mathbb{F} \text{ for every } j \}.$$

Convince yourself that this is indeed a vector space over $\mathbb{F}$ and $\mathbb{F}_0^\infty$ is a linear subspace of $\mathbb{F}^\infty$. This space is too big to carry out any useful analysis on. Nevertheless this space supports many norms.

We will consider a continuum of linear subspaces of $\mathbb{F}^\infty$, one for every $p > 0$; namely,

$$\ell^p \ := \ \{(x_j)_{j\ge 1} \in \mathbb{F}^\infty : \text{each } x_j \in \mathbb{F} \text{ and } \sum_{j=1}^\infty |x_j|^p < \infty\}$$

which we will call *little Lp*. It is easy to see that if $x = (x_j) \in \ell^p$ and $\alpha \in \mathbb{F}$, then $\alpha x = (\alpha x_j) \in \ell^p$ however it is not straightforward to see that if $x, y \in \ell^p$, then $x + y \in \ell^p$ but we will see that this is indeed the case (see Exercise 4 below for the case $0 < p < 1$). So ℓ^p is a linear subspace of $\mathbb{F}^\infty$ and a natural candidate for a norm on ℓ^p is[4]

$$\|x\|_p \ = \ \Bigl(\sum_{j=1}^\infty |x_j|^p\Bigr)^{1/p} \tag{p_3}$$

where it is easy to verify that $\|\cdot\|_p$ satisfies the first two properties of being a norm. However the triangle inequality only holds when $p \ge 1$ and the verification is not straightforward. We will prove this in the following section.

As in the finite dimensional case, there is an endpoint $p = \infty$ case. The linear subspace $\ell^\infty = \{x = (x_1, x_2, \ldots) \in \mathbb{F}^\infty : \sup_j |x_j| < \infty\}$ carries the natural norm $\|x\|_\infty := \sup_j |x_j|$. As in the finite dimensional case, it is straightforward to verify that the axioms of a norm are satisfied for $\|\cdot\|_\infty$ (check this!).

Let us now fix a subspace ℓ^p and although $\|\cdot\|_p$ is a natural norm on it, one could ask if there are other norms on ℓ^p. For $x \in \ell^p$, we have $\|x\|_q \le \|x\|_p$ for any $q \ge p$ (see Exercise 5 below) and so whenever $q \ge p$, $\|\cdot\|_q$ provides another norm on ℓ^p.

F. Let $C[0,1]$ be the space of all real or complex-valued continuous functions on $[0,1]$. Hence $C[0,1]$ has the structure of a vector space over $\mathbb{F}$. The basic norm on $C[0,1]$ is the uniform norm

$$\|f\|_{L^\infty} \ := \ \sup_{0\le x\le 1} |f(x)|$$

whose verification of the norm axioms is left as an exercise (see Exercise 7 below). As in the polynomial ring $\mathbb{F}[X]$ example, Example **D**, we can also consider the norms

$$\|f\|_{L^p} \ := \ \Bigl(\int_0^1 |f(x)|^p dx\Bigr)^{1/p}$$

for every $p \ge 1$. Proposition 2.7 below establishes the triangle inequality for $\|\cdot\|_{L^p}$. The verificiation of the other axioms for a norm are left as an exercise.

2.3. **Inequalities.**

We begin with a simple generalisation of the classical arithmetic-geometric mean inequality, $\sqrt{ab} \le (a+b)/2$ for nonnegative numbers a, b.

---

[4]We are using the same notation, $\|\cdot\|_p$, as in the finite dimensional case. This hopefully will not be confusing and it should be clear from the context what the notation means. In fact we might use this exact same notation when we discuss norms on $C[0,1]$ but we will try to avoid this. In a way, this is a progression from finite $\to$ countably infinite $\to$ uncountably infinite sequence spaces.

Lemma 2.4. *Let $A, B \geq 0$ and $0 \leq \theta \leq 1$. Then*

$$A^\theta B^{1-\theta} \ \leq\ \theta A + (1-\theta)B. \tag{AG}$$

In lecture we will give another proof of Lemma 2.4.

Proof. * We may suppose $B \neq 0$, $0 < \theta < 1$. By replacing A with AB, we see that it suffices to show $A^\theta \leq \theta A + (1-\theta)$ for any $A \geq 0$. Consider the function $f(x) = x^\theta - \theta x - (1-\theta)$. Then $f'(x) = \theta x^{\theta-1} - \theta$ which implies that $f(x)$ increases when $0 \leq x \leq 1$, it decreases when $x \geq 1$. Hence the continuous function f has a maximum at $x = 1$ which implies that $f(x) \leq f(1) = 0$ for $x \geq 0$ or $x^\theta \leq \theta x + (1-\theta)$ for all $x \geq 0$. □

We will use this lemma to establish a famous inequality, the so-called Hölder's inequality. When $1 < p < \infty$, we set $q = p/(p-1)$ so that $1/p + 1/q = 1$. We call q the conjugate exponent of p. When $p = 1$, we set $q = \infty$ and when $p = \infty$, we set $q = 1$ as the conjugate exponent.

Theorem 2.5. *(Hölder's inequality). For any two sequences $\{a_1, a_2, \ldots, a_n\}$ and $\{b_1, b_2, \ldots, b_n\}$ of nonnegative numbers, we have*

$$\sum_{j=1}^n a_j b_j \leq \Big(\sum_{j=1}^n a_j^p\Big)^{1/p} \cdot \Big(\sum_{j=1}^\infty b_j^q\Big)^{1/q} \tag{H}$$

for any $1 \leq p \leq \infty$ where q is the conjugate exponent of p.

Proof. Let us denote by $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ the two nonnegative sequences of n terms. Then we may express the inequality to be proved as $\sum_{j=1}^n a_j b_j \leq \|a\|_p \|b\|_q$. Without loss of generality, we may assume $\|a\|_p, \|b\|_q > 0$. Replacing a with $a/\|a\|_p$ and b by $b/\|b\|_q$, we may assume that $\|a\|_p = \|b\|_q = 1$. Applying the above lemma to $A = a_j^p$ and $B = b_j^q$, we have $a_j b_j \leq (1/p)a_j^p + (1/q)b_j^q$ for each j and so summing gives us

$$\sum_{j=1}^n a_j b_j \ \leq\ \frac{1}{p}\sum_{j=1}^n a_j^p + \frac{1}{q}\sum_{j=1}^n b_j^q = \frac{1}{p} + \frac{1}{q} = 1 = \|a\|_p \|b\|_q.$$

□

The inequality (H) immediately extends to infinite sequences $a = (a_j)_{j\geq 1}$ and $b = (b_j)_{j\geq 1}$ – why?

Theorem 2.6. *(Minkowski's inequality) For any two elements $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ of $\mathbb{F}^n$, we have*

$$\|x+y\|_p = \Big(\sum_{j=1}^n |x_j+y_j|^p\Big)^{1/p} \leq \Big(\sum_{j=1}^n |x_j|^p\Big)^{1/p} + \Big(\sum_{j=1}^n |y_j|^p\Big)^{1/p} = \|x\|_p + \|y\|_p \tag{M}$$

for any $1 \leq p \leq \infty$.

Proof. We may assume that $\sum_{j=1}^{\infty} |x_j + y_j|^p > 0$. Also we have

$$|x_j + y_j|^p \;\le\; |x_j||x_j + y_j|^{p-1} + |y_j||x_j + y_j|^{p-1}$$

and so applying Hölder's inequality, noting $q(p-1) = p$, we have

$$\sum_{j=1}^{n} |x_j + y_j|^p \le \Bigl(\sum_{j=1}^{n} |x_j|^p\Bigr)^{1/p} \Bigl(\sum_{j=1}^{n} |x_j + y_j|^p\Bigr)^{1/q} + \Bigl(\sum_{j=1}^{n} |y_j|^p\Bigr)^{1/p} \Bigl(\sum_{j=1}^{n} |x_j + y_j|^p\Bigr)^{1/q}$$

which upon dividing by $\sum_{j=1}^{n} |x_j + y_j|^p$, we conclude that $\|x + y\|_p \le \|x\|_p + \|y\|_p$. □

Again we can extend Minkowski's inequality (M) to infinite sequences $x = (x_j)_{j\ge 1}$ and $y = (y_j)_{j\ge 1}$. This is straightforward and shows that ℓ^p, when $p \ge 1$, is closed under addition.

Theorem 2.6 establishes the triangle inequality for the p-norms $\|\cdot\|_p$ on $\mathbb{F}^n$ or any finite dimensional vector space over $\mathbb{F}$. The extension to infinite sequences establishes the triangle inequality for the norm $\|\cdot\|_p$ on ℓ^p. In particular, this shows when $p \ge 1$, ℓ^p is indeed a linear subspace of $\mathbb{F}^\infty$.

The same proof which establishes Minkowski's inequality (M) for sequences also works for the analogous inequality for integrals.

Proposition 2.7. *(Minkowski's inequality for integrals[5]) For $f, g \in C[0,1]$, we have*

$$\Bigl(\int_0^1 |f(x) + g(x)|^p dx\Bigr)^{1/p} \le \Bigl(\int_0^1 |f(x)|^p dx\Bigr)^{1/p} + \Bigl(\int_0^1 |g(x)|^p dx\Bigr)^{1/p}. \quad (M - int)$$

Proof. We leave this as an exercise. □

2.8. **Metric space structure.**

Proposition 2.9. *Let $(X, \|\cdot\|)$ be a normed linear space. The distance function d defined by $d(x,y) := \|x - y\|$ satisfies the axioms of a metric. Hence the normed linear space $(X, \|\cdot\|)$ carries a metric space structure; that is, the pair (X, d) defines a metric space.*

Proof. The first axiom of a metric (i) $d(x,y) = \|x - y\| = 0$ only if $x = y$ follows immediately from the first axiom of a norm. The second axiom (ii) $d(x,y) = d(y,x)$ or $\|x - y\| = \|y - x\|$ follows from the second axiom of a norm since

$$\|y - x\| \;=\; \|-(x - y)\| \;=\; |-1|\|x - y\| \;=\; \|x - y\|.$$

Finally the last axiom (iii) $d(x,z) \le d(x,y) + d(y,z)$ follows from the triangle inequality for a norm. Indeed we have

$$d(x,z) = \|x - z\| = \|x - y + y - z\| \le \|x - y\| + \|y - z\| = d(x,y) + d(y,z).$$

□

[5]There is a more general inequaltiy which is commonly known Minkowski's inequality for integrals. See the Exercises.

Therefore every normed linear space is a metric space. This is wonderful since we can now use all the properties/notions/theorems which are valid for metic spaces in the setting of a normed linear space. For instance, the notion of a sequence $\{x_n\} \subset X$ converging to a point $x \in X$ (in symbols, $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ as $n \to \infty$) simply means that the sequence of nonnegative distances $d(x_n, x)$ converges to 0 (symbolically, $d(x_n, x) = \|x_n - x\| \to 0$ as $n \to \infty$) which has the precise following meaning: for every $\epsilon > 0$, there is an N such that $d(x_n, x) = \|x_n - x\| < \epsilon$ whenever $n \geq N$.

Furthermore, we can use the family of balls

$$B_r(x) \ := \ \big\{y \in X : \|y - x\| = d(y, x) < r\big\}$$

to arrive at the notions of *open* and *closed* sets.

Definition 2.10. Let $(X, \|\cdot\|)$ be a normed linear space. A set $G \subset X$ is **open** if for every $x \in G$, there is an $r > 0$ such that $B_r(x) \subset G$. A set $F \subset X$ is **closed** if $X \setminus F$ is open.

In other words, a set $G \subset X$ is open if no matter where we are inside G, there is a little *breathing room* – one can move a little bit in all directions without leaving G. Hence G does not contain any if its boundary points.[6] By definition F is closed if $X \setminus F$ is open but from your Honours Analysis course, we know that this is equivalent to the property that whenever a sequence $\{x_n\} \subset F$ converges to a point $x \in X$, then necessarily $x \in F$. In other words, *all limit points of F lie in F.*

A highly recommended exercise: show that every ball $B_r(x)$ is itself an open set.

The family of balls $\{B_r(x)\}$ has added structure in the normed linear space setting due to the underlying vector space structure. If X is a vector space and $E, F \subset X$ are any two subsets, we can form the sum $E + F := \{x + y : x \in E \text{ and } y \in F\}$ from vector addition. Furthermore, if $\alpha \in \mathbb{F}$ is a scalar, then we can form the scaled set $\alpha E := \{\alpha x : x \in E\}$ from scalar multiplication. We note that $B_r(x) = x + B_r(0) = x + rB_1(0)$ and so the entire collection of balls in a normed linear space is generated from the single *unit ball* $B_1(0) = \{x \in X : \|x\| < 1\}$, using the underlying vector space operations.

We will see that the unit ball (and hence all balls) is *convex.* But first the definition of convexity.

Definition 2.11. Let X be a vector space over the reals $\mathbb{R}$. A subset $C \subset X$ is said to be **convex** if whenever $x, y \in C$, then $\lambda x + (1 - \lambda)y \in C$ for all $\lambda \in (0, 1)$.

In other words, for every pair of points $x, y \in C$, the entire line segment $[x, y] := \{w = \lambda x + (1 - \lambda)y : \lambda \in [0, 1]\}$ connecting x to y must lie inside C. A simple exercise shows that if C is convex, then so is any translate $x + C$ and any dilate rC.

[6]By definition, a *boundary point* x of a set $S \subset X$ is a point such that for every $r > 0$, the punctured ball $B_r(x) \setminus \{x\}$ contains points in S and also it contains points in $X \setminus S$. More succinctly, the boundary of S is $\overline{S} \setminus \mathrm{int}(S)$

Lemma 2.12. *Let $(X, \|\cdot\|)$ be a normed linear space. The unit ball $B_1(0)$ is convex. Hence all balls $B_r(x)$ are convex.*

Proof. Suppose $x, y \in B_1(0)$ so that $\|x\|, \|y\| < 1$. Consider $w = \lambda x + (1-\lambda)y$ where $\lambda \in (0,1)$. Then

$$\|w\| = \|\lambda x + (1-\lambda)y\| \leq \|\lambda x\| + \|(1-\lambda)y\| = \lambda\|x\| + (1-\lambda)\|y\| < \lambda + (1-\lambda) = 1.$$

□

So all-in-all, a norm generates a family of balls $\{B_r(x)\}$ which in turn is generated by a single ball $B_1(0)$ which is a convex, open set. It turns out that we can reverse the process; starting with an open, convex set C, together with the properties of being bounded and symmetric, we can generate a norm $\|\cdot\|_C$ such that the unit ball with respect to this new norm is equal to C; that is $C = \{x \in X : \|x\|_C < 1\}$. We will examine this in more depth when we come to discuss *equivalence of norms.*

2.13. Completeness and Banach spaces.

Recall from your Honours Analysis course, a metric space (X, d) is said to be **complete** if every Cauchy sequence $\{x_n\} \subset X$ (in symbols, this means $d(x_n, x_m) \to 0$ as $m, n \to \infty$) converges to some point $x \in X$. Cauchy sequences can be used to detect whether a space X has any holes in it. If there was a hole $\circ$ in X, then one will be able to approach it from within X without every being able to reach it – this is what a Cauchy sequence can do. A complete metric space (X, d) is one which does not have any holes! This is a very good quality for a space to have. The presence of holes can be problematic.

In many problems in analysis, solving some sort of equation, say, a standard approach is to approximate the solution (or start with a reasonable guess for the solution) and then produce a method to produce a better approximation. Proceeding in this way, one produces a sequence of better and better approximations to the desired solution, and *hopefully!(?)* we end up with a Cauchy sequence. Now if (X, d) is complete, then this sequence of better and better approximations will then converge to the solution!

Definition 2.14. A normed linear space $(X, \|\cdot\|)$ is called a **Banach space** if the associated metric space (X, d) where $d(x, y) = \|x - y\|$ is a complete metric space.

We will tend to use the notation E instead of X when we are talking about Banach spaces

Examples

I. Any finite dimensional normed linear space $(E, \|\cdot\|)$ is a Banach space! We will come back to this when we discuss equivalence of norms and give a proof of this fact. Hence the only possible normed linear spaces which can have holes (that is, ones that are not Banach spaces) are infinite dimensional ones.

II. The space $(C[0,1], \|\cdot\|_{L^\infty})$ is a Banach space. This was shown in your Honours Analysis course. It is a good exercise to review and recall why this is the case.

III. The spaces $(C[0,1], \|\cdot\|_{L^p}), 1 \le p < \infty$ are **not** Banach spaces. Again you have already seen this and it is a good exercise to review it.

IV. The sequence spaces $(\ell^p, \|\cdot\|_p), 1 \le p \le \infty$ are all Banach spaces. Here is the proof: Let $\{x^n\} \subset \ell^p$ be a Cauchy sequence in $(\ell^p, \|\cdot\|_p)$. Hence for every n, $x^n = (x^n_j)_{j\ge 1}$ and our goal is to find an $x = (x_j)_{j\ge1} \in \ell^p$ such that $\|x^n - x\|_p \to 0$ as $n \to \infty$. A natural candidate for x suggests itself since for every fixed $j \ge 1$,

$$|x^m_j - x^n_j| \ \le \ \|x^m - x^n\|_p \ \to \ \infty \ \text{ as } \ m, n \ \to \infty,$$

implying that the sequence $\{x^n_j\}_{n\ge1}$ of scalars is a Cauchy sequence in $\mathbb{F}$ and hence converges. For each j, let us call the limit $x_j = \lim_{n\to\infty} x^n_j$. We need to show two things: (i) $x = (x_j)_{j\ge1} \in \ell^p$, and (ii) $\|x^n - x\|_p \to 0$ as $n \to \infty$.

Given any $\epsilon > 0$, there is an $N = N(\epsilon)$ such that $\|x^n - x^m\|_p \le \epsilon$ whenever $m, n \ge N$ (recall that $\{x^n\}$ is a Cauchy sequence in $(\ell^p, \|\cdot\|_p)$). Hence for any $J \ge 1$,[7]

$$\sum_{j=1}^{J} |x^n_j - x^m_j|^p \ \le \ \|x^n - x^m\|_p^p \ \le \ \epsilon^p \ \text{ for } \ n, m \ge N.$$

Since $\lim_{m\to\infty} x^m_j = x_j$ for every j, we have

$$\sum_{j=1}^{J} |x^n_j - x_j|^p \ = \ \lim_{m\to\infty} \sum_{j=1}^{J} |x^n_j - x^m_j|^p \ \le \ \epsilon^p$$

for every $n \ge N$ and every $J \ge 1$. Letting $J \to \infty$ shows that $x^n - x \in \ell^p$ and $\|x^n - x\|_p \le \epsilon$ whenever $n \ge N$. Since ℓ^p is a vector space, this implies (i) $x = x - x^N + x^N \in \ell^p$ and (ii) $\|x^n - x\|_p \to 0$ as $n \to \infty$.

V. From Exercise 5 below, we know that $\|x\|_q \le \|x\|_p$ whenever $q \ge p$ and hence $\|\cdot\|_q$ defines a norm on ℓ^p. The normed linear space $(\ell^p, \|\cdot\|_q)$ where $q > p$ is **not** a Banach space. This is worked out in two steps in Exercise 22.

Many of the results you know about complete metric spaces carry directly over to the setting of a normed linear space. For instance, we have the following result characterising when a linear subspace of a Banach space is itself a Banach space. A subspace Y of a normed linear space $(X, \|\cdot\|)$ can itself be thought of as a normed linear space $(Y, \|\cdot\|)$ simply by restricting the norm $\|\cdot\|$ on X to Y.

Lemma 2.15. *Let $(E, \|\cdot\|)$ be a Banach space and let Y be a linear subspace of X. Then $(Y, \|\cdot\|)$ is a Banach space if and only if Y is a closed, linear subspace of E.*

Proof. We leave this as a good review exercise of material from your Honours Analysis course. □

[7] the sum $\sum_{1\le j\le J}$ should be replaced by $\max_{1\le j\le J}$ when $p = \infty$; otherwise the proof is the same for the $p = \infty$ case.

Example Let $c_0 := \{x = (x_j)_{j\ge 1} \in \ell^\infty : x_j \to 0 \text{ as } j \to \infty\}$. It is easy to check that c_0 is a linear subspace of ℓ^∞. It is in fact a *closed* subspace of $(\ell^\infty, \|\cdot\|_\infty)$; see Exericse 20 below. Hence by Lemma 2.15, $(c_0, \|\cdot\|_\infty)$ is a Banach space.

Lemma 2.15 is the reason why closed linear subspaces will play an important role in the theory. Hence whenever we meet a linear subspace Y, it will be convenient to pass to the smallest closed, linear subspace containing Y, called the **closure** of Y and denoted by $\overline{Y}$.

Lemma 2.16. *Let Y be a linear subspace of a normed linear space $(X, \|\cdot\|)$. Then*

$$\bigcap_{Y\subset W \text{ closed subspace}} W \;=\; \bigcap_{Y\subset F \text{ closed set}} F \;=\; \big\{y \in X : \exists\{y_n\} \subset Y \text{ s.t. } y_n \to y\big\}. \quad (*)$$

The common quantity above is called the closure of Y and denoted by $\overline{Y}$.

Hence if Y is a linear subspace of a Banach space $(E, \|\cdot\|)$, then $(\overline{Y}, \|\cdot\|)$ is a Banach space.

Proof. In Honours Analysis, we learned that in a general metric space (X, d), we have for any subset $S \subset X$,

$$\bigcap_{S\subset F \text{ closed set}} F \;=\; \big\{y \in X : \exists\{y_n\} \subset S \text{ s.t. } y_n \to y\big\}$$

and the two common quantities is called the closure of S. Therefore it suffices to show that in the linear space setting, the first quantity in $(*)$ agrees with the other two.

Since every closed, linear subspace containing Y is a closed set containing Y, then in order to show equality of the first quantity in $(*)$ with the other two quantities, it suffices to show that the limit points of Y,

$$\mathcal{L} \;:=\; \{y \in X : \exists\ \{y_n\} \subset Y \text{ s.t. } y_n \to y\},$$

necessarily a closed set, is in fact a linear subspace. Let $y, w \in \mathcal{L}$. Then there are two sequences $\{y_n\}, \{w_n\} \subset Y$ such that $y_n \to y$ and $w_n \to w$ and in particular, $y_n + w_n \to y + w$. But $\{y_n + w_n\} \subset Y$ since Y is a linear subspace which shows that $y + w$ is a limit point of Y; that is, $y + w \in \mathcal{L}$. Similarly if $y \in \mathcal{L}$ and $\alpha \in \mathbb{F}$, we have $\alpha y \in \mathcal{L}$. Hence $\mathcal{L}$ is a closed, linear subspace of X. $\square$

Example Let $\mathbb{F}_0^\infty := \{x = \{x_j\}_{j\ge 1} \in \mathbb{F}^\infty : x_j = 0 \text{ for all but finitely many } j\}$ be the space of all *finite* sequences. It is easy to check that $\mathbb{F}_0^\infty$ is a linear subspace of ℓ^∞. We claim that the closure $\overline{\mathbb{F}_0^\infty}$ of $\mathbb{F}_0^\infty$ in $(\ell^\infty, \|\cdot\|_\infty)$ is c_0. Since c_0 is a closed subspace of $(\ell^\infty, \|\cdot\|_\infty)$ (see the Example after Lemma 2.15), it suffices to show that every $x \in c_0$ is a limit point of $\mathbb{F}_0^\infty$; that is, we need to find a sequence $\{x^n\}_{n\ge 1} \subset \mathbb{F}_0^\infty$ such that $\|x^n - x\|_\infty \to 0$ as $n \to \infty$. For our given $x = (x_j)_{j\ge 1} \in c_0$, simply take $x^n = (x_1, x_2, \ldots, x_n, 0, 0, \ldots) \in \mathbb{F}_0^\infty$. We have

$$\|x^n - x\|_\infty \;=\; \sup_{1\le j} |x^n_j - x_j| \;=\; \sup_{j>n} |x_j|$$

and since $x_j \to 0$ as $j \to \infty$, we have $\sup_{j>n} |x_j| \to 0$ as $n \to \infty$. See Exercise 21.

The space $\mathbb{F}_0^\infty$ is also a linear subspace of ℓ^p for any $p \geq 1$. The closure of $\mathbb{F}_0^\infty$ in $(\ell^p, \|\cdot\|_p)$ for $1 \leq p < \infty$ is the whole space; that is, $\overline{\mathbb{F}_0^\infty} = \ell^p$. See Exercise 23. Hence $\mathbb{F}_0^\infty$ is a **dense** subspace of ℓ^p for $p \geq 1$.

2.17. **Completion of normed linear spaces.**

When a normed linear space $(X, \|\cdot\|)$ is not a Banach space and so has holes in it, there is a formal procedure[8] to fill in the holes to arrive at a larger normed linear space $(E, \|\cdot\|')$ which is in fact a Banach space such that X sits inside E as a dense linear subspace with $\|x\| = \|x\|'$ whenever $x \in X$. We call $(E, \|\cdot\|')$ the **Banach space completion** of the normed linear space $(X, \|\cdot\|)$ – we will see later that there is essentially a unique completion of X.

In comes as no surprise that this formal procedure involves the Cauchy sequences of X since such sequences are precisely the objects which detect holes in our space. The procedure is carried out in the following steps. Exercise 24 below asks you to fill in details of the any claims made in the various steps without proof.

Step 1.* Let $C = \{\underline{x} = \{x_j\}_{j\geq 1} : \underline{x}$ is a Cauchy sequence in $X\}$ collect together all the Cauchy sequences of X. We introduce an equivalence relation $\sim$ on C by saying $\underline{x} \sim \underline{y}$ if $\|x_j - y_j\| \to 0$ as $j \to \infty$. We denote by $E = C/\sim$ the set of equivalence classes with respect to $\sim$. Furthermore, we denote an equivalence class by $[\underline{x}] \in E$ where $\underline{x} \in C$.

The space E carries a natural vector space structure, namely if $[\underline{x}]$, $[\underline{y}] \in E$ and $\alpha \in \mathbb{F}$, we set

$$[\underline{x}] + [\underline{y}] := [\{x_j + y_j\}_{j\geq 1}] \quad \text{and} \quad \alpha[\underline{x}] \; := \; [\{\alpha x_j\}_{j\geq 1}].$$

One needs to check that these definitions are well-defined (independent of the choice of representative of the equivalence class) and that the axioms of a vector space are satisfied.

Step 2.* For any $\underline{x} = \{x_j\}_{j\geq 1} \in C$, we have

$$|\|x_j\| - \|x_k\|| \;\leq\; \|x_j - x_k\| \;\to\; 0 \;\text{ as }\; j \to \infty$$

and so $\{\|x_j\|\}_{j\geq 1}$ is a Cauchy sequence of positive numbers. Using the completeness of the reals $\mathbb{R}$, we see that $\lim_{j\to\infty} \|x_j\|$ exists. We define a norm $\|\cdot\|'$ on E by

$$\|[\underline{x}]\|' \;:=\; \lim_{j\to\infty} \|x_j\|.$$

One needs to check that once again this definition is well-defined and satisfies the axioms of a norm.

Step 3.* We now identify each element x of X with the equivalence class $[\underline{x}]$ where $\underline{x} = \{x, x, x, \ldots\}$. Clearly $\underline{x}$ is a Cauchy sequence in X and hence $\underline{x} \in C$. We note that $\iota : x \to [\{x, x, x, \ldots\}]$ defines a linear embedding (injection) from X into E.

[8]The same procedure (with obvious modifications) works in a general metric space (X, ρ) which is not complete; there is a complete metric space $(\widetilde{X}, \rho')$ which contains X as a dense subset and $\rho'(x, y) = \rho(x, y)$ for any $x, y \in X$.

We leave the verification that the map ι is linear as an exercise. Furthermore if $\iota(x) = [\{x, x, \ldots\}] = 0$, then $\{x, x, x, \ldots\} \sim \{0, 0.0, \ldots\}$ and so $\|x\| = \|x_j - 0\| \to 0$, implying $x = 0$. If $x_j = x$ for all j, then $\|x\| = \lim_{j\to\infty} \|x_j\|$ and so $\|\iota(x)\|' = \|[\{x, x, \ldots\}]\|' = \lim_{j\to\infty} \|x\| = \|x\|$. This allows us to identify X as a linear subspace of E via ι AND $\|x\|' = \|x\|$ for every $x \in X$.

Step 4.* The normed linear space $(E, \|\cdot\|')$ is a Banach space. Suppose $\{[\underline{x}_n]\}_{n\geq 1}$ is a Cauchy sequence in $(E, \|\cdot\|')$. Write $\underline{x}_n = \{x_j^n\}_{j\geq 1}$ and consider the sequence $\underline{x} := \{x_n^n\}_{n\geq 1}$. We claim that $\underline{x} \in C$; that is, $\{x_n^n\}_{n\geq 1}$ is a Cauchy sequence in X. Let $\epsilon > 0$. Since $\{[\underline{x}_n]\}_{n\geq 1}$ is a Cauchy sequence in $(E, \|\cdot\|')$, we can find an N such that

$$\|[\underline{x}_m] - [\underline{x}_n]\|' = \lim_{j\to\infty} \|x_j^m - x_j^n\| \leq \epsilon/4$$

whenever $m, n \geq N$. Therefore there is a J such that $\|x_j^m - x_j^n\| \leq \epsilon/3$ whenever $m, n \geq N$ and $j \geq J$. Since $\{x_k^N\}_{k\geq 1}$ is a Cauchy sequence in $(X, \|\cdot\|)$, we can find a K such that $\|x_j^N - x_k^N\| \leq \epsilon/3$ for all $j, k \geq K$. Set $N' = \max(N, J, K)$. Then for $n, m \geq N'$, we have

$$\|x_n^n - x_n^m\| \leq \|x_n^n - x_n^N\| + \|x_n^N - x_m^N\| + \|x_m^N - x_n^m\| \leq \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon,$$

showing that $\underline{x} \in C$.

Finally we claim that

$$\|[\underline{x}_n] - [\underline{x}]\|' \;=\; \lim_{j\to\infty} \|x_j^n - x_j^j\| \;\to 0$$

as $n \to \infty$. Again let $\epsilon > 0$. From the above argument, we see that there is an N' such that $\|x_j^n - x_j^m\| \leq \epsilon$ whenever $j, m, n \geq N'$. In particular we have $\|x_j^n - x_j^j\| \leq \epsilon$ whenever $n, j \geq N'$. Now let $j \to \infty$ to conclude that

$$\|[\underline{x}_n] - [\underline{x}]\|' \;=\; \lim_{j\to\infty} \|x_j^n - x_j^j\| \;\leq\; \epsilon$$

whenever $n \geq N'$. This establishes the completeness of $(E, \|\cdot\|')$.

Step 5.* It remains to show that X is a dense linear subspace of $(E, \|\cdot\|')$; more precisely, we will show that $\iota(X)$ is dense in E. Let $[\underline{x}] \in E$. Then $\underline{x} = \{x_j\}_{j\geq 1}$ is a Cauchy sequence in X. Consider the sequence $\{\iota(x_j)\}$ in $\iota(X)$. We claim that

$$\|\iota(x_j) - [\underline{x}]\|' \;=\; \lim_{k\to\infty} \|x_j - x_k\| \;\to\; 0$$

as $j \to \infty$. But this follows immediately since $\{x_j\}$ is a Cauchy sequence in X.

Definition 2.18. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed linear spaces.

- A linear map $T : X \to Y$ is said to be an **isometry** if $\|Tx\|_Y = \|x\|_X$ for all $x \in X$.
- We say that that the two normed linear spaces X and Y are **isometrically isomorphic** is there is an isometry T from X onto Y. Note that $T^{-1} : Y \to X$ is automatically a surjective isometry.
- The **Banch space completion** of $(X, \|\cdot\|_X)$ is a pair, consisting of a Banach space $(Y, \|\cdot\|_Y)$ and an isometry $T : X \to Y$ such that $T(X)$ is dense subspace of Y.

The above 5 step procedure tells us that every normed linear space has a Banach space completion. We now discuss the uniqueness question.

Proposition 2.19. *Let* $\{T, (Y, \|\cdot\|_Y)\}$ *and* $\{S, (Z, \|\cdot\|_Z)\}$ *be two completions of* $(X, \|\cdot\|_X)$*. Then* Y *and* Z *are isometrically isomorphic.*

Proof. * Since $T : X \to T(X)$ is an isometry onto $T(X)$, then $T^{-1} : T(X) \to X$ is a surjective isometry. Hence $P = S \circ T^{-1} : T(X) \to S(X)$ is an isometry from $T(X)$ onto $S(X)$. Since $T(X)$ is a dense linear subspace of Y, there is a unique linear extension $f : Y \to Z$ of P which is an isometry (See Exercise 24). Note that $f \circ T = S$ since $f = P$ on $T(X)$, hence when $x \in X$, we have $Tx \in T(X)$ and so $f(Tx) = P(Tx) = S \circ T^{-1}(Tx) = S(x)$.

We need to show that f is onto. Changing the roles of Y and Z in the above argument and considering the isometry $Q := T \circ S^{-1}$ from $S(X)$ onto $T(X)$, we obtain an extension $g : Z \to Y$ of Q which is an isometry such that $g \circ S = T$.

Therefore $f \circ g(Sx) = f(Tx) = S(x)$ for all $x \in X$, implying that $f \circ g$ is the identity on the dense subspace $S(X)$ of Z. By the uniqueness part of Exericse 25, we see that $f \circ g(z) = z$ for all $z \in Z$ and this shows that f is onto. □

Given a normed linear space $(X, \|\cdot\|)$ which is not complete, it is an important endeavour to understand concretely what is the Banach space completion of X. When X already sits inside a Banach space E as a linear subspace, then by Lemma 2.15 and Proposition 2.19, the closure $\overline{X}$ of X in E is the Banach space completion of X and so we automatically have a concrete realisation of the elements of the completion of X. For example the completion of the space of finite sequences $\mathbb{F}_0^\infty$ with respect to the norm $\|\cdot\|_\infty$ is c_0, the space of sequences whose terms tend to zero. On the other hand, the completion of $\mathbb{F}_0^\infty$ with respect to the norms $\|\cdot\|_p,\ 1 \le p < \infty$ is ℓ^p.

However if we consider the space of continuous functions $C[0,1]$ with the L^p norms

$$\|f\|_{L^p} \ = \ \Big(\int_0^1 |f(x)|^p \, dx \Big)^{1/p} \quad \text{for} \ \ 1 \le p < \infty$$

which are all examples of incomplete spaces, it is a very interesting matter to understand concretely what are the elements of the completion. It turns out the completion of $C[0,1]$ with respect to $\|\cdot\|_{L^p}$ is the space of Lebesgue p-integrable functions which is usually denoted by $L^p[0,1]$. So in this case, the abstract completion we discussed above in terms of equivalence classes of Cauchy sequences of continuous functions can be realised as actual functions. These functions can be highly discontinuous (they can even be discontinuous at *every point in* $[0,1]$!). Such a description is not a prerequisite for this course but it is covered in the course Essentials of Analysis and Probability.

2.20. **Compact sets in normed linear spaces.**

As mentioned previously, it is an important endeavour to understand what are the compact sets in a given metric space. It is desirable to find simple properties which characterise the compact sets in a particular setting. From the discussion about completeness above, we motivated the property of completeness by solving some equation by successive approximations. We remarked that the approximations hopefully formed a Cauchy sequence. But this is often too difficult to verify (and may not be true). Instead it is much easier to verify that the sequence of approximations lies in some compact set, especially if we had some knowledge of what compact sets look like. If this is the case, then we can extract a subsequence which does indeed converge to some point in our compact set, arriving at our solution of the original equation!

Let us begin by recalling the definition of compact sets in a metric space.

Definition 2.21. A subset K of metric space (X, d) is **compact** if every open covering of K has a finite subcover.

Let us also recall some standard consequences of this definition.

Proposition 2.22. *A set K in a metric space (X, d) is compact if and only if every infinite sequence $\{x_n\}$ in K has a subsequence which converges to some point in K. Furthermore a compact set K is necessarily closed.*

We will recall soon the useful Heine-Borel theorem but before we do so, we make a definition and establish a proposition which will help put the theorem in context.

Definition 2.23. A set E in a normed linear space $(X, \|\cdot\|)$ is said to be **bounded** if there is an $R > 0$ such that $E \subset B_R(0)$.

Proposition 2.24. *If K is a compact subset of a normed linear space $(X, \|\cdot\|)$, then K is closed and bounded.*

Proof. Suppose K is compact. Then by Proposition 2.22, K must necessarily be a closed set.

Now let us consider the following open cover of K,

$$K \subset \bigcup_{x \in K} B_1(x).$$

Since K is compact, we can extract a finite subcover; that is, there exists a finite sequence $\{x_1, x_2, \ldots, x_N\} \subset K$ such that

$$K \subset \bigcup_{j=1}^{N} B_1(x_j).$$

We claim that $K \subset B_R(0)$ where $R = 1 + \max_{1 \le k \le N} \|x_k\|$. In fact if $x \in K$, then $x \in B_1(x_j)$ for some $1 \le j \le N$. Hence

$$\|x\| = \|x - x_j + x_j\| \le \|x - x_j\| + \|x_j\| < 1 + \max_{1 \le k \le N} \|x_k\| = R,$$

showing that $K \subset B_R(0)$. □

Now for the statement of the Heine-Borel theorem from your Honours Analysis course.

Theorem 2.25. *A set K in $(\mathbb{F}^n, \|\cdot\|_p)$ is compact if and only if K is closed and bounded.*

By Proposition 2.24, we know that a compact set in a normed linear space must necessarily be closed and bounded. The Heine-Borel theorem shows that in certain finite dimensional spaces, these two simple properties, being closed and bounded, in fact characterise compact sets. We will see momentarily that these two properties characterise compact sets in *any* finite dimensional normed linear space.

If you have only seen Theorem 2.25 for the case $p = 2$ (that is, only for the euclidean distance), this is fine since the general case follows from the $p = 2$ case. The reason for this is that all these norms are equivalent. See the discussion after the proof of Proposition 2.29 below.

Now let us examine compact sets in the setting of a general finite dimensional normed linear space X with the p-norm defined with respect to a basis $\mathcal{B} = \{e_1, \ldots, e_n\}$ as described in Example B. above; that is,

$$\|x\|_{p,\mathcal{B}} \;=\; \big(|x_1|^p + \cdots + |x_n|^p\big)^{1/p}$$

where $x = x_1 e_1 + \cdots + x_n e_n$. The linear isomorphism $T : (X, \|\cdot\|_{p,\mathcal{B}}) \to (\mathbb{F}^n, \|\cdot\|_p)$ given by $Tx = (x_1, x_2, \ldots, x_n)$ satisfies $\|Tx\|_p = \|x\|_{p,\mathcal{B}}$ for all $x \in X$ and so T gives a 1-1 correspondence between open sets in $(X, \|\cdot\|_{p,\mathcal{B}})$ and $(\mathbb{F}^n, \|\cdot\|_p)$. Hence from the definition of compact sets, it also gives a 1-1 correspondence between the compact sets of both these spaces. Therefore we have the following consequence of Theorem 2.25.

Corollary 2.26. *Let X be a finite dimensional vector space with basis $\mathcal{B}$. For any $1 \le p \le \infty$, a set K in $(X, \|\cdot\|_{p,\mathcal{B}})$ is compact if and only if K is closed and bounded.*

The situation in <u>infinite</u> dimensions is very different. It turns out that even the closed unit ball, $\overline{B_1(0)}$, the quintessential *closed and bounded* set, is **not** compact. See Theorem 2.33 below.

2.27. **Equivalence of norms.**

Given a vector space X over $\mathbb{F}$, we consider two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X. This gives rise to two normed linear spaces, $(X, \|\cdot\|_1)$ and $(X, \|\cdot\|_2)$.

Definition 2.28. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X are said to be **equivalent** if there is a constant A such that

$$\|x\|_1 \le A\|x\|_2 \quad \text{and} \quad \|x\|_2 \le A\|x\|_1 \quad \text{for all } x \in X. \tag{A}$$

As an exercise, show that the notion of equivalent norms is in fact an equivalence relation.

The two inequalities (A) above defining when two norms are equivalent can be reformulated in terms of balls. Let $B_r^1(x) = \{y \in X : \|y - x\|_1 < r\}$ denote the ball of radius $r > 0$ and centre x with respect to the norm $\|\cdot\|_1$. Similarly we define $B_r^2(x)$. We then have

$$B_r^1(x) \subset B_{Ar}^2(x) \quad \text{and} \quad B_r^2(x) \subset B_{Ar}^1(x) \tag{B}$$

for all $r > 0$ and all $x \in X$. Note that since $B_r^1(x) = x + rB_1^1(0)$ and $B_r^2(x) = x + rB_1^2(0)$, we see that (B) is equivalent to showing that

$$B_1^1(0) \subset B_A^2(0) \quad \text{and} \quad B_1^2(0) \subset B_A^1(0). \tag{B'}$$

We leave this equivalent formulation of the notion equivalence of norms as an exercise.

Although two equivalent norms are not the same and they can give different geometries on X, they do give the same class of convergent sequences, the same Cauchy sequences, the same open, closed and compact sets.

Proposition 2.29. *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two equivalent norms on X. Then*

- *If $\{x_n\}$ converges to x with respect to one of these norms, it converges with respect to the other norm.*
- *If $G \subset X$ is an open set with respect to one of these norms, it is an open set with respect ot the other norm.*
- *The normed linear space $(X, \|\cdot\|_1)$ is a Banach space if and only if $(X, \|\cdot\|_2)$ is a Banach space.*

Proof. Let A be the constant arising in the inequalities (A).

Suppose the $\{x_n\}$ converges to x with respect to $\|\cdot\|_1$. From the definition of sequential convergence, we know that given any $\epsilon > 0$, there is an N such that $\|x_n - x\|_1 < \epsilon/A$ whenever $n \geq N$. Therefore if $n \geq N$,

$$\|x_n - x\|_2 \;\leq\; A\,\|x_n - x\|_1 \;<\; A\,\epsilon/A \;=\; \epsilon$$

which implies that $\{x_n\}$ converges to x with respect to the norm $\|\cdot\|_2$.

Now suppose that $G \subset X$ is an open set with respect to $\|\cdot\|_1$. Our goal is to show that G is open with respect to $\|\cdot\|_2$; that is, given any $x \in G$, we need to find an $r > 0$ such that $B_r^2(x) \subset G$. Since G is open with respect to $\|\cdot\|_1$, we can find an $r > 0$ such that $B_{Ar}^1(x) \subset G$. Hence by (B), we have

$$B_r^2(x) \;\subset\; B_{Ar}^1(x) \;\subset\; G,$$

completing the second part of the proposition.

Suppose $(X, \|\cdot\|_1)$ is a Banach space and let $\{x_n\} \subset X$ be a Cauchy sequence with respect to $\|\cdot\|_2$; that is, $\|x_m - x_n\|_2 \to 0$ as $m, n \to \infty$. Since $\|x_m - x_n\|_1 \leq A\|x_m - x_n\|_2$, we see that $\{x_n\}$ is a Cauchy sequence with respect to $\|\cdot\|_1$. Hence

$\|x_n - x\|_1 \to 0$ as $n \to \infty$ for some $x \in X$. But $\|x_n - x\|_2 \le A\|x_n - x\|_1$ and so $x_n \to x$ with respect to $\|\cdot\|_2$ and this shows that $(X, \|\cdot\|_2)$ is a Banach space. □

Let us consider the space $\mathbb{F}^n$ with the various norms $\|\cdot\|_p$, $1 \le p \le \infty$ defined on it. From Exercise 5 below, we have the relation $\|x\|_q \le \|x\|_p$ whenever $q \ge p$. On the other hand, $\|x\|_1 \le n\|x\|_\infty$. In fact,

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n| \ \le \ n \max_{1\le j\le n} |x_j| \ = \ \|x\|_\infty.$$

Hence all the norms in the family $\{\|\cdot\|_p\}_{1\le p\le\infty}$ are equivalent. In particular the family of compact sets on $\mathbb{F}^n$ is the same with respect to any of the norms $\|\cdot\|_p$. So the Heine-Borel theorem for $p = 2$ implies that the compact sets in $(\mathbb{F}^n, \|\cdot\|_p)$ are precisely those sets which are closed and bounded.

We now come to important result.

Theorem 2.30. *Let X be a finite dimensional vector space over $\mathbb{F}$. Then all norms on X are equivalent.*

Hence on a finite dimensional vector space X, there is a *unique* class of convergent sequences and a *unique* class of open sets which come from any norm. So two norms on a finite dimensional space X must necessarily give rise to the same topology on X.

Proof. Let us fix a basis $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ on X and consider the norm

$$\|x\|_{\infty,\mathcal{B}} \ := \ \sup_{1\le j\le n} |x_j|$$

where $x = x_1e_1 + \cdots + x_ne_n$. Since equivalence of norms is an equivalence relation (see Exercise 16), it suffices to show that any other norm $\|\cdot\|$ on X is equivalent to $\|\cdot\|_{\infty,\mathcal{B}}$.

First we observe that if $x = x_1e_1 + \cdots + x_ne_n$,

$$\|x\| \le |x_1|\|e_1\| + |x_2|\|e_2\| + \cdots + |x_n|\|e_n\| \le A \max_{1\le j\le n} |x_j| = A\|x\|_{\infty,\mathcal{B}}$$

where $A = \sum_{j=1}^n \|e_j\|$. This gives us one of the inequalities in (A) which we need to establish. It also gives some useful information. For example, it shows that the function $f : (X, \|\cdot\|_{\infty,\mathcal{B}}) \to \mathbb{R}$ defined by $f(x) = \|x\|$ is a continuous function everywhere. In fact, if $x_n \to x$ in $(X, \|\cdot\|_{\infty,\mathcal{B}})$, then $\|x_n - x\| \le A\|x_n - x\|_{\infty,\mathcal{B}} \to 0$ as $n \to \infty$. Hence $\|x_n\| \to \|x\|$ (see Exericse 1) and so $f(x_n) \to f(x)$ as $n \to \infty$.

Now we use the key fact that the unit sphere $S_\infty := \{x \in X : \|x\|_{\infty,\mathcal{B}} = 1\}$ is a compact set in $(X, \|\cdot\|_{\infty,\mathcal{B}})$. This follows from Corollary 2.26 since S_∞ is closed and bounded. This implies that $\delta := \inf_{x\in S_\infty} \|x\| > 0$. In fact from the definition of infimum, there is a sequence $\{x_n\} \subset S_\infty$ such that $\|x_n\| \to \delta$. Since S_∞ is compact with respect to $\|\cdot\|_{\infty,\mathcal{B}}$, we can extract a subsequence $\{x_{n_k}\}$ which converges to some point $x \in S_\infty$; that is, $\|x_{n_k} - x\|_{\infty,\mathcal{B}} \to 0$ as $k \to \infty$. By the continuity of $f(x) = \|x\|$ on $(X, \|\cdot\|_{\infty,\mathcal{B}})$, we deduce that $\|x_{n_k}\| \to \|x\|$. Since $\|x_{n_k}\| \to \delta$ as well, we conclude that $\|x\| = \delta$. If $\delta = 0$, then $\|x\| = 0$ and hence $x = 0$ which is impossible since $x \in S_\infty$. Therefore $\delta > 0$.

For any $x \neq 0$, set $y := x/\|x\|_\infty \in S_\infty$ and so $\delta \leq \|y\|$ or $\delta\|x\|_\infty \leq \|x\|$ which gives the desired reverse inequality with $A = 1/\delta$. □

We have the immediate consequence of Theorem 2.30.

Corollary 2.31. *Let $(X, \|\cdot\|)$ be a finite dimensional normed linear space. Then $(X, \|\cdot\|)$ is a Banach space. Furthermore, if Y is a linear subspace of X, then Y is closed.*

Proof. The first part is a consequence of Theorem 2.30 and the fact that the metric space $(\mathbb{F}^n, d_2)$ is complete.

The second part is a consequence of the first part and Lemma 2.15. □

Using Theorem 2.30, we can now extend the Heine-Borel theorem on finite dimensional spaces, Corollary 2.26, from the special norm case to all norms.

Corollary 2.32. *Let $(X, \|\cdot\|)$ be a finite dimensional normed linear space. A subset $K \subset X$ is compact if and only if K is closed and bounded.*

We have arrived at an important result that states that Corollary 2.32 does not extend to infinite dimensions.

Theorem 2.33. *Let $(X, \|\cdot\|)$ be a normed linear space. Then $\overline{B_1(0)}$ is compact if and only if X is finite dimensional.*

The following proof simplifies in the setting of a Hilbert space. We will return to this later.

Proof. * If X is finite dimensional, the result follows from Corollary 2.32. Now suppose the X is infinite dimensional. Our goal is to construct an infinite sequence $\{e_n\}$ of unit vectors with the property that $\|e_m - e_n\| \geq 1/2$ whenever $m \neq n$. This then will give an infinite sequence in $\overline{B_1(0)}$ with no convergent subsequence and so $\overline{B_1(0)}$ cannot be compact.

We will construct this sequence by induction. Suppose that we have a finite sequence $\{e_1, e_2, \ldots, e_n\}$ of unit vectors with the property that $\|e_j - e_k\| \geq 1/2$ for all $1 \leq j \neq k \leq n$. Let $\mathcal{L}_n = \text{span}\{e_1, \ldots, e_n\}$. We will find a unit vector $e \notin \mathcal{L}_n$ such that $\|e - e_j\| \geq 1/2$ for all $1 \leq j \leq n$. If we can do this, then this process will produce the desired *infinite* sequence since X is assumed to be infinite dimensional and so $X \neq \mathcal{L}_n$ for any n.

Now fix n. Since $X \neq \mathcal{L}_n$, we can find an $f \in X \setminus \mathcal{L}_n$. Set $\delta = \inf\{\|f - y\| : y \in \mathcal{L}_n\}$. We claim that $\delta > 0$. Suppose $\delta = 0$. Then by the definition of the infimum, there is a sequence $\{y_k\} \subset \mathcal{L}_n$ such that $\|y_k - f\| \to 0$ and so $y_k \to f$. But by Corollary 2.31, the subspace $\mathcal{L}_n$ is closed and hence $f \in \mathcal{L}_n$ which is false. Hence $\delta > 0$. Therefore we can find a vector $y \in \mathcal{L}_n$ such that $\|y - f\| \leq 2\delta$. We claim the unit vector $e = (f-y)/\|f-y\|$ does the job.

First note that $e \notin \mathcal{L}_n$ since if it was, f would have to lie in $\mathcal{L}_n$ since $y \in \mathcal{L}_n$. Finally, we have for any $1 \leq k \leq n$,

$$\|e - e_k\| = \frac{1}{\|f-y\|}\|f - y - \|f-y\|e_k\| \geq \delta/(2\delta) = 1/2$$

since $y - \|f-y\|e_k \in \mathcal{L}_n$. □

An aside* As we have seen, a norm gives rise to a family of balls $\{B_r(0)\}$ which in turn is generated by a single open, convex and bounded set, the unit ball $B_1(0)$. We will now see that all equivalent norms are characterised by some open, convex, bounded (and symmetric) set.

Definition 2.34. A subset E of a normed linear space $(X, \|\cdot\|)$ is said to be **symmetric** if $x \in E \Leftrightarrow -x \in E$.

The notion of being symmetric only depends on the underlying vector space structure of X.

Now we will bring convexity into the discussion. Hence we will restrict ourselves to vector spaces with real scalars for the time being (recall that any vector space over $\mathbb{C}$ can be viewed as a vector space over $\mathbb{R}$ as well). We begin by observing that if $E \subset X$ is a nonempty, convex and symmetric set, then necessarily $0 \in E$. In fact, if $x \in E$, then $-x \in E$ and hence $[-x, x] = \{\lambda x - (1-\lambda)x : \lambda \in [0,1]\} \subset E$ and in particular, $0 = (1/2)x - (1/2)x \in E$. Furthermore, the unit ball $B_1(0)$ is a symmetic, open, convex set which is nonempty.

By (B), we see that if two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X are equivalent, then the unit ball $B_1^2(0)$ with respect to one of the norms is a bounded set with respect to the other norm. Of course, $B_1^2(0)$ is also open, convex and symmetric – these properties do not depend on the other norm $\|\cdot\|_1$.

Now suppose we start with a nonempty, bounded, symmetric, convex and open subset C of a real normed linear space $(X, \|\cdot\|)$. We will see that there is an *equivalent* norm $\|\cdot\|_C$ on X such that C is the unit ball with respect to $\|\cdot\|_C$; that is, $C = \{x \in X : \|x\|_C < 1\}$. This is consistent with the discussion above.

The proposed norm is given by

$$\|x\|_C \;:=\; \inf\{t > 0 : t^{-1}x \in C\} \qquad (MF)$$

whenever $x \neq 0$. We set $\|0\|_C = 0$. This is sometimes referred to as the *Minkowski functional* of C. Since C is open and $0 \in C$, there is an $r > 0$ such that $B_r(0) \subset C$. Hence for any $x \neq 0$, there is an $s > 0$ such that $sx \in B_r(0) \subset C$ (take any s with $0 < s < r/\|x\|$) and so $\|x\|_C$ is well-defined and finite.

Proposition 2.35. *Let C be a nonempty, open, symmetric and bounded subset of a real normed linear space $(X, \|\cdot\|)$. Then the functional $\|\cdot\|_C$ given in (MF) defines an equivalent norm on X. Furthermore $C = \{x \in X : \|x\|_C < 1\}$ is the unit ball with respect to $\|\cdot\|_C$.*

Proof. We begin by verifying that the axioms of a norm are satisfied by $\|\cdot\|_C$. First of all since C is bounded, there is an $R > 0$ such that $C \subset B_R(0)$. Hence if $\|x\|_C = 0$, then for every $t > 0$, $t^{-1}x \in C \subset B_R(0)$. This implies that $\|x\| < tR$ for every $t > 0$ which in turn implies that $\|x\| = 0$ and so $x = 0$.

Secondly from the definition of $\|\cdot\|_C$, we immediately see that $\|sx\|_C = s\|x\|_C$ for all $s > 0$. If $s < 0$, then since

$$t^{-1}sx \in C \qquad \Longleftrightarrow \qquad t^{-1}(-sx) \in C,$$

for all $t > 0$, we see that $\|sx\|_C = \|(-s)x\|_C = (-s)\|x\|_C = |s|\|x\|_C$. This establishes the second axiom for a norm.

For the triangle inequality, let $x, y \in X$ and consider any $s, t > 0$ such that $\|x\|_C < s$ and $\|y\|_C < t$. Set $u = s + t$. Hence $s^{-1}x, t^{-1}y \in C$ and since C is convex, we must have

$$u^{-1}(x + y) = \frac{s}{u}\big(s^{-1}x\big) + \frac{t}{u}\big(t^{-1}y\big) \ \in \ C$$

which implies $\|x + y\|_C \le u$ and so $\|x + y\|_C \le \|x\|_C + \|y\|_C$. This establishes the triangle inequality for $\|\cdot\|_C$.

Next, if $x \in C$, then since C is open, we can find an $\epsilon > 0$ such that $t^{-1}x \in C$ for $1 - \epsilon < t < 1$. Hence $\|x\|_C < 1$. On the other hand, suppose that $\|x\|_C < 1$. Then there exists a $t < 1$ such that $t^{-1}x \in C$. Since $0 \in C$, we know that

$$x \ = \ t(t^{-1}x) + (1 - t)0 \ \in \ C$$

due to the convexity of C. Therefore $C = \{x \in X : \|x\|_C < 1\}$ is the unit ball with respect to the norm $\|\cdot\|_C$.

Finally if we denote by $B_r^C(x)$ the ball of radius $r > 0$ and centre x with respect to the norm $\|\cdot\|_C$, then since C is bounded, we can find an $R > 0$ such that $B_1^C(0) = C \subset B_R(0)$. On the other hand, since $0 \in C$ and C is open, we can find an $r > 0$ such that $B_r(0) \subset C = B_1^C(0)$. This implies that $B_1(0) \subset B_{1/r}^C(0)$. By taking $A = \max(R, 1/r)$, we see that property (B') is satisfied and this says that $\|\cdot\|$ and $\|\cdot\|_C$ are equivalent norms. $\square$

EXERCISES

1. Let $(X, \|\cdot\|)$ be a normed linear space. Show that $\|x - y\| \ge |\|x\| - \|y\||$ holds for every $x, y \in X$. This inequality is equivalent to $\|x + y\| \ge |\|x\| - \|y\||$ for all $x, y \in X$. Why is this? Use this to show that $f : (X, \|\cdot\|) \to \mathbb{R}$ defined by $f(x) = \|x\|$ is continuous everywhere.

2. Show that for an n-tuple of positive numbers $a_1, \dots, a_n$, the inequality

$$\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \;\le\; a_1 + a_2 + \cdots + a_n$$

holds. Hint: try the $n = 2$ case first to get some idea of how to proceed. By letting $n \to \infty$, conclude that

$$\Big(\sum_{n=1}^{\infty} a_n^2\Big)^{1/2} \;\le\; \sum_{n=1}^{\infty} a_n$$

holds for all sequences $\{a_n\}$ of nonnegative numbers.

3. Consider the norms $\|x\|_p$ on $\mathbb{R}^n$. For every $x \in \mathbb{R}^n$, show that $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$.

4. For $0 < p < 1$, show that $(1+a)^p \le 1 + a^p$ holds for all $a > 0$. Hint: use calculus to prove an inequality $f(x) \le g(x)$ by examining the derivatives f', g'. Use this result to complete the proof that ℓ^p is indeed a subspace of $\mathbb{F}^\infty$ when $0 < p < 1$.

5. Show that $(a_1 + a_2 + \cdots + a_n)^p \le a_1^p + a_2^p + \cdots + a_n^p$ holds for $0 < p < 1$ and nonnegative numbers $a_1, \dots, a_n \ge 0$. Hint: consider the $n = 2$ case first and iterate. Next, extend this to infinite sequences; that is, show

$$\Big(\sum_{j=1}^{\infty} a_j\Big)^p \;\le\; \sum_{j=1}^{\infty} a_j^p \tag{$\le -p$}$$

holds for all nonnegative sequences $\{a_j\}$. Use this to give an alternative proof of the inequalities in Exercise 2 above. Finally use $(\le -p)$ to show that $\|x\|_q \le \|x\|_p$ holds for all $x \in \mathbb{F}^\infty$ whenever $q \ge p$.

6. Show that the triangle inequality fails for $\|\cdot\|_p$ on $\mathbb{R}^n$ when $0 < p < 1$. Hint: first consider the case $n = 2$ and examine the picture that $\{x : \|x\|_p = 1\}$ makes.

7. Let g be a nonnegative, continuous function on $[0, 1]$ whose integral $\int_0^1 g(x)dx = 0$. Show that this implies that $g(x) = 0$ for every $x \in [0, 1]$. Next, verify that the uniform norm verifies the axioms for a norm.

8. Verify that $B_r(x) = x + rB_1(0)$ in a normed linear space.

9. Let X be a vector space over the reals $\mathbb{R}$. If $C \subset X$ is convex, show that every translate $x + C$ and every dilate rC of C is convex. Furthermore, show that if $x_1, x_2, \dots, x_n \in C$, then $\mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_n x_n \in C$ where $\mu_j \ge 0$ and $\mu_1 + \cdots + \mu_n = 1$. Hint: Induct on n.

10. Show that every ball $B_r(x)$ in a normed linear space is an open set. Furthermore, show that $\overline{B_r(x)} := \{y \in X : \|y - x\| \le r\}$ is closed.

11. Show that the norm in (Lp) with $p = 1$ defines a norm on the space $\mathcal{P}_d$ or $\mathbb{F}[X]$.

12. Give the details of the proof of Proposition 2.7.

13. Show that (C) holds if and only if (B) holds. Warning: the implication $(B) \Rightarrow (C)$ requires a little care.

14.* Try to mimic the proof of Minkowski's inequality to establish the following inequality: for every $p \geq 1$,

$$\Big(\int_0^1 \Big[\int_0^1 |f(x,y)| dy\Big]^p dx\Big)^{1/p} \ \leq \ \int_0^1 \Big[\int_0^1 |f(x,y)|^p dx\Big]^{1/p} dy$$

holds for every continuous $F \in C([0,1] \times [0,1])$.

15. Let $(X, \|\cdot\|)$ be a normed linear space over the reals $\mathbb{R}$. Show that the unit ball (or for that matter, any ball) $B_1(0)$ is an open, convex, symmetric and bounded set in X which is nonempty.

16. Show that the notion of equivalence of norms is an equivlance relation. That is if we say $\|\cdot\|_1 \sim \|\cdot\|_2$ when the norms satisfy the inequalities in (A), then $\|\cdot\| \sim \|\cdot\|$ for all norms $\|\cdot\|$. Also if $\|\cdot\|_1 \sim \|\cdot\|_2$, then $\|\cdot\|_2 \sim \|\cdot\|_1$. Finally, if $\|\cdot\|_1 \sim \|\cdot\|_2$ and $\|\cdot\|_2 \sim \|\cdot\|_3$, then $\|\cdot\|_1 \sim \|\cdot\|_3$.

17. Let X be a 1-dimensional vector space over $\mathbb{F}$. Characterise all norms on X.

18. Every vector space X has a Hamel basis; that is, there is a collection $\{e_\alpha\}_{\alpha \in A}$ such that every $x \in X$ has a unique representation of a *finite* linear combination of the $\{e_\alpha\}_{\alpha \in A}$. Let us take this for granted.

Show that every vector space X over $\mathbb{F}$ supports some norm.

19. Show that $x \to \|x\|$ is a continuous function on a normed linear space $(X, \|\cdot\|)$.

20. Show that c_0 is closed in $(\ell^\infty, \|\cdot\|_\infty)$.

21. Suppose that $\{x_j\}_{j \geq 1}$ is a sequence of scalars such that $x_j \to 0$ as $j \to \infty$. Show that $\sup_{j>n} |x_j| \to 0$ as $n \to \infty$.

22. In Example **E**, we noted that $(\ell^p, \|\cdot\|_q)$ is a normed linear space whenever $q \geq p$. In fact $\|x\|_q \leq \|x\|_p$ whenever $p \leq q$ (see Exercise 5 above). By completing the following steps, show that when $p < q$, $(\ell^p, \|\cdot\|_q)$ is not a Banach space.

Step 1. Consider $x = (1/j^{1/p})_{j \geq 1}$ and the approximating sequence $\{x^n\}_{n \geq 1}$ where

$$x^n \ = \ (1, \ \frac{1}{2^{1/p}}, \ \frac{1}{3^{1/p}}, \ \ldots, \ \frac{1}{n^{1/p}}, \ 0, 0, 0, \ldots).$$

Show that $\{x^n\}_{n \geq 1}$ is a Cauchy sequence in $\ell^p, \|\cdot\|_q)$.

Step 2. Suppose that there is a $y = (y_j)_{j \geq 1} \in \ell^p$ such that $\|x^n - y\|_q \to 0$ as $n \to \infty$. Show that this implies that $y = x$, giving a contradition since $x \notin \ell^p$. Hint: use $|y_j - x_j^n| \leq \|y - x^n\|_q$ which holds for any $j \geq 1$.

23. Let $\mathbb{F}_0^\infty = \{y = (y_j)_{j\geq 1} : y_j = 0 \text{ for all but finitely many } j\}$ be the space of all *finite* sequences. Clearly $\mathbb{F}_0^\infty$ is a linear subspace of ℓ^p for any $1 \leq p \leq \infty$. Show that when $1 \leq p < \infty$, $\overline{\mathbb{F}_0^\infty} = \ell^p$. Hence $\mathbb{F}_0^\infty$ is a dense subspace of ℓ^p when $1 \leq p < \infty$.

24.* Fill in the details of the completion procedure of any normed linear space $(X, \|\cdot\|)$ given above. In particular, you should start with verifying that $\sim$ is indeed an equivalence relation on the space C of all Cauchy sequences of X. Recall that we say two Cauchy sequences $\underline{x} = \{x_j\}_{j\geq 1}$ and $\underline{y} = \{y_j\}_{j\geq 1}$ of X are equivalent if $\|x_j - y_j\| \to 0$ as $j \to \infty$. We denote this as $x \sim y$.

25. Let Y be a dense subpace of a normed linear space $(X, \|\cdot\|)$ and suppose that $T : Y \to Z$ is an isometry into a Banach space $(Z, \|\cdot\|_*)$. Show there exists a unique isometry $f : X \to Z$ such that $f(y) = T(y)$ for all $y \in Y$.

Hint: if $x \in X$, then there is a sequence $\{y_n\}_{n\geq 1} \subset Y$ such that $\|y_n - x\| \to 0$ as $n \to \infty$. Then show that $\{T(y_n)\}_{n\geq 1}$ is a Cauchy sequence in Z and hence converges to some point $z \in Z$. Define $f(x) = z$. You need to show that this definition does not depend on the approximating sequence $\{y_n\}$ to x.

3. Inner product spaces and Hilbert spaces

A very important example of a normed linear space is an *inner product space* which you studied in Honours Algebra. We have already encountered some inner product spaces; the examples $(\mathbb{F}^n, \|\cdot\|_2)$, $(\ell^2, \|\cdot\|_2)$ and $(C[0,1], \|\cdot\|_{L^2})$ are all examples of inner product spaces. In each case, the norm arises from an *inner product.* Some of the following will be a review of your Honours Algebra course.

Definition 3.1. Let X be a vector space over $\mathbb{F}$. An **inner product** on X is a map $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{F}$ which satisfies the following properties:

- $\langle x, x\rangle \geq 0$ for all $x \in X$ with equality holding if and only $x = 0$;
- $\langle x, y\rangle = \overline{\langle y, x\rangle}$ for all $x, y \in X$;
- $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$ for all $x, y, z \in X$ and all $\alpha, \beta \in \mathbb{F}$.

When the scalars are real, $\mathbb{F} = \mathbb{R}$, we have $\overline{\alpha} = \alpha$ for all $\alpha \in \mathbb{F}$ and so the three properties in this case simply say that $\langle \cdot, \cdot\rangle$ is a positive definite, symmetric bilinear form on X. When $\mathbb{F} = \mathbb{C}$, the second and third properties say that $x \to \langle x, y\rangle$ is linear for every $y \in X$ and $y \to \langle x, y\rangle$ is conjugate linear for every $x \in X$ (see Exercise 1 below); altogether the three properties in this case say that $\langle \cdot, \cdot\rangle$ is a positive definite, Hermitian form on X.

The pair $(X, \langle \cdot, \cdot\rangle)$ is called an **inner product space**.

The notion of an inner product gives rise to a norm, $\|x\| := \sqrt{\langle x, x\rangle}$, but before we verify this, let us look at some examples of inner product spaces.

3.2. **Examples.**

A. $\mathbb{F}^n$: for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, we define

$$\langle x, y\rangle \;=\; x_1\overline{y_1} + \cdots + x_n\overline{y_n}$$

and observe $\sqrt{\langle x, x\rangle} = \|x\|_2$. It is an easy but worthwhile exercise to verify the axioms of an inner product for this example.

E. ℓ^2: for $x = (x_j)_{j\geq 1}, y = (y_j)_{j\geq 1} \in \ell^2$, we define

$$\langle x, y\rangle \;=\; \sum_{j=1}^{\infty} x_j\overline{y_j}$$

and observe $\sqrt{\langle x, x\rangle} = \|x\|_2$. Note that the infinite series defining the inner product above converges by Hölder's inequality.

F. $C[0,1]$: for $f, g \in C[0,1]$, we set

$$\langle f, g\rangle \;=\; \int_0^1 f(x)\overline{g(x)}\, dx$$

and observe that $\sqrt{\langle f, f\rangle} = \|f\|_{L^2}$.

3.3. **The Cauchy-Schwarz inequality, the parallelogram law and the polarisation identity.**

As indicated above, the notion of an inner product gives rise to a norm, $\|x\| := \sqrt{\langle x, x\rangle}$. The following fundamental inequality, called the *Cauchy-Schwarz inequality*, will be used to show that $\|\cdot\|$ defined above is indeed a norm.

Theorem 3.4. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. For any $x \in X$, we set $\|x\| = \sqrt{\langle x, x\rangle}$. Then*

$$|\langle x, y\rangle| \ \le \ \|x\|\|y\| \tag{CS}$$

holds for any $x, y \in X$.

Observe that (CS) in the cases of the inner product $\langle\cdot,\cdot\rangle$ defined in the Examples **A.**, **E.** and **F.** above follow from Hölder's inequality.

Proof. We may assume x and y are both nonzero, otherwise both sides of (CS) are equal to zero (see Exercise 3 below). Replacing x with $x/\|x\|$ and y with $y/\|y\|$, we may assume $\|x\| = \|y\| = 1$.

Set $z = x - \langle x, y\rangle y$ and note that

$$\langle z, y\rangle = \langle x, y\rangle - \langle x, y\rangle\langle y, y\rangle = \langle x, y\rangle - \langle x, y\rangle = 0.$$

Thus

$$\langle x, x\rangle = \langle z + \langle x, y\rangle y, z + \langle x, y\rangle y\rangle = \langle z, z\rangle + \langle x, y\rangle\overline{\langle x, y\rangle}\langle y, y\rangle = \langle z, z\rangle + |\langle x, y\rangle|^2$$

and so

$$\|x\|^2 \ = \ \|z\|^2 + |\langle x, y\rangle|^2 \tag{P}$$

which shows $\|x\|^2 \ge |\langle x, y\rangle|^2$ or $|\langle x, y\rangle| \le \|x\|$, establishing the Cauchy-Schwarz inequality since $\|y\| = 1$. $\square$

There are hundreds of different proofs of the Cauchy-Schwarz inequality. The above proof is motivated by some intuition gained from examining vectors in $\mathbb{R}^2$. Recall from school geometry, if $x, y \in \mathbb{R}^2$ where y is a unit vector, then $\mathrm{P}_y(x) := \langle x, y\rangle y$ is the orthogonal projection of the vector x onto the line given by y. Hence the vectors $x, \mathrm{P}_y(x)$ and $z = x - \mathrm{P}_y(x)$ form the three sides of a right triangle where x is the hypotenuse. So $\mathrm{P}_y(x)$ and z are perpendicular, $\mathrm{P}_y(x) \perp z$ or $\langle z, \mathrm{P}_y(x)\rangle = 0$ and

$$\|x\|^2 \ = \ \|z\|^2 + \|\mathrm{P}_y(x)\|^2 \ = \ \|z\|^2 + |\langle x, y\rangle|^2$$

by the Pythagorean theorem. The proof of Theorem 3.4 simply realises that the two basic facts of plane geometry, namely $\mathrm{P}_y(x) = \langle x, y\rangle y$ and $z = x - \mathrm{P}_y x$ satisfy $\langle z, \mathrm{P}_y(x) y\rangle = 0$ and the corresponding Pythagorean theorem, holds in a general inner product space.

Equality holds in (CS) if and only if x and y are linearly dependent (see Exercise 4 below).

Proposition 3.5. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Then $\|x\| = \sqrt{\langle x, x\rangle}$ defines a norm on X.*

Proof. Suppose $\|x\| = 0$. Then $\langle x, x\rangle = 0$ and hence $x = 0$ by the first property of an inner product. Next we have

$$\|\alpha x\|^2 = \langle \alpha x, \alpha x\rangle = \alpha\langle x, \alpha x\rangle = \alpha\overline{\langle \alpha x, x\rangle} = \alpha\overline{\alpha}\overline{\langle x, x\rangle} = |\alpha|^2\langle x, x\rangle = |\alpha|^2\|x\|^2,$$

showing that the second property for a norm holds.

Finally we have

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle}$$
$$= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x, y\rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2,$$

implying the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$. In the above displayed equation, we used the Cauchy-Schwarz inequality in the form

$$\mathrm{Re}\langle x, y\rangle \ \leq \ |\langle x, y\rangle| \ \leq \ \|x\|\|y\|,$$

using the fact that $\mathrm{Re}\, z = x \leq |x| \leq \sqrt{x^2 + y^2} = |z|$ for any complex number $z = x + iy$. □

Notation: We write $x \perp y$ for two vectors $x, y \in X$ in an inner product space $(X, \langle\cdot,\cdot\rangle)$ if $\langle x, y\rangle = 0$ holds. In this case, we say x and y are **perpendicular** or **orthogonal**.

The statement (P) in the proof of the Cauchy-Schwarz inequaltiy is a special case of the Pythagorean theorem which we now state formally.

Proposition 3.6. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. If $x, y \in X$ are orthogonal, then*

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

Proof. From the proof of Proposition 3.5, we see that

$$\|x + y\|^2 \ = \ \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\,\langle x, y\rangle \qquad (S)$$

holds for any $x, y \in X$. Therefore if $\langle x, y\rangle = 0$, we obtain (π). □

Observe that if we replace y with $-y$ in (S), we get $\|x-y\|^2 = \|x\|^2+\|y\|^2-\mathrm{Re}\,\langle x, y\rangle$. Adding this to (S) gives us the so-called parallelogram law which is valid in any inner product space.

Proposition 3.7. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. For any $x, y \in X$, the* **parallelogram law**

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

holds.

By examining the identity above in the two dimensional euclidean setting, $(\mathbb{R}^2, \|\cdot\|_2)$, explain why it is called the parallelogram law.

One could ask whether the parallelogram law holds in a general normed linear space $(X, \|\cdot\|)$. The answer is no (see Exercise 5 below). In fact one can show that if the parallelogram law holds for all vectors in a normed linear space $(X, \|\cdot\|)$, then there is an inner product $\langle\cdot,\cdot\rangle$ on X such that $\|x\| = \sqrt{\langle x,x\rangle}$ for all $x \in X$. To begin to see this, let us assume our scalars are real, $\mathbb{F} = \mathbb{R}$. Then again replacing y with $-y$ in (S), we have the two equations

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x,y\rangle \quad \text{and} \quad \|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x,y\rangle$$

as before (but with Re dropped since $\mathrm{Re}\langle x,y\rangle = \langle x,y\rangle$ when our scalars are real). Instead of adding these two equations together to arrive at the parallelogram law, we take the difference and arrive at

$$4\langle x,y\rangle \;=\; \|x+y\|^2 - \|x-y\|^2 \tag{RP}$$

which is what we call the **real polarisation identity**. So if a real normed linear space $(X, \|\cdot\|)$ satisfies the parallelogram law, we define $\langle\cdot,\cdot\rangle$ by (RP) and note $\langle x,x\rangle = \|x\|^2$. The verification of the first two axioms of an inner product is straightforward but to establish the linearity of $x \to \langle x,y\rangle$ for any fixed $y \in X$ for this "inner product" is not straightforward and takes some effort. See Exercise 6 below.

Proposition 3.8. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space with complex scalars $\mathbb{C}$. Then*

$$4\langle x,y\rangle \;=\; \sum_{k=0}^{3} i^k \|x + i^k y\|^2 = \|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2.$$

The identity in Proposition 3.8 is called the **complex polarisation identity**. See Exercise 8 below for variants of the polarisation identity.

Proof. Exactly as in the derivation of (S), we have

$$\|x+i^k y\|^2 = \langle x + i^k y, x + i^k y\rangle = \langle x,x\rangle + \langle y,y\rangle + i^{-k}\langle x,y\rangle + i^k\langle y,x\rangle$$

and so

$$i^k\|x+i^k y\|^2 = i^k\langle x + i^k y, x + i^k y\rangle = i^k\langle x,x\rangle + i^k\langle y,y\rangle + \langle x,y\rangle + i^{2k}\langle y,x\rangle.$$

Adding these identities for $k = 0, 1, 2$ and 3 gives the desired result. We simply note that $\sum_{k=0}^3 i^k = \sum_{k=0}^3 i^{2k} = 0$. □

3.9. **Orthogonality and geometry: some basics.**

In the proof of the Cauchy-Schwarz inequality, we introduced the orthogonal projection $P_y(x) = \langle x,y\rangle y$ of a vector x onto the linear span of the unit vector y, Span(y) $:= \{\alpha y : \alpha \in \mathbb{F}\}$. The map $P_y : X \to \mathrm{Span}(y)$ is a linear map from X onto the one dimensional subspace $\mathrm{Span}(y)$ which has the properties that $P_y(x) \perp (x - P_y(x))$ and

$$\|x - P_y(x)\| \;=\; \inf\big\{\|x - \alpha y\| : \alpha \in \mathbb{F}\big\}.$$

The latter property says that the vector $P_y(x)$ is the closest vector to x amongst all the vectors in $\mathrm{Span}(y)$. To see this, note that

$$\|x-\alpha y\|^2 = \langle x-\alpha y, x-\alpha y\rangle = \|x\|^2 + |\alpha|^2 - \overline{\alpha}\langle x,y\rangle - \alpha\overline{\langle x,y\rangle} = \|x\|^2 - |\langle x,y\rangle|^2 + |\alpha - \langle x,y\rangle|^2$$

which is clearly minimised when $\alpha = \langle x,y\rangle$. This calculation also shows that $P_y(x)$ is the *unique* closest point to $\mathrm{Span}(y)$.

The above formula for $\|x - \alpha y\|^2$ extends readily to higher dimensional subspaces. First recall that if Y is a finite dimensional subspace of an inner product space $(X, \langle\cdot,\cdot\rangle)$, then it has an **orthonormal normal basis** or ONB; that is $Y = \mathrm{Span}(e_1, e_2, \ldots, e_n)$ where $e_j \perp e_k$ for any $j \neq k$ and $\|e_j\| = 1$ (physicists often write this as $\langle e_j, e_k\rangle = \delta_{j,k}$). A simple formula to verify but which is extremely useful is the following: if

$$y \ = \sum_{j=1}^{n} \alpha_j e_j, \quad \text{then} \quad \|y\|^2 \ = \ \sum_{j=1}^{n} |\alpha_j|^2. \tag{$\perp$}$$

Theorem 3.10. *Let Y be an n dimensional subspace of an inner product space $(X, \langle\cdot,\cdot\rangle)$ with ONB $\{e_j\}_{j=1}^n$. Then the linear map $P_Y : X \to Y$ given by $P_Y(x) = \sum_{j=1}^n \langle x, e_j\rangle e_j$ has the following properties.*

(a) $x - P_Y(x) \perp y$ *for all* $y \in Y$*;*
(b) $\sum_{j=1}^n |\langle x, e_j\rangle|^2 \ \leq \ \|x\|^2$*. This is* **Bessel's inequality** *and follows from the identity*

$$\|x - P_Y(x)\|^2 \ = \|x\|^2 - \sum_{j=1}^{n} |\langle x, e_j\rangle|^2 \ = \ \|x\|^2 - \|P_Y(x)\|^2;$$

(c) $P_Y(x)$ *is the unique closest point in* Y *to* x*; that is,*

$$\|x - P_Y(x)\| \ = \ \min\big\{\|x - y\| : y \in Y\big\}$$

and if $y \in Y$ *is such that* $\|x - y\| = \|x - P_Y(x)\|$*, then* $y = P_Y(x)$*.*

We call the map P_Y the **orthogonal projection** of X onto Y.

Proof. For any $y = \sum_{j=1}^n \alpha_j e_j \in Y$, we have

$$\langle x - P_Y(x), \ \sum_{j=1}^{n} \alpha_j e_j\rangle = \sum_{j=1}^{n} \overline{\alpha}_j \langle x, e_j\rangle \ - \sum_{j,k=1}^{n} \langle x, e_k\rangle \overline{\alpha}_j \langle e_k, e_j\rangle = 0$$

where the last equality follows from the orthonormal property of the basis vectors $\{e_j\}$.

In particular since $P_Y(x) \in Y$, we have $x - P_Y(x) \perp P_Y(x)$ and so (π) implies $\|x - P_Y(x)\|^2 = \|x\|^2 - \|P_Y(x)\|^2$. From $(\perp)$, we have $\|P_Y(x)\|^2 = \sum_{j=1}^n |\langle x, e_j\rangle|^2$ which in turn shows

$$0 \ \leq \ \|x - P_Y(x)\|^2 = \|x\|^2 - \sum_{j=1}^{n} |\langle x, e_j\rangle|^2 \ \text{ and so } \ \sum_{j=1}^{n} |\langle x, e_j\rangle|^2 \ \leq \ \|x\|^2.$$

For any $y = \sum_{j=1}^n \alpha_j e_j \in Y$, we have $\|x - y\|^2 =$

$$\|x\|^2 - \sum_{j=1}^{n} \overline{\alpha_j} \langle x, e_j\rangle - \sum_{j=1}^{n} \alpha_j \overline{\langle x, e_j\rangle} + \sum_{j=1}^{n} |\alpha_j|^2 = \|x\|^2 - \sum_{j=1}^{n} |\langle x, e_j\rangle|^2 + \sum_{j=1}^{n} |\alpha_j - \langle x, e_j\rangle|^2$$

which is minimised when $\alpha_j = \langle x, e_j\rangle$ for all $1 \leq j \leq n$. Furthermore if $y \in Y$ satisfies $\|x - y\| = \|x - P_Y(x)\|$, then $y = P_Y(x)$. $\square$

From Theorem 3.10, we see that $x \in Y$ if and only if $x = P_Y(x)$. Hence if $x \notin Y$, then $e := (x - P_y(x))/\|x - P_Y(x)\|$ is a unit vector orthogonal to every e_j – compare this with the Gram-Schmidt process from Honours Algebra where an orthonormal basis was constructed from a given basis.

Examples

(A) If X is an infinite dimensional inner product space, then the procedure indicated in Theorem 3.10 (or equivalently, the Gram-Schmidt process) produces an infinite sequence $\{e_j\}_{j\geq 1}$ of orthonormal vectors. Hence $\{e_j\}_{j\geq 1} \subset \overline{B_1(0)}$ and when $j \neq k$,

$$\|e_j - e_k\|^2 \ = \ \langle e_j - e_k, e_j - e_k \rangle = \langle e_j, e_j \rangle + \langle e_k, e_k \rangle = 2 \tag{$\sqrt{2}$}$$

implying that $\|e_j - e_k\| = \sqrt{2}$. Therefore there is no convergent subsequence of $\{e_j\}_{j\geq 1}$ and hence the closed unit ball $\overline{B_1(0)}$ is NOT a compact set. Earlier we proved that this is the case in any infinite dimensional normed linear space but indicated the construction of an infinite, uniformly separated sequence of unit vectors simplifies in the setting of an inner product space. Note that in the normed linear space setting, we were only able to achieve a uniform separation of $\|e_j - e_k\|$ by some distance strictly less than 1 but here, we are able to achieve $\sqrt{2}$ separation.

(B) Let $X = C[0,1]$ be the space of complex-valued continuous functions on $[0,1]$ with the inner product

$$\langle f, g \rangle \ = \ \int_0^1 f(x)\overline{g(x)}\, dx$$

as described in Example **F** above. We introduce the complex exponentials $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ given by

$$e^{2\pi inx} = \cos(2\pi nx) + i\sin(2\pi inx)$$

which nicely combine the basic trigonometric functions cosine and sine at a given frequency n into a single complex number, geometrically representing a point on the unit circle in the complex plane. If we denote the complex exponential displayed above as e_n, then we see that

$$\langle e_m, e_n \rangle \ = \ \int_0^1 e^{2\pi mx} e^{-2\pi inx} dx \ = \ \int_0^1 e^{2\pi i(m-n)x} dx$$

$$= \ \int_0^1 \cos(2\pi(m-n)x)dx + i\int_0^1 \sin(2\pi(m-n)x)dx \ = \ 0$$

whenever $m \neq n$ and $\langle e_m, e_n \rangle = 1$ if $m = n$; hence $\{e_n\}_{n\in\mathbb{Z}}$ forms an orthonormal sequence of vectors in $(C[0,1], \langle \cdot, \cdot \rangle)$. For any N, consider

$$Y \ = \ \mathrm{Span}(\{e_j\}_{j=-N}^N) \ = \ \Big\{ \sum_{j=-N}^{N} a_j e^{2\pi ijx} : a_j \in \mathbb{C} \Big\}$$

which is the finite dimensional subspace of **trigonometric polynomials** of degree at most N. The collection $\{e^{2\pi inx}\}_{|n|\leq N}$ is an ONB for Y.

Theorem 3.10 implies that for any $f \in C[0,1]$, the best approximation to f in the square mean

$$\int_0^1 |f(x) - \sum_{j=-N}^{N} \widehat{f}(j) e^{2\pi i j x}|^2 \, dx \ \le \ \int_0^1 |f(x) - \sum_{j=-N}^{N} a_j e^{2\pi i j x}|^2 \, dx$$

among all trigonometric polynomials of degree at most N is given by the Nth Fourier partial sum $S_N(f)(x) = \sum_{|n|\le N} \widehat{f}(n) e^{2\pi n x}$ where

$$\widehat{f}(n) \ := \ \langle f, e_n \rangle \ = \ \int_0^1 f(x) e^{-2\pi i n x} \, dx$$

are the Fourier coefficients of f. For this reason, in a general inner product space $(X, \langle \cdot, \cdot \rangle)$, if $\{e_\alpha\}_{\alpha \in A}$ is an orthonormal family of vectors, we call the inner products $\widehat{x}(\alpha) := \langle x, e_\alpha \rangle$ the **Fourier coefficients** of $x \in X$ with respect to the family $\{e_\alpha\}_{\alpha \in A}$.

(C) One can use Theorem 3.10 to compute

$$\inf_{a_0, a_1, \ldots, a_{n-1} \in \mathbb{R}} \int_0^1 |x^n - a_{n-1}x^{n-1} - \cdots a_1 x - a_0|^2 dx.$$

Of course we could multiply out the square in the integrand and integrate to realise the integral as a complicated polynomial $f(a_0, a_1, \ldots, a_{n-1})$ in n variables of degree n. By staring at this polynomial, it is not even clear that the infimum is achieved. let alone how to compute the minimum.

Instead we use Theorem 3.10 in the following way. Consider the $n+1$ dimensional inner product space $(\mathcal{P}_n, \| \cdot \|_{L^2})$ of real polynomials of degree at most n with the inner product that induces the L^2 norm (see Example **F** above). Let $Y = \{p(x) = \sum_{j=0}^{n-1} a_j x^j : a_j \in \mathbb{R}\}$ be the n dimensional subspace of polynomials of degree at most $n-1$.

In order to compute the orthogonal projection onto Y, we need to find an ONB $\{e_1(x), e_2(x), \ldots, e_n(x)\}$ for Y. This can be done using Theorem 3.10 (or Gram-Schmidt); for example in the $n = 2$ case, to compute an ONB for $Y = \{ax + b : a, b \in \mathbb{R}\}$, we start with $e_1(x) \equiv 1$ and consider the one dimensional space $Z = \text{Span}(1) = \{b : b \in \mathbb{R}\}$ of constant polynomials. Then $P_Z(x) = \langle x, 1 \rangle 1 = 1/2$ is the orthogonal projection of the polynomial x onto Z. According to Theorem 3.10, $x - P_Z(x) = x - 1/2$ is orthogonal to all elements in Z. Hence by taking $e_2(x) = (x - 1/2)/\|x - 1/2\| = \sqrt{12}(x - 1/2)$, we arrive at an orthonormal basis $\{e_1, e_2\}$ for Y.

According to Theorem 3.10, the polynomial $p \in Y$ which achieves the minimum in the above integral is given by the orthogonal projection $P_Y(f)(x) = \sum_{j=0}^{n-1} \langle f, e_j \rangle e_j(x)$ where $f(x) = x^n$. In the $n = 2$ case, we have $P_Y(f)(x) = x - 1/6$ and so we should take $a_1 = 1$ and $a_0 = -1/6$ in the corresponding integral to compute the infimum.

3.11. **Hilbert spaces.**

As we have already discussed above, any inner product space $(X, \langle\cdot,\cdot\rangle)$ can be thought of as a normed linear space with the norm $\|x\| = \sqrt{\langle x, x\rangle}$ induced from the inner product and hence X has a metric space structure (X, d) with $d(x, y) = \|x-y\| = \sqrt{\langle x-y, x-y\rangle}$. When this metric is complete, we call the corresponding inner product space a **Hilbert space**. In particular all Hilbert spaces are examples of Banach spaces.

We tend to denote Hilbert spaces by H.

Examples Let us review the examples of inner product spaces we listed at the outset of the section.

A. The space $\mathbb{F}^n$ with the inner product

$$\langle x, y\rangle \; = \; x_1\overline{y_1} + \cdots + x_n\overline{y_n}$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is a Hilbert space since it is Banach space, as observed previously. In fact all finite dimensional inner product spaces are Hilbert spaces.

E. The space ℓ^2 with the inner product

$$\langle x, y\rangle \; = \; \sum_{j=1}^{\infty} x_j\overline{y_j}$$

where for $x = (x_j)_{j\geq 1}, y = (y_j)_{j\geq 1} \in \ell^2$ is a Hilbert space since we have seen that it defines a Banach space.

F. The space $C[0, 1]$ with the inner product

$$\langle f, g\rangle \; = \; \int_0^1 f(x)\overline{g(x)}\, dx$$

where $f, g \in C[0, 1]$ is an example of an inner product space which is not a Hilbert space.

Exactly as in the case of normed linear spaces, we can consider the **Hilbert space completion** of an inner product space (see Exercise 11 below). The Hilbert space completion of Example **F** above is the space of Lebesgue square integrable functions $L^2[0, 1]$ which we mentioned briefly in the previous section.

Now let $(X, \langle\cdot,\cdot\rangle)$ be a Hilbert space which is infinite dimensional. The procedure indicated in Theorem 3.10 (or equivalently, using the Gram-Schmidt process – see Exercise 12 below) produces an infinite sequence $\{e_1, e_2, \ldots\}$ of orthonormal vectors in X. The subspaces $Y_n := \mathrm{Span}(e_1, e_2, \ldots, e_n\}$ are strictly increasing whose union $Y := \cup_n Y_n$ is a linear subspace of X. It may or may not be the case that the closure $H = \overline{Y}$ of Y is equal to X. Nevertheless with the inner product from X, the space H is itself a Hilbert space.

Let $x \in H$. By Theorem 3.10, we have

$$\sum_{j=1}^{n} |\langle x, e_j \rangle|^2 \;\le\; \|x\|^2 \quad \text{for all} \;\; n.$$

Letting $n \to \infty$, we see that the series $\sum_{j=1}^{\infty} |\langle x, e_j \rangle|^2$ converges and Bessel's inequality

$$\sum_{j=1}^{\infty} |\langle x, e_j \rangle|^2 \;\le\; \|x\|^2 \qquad (B \le)$$

holds for all $x \in H$. As a consequence, we see that

$$y_n \;:=\; P_{Y_n}(x) \;=\; \sum_{j=1}^{n} \langle x, e_j \rangle e_j$$

defines a Cauchy sequence in H; in fact for any $m < n$,

$$\|y_m - y_n\|^2 \;=\; \|P_{Y_m}(x) - P_{Y_n}(x)\|^2 \;=\; \sum_{j=m+1}^{n} |\langle x, e_j \rangle|^2$$

and this tends to 0 as $m, n \to \infty$. Hence $P_{Y_n}(x)$ converges to some $y \in H$.

We claim that $y = x$. In fact, from the continuity of $w \to \langle w, e_j \rangle$ (see Exercise 14), we have

$$\langle y, e_j \rangle \;=\; \lim_{n\to\infty} \langle P_{Y_n}, e_j \rangle \;=\; \lim_{n\to\infty} \sum_{k=1}^{n} \langle x, e_k \rangle \langle e_k, e_j \rangle \;=\; \langle x, e_j \rangle,$$

using the orthogonality properties of $\{e_j\}$. But this implies $\langle y - x, e_j \rangle = 0$ for all $j \ge 1$, hence $\langle y - x, w \rangle = 0$ for every $w \in \cup_n Y_n$. For any $w \in H$, there is a sequence $\{w_j\} \subset \cup_n Y_n$ such that $w_j \to w$ and so $\langle y - x, w_j \rangle \to \langle y - x, w \rangle$ but $\langle x - y, w_j \rangle = 0$ for all j and so $\langle x - y, w \rangle = 0$. In particular, $\langle y - x, y - x \rangle = 0$ or $x = y$.

Therefore

$$\sum_{j=1}^{\infty} \langle x, e_j \rangle e_j \;=\; \lim_{n\to\infty} \sum_{j=1}^{n} \langle x, e_j \rangle e_j \;=\; \lim_{n\to\infty} P_{Y_n}(x) \;=\; x$$

and this representation of x is unique. Siince $\|P_{Y_n} x\|^2 \;=\; \sum_{j=1}^{n} |\langle x, e_j \rangle|^2$ and $P_{Y_n} x \to x$ as $n \to \infty$, we see that Bessel's inequality $(B \le)$ is in fact an equality:

$$\|x\|^2 \;=\; \sum_{j=1}^{\infty} |\langle x, e_j \rangle|^2 \quad \text{for all } x \in H \qquad (P =)$$

which is known as **Parseval's identity**. This shows that the orthonormal set $\{e_j\}_{j \ge 1}$ is in some sense a basis, albeit an infinite one, for H; that is, every element x in the Hilbert space H has a unique expansion as an infinite linear combination of the $\{e_j\}_{j \ge 1}$. We will return to this discussion after we develop the notion of orthogonality from a more robust perspective.

3.12. **Orthogonality and geometry: a more advanced perspective.**

Although the inner product space $(X, \langle \cdot, \cdot \rangle)$ in Theorem 3.10 is not assumed to be a Hilbert space, the subspace Y which we are orthogonally projecting vectors onto is assumed to be finite dimensional and hence $(Y, \langle \cdot, \cdot \rangle)$ is a Hilbert space and in particular closed (this is just the inner product analogue of Corollary 2.31). One of the main conclusions from Theorem 3.10 is the existence of a unique closest point in Y to any given point in X, but it turns out that in the context of a Hilbert space, this conclusion is true for a much richer class of sets than closed, linear subspaces. It remains valid for any closed, convex set in a Hilbert space.

Theorem 3.13. *Let $(H, \langle , \rangle)$ be a Hilbert space and $C \subset H$ a closed, convex set. Then for every $x \notin C$, there exists a unique vector $y_0 \in C$ such that*

$$\|x - y_0\| \; = \; \inf\big\{ \|x - y\| : y \in C \big\}.$$

Of course any linear subspace is a convex set.

Proof. By considering the closed, convex set $C - x = \{y - x : y \in C\}$, we may assume $x = 0 \notin C$ and then show there exists a unique $y_0 \in C$ such that $\|y_0\| = \inf\{\|y\| : y \in C\}$. Let us denote this infimum by δ. For any $y, z \in C$, we have $(y+z)/2 \in C$ by the convexity of C and hence $\|(y+z)/2\| \ge \delta$. By the parallellogram law,

$$\Big\|\frac{y-z}{2}\Big\|^2 \; = \; \frac{1}{2}\|y\|^2 + \frac{1}{2}\|z\|^2 - \Big\|\frac{y+z}{2}\Big\|^2 \; \le \; \frac{1}{2}\|y\|^2 + \frac{1}{2}\|z\|^2 - \delta^2.$$

If both y and z satisfy $\|y\| = \|z\| = \delta$, then $\|y - z\|^2 \le 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$, implying that $y = z$ and this establishes the uniqueness part.

For the existence of an element $y_0 \in C$ such that $\|y_0\| = \delta$, let $\{y_n\}$ be a sequence in C such that $\|y_n\| \to \delta$ as $n \to \infty$. From above, we have

$$\|y_m - y_n\|^2 \; \le \; 2\|y_m\|^2 + 2\|y_n\|^2 - 4\delta^2 \quad \longrightarrow \quad 2\delta^2 + 2\delta^2 - 4\delta^2 \; = \; 0$$

as $m, n \to \infty$. Hence $\{y_n\}$ is a Cauchy sequence in the Hilbert space H and so there is some $y_0 \in H$ such that $y_n \to y_0$ as $n \to \infty$. But C is closed and so $y_0 \in C$. Furthermore, $\|y_n - y_0\| \to 0$ implies $\|y_n\| \to \|y_0\|$ as $n \to \infty$. Since $\|y_n\| \to \delta$, we have $\|y_0\| = \delta$. □

We will apply Theorem 3.13 when the closed, convex set is a closed, linear subspace M to establish a significant generalisation of Theorem 3.10. But first we define the orthogonal complement to a given set of vectors in an inner product space.

Definition 3.14. Let S be any set of vectors in an inner product space $(X, \langle \cdot, \cdot \rangle)$. The **orthogonal complement** to S is defined as

$$S^{\perp} \; = \; \big\{ x \in X : x \perp y \;\; (\text{or } \langle x, y \rangle = 0) \;\; \text{for all } y \in S \big\}.$$

Not only is the orthogonal complement $S^{\perp}$ a linear subspace of X (something which is easily verified), it is also closed (see Exercise 14). Furthermore, the only possible vector which can be in S and $S^{\perp}$ is the zero vector 0, assuming $0 \in S$. In fact if

$x \in S \cap S^\perp$, then $\langle x, x\rangle = 0$, implying $x = 0$. In particular if S is a linear subspace, then $S \cap S^\perp = \{0\}$.

Theorem 3.15. *Let H be a Hilbert space and M a proper, closed, linear subspace of H. There exist linear maps $P_M : H \to M$ and $Q_M : H \to M^\perp$ such that for every $x \in H$,*

$$x \;=\; P_M\, x \;+\; Q_M\, x.$$

Furthermore this representation is unique; this is, if $x = y + z$ where $y \in M$ and $z \in M^\perp$, then $y = P_M\, x$ and $z = Q_M\, x$.

The map P_M is called the **orthogonal projection** of H onto M. See Exercise 17 for further properties of P_M.

Since $M = M^{\perp\perp}$ (see Exercise 15), we could equally well apply Theorem 3.15 to $M^\perp$ and so we see that $Q_M = P_{M^\perp}$. Note that since M is closed, then $(M, \langle\cdot,\cdot\rangle)$ is a Hilbert space. In fact this is all we need; Theorem 3.15 remains true in a general inner product space as long as we assume the subspace $(M\langle\cdot,\cdot\rangle)$ is a Hilbert space. See Exercise 16. In this sense, Theorem 3.15 is a strict generalisation of Theorem 3.10 via Corollary 2.31.

Proof. Given any $x \in H$, we can use Theorem 3.13 to define $P_M\, x$ as the unique vector in M which is closest to x. We define $Q_M\, x := x - P_M\, x$ and claim that $Q_M\, x$ lies in $M^\perp$. In fact for any $y \in M$, consider the function $f(t) := \|x - (P_M\, x + ty)\|^2$ of a real variable t which has its unique minimum at $t = 0$ by the definition of $P_M\, x$. Hence $f'(0) = 0$. Expanding the expression for f in the usual way

$$f(t) \;=\; \|x - P_M\, x\|^2 + t^2\|y\|^2 - 2t\mathrm{Re}\langle x - P_M\, x, y\rangle,$$

we see that $f'(0) = 2\mathrm{Re}\langle x - P_M\, x, y\rangle$, implying that $\mathrm{Re}\langle x - P_M\, x, y\rangle = 0$ for all $y \in M$. This in turn implies that $Q_M\, x = x - P_M\, x \in M^\perp$ since $\mathrm{Im}\langle x - P_M\, x, y\rangle = \mathrm{Re}\langle x - P_M\, x, iy\rangle$ and $y \in M$ if and only if $iy \in M$.

Next, we note that if $x = y + z$ where $y \in M$ and $z \in M^\perp$, then $P_M\, x - y = z - Q_M\, x \in M \cap M^\perp$. Since $M \cap M^\perp = \{0\}$, we have $y = P_M\, x$ and $z = Q_M\, x$ which gives the uniqueness statement. Hence $P_M\, x$ is the unique vector in M such that $x - P_M\, x \in M^\perp$.

Therefore since $\alpha P_M\, x_1 + \beta P_M\, x_2 \in M$ for any $x_1, x_2 \in H$ and $\alpha, \beta \in \mathbb{F}$ and

$$\alpha x_1 + \beta x_2 - (\alpha P_M\, x_1 + \beta P_M\, x_2) = \alpha(x_1 - P_M\, x_1) + \beta(x_2 - P_M\, x_2) \in M^\perp,$$

we see that $\alpha P_M\, x_1 + \beta P_M\, x_2 = P_M(\alpha x_1 + \beta x_2)$, showing that P_M is a linear map. Since $Q_M = I - P_M$ where $I : H \to H$ is the identity operator, then Q_M is linear as well. $\square$

When M is a proper, closed, linear subspace of H, the conclusion of Theorem 3.15 can be succinctly written as $H = M \oplus M^\perp$; that is, H is the direct sum of M and its orthogonal complement $M^\perp$.

3.16. **Orthonormal bases.**

We pick up the discussion at the end of section **Hilbert spaces**. Let us fix an infinite, orthonormal sequence $\{e_j\}_{j\geq 1}$ of vectors in a Hilbert space H. There we proved that for all $x \in \overline{\mathrm{Span}(\{e_j\}_{j\geq 1})}$, then

$$x \;=\; \sum_{j=1}^{\infty} \langle x, e_j\rangle e_j \quad \text{and consequently,} \;\; \|x\|^2 \;=\; \sum_{j=1}^{\infty} |\langle x, e_j\rangle|^2.$$

Recall the notation $\widehat{x}(j) = \langle x, e_j\rangle$, the Fourier coefficients of x with respect to $\{e_j\}$.

Theorem 3.17. *The following are equivalent.*

(a) For every $x \in H$, if $\widehat{x}(j) = 0$ for all $j \geq 1$, then $x = 0$. In other words, the sequence $\{e_j\}_{j\geq 1}$ is a **maximal orthonormal** *family of vectors (the only vector which is orthogonal to every e_j is the zero vector);*

(b) $H \;=\; \overline{\mathrm{Span}(\{e_j\}_{j\geq 1})}$;

(c) For every $x \in H$, $x \;=\; \sum_{j=1}^{\infty} \widehat{x}(j)e_j \;=\; \lim_{n\to\infty} \sum_{j=1}^{n} \widehat{x}(j)e_j$;

(d) For every $x, y \in H$, $\langle x, y\rangle \;=\; \sum_{j=1}^{\infty} \widehat{x}(j)\overline{\widehat{y}(j)}$;

(e) For every $x \in H$, Parseval's identity $\|x\|^2 = \sum_{j=1}^{\infty} |\widehat{x}(j)|^2$ holds.

Implicit in Theorem 3.17 is the assumption that the Hilbert space H is infinite dimensional. However it is straightforward to verify that when $\dim(H) < \infty$ and $\{e_j\}_{j=1}^{n}$ is an ONB for H, then the conditions (a)-(e) are automatically satisfied (see Exercise 18).

Any orthonormal sequence $\{e_j\}$ which satisfies any (and hence all) of the conditions (a)-(e) in Theorem 3.17 is called an **orthonormal basis** for H.

Proof. Suppose $M := \overline{\mathrm{Span}(\{e_j\}_{j\geq 1})} \neq H$. Then since $H = M \oplus M^{\perp}$, we have $M^{\perp} \neq \{0\}$ and so there exists a nonzero $x \in M^{\perp}$ and hence $\widehat{x}(j) = \langle x, e_j\rangle = 0$ for all j. Hence (a) cannot hold.

The implications $(b) \Rightarrow (c)$ has already been established; see the outset of this subsection.

Suppose now that for every $x \in H$, $x = \lim_{N\to\infty} x_N$ where $x_N = \sum_{j=1}^{N} \widehat{x}(j)\rangle e_j$. Hence for any $x, y \in H$, $\langle x_N, y_N\rangle \to \langle x, y\rangle$ (see Exercise 14) and so

$$\langle x, y\rangle \;=\; \lim_{N\to\infty} \sum_{j=1}^{N}\sum_{k=1}^{N} \widehat{x}(j)\overline{\widehat{y}(k)}\langle e_j, e_k\rangle \;=\; \lim_{N\to\infty} \sum_{j\geq 1}^{N} \widehat{x}(j)\overline{\widehat{y}(j)} \;=\; \sum_{j=1}^{\infty} \widehat{x}(j)\overline{\widehat{y}(j)}.$$

Applying the above to $y = x$ gives the implication $(d) \Rightarrow (e)$.

Finally suppose that Parseval's identity holds. Let $x \in H$ be such that $\widehat{x}(j) = \langle x, e_j\rangle = 0$ for all $j \geq 1$. Then $\|x\|^2 = \sum_{j\geq 1} |\widehat{x}(j)|^2 = 0$ and so $x = 0$. ☐

Let $\{e_j\}$ be an orthonormal basis for H. As we mentioned immediately after the statement of Theorem 3.17, when H is finite dimensional, say $\dim(H) = n$, then conditions (a)-(e) automatically hold. Hence the map $T : H \to \mathbb{F}^n$ defined by $Tx = (\widehat{x}(1), \ldots, \widehat{x}(n))$, recalling $x = \sum_{j=1}^n \widehat{x}(j)e_j$, is not only a linear isomorphism but it is also an isometry. Furthermore, the finite dimension version of part (d) in Theorem 3.17 states that the map T preserves the inner product structure of both Hilbert spaces; that is

$$\langle x, y\rangle \ = \ \sum_{j=1}^n \widehat{x}(j)\overline{\widehat{y}(j)} \ = \ \langle (\widehat{x}(j))_{j=1}^n, (\widehat{y}(j))_{j=1}^n \rangle_2.$$

In this case, we say that T is a **Hilbert space isomorphism**. Hence from a Hilbert space perspective,

$$\boxed{\text{when } \dim(H) = n, \quad (H, \langle\cdot,\cdot\rangle) \text{ and } (\mathbb{F}^n, \langle\cdot,\cdot\rangle_2) \text{ are indistinguishable.}}$$

When H is infinite dimensional (that is, when the sequence $\{e_j\}$ is infinite), the map $T : H \to \ell^2$ defined by $Tx = (\widehat{x}(j))_{j\geq 1}$ is an isometry according to Theorem 3.17. Furthermore, part (d) in Theorem 3.17 states that the map T preserves the inner product structure of both Hilbert spaces; that is

$$\langle x, y\rangle \ = \ \sum_{j=1}^\infty \widehat{x}(j)\overline{\widehat{y}(j)} \ = \ \langle (\widehat{x}(j))_{j\geq 1}, (\widehat{y}(j))_{j\geq 1} \rangle_{\ell^2}.$$

However, unlike the finite dimensional case, it is not immediately clear that T is an onto map.

Proposition 3.18. *The map $T : H \to \ell^2$ defined by $Tx = (\widehat{x}(j))_{j\geq 1}$ is onto.*

Proof. Suppose $(a_j)_{j\geq 1} \in \ell^2$, then $x_n := \sum_{j=1}^n a_j e_j$ is a Cauchy sequence in H since

$$\|x_n - x_m\|^2 \ = \ \sum_{j=m+1}^n |a_j|^2 \ \to \ 0 \ \text{ as } m, n \to \infty.$$

Therefore $x_n \to x$ for some $x \in H$ or in other words, $\sum_{j=1}^\infty a_j e_j = x$. Again using the continuity of the inner product $\langle\cdot,\cdot\rangle$, we see that

$$\widehat{x}(j) \ = \ \langle \lim_{n\to\infty} \sum_{k=1}^n a_k e_k, e_j\rangle \ = \ \lim_{n\to\infty} \sum_{k=1}^n a_k \langle e_k, e_j\rangle \ = \ a_j$$

by the orthogonality properties of $\{e_j\}$. This establishes that T is onto. $\square$

Hence T is a Hilbert space isomorphism and so once again, if H **is infinite dimensional and has an orthonormal basis**,

$$\boxed{(H, \langle\cdot,\cdot\rangle) \text{ and } (\ell^2, \langle\cdot,\cdot\rangle_{\ell^2}) \text{ are indistinguishable as Hilbert spaces.}}$$

A natural question arises: does every Hilbert space H contain an orthonormal basis? The answer is no but before we discuss this further, let us examine a couple examples.

Examples

I. Consider the Hilbert space $(\ell^2, \|\cdot\|_2)$ with the sequence of vectors $\{e_n\}_{n\geq 1}$ where e_n is the vector $(x_j)_{j\geq 1}$ with $x_n = 1$ and $x_j = 0$ for all $j \neq n$; that is,

$$e_n \;:=\; (0, 0, \ldots, 0, 1, 0, 0, \ldots)$$

where the isolated entry 1 is in the nth place. We claim that $\{e_n\}_{n\geq 1}$ is an orthnormal basis for ℓ^2. In fact, if $x = (x_j)_{j\geq 1} \in \ell^2$, then $\widehat{x}(j) = \langle x, e_j\rangle = x_j$. If these Fourier coefficients vanishes for all $j \geq 1$, then $x = 0$ showing that condition (a) of Theorem 3.17 holds (and hence (b)-(e) also hold) and therefore $\{e_n\}_{n\geq 1}$ is an ONB for ℓ^2.

II. Let us consider the completion of $C[0,1]$ with the inner product given by

$$\langle f, g\rangle \;=\; \int_0^1 f(x)\overline{g(x)}\,dx.$$

Previously we called this completion $L^2[0,1]$ and for us, unless you have taken Essentials of Analysis and Probability, the elements of $L^2[0,1]$ are equivalence classes of Cauchy sequences of continuous functions. Nevertheless, our original space of continuous functions $C[0,1]$ sits inside $L^2[0,1]$ as a dense linear subspace. In Example (A) after the proof of Theorem 3.10 we introduced the infinite orthonormal sequence $\{e_n(x)\}_{n\in\mathbb{Z}} \subset C[0,1]$ where

$$e_n(x) \;:=\; e^{2\pi i n x} \;=\; \cos(2\pi n x) + i\sin(2\pi n x).$$

The sequence $\{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$ in fact forms an orthonormal basis for $L^2[0,1]$. To show this requires a bit of work and we will not do so here. This will be shown in the course Fourier Analysis.

III. We begin with the a certain linear subspace of the space $C(\mathbb{R})$ of complex-valued, continuous functions on $\mathbb{R}$; namely,

$$\mathcal{F} \;:=\; \Big\{\sum_j a_j e^{i\lambda_j x} : (a_j)_{j\geq 1} \in \mathbb{C}_0^\infty \text{ and } \{\lambda_j\} \text{ a distinct sequence of reals } \mathbb{R}\Big\}.$$

Hence the sums $f(x) = \sum_j a_j e^{i\lambda_j x}$ defining elements in $\mathcal{F}$ are all finite. Next, we define an inner produce on $\mathcal{F}$ as follows: if f and g are two elements of $\mathcal{F}$, we define

$$\langle f, g\rangle \;:=\; \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(x)\overline{g(x)}\,dx.$$

Before we think about whether or not the above formula defines an inner product on $\mathcal{F}$, we should verify that the limit defining the proposed inner product exists. In fact if $f(x) = \sum_j a_j e^{i\lambda_j x}$ and $g(x) = \sum_k b_k e^{i\mu_k x}$, then

$$\frac{1}{2T}\int_{-T}^{T} f(x)\overline{g(x)}\,dx = \sum_{j,k} a_j\overline{b_k}\, \frac{1}{2T}\int_{-T}^{T} e^{i(\lambda_j-\mu_k)x}\,dx.$$

If $\lambda_j \neq \mu_k$, then

$$\frac{1}{2T}\int_{-T}^{T} e^{i(\lambda_j-\mu_k)x}\,dx \;=\; \frac{\sin((\lambda_j-\mu_k)x)}{T(\lambda_j-\mu_k)} \;\longrightarrow\; 0 \text{ as } T\to\infty \qquad (\lambda)$$

whereas if $\lambda_j = \mu_k$, then $(2T)^{-1} \int_{-T}^{T} e^{i(\lambda_j - \mu_k)x} dx = 1$ for all $T > 0$. Hence the limit defining the proposed inner product above exists and

$$\langle f, g \rangle \;:=\; \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x)\overline{g(x)}\, dx \;=\; \sum_{j:\lambda_j = \mu_k \text{ for some } \mu_k} |\lambda_j|^2.$$

Let us now check that the axioms of an inner product are satisfied. First of all, from the above expression for $\langle f, g\rangle$, we have $\langle f, f\rangle = \sum_j |\lambda_j|^2$ and so if this vanishes, then $\lambda_j = 0$ for all j and hence $f = 0$. Next, since

$$\langle f, g\rangle_T \;:=\; \int_{-T}^{T} f(x)\overline{g(x)}\, dx \;=\; \overline{\langle g, f\rangle_T}$$

for all $T > 0$, we see that $\langle f, g\rangle = \overline{\langle g, f\rangle}$ holds for all $f, g \in \mathcal{F}$. Finally if $f, g, h \in \mathcal{F}$, then

$$\langle \alpha f + \beta g, h\rangle \;=\; \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [\alpha f(x) + \beta g(x)]\overline{h(x)}\, dx \;=$$

$$\alpha \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(x)\overline{h(x)}\, dx + \beta \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g(x)\overline{h(x)}\, dx = \alpha\langle f, h\rangle + \beta\langle g, h\rangle,$$

verifying the last axiom of an inner product.

The Hilbert space completion of $\mathcal{F}$ with respect to the above inner product is called the space of **almost periodic functions** which we denote by $AP(\mathbb{R})$. The theory of almost periodic functions is fascinating but its development would require a whole course devoted to it. Almost periodic functions are important in celestial mechanics since the orbits of the planets in our solar system are not perfectly periodic, but they are *almost periodic*.

The computation in (λ) shows that the family $\{e_\lambda := e^{i\lambda x}\}_{\lambda\in\mathbb{R}}$ is an orthonormal set of vectors in the space $AP(\mathbb{R})$ of almost periodic functions which is indexed by the *uncountable* set of reals $\mathbb{R}$. In particular by ($\sqrt{2}$) above, we have $\|e_\lambda - e_\mu\| = \sqrt{2}$ for any $\lambda \neq \mu$ and so the uncountable family of balls $\{B_1(e_\lambda)\}_{\lambda\in\mathbb{R}}$ is a pairwise disjoint collection of sets. Therefore there cannot exist a countable set S which is dense in $AP(\mathbb{R})$ (otherwise, there would be an element of S in the ball $B_1(e_\lambda)$ for every $\lambda \in \mathbb{R}$ which is clearly impossible!). In the next proposition, we will see that this precludes $AP(\mathbb{R})$ from possessing an orthonormal basis.

We now address the question when a Hilbert space possesses an orthonormal basis by which we mean a finite $\{e_j\}_{j=1}^n$ or a *countable* (that is, a infinite sequence) $\{e_j\}_{j=1}^\infty$ of vectors satisfying properties (a)-(e) from Theorem 3.17.

Proposition 3.19. *Let H be a Hilbert space. Then H possesses an orthonormal basis ONB if and only if there is countable set $S \subset H$ of vectors which is dense in H.*

When a Hilbert space has an ONB or equivalently, when it has countable dense set, we call it a **separable Hilbert space**.

Proof. * The proof depends on two facts: ① the scalar field $\mathbb{F}$ contains a countable dense set which is the rationals $\mathbb{Q}$ when $\mathbb{F} = \mathbb{R}$ and the set $\mathbb{Q} + i\mathbb{Q} = \{r + is \in \mathbb{C} : r, s \in \mathbb{Q}\}$ when $\mathbb{F} = \mathbb{C}$ AND ② a countable union of countable sets is countable.

Suppose H has an orthonormal basis $\{e_j\}$ which is either a finite or infinite sequence of vectors. Let us set $D = \mathbb{Q}$ or $\mathbb{Q} + i\mathbb{Q}$. The set of vectors

$$S = \Big\{ x = \sum_{j=1}^{N} a_j e_j \text{ for some } N \text{ and } \{a_j\} \subset D \Big\}$$

in H is countable since there is a one to one correspondence between S and D^n when $\#\{e_j\} = n$ for some n and $\mathbb{D}_0^\infty$, the space of finite sequences of D, when $\#\{e_j\} = \infty$. We claim that S is dense in H. Consider any $x \in H$ and $\epsilon > 0$. By property (a) of Theorem 3.17, we can find a $y \in \mathrm{Span}(\{e_j\})$ (that is, $y = \sum_{j=1}^N b_j e_j$ for some N and $\{b_j\} \subset \mathbb{F}$) such that $\|x - y\| < \epsilon/2$. By the density of D in $\mathbb{F}$, we can find $\{a_j\} \subset D$ such that $|a_j - b_j| < \epsilon/2N$ for every $1 \le j \le N$. Consider $z = \sum_{j=1}^N a_j e_j \in S$ and note that $y - z = \sum_{j=1}^N (b_j - a_j) e_j$ so that

$$\|y - z\| \;\le\; \sum_{j=1}^{N} |b_j - a_j| \|e_j\| \;\le\; \sum_{j=1}^{N} \epsilon/2N \;=\; \epsilon/2$$

and hence $\|x - z\| \le \|x - y\| + \|y - z\| \le \epsilon/2 + \epsilon/2 = \epsilon$, showing that the countable set S is dense in H.

Next suppose there exists a countable subset $S = \{x_j\}_{j \ge 1}$ of H which is dense. From the Gram-Schmidt process (see Exercise 12 below), we can construct a orthonormal sequence $\{e_j\}$ whose linear span is the same as the linear span of S. Hence

$$\overline{S} \subset \overline{\mathrm{Span}(\{x_j\})} = \overline{\mathrm{Span}(\{e_j\})}$$

and since $\overline{S} = H$, we see that condition (a) of Theorem 3.17 holds for the orthonormal sequence $\{e_j\}$ and therefore H possesses an ONB. □

EXERCISES

1. Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. From the axioms of an inner product, show that

$$\langle x, \alpha y + \beta z \rangle \;=\; \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle \quad \text{for all } \; x, y, z \in X \; \text{ and } \alpha, \beta \in \mathbb{F}.$$

In particular, we have $\langle x, \alpha y \rangle = \overline{\alpha} \langle x, y \rangle$.

2. Verify the axioms of an inner product for $\langle \cdot, \cdot \rangle$ defined in Examples **A**, **E** and **F** above.

3. Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. Show that $\langle 0, y \rangle = 0$ for all $y \in X$.

4. By examining the proof of Theorem 3.4, show that equality holds in the Cauchy-Schwarz inequality if and only if x and y are linearly dependent; that is, $x = \alpha y$ for some $\alpha \in \mathbb{F}$ if $y \neq 0$.

5. Show that the parallelogram law does not hold in $(C[0,1], \|\cdot\|_{L^1})$.

6.* Let $(X, \|\cdot\|)$ by a normed linear space with real scalars $\mathbb{R}$ satisfying the parllelogram law $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$. Define $\langle\cdot,\cdot\rangle$ by the formula

$$\langle x, y\rangle \;=\; \frac{1}{4}\big[\|x+y\|^2 - \|x-y\|^2\big].$$

Show that $\langle\cdot,\cdot\rangle$ satisfies the axioms of an inner product by completing the following steps.

Step 1: Verify the first two axioms of an inner product for $\langle\cdot,\cdot\rangle$.

Step 2: Use the parellogram law to show that

$$2\|x+z\|^2 + 2\|y\|^2 \;=\; \|x+y+z\|^2 + \|x-y+z\|^2$$

and show that this implies

$$\begin{aligned}\|x+y+z\|^2 &= 2\|x+z\|^2 + 2\|y\|^2 - \|x-y+z\|^2 \\ &= 2\|y+z\|^2 + 2\|x\|^2 - \|y-x+z\|^2\end{aligned}$$

Step 3: From the previous step, conclude that

$$\|x+y+z\|^2 = \|x\|^2 + \|y\|^2 + \|x+z\|^2 + \|y+z\|^2 - \frac{1}{2}\|x-y+z\|^2 + \frac{1}{2}\|y-x+z\|^2$$

and also

$$\|x+y-z\|^2 = \|x\|^2 + \|y\|^2 + \|x-z\|^2 + \|y-z\|^2 - \frac{1}{2}\|x-y-z\|^2 + \frac{1}{2}\|y-x-z\|^2.$$

Deduce that $\langle x+y.z\rangle = \langle x,z\rangle + \langle y,z\rangle$.

Step 4: Establish $\langle\lambda x, y\rangle = \lambda\langle x,y\rangle$ for all $\lambda \in \mathbb{Q}$ and $x, y \in X$. (First do this for $\lambda \in \mathbb{N}$ via the previous steps, next for $\lambda = -1$, then for $\lambda \in \mathbb{Z}$ and finally for the general case.)

Step 5: Show that the function $t \to \langle tx, y\rangle$ is a continuous function for every fixed $x, y \in X$. See Exercise 19 from the previous section. Conclude that $\langle\lambda x, y\rangle = \lambda\langle x, y\rangle$ holds for all $\lambda \in \mathbb{R}$.

7.* Extend Exercise 6 to complex inner product spaces satisfying the parallelogram law, using the polarisation identity in Proposition 3.8.

8. Consider the primitive nth root of unity $\omega = e^{2\pi i/n}$ where $n \geq 3$. Let $(X, \langle\cdot,\cdot\rangle)$ be a complex inner product space. Show that

$$n\langle x, y\rangle \;=\; \sum_{k=0}^{n-1} \omega^k \|x + \omega^k y\|^2.$$

Hint: Use the fact that $\sum_{k=0}^{n-1} \omega^k = 0$. You may also find $\sum_{k=0}^{n-1} \omega^{2k} = 0$ useful.

Furthermore, show that

$$\langle x, y\rangle \;=\; \int_0^1 e^{2\pi i\theta}\|x + e^{2\pi i\theta}y\|^2 \; d\theta.$$

9. Verify the simple orthogonality relation given in $(\perp)$.

10. Compute the following infimum:

$$\inf_{a,b,c\in\mathbb{R}} \int_0^\infty |x^3 - a - bx - cx^2|^2 e^{-x}\, dx.$$

11.* Go through the completion process of a normed linear space but now in the context of an inner product space. Pay particular attention to how one defines the natural inner product $\langle\cdot,\cdot\rangle'$ on H which makes $(H, \langle\cdot,\cdot\rangle')$ into a Hilbert space.

12. Review (or deduce from Theorem 3.10) the Gram-Schmidt process: given a finite or infinite sequence of vectors $\{v_j\}$ in an inner product space $(X, \langle\cdot,\cdot\rangle)$, construct a orthonormal sequence of vectors $\{e_j\}$ whose linear span is the same as the linear span of $\{v_j\}$.

13. Let H be a Hilbert space, M a closed linear subspace and $x_0 \in H \setminus M$. Show that

$$\min_{x\in M} \|x - x_0\| \;=\; \max_{y\in M^\perp, \|y\|=1} |\langle x_0, y\rangle|.$$

14. Show that in any inner produce space $(X, \langle\cdot,\cdot\rangle)$, the map

$$x \longrightarrow \langle x, y\rangle$$

is continuous for every fixed $y \in X$. Use this to show that the orthogonal subspace $S^\perp$ of any set $S \subset X$ is a closed, linear subspace of X. Furthermore, show that if $x_n \to x$ and $y_n \to y$ in X, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.

15. Let S by any collection of vectors in an inner product space X. Show that $S \subset S^{\perp\perp}$ and $S^{\perp\perp} = \overline{\mathrm{Span}(S)}$. Hence if S is a closed, linear subspace of X, then $S = S^{\perp\perp}$.

16. Let M be a linear subspace of an inner product space $(X, \langle\cdot,\cdot\rangle)$ and suppose that $(M, \langle\cdot,\cdot\rangle)$ is a Hilbert space. Show that M is necessarily a closed linear subspace in X and that Theorem 3.15 holds in this setting.

17. Let H be a Hilbert space, M a closed linear subpace and $P_M : H \to M$ the orthogonal projection onto M. Show that the range of P_M is M; that is, $P_M(H) = M$. Also show that $P_M^2 = P_M$ and $\ker(P_M) = M^\perp$.

18. Let H be a finite dimensional Hilbert space with ONB $\{e_j\}_{j=1}^n$. Show that conditions (a)-(e) automatically hold for $\{e_j\}_{j=1}^n$.

19.* Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and let $\{e_\alpha\}_{\alpha\in A}$ be an orthonormal family in X; that is, if $\alpha \neq \beta$, then $e_\alpha \perp e_\beta$. We call $\{e_\alpha\}_{\alpha\in A}$ a **maximal family**

of orthnormal vectors if whenever $x \in H$ satisfies $\langle x, e_\alpha \rangle = 0$ for all $\alpha \in A$, then $x = 0$. From Bessel's inequality written down in part (b) in Theorem 3.10, we see that

$$\sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 \ \le \ \|x\|^2$$

for any finite subset $F \subset A$. Hence

$$\sup_F \sum_{\alpha \in F} |\langle x, e_\alpha \rangle|^2 \ \le \ \|x\|^2$$

where the supremum is taken over all finite subsets F of A. It is tempting to define $\sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2$ by the expression on the above left hand side and we will do so.

Now suppose that X is a Hilbert space. Show that $\{e_\alpha\}_{\alpha \in A}$ is a maximal family of orthonormal vectors if and only if for every $x \in H$, $\|x\|^2 = \sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2$.

4. Continuous linear operators

Having discussed the theory and examples of normed linear spaces/Banach spaces and inner product spaces/Hilbert spaces, we now turn to the natural maps between these spaces.

Suppose X and Y are two vector spaces, then the *linear* maps $T : X \to Y$ between them[9] are precisely those which preserve the underlying algebraic (vector space) operations; that is, for arbitrary $x_1, x_2 \in X$ and arbitrary scalars α and β,

$$T(\alpha x_1 + \beta x_2) = \alpha T x_1 + \beta T x_2$$

and so T maps linear combination $\alpha x_1 + \beta x_2$ in X to the corresponding linear combination $\alpha T x_1 + \beta T x_2$ in Y.

Next we suppose that X and Y are two metric spaces. Then the continuous maps $T : X \to Y$ between them are precisely those maps which preserve convergent sequences; that is, if $x_n \to x$, then $T(x_n) \to T(x)$ and so T maps the convergent sequence $x_n \to x$ in X to the convergent sequence $T(x_n) \to T(x)$ in Y.

Now suppose that $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ are two normed linear spaces,[10] so X and Y are both vector spaces and metric spaces. Hence continuous, linear maps $T : X \to Y$ between them are precisely the maps which preserve the algebraic operations of the underlying vector spaces AND preserve the classes of convergent sequences.

We now introduce a very important definition.

Definition 4.1. Let $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ be two normed linear spaces and let $T : X \to Y$ be a linear map. Then T is **bounded** if there exists a nonnegative real number M such that

$$\|Tx\| \le M\|x\| \tag{1}$$

for all $x \in X$. In such a case we say that T is a **bounded linear operator** between X and Y.

Exercise. Let $M > 0$ be fixed. Show that the following statements about a linear map $T : X \to Y$ between normed spaces are equivalent:

(i) For all $x \in X$, $\|Tx\| \le M\|x\|$.
(ii) For all x with $\|x\| \le 1$, $\|Tx\| \le M$.
(iii) For all x with $\|x\| = 1$, $\|Tx\| \le M$.

---

[9]When discussing linear maps between two vector spaces, we are tacitly assuming that the underlying scalar fields are the same, either both are the real field $\mathbb{R}$ or both are the complex field $\mathbb{C}$.

[10]We should probably tag the norms as $\|\cdot\|_X$ and $\|\cdot\|_Y$ in order to distinguish them but this introduces cumbersome subscripts and we prefer to suppress them. Please keep in mind that the norms on X and Y are different and it should be clear from the context which norm is which.

(iv) For all $x \neq 0$, $\frac{\|Tx\|}{\|x\|} \leq M$.

Consequently the three numbers

$$\sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}, \quad \sup_{\|x\|=1} \|Tx\| \quad \text{and} \quad \sup_{\|x\| \leq 1} \|Tx\|$$

coincide. We denote their common value by $\|T\|$ and we call it the **operator norm** of T with respect to the two norms on X and Y. We will see momentarily that this defines a norm on the vector space of all bounded linear operators between X and Y. Again we should tag this norm to distinguish it from the norms on X and Y but we shall not do so. Note that

$$\|Tx\| \leq \|T\| \|x\|$$

for all $x \in X$ and here all three norms are on display.

Theorem 4.2. *Let $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ be two normed linear spaces and let $T : X \to Y$ be a linear map. Then the following are equivalent:*

(a) T is continuous at 0; that is, whenever $x_n \to 0$, then $Tx_n \to T0 = 0$.

(b) T is continuous at every point $x \in X$; that is, whenever $x_n \to x$, then $Tx_n \to Tx$.

(c) T is bounded.

Proof. Clearly (b) implies (a). Suppose that T is continuous at 0 and $x_n \to x$. Then $x_n - x \to 0$ and hence $T(x_n - x) \to 0$. But $Tx_n - Tx = T(x_n - x)$ and so $Tx_n \to Tx$, showing that T is continuous at x. Therefore (a) and (b) are equivalent.

Now suppose (c) holds and $x_n \to 0$. Since $\|Tx_n\| \leq \|T\| \|x_n\|$ and $\|T\|$ is finite, we see that $Tx_n \to 0$ and so T is continuous at 0. In the reverse direction, suppose T is not bounded. Then for every $n \geq 1$, we can find an x_n with $\|x_n\| \leq 1$ such that $\|Tx_n\| \geq n$. Set $y_n = (1/n)x_n$ and note that $\|y_n\| = \|x_n\|/n \leq 1/n \to 0$ as $n \to \infty$. On the other hand, $\|Ty_n\| = \|Tx_n\|/n \geq 1$ for all n. This shows that T is not continuous at 0 and hence (a) imples (c), completing the proof of the theorem. □

Exercise. Show (a) implies (c) directly using the $\epsilon - \delta$ definition of continuity.

If $T : X \to Y$ is a bounded linear operator then the set of all M such that (1) holds is nonempty and bounded below by 0 – and therefore has an infimum.

Exercise. Show that if $T : X \to Y$ is a bounded linear operator then

$$\|T\| = \min\{M : \|Tx\| \leq M\|x\| \text{ for all } x \in X\},$$

so that $\|T\|$ is the "best" M for which (1) holds for all $x \in X$.

We will see that often it is not difficult to show that a linear operator is continuous or bounded by using (1) – BUT it is often very difficult to compute its operator norm $\|T\|$ exactly.

Given two normed linear spaces X and Y, we denote by $\mathcal{L}(X, Y)$ the collection of all bounded linear operators between X and Y. When $X = Y$, we denote $\mathcal{L}(X, X)$ simply as $\mathcal{L}(X)$. The collection $\mathcal{L}(X, Y)$ is clearly a vector space in its own right (over the same scalar field) but it is also a normed linear space.

Proposition 4.3. *The operator norm $\|T\|$ of a bounded, linear operator $T \in \mathcal{L}(X, Y)$ is indeed a norm.*

Proof. Clearly $\|T\| \geq 0$ for every $T \in \mathcal{L}(X, Y)$. Suppose that $\|T\| = 0$. Then for every $x \in X$ with $x \neq 0$, $\|Tx\| = 0$ and so $Tx = 0$. Hence T is the zero operator, $T = 0$.

Next let $T \in \mathcal{L}(X, Y)$ and $\alpha \in \mathbb{F}$. Then $\|(\alpha T)x\| = \|\alpha Tx\| = |\alpha|\|Tx\|$ and so

$$\sup_{\|x\|\leq 1} \|(\alpha T)x\| \;=\; |\alpha| \sup_{\|x\|\leq 1} \|Tx\| \;=\; |\alpha|\|T\|$$

from the definition of the operator norm. Hence $\|\alpha T\| = |\alpha|\|T\|$.

Finally suppose $S, T \in \mathcal{L}(X, Y)$. Then

$$\|(S + T)x\| = \|Sx + Tx\| \leq \|Sx\| + \|Tx\| \leq [\|S\| + \|T\|]\|x\|$$

and so (1) holds with $M = \|S\| + \|T\|$. Since $\|S + T\|$ is the smallest constant M in (1), we see that $\|S + T\| \leq \|S\| + \|T\|$, establishing the triangle inequality for the operator norm. □

It is interesting to note that in the above proof, the verification that the operator norm $\|T\|$ is a norm relies mainly on the norm properties for the norm $\|\cdot\|$ on Y. Even when X and Y are given explicitly, the matter of characterising $\mathcal{L}(X, Y)$ can be very difficult – or even essentially impossible.

For some examples of bounded linear operators see Workshop 3.

The next result asserts that if Y is complete, then $\mathcal{L}(X, Y)$ is also complete.

Theorem 4.4. *Let X and Y be two normed linear spaces. If Y is a Banach space, then $\mathcal{L}(X, Y)$ is also a Banach space.*

Proof. Suppose $\{T_n\} \subseteq \mathcal{L}(X, Y)$ is a Cauchy sequence. So, given any $\epsilon > 0$ there is an N such that $m, n \geq N$ implies $\|T_m - T_n\| < \epsilon$.

So for all $x \in X$ we have that for $m, n \geq N$,

$$\|T_m x - T_n x\| \leq \|T_m - T_n\|\|x\| < \epsilon\|x\|. \tag{2}$$

Thus for all $x \in X$, the sequence $\{T_n x\}$ is Cauchy in Y and hence, by completeness of Y, converges to some member – let us call it Tx – of Y.

Then:

- T is linear – we leave this as an exercise.
- T is bounded – for $\|Tx\| = \lim_{n\to\infty} \|T_n x\| \le \sup_n \|T_n\|\|x\|$ and $\{\|T_n\|\}$ is bounded as any Cauchy sequence is bounded (– from Honours Analysis).
- $T_n \to T$ – let $m \to \infty$ in (2) to conclude that for $n \ge N$

  $$\|(T - T_n)x\| \le \epsilon\|x\|.$$

  This implies that $\|T - T_n\| \to 0$ as $n \to \infty$, and so $\mathcal{L}(X, Y)$ is complete.

$\square$

Note that the fact that T is bounded actually follows from the third bullet point, bypassing the second one. The converse statement (that if $\mathcal{L}(X, Y)$ is complete, so is Y), is also true, but is harder to prove, and we omit it for now.

The next result is especially useful when we are dealing with linear operators on spaces like $L^p([0,1])$ where we have a particularly helpful dense linear subspace such as $C([0,1])$.

Proposition 4.5 (Extension from a dense subspace)**.** *Suppose X is a normed linear space and Y is a Banach space. Suppose that X' is a dense linear subspace of X. Suppose that $T \in \mathcal{L}(X', Y)$. Then there is a unique $\tilde{T} \in \mathcal{L}(X, Y)$ such that $\tilde{T}|_{X'} = T$. Moreover $\|\tilde{T}\| = \|T\|$. Furthermore, if T is an isometry, so is $\tilde{T}$.*

Proof. The first job is to define $\tilde{T}x$ for each $x \in X$. All the know is that there is a sequence x_n with $x_n \in X'$ such that $x_n \to x$ in X. So it makes sense to try $\tilde{T}x := \lim_{n\to\infty} Tx_n$. But why does this limit exist? Well, (Tx_n) is Cauchy in Y since $\|Tx_n - Tx_m\| \le \|T\|\|x_m - x_n\|$. So completeness of Y tells us that $\lim_{n\to\infty} Tx_n$ does indeed exist, and so we can declare $\tilde{T}x := \lim_{n\to\infty} Tx_n$ to be this member of Y.

However, we could have started with a *different* sequence converging to x, and so we need to know we'd arrive at the same answer – that is, that $\tilde{T}x$ is *well-defined.* So suppose $\xi_n \to x$ with $\xi_n \in X'$. We need to show that $\lim_{n\to\infty} T\xi_n = \lim_{n\to\infty} Tx_n$ (both limits exist by the argument of the previous paragraph). But $\|T\xi_n - Tx_n\| \le \|T\|\|x_n - \xi_n\|$ and this goes to zero since x_n and ξ_n have the same limit x. So indeed $\lim_{n\to\infty} T\xi_n = \lim_{n\to\infty} Tx_n$ as required.

Next we need to show $\tilde{T}$ is linear. We leave this as an exercise.

To establish $\tilde{T}|_{X'} = T$ we need to see that if $x \in X'$, then $\tilde{T}x = Tx$. Simply take $x_n = x$ for all n. This gives a sequence in X' converging to x. By defnition of $\tilde{T}$, we have $\tilde{T}x = \lim_{n\to\infty} Tx_n = Tx$.

Next we need to see that $\tilde{T}$ is bounded and that $\|\tilde{T}\| = \|T\|$. Now we certainly have that $\|T\| \leq \|\tilde{T}\|$ as for $\tilde{T}$ we are sup'ing over a larger set. On the other hand, for $x \in X$, $\|\tilde{T}x\| = \lim_{n\to\infty} \|Tx_n\| \leq \|T\| \lim_{n\to\infty} \|x_n\| = \|T\|\|x\|$ so that indeed $\tilde{T}$ is bounded and $\|\tilde{T}\| \leq \|T\|$, giving $\|\tilde{T}\| = \|T\|$. Notice that if T is an isometry, we get $\|\tilde{T}x\| = \lim_{n\to\infty} \|Tx_n\| = \lim_{n\to\infty} \|x_n\| = \|x\|$, showing that $\tilde{T}$ is also an isometry.

Finally, if $\tilde{\tilde{T}}$ is *any* member of $\mathcal{L}(X, Y)$ such that $\tilde{\tilde{T}}|_{X'} = T$ and if $x \in X$ with $x_n \to x$ and $x_n \in X'$ for all n, we have that $\tilde{\tilde{T}}x = \lim_{n\to\infty} \tilde{\tilde{T}}x_n = \lim_{n\to\infty} Tx_n = \tilde{T}x$, showing the uniqueness of the operator $\tilde{T}$. □

Let us recall the notions of the **kernel** and **image** of a linear operator $T : X \to Y$ between vector spaces X and Y. The **kernel** of T is the linear subspace $\ker(T) := \{x \in X : Tx = 0\}$ of X and the **image** of T is the linear subspace

$$\mathrm{im}(T) = T(X) \ := \ \big\{y \in Y : \text{there exists an } x \in X \text{ such that } Tx = y\big\}$$

of Y. In the linear category, T being $1-1$ or injective is equivalent to the kernel $\ker(T)$ being the zero subspace $\{0\}$; and T being onto or surjective is equivalent to the image $T(X)$ being equal to Y, i.e. $T(X) = Y$. Recall from Honours Algebra (or even first year Linear Algebra) that when $X = Y$ is finite dimensional, then T being injective is equivalent to T being surjective. *This is no longer the case when $X = Y$ is infinite dimenisional – see Example A below.*

Now let us consider a linear map $T : X \to Y$ between normed linear spaces X and Y. Previously we have discussed/motivated why *closed* linear subspaces are important.

Proposition 4.6. *Suppose $T \in \mathcal{L}(X, Y)$ is a bounded, linear operator between two normed linear spaces X and Y. Then the kernel of T, $\ker(T)$ is a* closed *subspace.*

Proof. Let $\{x_n\} \subseteq \ker(T)$ be a sequence in the kernel of T which converges to an element $x \in X$, $x_n \to x$. Hence $Tx_n = 0$ for all n and $Tx_n \to Tx$ since T is continuous. Therefore $x \in \ker(T)$ and so the kernel of T is closed. □

It is **not** necessarily the case that the image $T(X)$ of a continuous, linear map $T : X \to Y$ between normed linear spaces is a closed subspace. For instance, take any normed linear space X which is not a Banach space; for example $(C[0, 1], \|\cdot\|_{L^2})$. Consider the Banach space completion Y of X so that X sits inside Y as a dense, linear subpace, $\overline{X} = Y$ AND $X \neq Y$. Let $I : X \to Y$ be the identity or embedding map, $Ix = x$. Then $I(X) = X$ is not closed since its closure is all of Y and in particular, its closure is not equal to itself.

4.7. **Linear operators on finite dimensional spaces.**

It turns out that if X is finite-dimensional, then *every* linear operator $T : X \to Y$ into any normed linear space Y is necessarily continuous. But even when both X

and Y are finite dimensional, the computation of the operator norm $\|T\|$, which depends on the choice of norms on the underlying spaces X and Y, is not always an easy task. (When the norms come from an inner product $\langle \cdot, \cdot \rangle$, then there are certain situations when we can compute the operator norm $\|T\|$ more easily. See below.)

Proposition 4.8. *Let T be a linear operator between two normed linear spaces X and Y where X is finite dimensional. Then T is automatically continuous.*

Proof. We denote the norm on X by $\|\cdot\|_X$ and the norm on Y by $\|\cdot\|_Y$. Since X is finite dimensional, then X has a basis $\mathcal{B} = \{e_1, \ldots, e_n\}$. Besides the given norm $\|\cdot\|_X$ on X, we will use the norm $\|\cdot\|_{\mathcal{B},1}$; that is $\|x\|_{\mathcal{B},1} = \sum_{j=1}^n |x_j|$ where $x = x_1 e_1 + \cdots + x_n e_n$ is the unique representation of x with respect to the basis $\mathcal{B}$. Since all norms are equivalent on a finite dimensional vector space, there is a constant A such that $\|x\|_{\mathcal{B},1} \le A\|x\|_X$ for all $x \in X$. Hence

$$\begin{aligned} \|Tx\|_Y &= \|x_1 Te_1 + \cdots + x_n Te_n\|_Y \le |x_1|\|Te_1\|_Y + \cdots + |x_n|\|Te_n\|_Y \\ &\le \max_{1\le j\le n} \|Te_j\|_Y \, \|x\|_{\mathcal{B},1} \le A \max_{1\le j\le n} \|Te_j\|_Y \, \|x\|_X \end{aligned}$$

which shows that (1) is satisfied with $M = A\max_{1\le j\le n} \|Te_j\|_Y$. Hence T is bounded and therefore continuous by Theorem 4.2. □

Now suppose both X and Y are finite dimensional, say $\dim(X) = n$ and $\dim(Y) = m$. Let $T : X \to Y$ be a linear operator. No matter which norms we consider on X and Y, say $\|\cdot\|_X$ on X and $\|\cdot\|_Y$ on Y, Proposition 4.8 implies that $T : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ is bounded; that is, $\|T\|$ is finite. However, even in this finite dimensional case, determining exactly what is the operator norm $\|T\| = \sup_{\|x\|_X \le 1} \|Tx\|_Y$, which clearly depends on both the underlying norms involved, can be very difficult. The upper bound $A\max_{1\le j\le n} \|Tf_j\|_Y$ for $\|T\|$ given in the proof of Proposition 4.8 is typically a very bad approximation to the actual value of $\|T\|$.

There is a general lower bound for $\|T\|$ in terms of the eigenvalues of T when $X = Y$ (so that $m = n$) with a common norm $\|\cdot\| = \|\cdot\|_X = \|\cdot\|_Y$. Recall from Honours Algebra that $\lambda \in \mathbb{F}$ is an **eigenvalue** of T if there exists a nonzero $x \in X$ (called an **eigenvector**) such that $Tx = \lambda x$. Hence

$$|\lambda|\|x\| = \|\lambda x\| = \|Tx\| \le \|T\|\|x\| \ \text{ implying } \quad \sup_{\text{eigenvalues } \lambda} |\lambda| \le \|T\|. \qquad (SR)$$

When the scalars are real, $\mathbb{F} = \mathbb{R}$, a linear operator may not have any eigenvalues; for example, when $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $T(x, y) = (-y, x)$. In fact, if A is the matrix representing T with respect to the usual bases,[11] then solutions of

[11]From the first year Linear Algebra course (explained in more detail in the Honours Algebra course), we know that with respect to a pair a bases, one for X and one for Y, the linear operator T can be represented by an $m \times n$ matrix. If fact, if $\mathcal{B} = \{e_1, \ldots, e_n\}$ is a basis for X and $\mathcal{C} = \{f_1, \ldots, f_m\}$ is a basis for Y, then the $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

the equation $\det(\lambda I - A) = 0$ give us the eigenvalues of T. For our example $T(x,y) = (-y,x)$, we have

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and so} \quad 0 \ = \ \det\bigl(\lambda I - A\bigr) \ = \ \lambda^2 + 1$$

has no real solutions.

However when the scalars are complex, $\mathbb{F} = \mathbb{C}$, then the polynomial equation $\det(\lambda I - A) = 0$ has precisely n solutions, giving us *all* the eigenvalues of T, counted with multiplicity. In this case, the general lower bound

$$\rho(T) := \sup |\lambda| \leq \|T\|$$

given in (SR) as the supremum of the eigenvalues is called the **spectral radius** of T. Note that the spectral radius $\rho(T)$ is independent of the norm $\|\cdot\|$ of the underlying space X and this is why we call it a *general* lower bound. As we shall now see, the spectral radius is more closely related to the operator norm $\|T\|$ when the norm $\|\cdot\|$ arises from an inner product $\langle \cdot, \cdot \rangle$; that is $\|x\| = \sqrt{\langle x, x \rangle}$.

Indeed, when the norm arises from an inner product $\langle \cdot, \cdot \rangle$, and T is **self-adjoint**, then $\rho(T) = \|T\|$. To see this, recall from your Honours Algebra course, T is self-adjoint with respect to the inner product if $\langle Tx, y \rangle = \langle x, Ty \rangle$ for all $x, y \in X$. In this case, all the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ are real and we can find an orthonormal basis $\{e_1, \ldots, e_n\}$ consisting entirely of eigenvectors of T, $Te_j = \lambda_j e_j, 1 \leq j \leq n$ (this was Theorem 5.3.9 from Honours Algebra – we shall be generalising this to a general Hilbert space later). Hence for any $x \in X$, we have $x = \sum_{j=1}^n \langle x, e_j \rangle e_j = \sum_{j=1}^n \widehat{x}(j) e_j$ and so $Tx = \sum_{j=1}^n \lambda_j \widehat{x}(j) e_j$, implying

$$\|Tx\|^2 = \|\sum_{j=1}^n \lambda_j \widehat{x}(j) e_j\|^2 = \sum_{j=1}^n |\lambda_j \widehat{x}(j)|^2 \leq \rho(T)^2 \sum_{j=1}^n |\widehat{x}(j)|^2 = \rho(T)^2 \|x\|^2$$

and so $\|T\| \leq \rho(T)$. Therefore $\rho(T) = \|T\|$ in this case.

Even when T is not self-adjoint, that is, when T is not equal to its **adjoint** T^* (which has the defining property $\langle Tx, y \rangle = \langle x, T^* y \rangle$ for all $x, y \in X$), nevertheless T^*T is self-adjoint since $\langle T^*Tx, y \rangle = \langle Tx, Ty \rangle = \langle x, T^*Ty \rangle$. The eigenvalues of T^*T are called the **singular values** of T. Hence $\rho(T^*T) = \|T^*T\|$ and we will see later that $\|T^*T\| = \|T\|^2$ which tells us that we can compute the operator norm $\|T\|$ of a general linear operator on finite dimensional complex inner product space in terms of its singular values.

Note that the adjoint T^* (and hence the notion of being self-adjoint, $T = T^*$) depends on the underlying inner product $\langle \cdot, \cdot \rangle$ which in turn defines the norm $\|\cdot\| =$

---

represents T as follows. The coefficients $\{a_{jk}\}$ defining A satisfy $Te_j = a_{1j} f_1 + \cdots + a_{mj} f_m$ for each $1 \leq j \leq n$. If $x = \sum_{j=1}^n x_j e_j$, then Tx" = "Ax in the sense that

$$Ax = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \ = \ \bigl(\sum_{j=1}^n a_{1j} x_j, \ldots, \sum_{j=1}^n a_{mj} x_j\bigr)$$

and $Tx = \sum_{k=1}^m \bigl[\sum_{j=1}^n a_{kj} x_j\bigr] f_k$.

$\sqrt{\langle \cdot, \cdot \rangle}$ used in the definition of the operator norm $\|T\|$. And although the spectral radius for a given operator does not depend on the underlying inner product, the quantity $\rho(T^*T)$ does depend on the inner product since it depends on the adjoint T^*.

4.9. **Examples in infinite dimensional spaces.**

(A) **Shift operators.** The *left shift operator* $L : \ell^p \to \ell^p$ and the *right shift operator* $R : \ell^p \to \ell^p$ are defined by

$$Lx = (x_2, x_3, \ldots) \text{ and } Rx = (0, x_1, x_2, \ldots) \text{ where } x = (x_j)_{j \geq 1} \in \ell^p.$$

These examples are simple enough that we are able to compute the operator norms exactly.

Since $\|Rx\|_p^p = 0 + \sum_{j \geq 1} |x_j|^p = \|x\|_p^p$, we see that R is bounded with respect to the $\| \cdot \|_p$ norm and $\|R\| = 1$. Furthermore, $\|Rx\|_p = \|x\|_p$ for all $x \in \ell^p$ and hence R is an isometry but R is not onto since $e = (1, 0, 0, \ldots)$ is clearly not in the image of R. However R is injective since if $Rx = 0$, then $x_1 = x_2 = \cdots = 0$ and so $x = 0$.

Similarly, $\|Lx\|_p^p = \sum_{j \geq 2} |x_j|^p \leq \sum_{j \geq 1} |x_j|^p = \|x\|_p^p$ so that $\|Lx\|_p \leq \|x\|_p$ for all $x \in \ell^p$ showing that L is bounded and $\|L\| \leq 1$. However for $f = (0, 1, 0, 0, \ldots)$, we have $\|Lf\|_p = 1 = \|f\|_p$ and so in fact, $\|L\| = 1$. In constrast to R, L is onto but not 1-1; in fact, if $y = (y_j)_{j \geq 1} \in \ell^p$, set $x = (0, y_1, y_2, \ldots) \in \ell^p$ and note that $Lx = y$. Also any vector $x \in \ell^p$ of the form $x = (\alpha, 0, 0, \ldots)$ has the property $Lx = 0$, showing that L is not 1-1.

(B) **Integral operators.** Let T be a linear operator from $C[0, 1]$ into itself given by

$$Tf(x) \ = \ \int_0^1 K(x, y) f(y) \, dy$$

where $K \in C([0, 1] \times [0, 1])$. The continuous function K is called the **kernel** of the linear operator T which is an example of an **integral operator**. For $1/p + 1/q = 1$, we have

$$|Tf(x)| \ \leq \ \Bigl(\int_0^1 |K(x, y)|^q dy \Bigr)^{1/q} \|f\|_{L^p}$$

by Hölder's inequality and so for any $1 \leq r \leq \infty$,

$$\|Tf\|_{L^r} \ \leq \ \Bigl[\int_0^1 \Bigl(\int_0^1 |K(x, y)|^q dy \Bigr)^{r/q} dx \Bigr]^{1/r} \|f\|_{L^p} \ =: \ A_{p,r,K} \|f\|_{L^p},$$

showing that T is bounded from $(C[0, 1], \| \cdot \|_{L^p})$ to $(C[0, 1], \| \cdot \|_{L^r})$ whose operator norm $\|T\|_{p,r,K}$ is bounded above by the mixed $L^r(L^p)$ norm of K given as $A_{p,r,K}$.

Again, it is rare that $\|T\|_{p,r,K} = A_{p,r,K}$, but the special case $p = r = 2$ (which of course is the case where the underlying norms are the same and come from an inner product) is of particular interest. In this case $A_{2,2,K} = \|K\|_{L^2([0,1] \times [0,1])}$ is the L^2 norm of the kernel K over the unit square $[0, 1] \times [0, 1]$. This L^2 norm of the kernel K is called the **Hilbert-Schmidt norm** of T.

The vectors $\{e_n = e^{2\pi inx}\}_{n\in\mathbb{Z}}$ form an orthonormal sequence in the complex Hilbert space $L^2[0,1]$, the Hilbert space completion of $(C[0,1], \|\cdot\|_{L^2})$. In the previous chapter we claimed that in fact $\{e_n\}$ is an ONB for $L^2[0.1]$. This takes a little work to establish and we will do so in the Fourier Analysis course next semester. Here we will take it for granted that $\{e_n\}_{n\in\mathbb{Z}}$ does indeed form an ONB.

In particular, for any $f \in C[0,1]$, $\|f\|_{L^2}^2 = \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2$ where

$$\widehat{f}(n) \;=\; \langle f, e_n\rangle \;=\; \int_0^1 f(x)e^{-2\pi inx}\, dx.$$

For $K \in C([0,1]\times[0,1])$, set $K_x(y) := K(x,y)$ and since $K_x \in C[0,1]$, we have

$$Te_{-n}(x) = \int_0^1 K(x,y)e^{-2\pi inx}\, dy = \langle K_x, e_n\rangle \;\text{ and since } Te_{-n} \in C[0,1],$$

$$\int_0^1 \Big[\int_0^1 |K(x,y)|^2\, dy\Big] dx \;=\; \int_0^1 \Big[\sum_{n\in\mathbb{Z}} |\langle K_x, e_n\rangle|^2\Big] dx \;=\; \sum_{n\in\mathbb{Z}} \int_0^1 |Te_{-n}(x)|^2 dx$$

from which we see that $\|K\|_{L^2}^2 = \sum_{n\in\mathbb{Z}} \|Te_n\|^2$.

Definition 4.10. Let H and K be separable Hilbert spaces. An operator $T \in \mathcal{L}(H,K)$ is called a **Hilbert-Schmidt operator** if there exists an orthonormal basis $\{e_n\}_{n\geq 1}$ of H such that

$$\|T\|_{HS}^2 \;:=\; \sum_{n=1}^{\infty} \|Te_n\|^2 \;<\; \infty.$$

The "norm" $\|T\|_{HS}$ is called the **Hilbert-Schmidt norm** of T, and in principle it depends on the choice of orthonormal basis $\{e_n\}$. However we will later show that the quantity

$$\sum_{n=1}^{\infty} \|Te_n\|^2$$

is *independent* of the choice of orthonormal basis $\{e_n\}$.

Exercise. Show that if $\{f_m\}_{m\geq 1}$ is any orthonormal basis of K, then

$$\|T\|_{HS}^2 \;= \sum_{n=1}^{\infty} \|Te_n\|^2 = \sum_{m,n=1}^{\infty} |\langle Te_n, f_m\rangle|^2. \tag{3}$$

Given two separable Hilbert spaces H and K, we denote by $\mathcal{HS}(H,K)$ the collection of Hilbert-Schmidt operators between H and K.

Exercise. Show that $\mathcal{HS}(H,K)$ is a linear subspace of $\mathcal{L}(H,K)$ and $\|T\|_{HS}$ defines a norm on $\mathcal{HS}(H,K)$, making it into a normed linear space.

Proposition 4.11. *Let H and K be two separable Hilbert spaces and let $T \in \mathcal{HS}(H,K)$ be a Hilbert-Schmidt operator. Then $\|T\| \leq \|T\|_{HS}$.*

This shows that the Hilbert-Schmidt norm $\|T\|_{HS}$ provides an upper bound for the operator norm $\|T\|$.

Proof. Writing $x \in H$ as $\lim_{N\to\infty} x_N$ where $x_N = \sum_{n=1}^N \langle x, e_n\rangle e_n$, we can use the continuity of T to see that $Tx = \lim_{N\to\infty} Tx_N$ and so

$$|\langle Tx, f_m\rangle|^2 \ = \ |\lim_{N\to\infty} \langle Tx_N, f_m\rangle|^2 \ = \ |\sum_{n=1}^{\infty} \langle x, e_n\rangle \langle Te_n, f_m\rangle|^2 \ \le$$

$$\sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \sum_{n=1}^{\infty} |\langle Te_n, f_m\rangle|^2 \ = \ \|x\|^2 \sum_{n=1}^{\infty} |\langle Te_n, f_m\rangle|^2.$$

The single inequality above follows from the Cauchy-Schwarz inequality. Summing over $m \ge 1$, shows that $\|Tx\|^2 = \sum_{m\ge 1} |\langle Tx, f_m\rangle|^2 \le \|T\|_{HS}^2 \|x\|^2$ and hence, by (3), $\|T\| \le \|T\|_{HS}$. □

(C) **Finite-rank operators.** Here we will explore linear operators $T : X \to Y$ between normed linear spaces where the image of T, $T(X)$, is finite-dimensional. Such operators are called **finite-rank operators**. Earlier we explored the situation when X is finite dimensional. Finite-rank operators will be important when we come to discuss compact operators but they will also motivate the discussion on dual spaces.

First suppose that X and Y are only vector spaces, $T : X \to Y$ is a *any* map (not necessarily linear) and Y is finite-dimensional. Let $\mathcal{B} = \{f_1, \ldots, f_m\}$ be a basis for Y so that for every $x \in X$, we have $Tx = c_1(x) f_1 + \cdots + c_m(x) f_m$ for a unique choice of scalars $(c_1(x), \ldots, c_m(x)) \in \mathbb{F}^m$. This defines m maps $c_j : X \to \mathbb{F}$, $1 \le j \le m$ which are called the **coordinate maps** of T with respect to the basis $\mathcal{B}$. If in addition T is linear, then the coordinate maps are also linear; in fact $T(\alpha x_1 + \beta x_2) = \alpha T x_1 + \beta T x_2$ can be written as

$$\sum_{j=1}^m c_j(\alpha x_1 + \beta x_2) f_j \ = \ \alpha \sum_{j=1}^m c_j(x_1) f_j + \beta \sum_{j=1}^m c_j(x_2) f_j$$

or

$$\sum_{j=1}^m \big[c_j(\alpha x_1 + \beta x_2) - \alpha c_j(x_1) - \beta c_j(x_2)\big] f_j \ = \ 0,$$

implying that $c_j(\alpha x_1 + \beta x_2) = \alpha c_j(x_1) + \beta c_j(x_2)$ for every $1 \le j \le m$ by the linear independence of the vectors $\{f_1, \ldots, f_m\}$.

We go one step further and suppose that X and Y are also normed linear spaces.

Proposition 4.12. ** Suppose that $T : X \to Y$ is a linear map between two normed linear spaces X and Y, with Y finite-dimensional, and $c_j : X \to \mathbb{F}, 1 \le j \le m$ are the linear coordinate maps of T with respect to a basis $\mathcal{B} = \{f_j\}_{j=1}^m$ of Y. Then T is continuous if and only if each c_j is continuous.*

Proof. * Suppose that $x_n \to 0$ in X.

If for each $1 \le j \le m$, $|c_j(x_n)| \to 0$ as $n \to \infty$, then $\sum_{j=1} a_j |c_j(x_n)| \to 0$ as $n \to \infty$ for any choice of nonnegative $a_j \ge 0$. Hence

$$\|Tx_n\| = \|\sum_{j=1}^{m} c_j(x_n) f_j\| \le \sum_{j=1}^{m} |c_j(x_n)| \|f_j\| \;\to\; 0 \;\text{ as } n \to \infty,$$

showing that $Tx_n \to 0$.

On the other hand, suppose $Tx_n \to 0$ as $n \to \infty$.

Since all norms on Y are equivalent, we can find a constant A such that $\|y\|_{\mathcal{B},\infty} \le A\|y\|$ for all $y \in Y$. Recall that $\|y\|_{\mathcal{B},\infty} = \max_{1\le j\le m} |y_j|$ where $y = \sum_{j=1}^{m} y_j f_j$. Hence for any $1 \le j \le m$,

$$|c_j(x_n)| \;\le\; \max_{1\le j\le m} |c_j(x_n)| \;=\; \|Tx_n\|_{\mathcal{B},\infty} \;\le\; A\|Tx_n\| \;\to\; 0 \;\text{ as } n \to \infty,$$

showing that $c_j(x_n) \to 0$ as $n \to \infty$. □

Proposition 4.12 shows that the problem of understanding finite-rank operators in $\mathcal{L}(X, Y)$ with Y finite-dimensional reduces to understanding the continuous linear maps from X to the scalar field $\mathbb{F}$. This is a topic to which we will turn soon.

Now let us suppose both $X = H$ and $Y = K$ are separable Hilbert spaces, K is infinite-dimensional, and $T \in \mathcal{HS}(H, K)$. Then there is an orthonormal basis $\{e_j\}_{j\ge1}$ for H, and by (3) we have $\|T\|_{HS}^2 = \sum_{n\ge1} \|Te_n\|^2 < \infty$.

For every $x \in H$, we have $x = \sum_{n\ge1} \langle x, e_n\rangle e_n$ and by the continuity of T,

$$Tx \;=\; T\Big(\lim_{N\to\infty} \sum_{n=1}^{N} \langle x, e_n\rangle e_n\Big) \;=\; \lim_{N\to\infty} \sum_{n=1}^{N} T\big(\langle x, e_n\rangle e_n\big) \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle Te_n.$$

We define a sequence of finite-rank operators $T_N : H \to K$ by $T_N x := \sum_{n=1}^{N} \langle x, e_n\rangle Te_n$. Then $Tx - T_N x = \sum_{n\ge N+1} \langle x, e_n\rangle Te_n$ and so by the Cauchy-Schwarz inequality,

$$\|Tx - T_N x\| \;\le\; \sum_{n=N+1}^{\infty} |\langle x, e_n\rangle| \|Te_n\| \;\le\; \sqrt{\sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2} \sqrt{\sum_{n=N+1}^{\infty} \|Te_n\|^2}.$$

Hence $\|Tx - T_N x\| \le \sqrt{\sum_{n\ge N+1} \|Te_n\|^2}\; \|x\|$ implying

$$\|T - T_N\| \;\le\; \sqrt{\sum_{n=N+1}^{\infty} \|Te_n\|^2} \;\;\to\;\; 0 \;\text{ as } N \to \infty.$$

We record this observation as a proposition.

Proposition 4.13. *Let H and K be two separable Hilbert spaces and suppose $T \in \mathcal{HS}(H, K)$. Then there exists a sequence $\{T_N\} \subseteq \mathcal{L}(H, K)$ of finite-rank operators such that $T_N \to T$ in $\mathcal{L}(H, K)$; that is, $\|T_N - T\| \to 0$.*

(D) **An unbounded linear map.** Here we will consider a linear operator $T : X \to Y$ between two explicit normed linear spaces X and Y where T is **not** bounded or equivalently, continuous. A different example was considered in lecture.

Consider the normed linear space $(C[0,1], \|\cdot\|_{L^1})$ and the linear operator $T : C[0,1] \to \mathbb{F}$ defined by $Tf = f(0)$. We claim that T is not bounded. If fact, consider the sequence

$$f_n(x) = \begin{cases} 1 & 0 \le x \le 1 - 1/n \\ -n(x-1) & 1 - 1/n < x \le 1 \end{cases} \in C[0,1]$$

where $\|f_n\|_{L^1} = 1/2n$, implying $f_n \to 0$ in $(C[0,1], \|\cdot\|_{L^1})$. However $Tf_n = f_n(0) = 1$ for all n, showing that T is not continuous and so it is not bounded.

A natural question arises at this point: given any two normed linear spaces X and Y where X is infinite dimensional, does there always exist a linear map $T : X \to Y$ such that T is NOT continuous? See the Exercises.

EXERCISES

1. Show that if T is a bounded linear operator between two normed linear spaces X and Y, then

$$\sup_{\|x\| \le 1} \|Tx\| = \sup_{\|x\| < 1} \|Tx\| = \sup_{\|x\| = 1} \|Tx\|$$

and this common value is the operator norm $\|T\|$.

2. Suppose that if T is a bounded linear operator between two normed linear spaces X and Y. Show that $\|T\| = \min\{M : \|Tx\| \le M\|x\| \text{ for all } x \in X\}$.

3. Let H and K be two separable Hilbert spaces. Show that $(\mathcal{HS}(H,K), \|\cdot\|_{HS})$ is a normed linear space. Hint: use (3) to show that $\mathcal{HS}(H,K)$ is a vector space and $\|\cdot\|_{HS}$ is a norm on $\mathcal{HS}(H,K)$.

4. Consider the left and right shift operators on ℓ^p considered in Example A. Find the kernels and images of both operators and show that the images are closed subspaces.

5. For every $n \ge 1$, define $T : \ell^2 \to \ell^2$ by $T_n x = (x_1, x_2, \ldots, x_n, 0, 0, \ldots)$ where $x = (x_j)_{j \ge 1} \in \ell^2$. Show that $T_n \in \mathcal{L}(\ell^2)$ and $\|T_n\| = 1$ for every $n \ge 1$.

For every $x \in \ell^2$, show $T_n x \to x$ in ℓ^2; that is, $\|T_n x - x\|_2 \to 0$ as $n \to \infty$. Hence $T_n \to I$ *pointwise* where $I : \ell^2 \to \ell^2$ is the identity operator.

Finally show that $\|T_n - T_m\| = 1$ for every $n \ne m$. Hence $\{T_n\} \subseteq \mathcal{L}(\ell^2)$ does not converge in the operator norm to any operator.

6. Let H be a Hilbert space and let M be a closed subspace. Show that the orthogonal projection $P_M : H \to H$ is a bounded linear operator and calculate its norm.

7. Show that if X, Y and Z are normed linear spaces and if $T \in \mathcal{L}(X, Y)$ and $S \in \mathcal{L}(Y, Z)$ then $ST \in \mathcal{L}(X, Z)$ and $\|ST\| \leq \|S\|\|T\|$. Deduce that $\mathcal{L}(X)$ is an algebra and that $\|T^n\|^{1/n} \leq \|T\|$ for all n. Show that $\|T^n\|^{1/n} \to \inf_m \|T^m\|^{1/m}$ as $n \to \infty$. Finally, if X is complete, and if $\lim_{n\to\infty} \|T^n\|^{1/n} < 1$, show that $I - T$ is invertible in $\mathcal{L}(X)$.

8. Let H and K be two Hilbert spaces and $M \subseteq H$ a proper closed linear subspace of H which contains a non-zero vector. Hence M with the inner product from H is a Hilbert space in its own right. Suppose that $T : M \to K$ is a bounded linear map with operator norm A; that is $T \in \mathcal{L}(M, K)$ and

$$A = \sup_{x \in M : \|x\| \leq 1} \|Tx\|.$$

Show that there exists an $S \in \mathcal{L}(H, K)$ such that $Sx = Tx$ for all $x \in M$ and $\|S\| = A$.

Hint: Write $H = M \oplus M^{\perp}$.

9. Let X and Y be two normed linear spaces and suppose $M \subset X$ is a dense linear subspace of X. Suppose $T, S : X \to Y$ are two bounded, linear operators such that $Tx = Sx$ for all $x \in M$. Show that $T = S$.

10. Let H be a separable Hilbert space with an orthonormal set $\{e_j\}_{j=1}^{\infty}$. Let $\{\lambda_j\}_{j\geq 1} \subset \mathbb{F}$ and $X = \mathrm{Span}(\{e_j\}_{j\geq 1})$.

(a) Show that there exists a unique linear operator $T : X \to H$ such that $Te_j = \lambda_j e_j$ for every $j \geq 1$.
(b) Suppose that $A := \sup_j |\lambda_j| < \infty$. Show that there exists an operator $S \in \mathcal{L}(H)$ such that $Sx = Tx$ for all $x \in X$ and that $\|S\| = A$.
(c) When is such an S unique?

11. Let X be a real normed space and let $T : X \to \mathbb{R}$ be linear. Show that T is continuous if and only if $\ker(T)$ is closed.

19.* This exercise shows that for every infinite dimensional normed linear space $(X, \|\cdot\|)$, we can find a linear map $\phi : X \to \mathbb{F}$ which is not continuous; that is, $\phi \notin X^*$. First show that X possesses an infinite sequence $M = \{e_n\}_{n\geq 1}$ of linearly independent, unit vectors.

Next we need the fact that M can be extended to a Hamel basis $\mathcal{B}$; that is, $M \subset \mathcal{B}$ and every $x \in X$ has a unique representation $x = \sum_j x_j b_j$ as a *finite* linear combination of vectors $\{b_j\}$ in $\mathcal{B}$. Let us assume this.

For each $n \geq 1$, define for any $x = \sum_j x_j b_j$,

$$\phi_n(x) = \begin{cases} x_j & \text{if } b_j = e_n \\ 0 & \text{otherwise} \end{cases}.$$

Now define $\phi : X \to \mathbb{F}$ by $\phi(x) := \sum_{n \geq 1} n\phi_n(x)$. Show that ϕ is well-defined, linear but NOT continuous.

Can you extend this to find a linear map $T : X \to Y$ which is not continuous whenever X is infinite dimensional?

5. Duality

Although it is generally very difficult to characterise $\mathcal{L}(X,Y)$ even when the normed spaces X and Y are given explicitly, it is often relatively straightforward to understand matters when the target space Y consists of the scalars. So we now study the space $\mathcal{L}(X,\mathbb{F})$ and we shall see that in many cases we can explicitly calculate $\|T\|$ for $T \in \mathcal{L}(X,\mathbb{F})$. For these and many other reasons the spaces $\mathcal{L}(X,\mathbb{F})$ occupy a special place in the theory and have a special name.

Definition 5.1. Let X be a normed linear space over the field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. The **dual space** to X is the space $\mathcal{L}(X,\mathbb{F})$, i.e. the space of all continuous linear maps from X into $\mathbb{F}$. The dual space $\mathcal{L}(X,\mathbb{F})$ is denoted by X^*. A member T of X^* is called a **bounded linear functional** on X.

Note that by Theorem 4.4, the dual space X^* is complete *whether or not* X is. At this point we have to admit the possibility that X^* might consist only of the trivial space $\{0\}$. In fact, this possibility does not occur for non-trivial X, but this is harder to prove – it is a consequence of the famous Hahn–Banach theorem – and so we omit it in general. However, for the case of *Hilbert spaces* H we can see directly that H^* is nontrivial provided H is. See below.

The definition of the dual space parallels the theory of dual spaces in linear algebra where the dual of a finite-dimensional vector space X is simply the collection of linear maps from X into the base field. One of the great results of duality theory in linear algebra is that X^{**} is *canonically* isomorphic to X and we'll be exploring to what extent this also holds in the infinite-dimensional setting. If X is $\mathbb{F}^n$ thought of as column vectors, then X^* is $\mathbb{F}^n$ thought of as row vectors, and the action of $T \in X^*$ on $x \in X$ is given by matrix multiplication.

First we need to explore some examples.

Proposition 5.2. *Let* $1 < p < \infty$. *The dual of* ℓ^p *is isometrically isomorphic to* ℓ^q, *where* $1/p + 1/q = 1$.

Before we prove this it may be helpful to recall Hölder's inequality stating that

$$\sum_j |x_j y_j| \leq \|x\|_p \|y\|_q.$$

Moreover $y \in \ell^q$ if and only if $\sup_{\|x\|_p \leq 1} |\sum_j x_j y_j| < \infty$, in which case

$$\|y\|_q = \sup_{\|x\|_p \leq 1} |\sum_j x_j y_j|.$$

Also, recall that the sgn function is defined on $\mathbb{C}$ by $\operatorname{sgn}(z) = |z|/z$ when $z \neq 0$ and $\operatorname{sgn}(0) = 0$. (So, for $t \in \mathbb{R}$, $\operatorname{sgn}(t) = 1$ if $t > 0$, $\operatorname{sgn}(t) = -1$ if $t < 0$, and $\operatorname{sgn}(0) = 0$.)

Proof. We have already seen in Workshop 3 that if $\xi \in \ell^q$, then the linear map $T_\xi : \ell^p \to \mathbb{F}$ given by

$$T_\xi x = \sum_{j=1}^{\infty} x_j \xi_j$$

defines a bounded linear operator on ℓ^p and that $\|T_\xi\| = \|\xi\|_q$.

It is easy to see that the map $T : \xi \mapsto T_\xi$ is a linear map from ℓ^q into $(\ell^p)^*$ and coupled with the above this tells us that $T : \ell^q \to (\ell^p)^*$ is a linear isometry (and in particular is injective). To complete the proof we therefore need to show that T is surjective. That is, we need to show that given any $L \in (\ell^p)^*$, there is a $\xi \in \ell^q$ such that $L = T_\xi$. How do we find this ξ? Our only information is what L does to individual vectors in ℓ^p and it makes sense to consider what L does to e_j

Indeed, for $j \geq 1$ let $\xi_j = Le_j$. We claim that $\xi = (\xi_j) \in \ell^q$. To see this we can either use the criterion for membership of ℓ^q mentioned above, or we can argue as follows. Consider

$$L(|\xi_1|^{q-1}\mathrm{sgn}(\xi_1), \dots, |\xi_N|^{q-1}\mathrm{sgn}(\xi_N), 0, 0, \dots) = \sum_{j=1}^{N} |\xi_j|^{q-1}\mathrm{sgn}(\xi_j)\xi_j = \sum_{j=1}^{N} |\xi_j|^q$$

$$\leq \|L\|(\sum_{j=1}^{N}[||\xi_j|^{q-1}\mathrm{sgn}(\xi_j)|]^p)^{1/p} = \|L\|(\sum_{j=1}^{N}|\xi_j|^q)^{1/p},$$

using the fact that $p(q-1) = q$. Rearranging, we get

$$(\sum_{j=1}^{N}|\xi_j|^q)^{1/q} \leq \|L\|$$

and since this is independent of N we conclude that $\xi \in \ell^q$ and that $\|\xi\|_q \leq \|L\|$.

Finally, we need to see that $T_\xi = L$, i.e. $T_\xi x = Lx$ for all $x \in \ell^p$. Now if $x = \sum_j x_j e_j$ is a *finite* linear combination of e_j's, then $Lx = \sum_j x_j Le_j = \sum_j x_j \xi_j = T_\xi x$ so that L and T_ξ coincide on the dense linear subspace of ℓ^p consisting of finite linear combinations of e_j's. So by Proposition 4.5 $L = T_\xi$ (since $\mathbb{F}$ is complete). $\square$

In particular we see that the dual space to the Hilbert space ℓ^2 is isometrically isomorphic to itself. We shall soon see something very similar for all Hilbert spaces.

Exercises. 1. The dual of ℓ^1 is isometrically isomorphic to ℓ^∞.

2. The dual of c_0 is isometrically isomorphic to ℓ^1.

3. There is an isometric imbedding of ℓ^1 into the dual of ℓ^∞.

It is not hard to see that for $1 \leq p < \infty$ and $1/p + 1/q = 1$, the space L^q imbeds isometrically into $(L^p)^*$. In fact this isometry is also surjective – the fact that every linear functional on L^p arises from an L^q function is harder to prove and is addressed in courses on measure theory. We will use the identification of $(L^p)^*$ with

L^q when $1 \leq p < \infty$ freely below. When $p = 2$, we can (more or less) obtain it directly, as we shall now see.

The dual of a Hilbert space. We now study the dual space of an arbitrary Hilbert space H. For the time being we work in the setting of *complex* Hilbert spaces – the real case is a little less subtle. You may wish to refer back to Question 1 of Workshop 3 at this point. Indeed, in that question we showed that if H is an inner product space and if $y \in H$, then the map $T_y : H \to \mathbb{C}$ given by

$$T_y : x \mapsto \langle x, y \rangle$$

defines a member of H^* and that $\|T_y\| = \|y\|$. It is natural to expect the map $y \mapsto T_y$ to be a linear map from H into H^* – but when we are working over the complex numbers this simply isn't so. Indeed, since

$$T_{\alpha_1 y_1 + \alpha_2 y_2} x = \langle x, \alpha_1 y_1 + \alpha_2 y_2 \rangle = \overline{\alpha_1} \langle x, y_1 \rangle + \overline{\alpha_2} \langle x, y_2 \rangle = \overline{\alpha_1} T_{y_1} x + \overline{\alpha_2} T_{y_2} x,$$

we have

$$T_{\alpha_1 y_1 + \alpha_2 y_2} = \overline{\alpha_1} T_{y_1} + \overline{\alpha_2} T_{y_2}$$

and so the map $T_y : H \to \mathbb{C}$ is *conjugate-linear* instead of linear. There is simply no way to avoid this in the setting of general complex Hilbert spaces H. This is because there is no natural notion of a "complex conjugate" map $x \to \overline{x}$ from H to itself. (In the case of *specific* complex Hilbert spaces such as ℓ^2 or L^2 there is such a notion – what is it?)

Theorem 5.3 (F. Riesz representation theorem)**.** *Let H be a complex Hilbert space. Then there is a conjugate-linear isometric isomorphism from H onto H^* given by*

$$T : y \mapsto T_y$$

where $T_y x = \langle x, y \rangle$.

Proof. We have already seen everything except the surjectivity of T. Let $L \in H^*$ with $L \neq 0$. We want to show that $L = T_y$ for some $y \in H$. Now $K := \ker L$ is a proper closed linear subspace of H and so by Theorem 3.15 we can write

$$H = K \oplus K^\perp.$$

Pick $z \in K^\perp \setminus \{0\}$; it's pretty clear that our y has to be a scalar multiple of z. Consider $L(x)z - L(z)x$. If we apply L to this vector we get 0, hence $L(x)z - L(z)x \in K$ for all $x \in H$. So if we take its inner product with $z \in K^\perp$ we get zero:

$$L(x)\langle z, z \rangle - L(z)\langle x, z \rangle = 0$$

for all $x \in H$. Rearranging, this gives

$$L(x) = \langle x, z \rangle \frac{L(z)}{\|z\|^2} = \langle x, y \rangle$$

where $y = z\overline{L(z)}/\|z\|^2$. So $L = T_y$ for this y and therefore T is surjective. (Alternatively, set $y = \alpha z$ and check that for a suitable value of α, $L = T_y$.) $\square$

Remark. When the scalar field is $\mathbb{R}$, then the map T is linear instead of conjugate-linear.

Examples – The dual of ℓ^2. Let H be the Hilbert space ℓ^2 over the complex numbers $\mathbb{C}$. Then for $y \in \ell^2$ the mapping $y \mapsto T_y$ of Theorem 5.3 is given by

$$T_y(x) = \sum_{j=1}^{\infty} x_j \overline{y}_j.$$

Note the complex conjugate over y_j. If instead H is the Hilbert space ℓ^2 over the real numbers $\mathbb{R}$, then

$$T_y(x) = \sum_{j=1}^{\infty} x_j y_j$$

without the complex conjugate. Similarly for L^2.

Adjoints and transposes: i) Adjoints.

Recall that if $T : V \to W$ is a linear transformation between finite-dimensional inner product spaces V and W, then there is a unique linear transformation $T^* : W \to V$ with the property that

$$\langle Tx, y \rangle = \langle x, T^* y \rangle \text{ for all } x \in V, y \in W.$$

The linear transformation T^* is called the adjoint of T. If the matrix of T is (a_{ij}) then the matrix of T^* is $(\overline{a_{ji}})$. Our next purpose is to extend this notion to bounded linear operators on a Hilbert space.

Theorem 5.4. *Let H and K be Hilbert spaces, and let $T \in \mathcal{L}(H, K)$. For each $y \in K$ there is a unique $T^* y \in H$ such that*

$$\langle Tx, y \rangle = \langle x, T^* y \rangle \tag{4}$$

for all $x \in H$. The map $y \mapsto T^* y$ is a linear map from K to H, is bounded and satisfies $\|T^*\| = \|T\|$. We call the map $T^* \in \mathcal{L}(K, H)$ the **adjoint** *of T. The adjoint of T^* is T.*

Proof. For $y \in K$ consider the map

$$f_y : x \mapsto \langle Tx, y \rangle.$$

This is a bounded linear functional on H with norm at most $\|T\|\|y\|$. So by the Riesz representation theorem, Theorem 5.3, there is a unique $T^* y \in H$ such that

$$f_y(x) = \langle x, T^* y \rangle \text{ for all } x \in H,$$

and $\|T^* y\| = \|f_y\| \le \|T\|\|y\|$. The map $y \mapsto T^* y$ is linear because

$$\begin{aligned} \langle x, T^*(\alpha y_1 + \beta y_2) \rangle &= \langle Tx, \alpha y_1 + \beta y_2 \rangle = \overline{\alpha} \langle Tx, y_1 \rangle + \overline{\beta} \langle Tx, y_2 \rangle \\ &= \overline{\alpha} \langle x, T^* y_1 \rangle + \overline{\beta} \langle x, T^* y_2 \rangle = \langle x, \alpha T^* y_1 + \beta T^* y_2 \rangle, \end{aligned}$$

and is bounded since $\|T^* y\| \le \|T\|\|y\|$; this also shows that $\|T^*\| \le \|T\|$.

It remains to see that $\|T^*\| = \|T\|$ and that $T^{**} = T$. We first see the latter.

Applying the defining equation (4) to T^* in place of T we see that $\langle T^*y, x\rangle = \langle y, T^{**}x\rangle$ for all $x \in H$ and $y \in K$; since we also have $\langle T^*y, x\rangle = \langle y, Tx\rangle$ for all x and y we conclude that

$$\langle y, Tx - T^{**}x\rangle = 0$$

for all $x, y \in H$. Taking $y = Tx - T^{**}x$ gives $\|Tx - T^{**}x\| = 0$ for all x, thus $T^{**} = T$.

Finally, since $\|T^*\| \leq \|T\|$ we have

$$\|T\| = \|T^{**}\| \leq \|T^*\| \leq \|T\|$$

and all these quantities coincide. □

Proposition 5.5. *Let H and K be Hilbert spaces, and let $T \in \mathcal{L}(H, K)$. Then*

$$\|T^*T\| = \|TT^*\| = \|T\|^2 = \|T^*\|^2.$$

Proof. We have already seen the final equality in Theorem 5.4. By symmetry between T and T^* it suffices to show that $\|T^*T\| = \|T\|^2$. Now we always have (see the exercises to the previous section) $\|ST\| \leq \|S\|\|T\|$ so that $\|T^*T\| \leq \|T\|^2$. On the other hand,

$$\|Tx\|^2 = \langle Tx, Tx\rangle = \langle x, T^*Tx\rangle \leq \|x\|\|T^*Tx\|$$

so that for $\|x\| \leq 1$ we have

$$\|Tx\|^2 \leq \|T^*Tx\| \leq \|T^*T\|.$$

Therefore

$$\|T\|^2 \leq \|T^*T\|$$

and we are done. □

The theory of duality is of great importance not only theoretically but also practically, in showing boundedness of concrete operators between concrete Hilbert spaces. Suppose we wish to show that an operator T given by an expression of the form

$$Tf(x) = \int_0^1 K(x, y)f(y)dy$$

is bounded on $L^2([0, 1])$. (It may not even be the case that the expression for T makes formal sense for arbitrary $f \in L^2$, but suppose that it does make formal sense at least for f belonging to some dense subspace of L^2 such as $C([0, 1])$. Now the adjoint T^*, at least formally, will have to be given by an expression of the form

$$T^*g(x) = \int_0^1 \overline{K(y, x)}g(y)dy.$$

Now it may well be that one can prove directly that $T^* \in \mathcal{B}(L^2)$, and, since $\|T\| = \|T^*\|$ for *bounded* linear operators we will have shown that $\|T\| = \|T^*\|$ *provided* that $\|T\| < \infty$. It is then usually a matter of routine analysis (working with the dense subspace of L^2 on which T is well-defined) to finesse the situation and conclude that T is indeed bounded.

In order to see that T is bounded between H and K, a good additional strategy will often be to try to see that its formal dual T^* satisfies either $T^*T \in \mathcal{B}(H)$ or $TT^* \in \mathcal{B}(K)$, and thence to deduce that indeed T is bounded between H and K. Surprisingly, it is often very easy to verify *one* of these two statements directly but almost impossible to verify the other one (although equivalent!) directly. There is an example illustrating these ideas in the exercises below.

In a related scenario, sometimes one knows that an operator T is bounded, but one needs good estimates on how big its norm is. It will often be the case that we can get a good estimate on $\|T^*\|$ or $\|T^*T\|$ with relative ease; the theory of duality allows us to deduce good bounds for $\|T\|$ also.

Adjoints and transposes: ii) Transposes.

We have made strong use of the inner product structure in our discussion of adjoints. How far can we develop a similar theory in the context of bounded linear operators on general *normed* spaces? Once again we take our cue from the finite-dimensional setting. If $A = (a_{ij})$ is an $m \times n$ matrix, its *transpose* A' is the $n \times m$ matrix (a_{ji}), and it satisfies

$$\mathbf{y}^T A\mathbf{x} = \mathbf{x}^T A'\mathbf{y}$$

for all column vectors $\mathbf{x} \in \mathbb{F}^n$ and $\mathbf{y} \in \mathbb{F}^m$.

Let X and Y be normed spaces and let $T \in \mathcal{L}(X, Y)$. We want to define T' as a member of $\mathcal{L}(Y^*, X^*)$. So for $g \in Y^*$ – that is, $g : Y \to \mathbb{F}$ a bounded linear operator – we want $T'g$ to be a member of X^*, that is, to be a bounded linear operator from X to $\mathbb{F}$. In particular we have to define, for each $x \in X$ and each bounded linear $g : Y \to \mathbb{F}$, a member $(T'g)(x)$ of $\mathbb{F}$ by combining x, T and g in some way. Now there is only one sensible way to do this: x gets mapped to $Tx \in Y$ by T, and this gets mapped to $g(T(x)) \in \mathbb{F}$ by g. So we define $T'g$ by its action on X via $(T'g)(x) = g(T(x))$.

Note that $g(T(\alpha_1 x_1 + \alpha_2 x_2)) = \alpha_1 g(T(x_1)) + \alpha_2 g(T(x_2))$ so that $T'g$ is indeed a linear map on X and moreover it is bounded since $|T'g(x)| = |g(Tx)| \le \|g\|\|T\|\|x\|$, showing that $\|T'g\| \le \|g\|\|T\|$. So the following is a good definition:

Definition 5.6. Let X and Y be normed spaces and let $T \in \mathcal{L}(X, Y)$. We define the **transpose** of T, denoted by T' as the map $T' : Y^* \to X^*$ given by

$$(T'g)(x) = g(T(x)).$$

Proposition 5.7. *Let X and Y be normed spaces and let $T \in \mathcal{L}(X, Y)$. Then the transpose map $T' : Y^* \to X^*$ is a bounded linear operator and moreover $\|T'\| \le \|T\|$.*

Proof. We have already seen that T' is well-defined as a map from Y^* to X^*. We still need to see it is linear, that it is bounded and that $\|T'\| \le \|T\|$.

For linearity consider g_1 and g_2 in Y^* and α_1 and α_2 in $\mathbb{F}$. We need to see $T'(\alpha_1 g_1 + \alpha_2 g_2) = \alpha_1 T' g_1 + \alpha_1 T' g_1$. To do this we apply both sides to an $x \in X$ and use the defining property $(T'g)(x) = g(T(x))$. We leave the details to the reader.

For boundedness, we have already seen that $\|T'g\| \leq \|g\|\|T\|$ for all $g \in Y^*$, so taking the sup over all g with $\|g\| \leq 1$ gives the desired inequality $\|T'\| \leq \|T\|$. $\square$

For any normed space X and any $f \in X^*$ we have $\|f\| = \sup_{\|x\|\leq 1} |f(x)|$ by the definition of the norm on X^*. It turns out that it is also true that for every $x \in X$,

$$\|x\| = \sup_{f\in X^*,\, \|f\|\leq 1} |f(x)|. \tag{5}$$

Since $|f(x)| \leq \|f\|\|x\|$ it is trivial that $\sup_{\|f\|\leq 1} |f(x)| \leq \|x\|$ and the reverse inequality $\|x\| \leq \sup_{\|f\|\leq 1} |f(x)|$ follows from the Hahn–Banach theorem, and we will not prove it in general. But in the setting of Hilbert spaces it follows from Theorem 5.3, and we can also easily see it directly when X is any of the spaces c_0, ℓ^p for $1 \leq p < \infty$ or $C([0,1])$, L^p for $1 \leq p < \infty$.

Exercise. Show that if X is any of the spaces c_0, ℓ^p for $1 \leq p < \infty$, or $C([0,1])$, L^p for $1 \leq p < \infty$ we have

$$\|x\| = \sup_{f\in X^*,\, \|f\|\leq 1} |f(x)|.$$

Proposition 5.8. *Suppose X and Y are normed spaces and $T \in \mathcal{L}(X,Y)$. If Y satisfies $\|y\| = \sup_{g\in Y^*,\, \|g\|\leq 1} |g(y)|$ for all $y \in Y$, then $\|T'\| = \|T\|$.*

Proof. We already know that $\|T'\| \leq \|T\|$ so we want to show $\|T\| \leq \|T'\|$. Let $x \in X$. So $Tx \in Y$ and therefore, by hypothesis,

$$\begin{aligned}\|Tx\| &= \sup_{g\in Y^*, \|g\|\leq 1} |g(Tx)| = \sup_{g\in Y^*, \|g\|\leq 1} |T'(g)(x)| \\ &\leq \sup_{g\in Y^*, \|g\|\leq 1} \|T'(g)\|\|x\| = \|T'\|\|x\|.\end{aligned}$$

$\square$

In many practical applications, it may be difficult to compute or even estimate $\|T\|$ directly, but computation of $\|T'\|$ is relatively straightforward. So the equality of norms $\|T'\| = \|T\|$ is very useful in practice. Suppose we wish to show that an operator T given by an expression of the form

$$Tf(x) = \int K(x,y)f(y)dy$$

is bounded on some L^p space. (It may not even be the case that the expression for T makes formal sense for arbitrary $f \in L^p$, but suppose that it does make formal sense at least for f belonging to some dense subspace of L^p.) Now the transpose T', at least formally, will have to be given by an expression of the form

$$T'g(x) = \int K(y,x)g(y)dy.$$

Now it may well be that one can prove directly that $T' \in \mathcal{B}(L^q)$ where $1/p+1/q = 1$, and, accepting for the moment that $\|T\| = \|T'\|$ for bounded linear operators we will have shown that $\|T\| = \|T'\|$ *provided* that $\|T\| < \infty$. It is then usually a matter of routine analysis (working with the dense subspace of L^p on which T is well-defined) to finesse the situation and conclude that T is indeed bounded.

Reflexivity. If X is any normed space and $x \in X$, we can make x act on X^* by considering the map

$$\tilde{x} : f \mapsto f(x).$$

Now it is clear that this is a linear map, and it is bounded since $|f(x)| \le \|f\|\|x\|$, and this also shows that $\|\tilde{x}\| \le \|x\|$. Thus $\tilde{x}$ is a member of the dual of X^*, i.e. $\tilde{x} \in X^{**}$. Now the map $\iota : x \mapsto \tilde{x} = \iota(x)$ is linear – for $\iota(\alpha x + \beta y)(f) = f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = \alpha\iota(x)(f) + \beta\iota(x)(f)$. Moreover, by (5), we also know that $\|\tilde{x}\| = \|x\|$ (by the Hahn–Banach theorem in general) so that ι is an isometry of X into X^{**}.

Definition 5.9. The Banach space X is **reflexive** if the map $\iota : X \to X^{**}$ is surjective.

Exercise. Any Hilbert space is reflexive. Any ℓ^p or $L^p([0,1])$ for $1 < p < \infty$ is reflexive. The spaces c_0 and $C([0,1])$ are **not** reflexive.

Remark. Suppose that X and Y are reflexive spaces and that $T \in \mathcal{L}(X, Y)$. Then $T'' = \iota_Y \circ T \circ \iota_X^{-1}$.

Remark. It's possible for ι not to be surjective – so that X is not reflexive – but nonetheless for X and X^{**} to be isometrically isomorphic. Look up the "James space" on Wikipedia.

EXERCISES

1. Let X be a normed space. Show that $\mathcal{L}(\mathbb{F}, X)$ is isometrically isomorphic to X.

2. (a) Show that the dual of ℓ^1 is isometrically isomorphic to ℓ^∞.

(b) Show that the dual of c_0 is isometrically isomorphic to ℓ^1.

(c) Show that there is an isometric imbedding of ℓ^1 into the dual of ℓ^∞.

(d) Why does your argument for part (a) fail to show that $(\ell^\infty)^*$ is isometrically isomorphic to ℓ^1?

(e) Look up "Banach limit" on Wikipedia. What does this tell us about the isometric imbedding of part (c)?

3. Suppose $1 < p < \infty$ and let $g \in L^q$. Show that the map $T_g : f \mapsto \int fg$ defines a member of $(L^p)^*$ and that $\|T_g\| = \|g\|_q$.

4. Let U, V and W be complex vector spaces. Suppose that $T : U \to V$ and $S : V \to W$ are conjugate-linear maps. Show that $ST : U \to W$ is linear. Deduce

that if H is a complex Hilbert space there is a natural *linear* isometric isomorphism from H onto H^{**}.

5. Show that the mapping $T \mapsto T^*$ is conjugate-linear. Show that if H, K and M are Hilbert spaces, $T \in \mathcal{L}(H,K)$ and $S \in \mathcal{L}(K,M)$, then $(ST)^* = T^*S^*$.

6. Show that if X is any of the spaces c_0, $C([0,1])$, ℓ^p or L^p for $1 \leq p < \infty$, we have $\|x\| = \sup_{f\in X^*,\, \|f\|\leq 1} |f(x)|$.

7. Let H and K be Hilbert spaces and suppose $T \in \mathcal{L}(H,K)$. What is the relation between T^* and T'?

8. Prove that the map ι of Definition 5.9 has norm exactly 1 *without* using $\|\tilde{x}\| = \|x\|$ for all $x \in X$.

9. Every Hilbert space is reflexive. The spaces ℓ^p or $L^p([0,1])$ for $1 < p < \infty$ is reflexive. The spaces c_0 and $C([0,1])$ are **not** reflexive.

10. Show that if $T \in \mathcal{L}(H,K)$ then ker $T^* = (\mathrm{im} T)^\perp$ and that $(\ker T^*)^\perp = \overline{\mathrm{im} T}$.

11. For $f \in L^2(0,\infty)$ let
$$Tf(s) = \frac{1}{s}\int_0^s f(t)dt.$$
The aim of this question is to show that T defines an operator in $\mathcal{L}(L^2(0,\infty))$ and that $\|T\| = 2$.

(a) Write down a formula for the formal adjoint T^*g of T applied to $g \in C_c((0,\infty))$.

(b) Calculate T^*T formally, and show that $T^*T = T + T^*$.

(c) Show that $T \in \mathcal{L}(L^2(0,\infty))$ and show that $\|T\| \leq 2$.

(d) Show that $\|T\| = 2$.

(e) Fill in any details to ensure the argument is rigorous.

12. Let H be a Hilbert space and suppose $S, T \in \mathcal{L}(H)$ satisfy $S^*T = 0 = ST^*$. Show that $(S+T)(S+T)^* = SS^* + TT^*$ and more generally that $\{(S+T)(S+T)^*\}^n = (SS^*)^n + (TT^*)^n$ when $n \in \mathbb{N}$ is a power of 2. Deduce that $\|S+T\|^{2n} \leq \|S\|^{2n} + \|T\|^{2n}$ when n is a power of 2, and hence that $\|S+T\| \leq \max\{\|S\|, \|T\|\}$.

13. Let X be a normed space. Suppose that X^* is separable. Show that X is separable. Deduce that $(L^\infty)^* \neq L^1$. (**Hint:** Let (f_n) be a countable dense subset of X^*, and let $x_n \in X$ satisfy $\|x_n\| = 1$ and $|f_n(x_n)| \geq \|f_n\|/2$. Show that the set of all linear combinations of $\{x_n\}$ with rational coefficients is dense in X.)

6. Compact operators

Definition 6.1. Let X and Y be normed spaces and let $T \in \mathcal{L}(X, Y)$. Then T is said to be **compact** if $\overline{T(\{x : \|x\| \leq 1\})}$ is a compact subset of Y.

Recall that for a subset of a metric space to be compact it is necessary that it be closed and bounded; taking the closure of $T(\{x : \|x\| \leq 1\})$ ensures the former and the boundedness of T ensures the latter. So $\overline{T(\{x : \|x\| \leq 1\})}$ is at least a candidate for being a compact subset of Y.

Next recall that in a finite-dimensional space, closed and bounded sets are compact (the Heine–Borel theorem), so that if T is a finite-rank operator it is automatically compact. On the other hand, if X is infinite-dimensional the identity opertor on X is not compact. This is because we have already seen (see Theorem 2.33) that the closed unit ball of a normed space X is compact if and only if X is finite-dimensional.

Exercise. Let H be a separable Hilbert space with an orthonormal basis $\{e_j\}$, $\lambda \in \ell^\infty$ and $T \in \mathcal{L}(H)$ be defined by $T(\sum_j x_j e_j) = \sum_j \lambda_j x_j e_j$. If T is compact then $\lambda \in c_0$.

Are there interesting compact operators which are not finite-rank? For example, is $\lambda \in c_0$ a sufficient condition for T of the previous exercise to be compact? The answer is yes, and it is a consequence of the next result.

Theorem 6.2. *Suppose that X is a normed space, Y is a Banach space and that $T_j \in \mathcal{L}(X, Y)$ is a compact operator. Suppose there is a $T \in \mathcal{L}(X, Y)$ such that $\|T_j - T\| \to 0$ as $j \to \infty$. Then T is compact.*

Continuing the previous exercise, let H be a separable Hilbert space with an orthonormal basis $\{e_j\}$, $\lambda \in \ell^\infty$ and $T_k \in \mathcal{L}(H)$ be defined by $T_k(\sum_{j=1}^\infty x_j e_j) = \sum_{j=1}^k \lambda_j x_j e_j$. Then each T_k, being finite-rank, is compact, and if we let $T \in \mathcal{L}(H)$ be defined by $T(\sum_j x_j e_j) = \sum_j \lambda_j x_j e_j$, we have

$$\|T_k x - Tx\|^2 = \|\sum_{j=k+1}^\infty \lambda_j x_j e_j\|^2 = \sum_{j=k+1}^\infty |\lambda_j x_j|^2 \leq \sup_{j>k} |\lambda_j|^2 \sum_{j=k+1}^\infty |x_j|^2 \leq \sup_{j>k} |\lambda_j|^2 \|x\|^2$$

so that if we assume $\lambda \in c_0$, $\|T_k - T\| \leq \sup_{j\,:\,j>k} |\lambda_j|$ which goes to 0 as $k \to \infty$ since $\lambda_j \to 0$ as $j \to \infty$.

Theorem 6.2 implies in particular that if $T_j \in \mathcal{L}(X, Y)$ is finite-rank and if $\|T_j - T\| \to 0$ as $j \to \infty$, then T is compact. A natural question at this point is therefore whether any compact operator must be the norm limit of a sequence of finite-rank operators. See Exercise 6 below for further discussion.

The next lemma is very useful when we are dealing with compact operators, in particular in the proof of Theorem 6.2.

Lemma 6.3. *Let X and Y be normed spaces and suppose $T \in \mathcal{L}(X,Y)$. Then T is compact if and only if for every bounded sequence $(x_n) \subseteq X$, the sequence $(Tx_n) \subseteq Y$ has a convergent subsequence.*

And for this lemma we need to recall the following fact about metric spaces, proved in one of the workshops in Honours Analysis:

Proposition 6.4. *Let (X,d) be a metric space. Then X is compact if and only if every sequence in X has a convergent subsequence.*

Proof. (of Lemma 6.3). Let us suppose T is compact. Let $(x_n) \subseteq X$ be a bounded sequence. Remove any instances of $x_n = 0$ if necessary. Then the sequence $T(x_n)/\|x_n\|$ is in $T(\{x : \|x\| \leq 1\}) \subseteq \overline{T(\{x : \|x\| \leq 1\})}$ and so by Proposition 6.4 there is a subsequence $T(x_{n_k})/\|x_{n_k}\|$ convergent to some $y \in \overline{T(\{x : \|x\| \leq 1\})}$. Now, $(\|x_{n_k}\|)$ being a bounded sequence in $\mathbb{R}$, there is a further subsequence $(\|x_{n_{k_j}}\|)$ which converges to some $\lambda \in \mathbb{R}$. So $\Big(T(x_{n_{k_j}})/\|x_{n_{k_j}}\|\Big) \times \|x_{n_{k_j}}\| = T(x_{n_{k_j}})$ converges to $y\lambda \in Y$ as $j \to \infty$ as required.

For the converse, suppose now that for every bounded sequence (x_n) in X, (Tx_n) has a convergent subsequence. Let $y_n \in \overline{T(\{x : \|x\| \leq 1\})}$ and suppose that x_n with $\|x_n\| \leq 1$ is such that $\|Tx_n - y_n\| < 1/n$. By hypothesis there is a subsequence (x_{n_k}) such that Tx_{n_k} converges to some $y \in \overline{T(\{x : \|x\| \leq 1\})}$. Since $\|Tx_n - y_n\| \to 0$ as $n \to \infty$, $y_{n_k} \to y$ as $k \to \infty$. Thus y_{n_k} converges to $y \in \overline{T(\{\|x\| \leq 1\})}$. So by Proposition 6.4, T is compact. $\square$

Proof. (of Theorem 6.2) Let (x_n) be a bounded sequence in X. Indeed, without loss of generality, assume that $\|x_n\| \leq 1$ for all n. We wish to show that Tx_n has a convergent subsequence. We do this by a "diagonal argument".

First of all, as T_1 is compact, there is a subsequence of (x_n), which we relabel as $(x_1^1, x_2^1, x_3^1, \dots)$, such that $T_1 x_n^1 \to y^1$ as $n \to \infty$ for some $y^1 \in Y$.

Secondly, since T_2 is compact, there is a subsequence of (x_n^1), which we relabel as $(x_1^2, x_2^2, x_3^2, \dots)$, such that $T_2 x_n^2 \to y^2$ as $n \to \infty$ for some $y^2 \in Y$.

Continuing, since T_j is compact, there is a subsequence of (x_n^{j-1}), which we relabel as $(x_1^j, x_2^j, x_3^j, \dots)$, such that $T_j x_n^j \to y^n$ as $n \to \infty$ for some $y^n \in Y$.

Define $\xi_1 := x_1^1, \xi_2 := x_2^2, \xi_3 := x_3^3$ and $\xi_n := x_n^n$. (Hence the term "diagonal argument".) Then $(\xi_1, \xi_2, \dots)$ is a subsequence of (x_n). The crucial observation is that if $m \geq n$, then ξ_m appears in the sequence $(x_1^n, x_2^n, x_3^n, \dots, x_n^n, \dots)$ in some place after $x_n^n = \xi_n$.

Now we claim that $(T\xi_n)$ converges. We clearly have to use the hypothesis that Y is complete so it makes sense to show that $(T\xi_n)$ is Cauchy, and then we will be done.

Let $\epsilon > 0$. Pick N such that $\|T_N - T\| < \epsilon/3$. Then, whatever the values of m and n,

$$\|T\xi_m - T\xi_n\| \leq \|T\xi_m - T_N\xi_m\| + \|T_N\xi_m - T_N\xi_n\| + \|T_N\xi_n - T\xi_n\|$$

$$\leq \|T - T_N\|\|\xi_m\| + \|T_N\xi_m - T_N\xi_n\| + \|T - T_N\|\|\xi_n\| \leq \|T_N\xi_m - T_N\xi_n\| + 2\epsilon/3.$$

Now, for N fixed, the sequence $(T_N x_n^N)_n$ is convergent and hence Cauchy, so there is a K such that $m, n \geq K$ implies $\|T_N x_m^N - T_N x_n^N\| < \epsilon/3$. So if $m, n \geq \max\{K, N\}$, ξ_m and ξ_n appear in the sequence $(x_j^N)_j$ in some place after x_N^N, and so $\|T_N\xi_m - T_N\xi_n\| < \epsilon/3$.

Combining this with the previous paragraph, we have that for $m, n \geq \max\{K, N\}$, $\|T\xi_m - T\xi_n\| < \epsilon$, and we are done. □

Theorem 6.2, together with the discussion of whether any compact operator is a norm-limit of finite-rank operators, describes the interaction of compactness and limits, that is, the **analytical properties** of the class of compact operators. In view of our development of the theory, it is also natural to ask how the class of compact operators interacts with algebraic properties, and, furthermore, how it acts with respect to duality.

Algebraic properties of the class of compact operators. Suppose that X, Y and Z are normed spaces and suppose $T \in \mathcal{L}(X, Y)$ and $S \in \mathcal{L}(Y, Z)$. It is easy to see that if T is finite-rank then $ST \in \mathcal{L}(X, Z)$ is finite-rank, and that if S is finite-rank then that $ST \in \mathcal{L}(X, Z)$ is finite-rank. It is natural to ask if "finite-rank" here can be replaced by compact. This is true – see Exercise 4 below.

Duality properties of the class of compact operators. One of the deeper results in elementary linear algebra is the following: if A is an $m \times n$ matrix over the field $\mathbb{F}$, then its row rank equals its column rank. This means that the dimension of the space spanned by the rows of A equals the dimension of the space spanned by the columns of A. Applying this to a linear transformation $T : V \to W$ between finite-dimensional vector spaces we deduce that the two spaces $\mathrm{im}(T)$ and $\mathrm{im}(T')$ have the same dimension. (Here and below, T' denotes the transpose of T.) Suppose now that X and Y are normed spaces and that $T \in \mathcal{L}(X, Y)$ is finite-rank. Can we deduce that $T' : Y^* \to X^*$ is finite-rank and that the two spaces $\mathrm{im}(T)$ and $\mathrm{im}(T')$ have the same dimension? Finally, suppose that X and Y are normed spaces and that $T \in \mathcal{L}(X, Y)$. If T is compact can we deduce that T' is compact? If T' is compact can we deduce that T is compact? Both of these are true, and the statement that T is compact if and only if T' is compact is called Schauder's theorem. In Exercise 7 we'll prove this in the case of Hilbert spaces where it's a bit easier.

Algebraic-analytic properties of the class of compact operators. We can also ask what happens when we mix algebraic and analytical properties. For example, if $T \in \mathcal{L}(X, Y)$ is compact and $S_j \in \mathcal{L}(Y, Z)$ has the property that for all $y \in Y$, $S_j y \to Sy$ as $j \to \infty$, can we conclude that $\|S_j T - ST\| \to 0$ as $j \to \infty$? (See Exercise 5 below.) More generally still, if $T_j \in \mathcal{L}(X, Y)$ are compact and $\|T_j - T\| \to 0$, and if S_j are as before, can we conclude that ST is compact [yes, using Theorem

6.2 and the algebraic properties of compactness] and that $\|S_jT_j - ST\| \to 0$? See Exercise 8 below. (What about TS_j and T_jS_j instead of S_jT and S_jT_j (where now of course $S_j \in \mathcal{L}(Z, X)$?))

The proofs of these results are surprisingly intermingled.

6.5. **Hilbert–Schmidt Operators.** If we are working in the context of bounded linear operators between normed spaces X and Y we have the containments

$$\text{finite-rank operators} \subseteq \text{compact operators} \subseteq \text{ bounded operators.}$$

If however we are working in the context of bounded linear operators between Hilbert spaces H and K we have the extra category of **Hilbert–Schmidt** operators which fits between the finite-rank ones and the compact ones:

$$\text{finite-rank} \subseteq \text{Hilbert–Schmidt} \subseteq \text{compact} \subseteq \text{ bounded.}$$

We introduced the class of Hilbert–Schmidt operators above in Section 4.9, Example (B), as the class of bounded linear operators $T \in \mathcal{L}(H, K)$ for which there exists an orthonormal basis[12] $\{e_n\}_{n\geq 1}$ of H such that $\sum_{n=1}^{\infty} \|Te_n\|^2 < \infty$. However we did not show there that the quantity $\sum_{n=1}^{\infty} \|Te_n\|^2$ is independent of the choice of orthonormal basis.

Lemma 6.6. *Suppose $T \in \mathcal{L}(H, K)$ where H and K are separable Hilbert spaces. Suppose $\{e_n\}_{n\geq 1}$ is an orthonormal basis of H and $\{f_m\}_{m\geq 1}$ is an orthononormal basis of K. Then*

$$\sum_{n=1}^{\infty} \|Te_n\|^2 = \sum_{m=1}^{\infty} \|T^* f_m\|^2.$$

Proof. Since $\{f_m\}_{m\geq 1}$ is an orthononormal basis of K we have

$$\|Te_n\|^2 = \sum_m |\langle Te_n, f_m\rangle|^2 = \sum_m |\langle e_n, T^* f_m\rangle|^2$$

by the defining property of the adjoint T^*. Therefore

$$\sum_n \|Te_n\|^2 = \sum_n \sum_m |\langle e_n, T^* f_m\rangle|^2 = \sum_m \sum_n |\langle e_n, T^* f_m\rangle|^2 = \sum_m \|T^* f_m\|^2.$$

□

Corollary 6.7. *Let H and K be separable Hilbert spaces. If $T \in \mathcal{L}(H, K)$ is a Hilbert–Schmidt operator, the quantity $\sum_n \|Te_n\|^2$ is independent of the choice of orthonormal basis $\{e_n\}_{n\geq 1}$ for H. Hence the* **Hilbert–Schmidt norm** *of T, $\|T\|_{HS}$ given by*

$$\|T\|_{HS}^2 = \sum_n \|Te_n\|^2$$

[12] We are assuming H is separable here.

whenever $\{e_n\}_{n\geq 1}$ *is an orthonormal basis for* H*, is well-defined on the class of Hilbert–Schmidt operators. Moreover if* T *is Hilbert–Schmidt,* T^* *is also Hilbert–Schmidt and* $\|T^*\|_{HS} = \|T\|_{HS}$.

Actually, we do not need to know *a priori* that T is bounded before we consider its Hilbert–Schmidt norm. Indeed, a slight strengthening of the argument of Proposition 4.11 shows that if $T : H \to K$ is linear and has finite Hilbert–Schmidt norm, then it is bounded and $\|T\| \leq \|H\|_{HS}$:

Proposition 6.8. *Suppose that* H *and* K *are separable Hilbert spaces and that* $T : H \to K$ *is a linear map. If* $\{e_n\}$ *is an orthonormal basis of* H *such that* $\sum_n \|Te_n\|^2 < \infty$ *then* $T \in \mathcal{L}(H, K)$ *and* $\|T\| \leq \left(\sum_n \|Te_n\|^2\right)^{1/2}$.

Proof. Let $x \in H$ be a *finite* linear combination of the e_j's: so $x = \sum_j \alpha_j e_j$ and $\|x\|^2 = \sum_j |\alpha_j|^2$. Then, by the triangle inequality followed by Cauchy–Schwarz,

$$\|Tx\| = \|\sum_j \alpha_j Te_j\| \leq \sum_j |\alpha_j|\|Te_j\|$$

$$\leq \left(\sum_j |\alpha_j|^2\right)^{1/2} \left(\sum_j \|Te_j\|^2\right)^{1/2} = \|x\|\|T\|_{HS}.$$

So T, when restricted to the dense linear subspace of H consisting of finite linear combination of the e_j's, is bounded, with norm $\|T\| \leq \|T\|_{HS}$. By Proposition 4.5, T is bounded from H to K and its operator norm satisfies the same inequality. □

If $T \in \mathcal{L}(H, K)$ is finite-rank it is easy to see that T must be Hilbert–Schmidt. It is also true that if $T \in \mathcal{L}(H, K)$ is Hilbert–Schmidt, then T is compact, (see Exercise 12 below.) By Proposition 4.13 every Hilbert–Schmidt operator is the norm-limit of a sequence of finite-rank operators. (This also follows from the previous remark and Exercise 6 below.)

We saw previously in Section 4.9, Example (B), that every integral operator $T \in \mathcal{L}(L^2)$ of the form

$$Tf(x) = \int K(x, y)f(y)dy$$

where

$$\int\int |K(x, y)|^2 dxdy < \infty$$

is a Hilbert–Schmidt operator and that $\|T\|^2_{HS} = \int\int |K(x, y)|^2 dxdy$. The calculation leading to this had nothing to do with the specific orthonormal basis we used in the discussion in Section 4.9.

One can show that *every* Hilbert–Schmidt operator has this form if we permit ourselves to suitably interpret "integrals" as being sums. Indeed, given orthonormal bases $\{e_i\}$ for H and $\{f_j\}$ for K, there are, by the discussion following Proposition 3.18, isometric isomporhisms $U_H : H \to \ell^2_H$ and $U_K : K \to \ell^2_K$ so that if $T \in \mathcal{L}(H, K)$ is a Hilbert–Schmidt operator then $\tilde{T} : \ell^2_H \to \ell^2_K$ given by $\tilde{T} = U_K \circ T \circ U_H^{-1}$

is the operator on ℓ^2 given by the matrix $\alpha_{ij} = \langle Te_i, f_j \rangle$. For such an operator its Hilbert–Schmidt norm is simply $\sum_{ij} |\alpha_{ij}|^2$, and so $\|\tilde{T}\|_{HS}^2 = \sum_{ij} |\alpha_{ij}|^2 = \|T\|_{HS}^2$. We describe this by saying that the operators T and $\tilde{T}$ are **unitarily equivalent.**

EXERCISES

1. Let H be a separable Hilbert space with an orthonormal basis $\{e_j\}$, $\lambda \in \ell^\infty$ and $T \in \mathcal{L}(H)$ be defined by $T(\sum_j x_j e_j) = \sum_j \lambda_j x_j e_j$. Show that if T is compact then $\lambda \in c_0$. **Hint:** Use Lemma 6.3.

2*. Let $T \in \mathcal{L}(H, Y)$ be a compact operator between a Hilbert space H and a normed space Y, and let $\{e_n\}$ be an orthonormal sequence in H. Show that $Te_n \to 0$ as $n \to \infty$.

3. Let T be the operator on $\ell^2 \oplus \ell^2$ defined by $T(x, y) = (0, x)$. Show that $T^2 = 0$ but T is not compact.

4. Let X, Y and Z be normed spaces and suppose $T \in \mathcal{L}(X, Y)$ and $S \in \mathcal{L}(Y, Z)$.

a) If T is compact show that $ST \in \mathcal{L}(X, Z)$ is compact.

b) If S is compact show that $ST \in \mathcal{L}(X, Z)$ is compact.

c) Show that if $T \in \mathcal{L}(X, Y)$ is compact, then it cannot have a bounded inverse $S \in \mathcal{L}(Y, X)$.

5. Let X, Y and Z be normed spaces. If $T : X \to Y$ is compact, $S_n, S \in \mathcal{L}(Y, Z)$ satisfy $S_n y \to Sy$ as $n \to \infty$ for all $y \in Y$, and $\{\|S_n\|\}$ is bounded, show that $\|S_n T - ST\| \to 0$ as $n \to \infty$.

6. Show that if $T : X \to H$ is compact where X is a normed space and H is a separable Hilbert space, then T is the norm-limit of a sequence of finite-rank operators. (**Hint:** Use the previous question.)[13]

7. Let H and K be Hilbert sapaces. If $T \in \mathcal{L}(H, K)$ and T^*T is compact, then T is compact. (**Hint:** If T^*Tx_n is convergent, show Tx_n is Cauchy.) Deduce that T is compact if and only if T^* is compact.

8. If $T_n : X \to Y$ is compact, $\|T_n - T\| \to 0$ as $n \to \infty$, $S_n, S \in \mathcal{L}(Y, Z)$ satisfy $S_n y \to Sy$ as $n \to \infty$ for all $y \in Y$, and $\{\|S_n\|\}$ bounded, show that $\|S_n T_n - ST\| \to 0$ as $n \to \infty$.

9. Let S_n be a bounded sequence in $\mathcal{L}(H)$ such that for all $x, y \in H$, $\lim_{n\to\infty} \langle S_n x, y \rangle = 0$. Let $T \in \mathcal{L}(H)$ be compact. Show that $\|TS_nT\| \to 0$ as $n \to \infty$. (**Hint:** Use the ideas from Problem 5.)

[13]The corresponding result when H is replaced by a Banach space is false. This is harder and due to Enflo.

10. Let H and K be separable Hilbert spaces. Show that if $T \in \mathcal{L}(H, K)$ is finite-rank then it is Hilbert–Schmidt.

11. Let H and K be separable Hilbert spaces. Show that if $T \in \mathcal{L}(H, K)$ is Hilbert–Schmidt and if S is bounded then ST and TS are Hilbert–Schmidt.

12. Let H and K be separable Hilbert spaces. Show that if $T \in \mathcal{L}(H, K)$ is Hilbert–Schmidt then T is compact. **Hint:** Use Proposition 4.13 and Theorem 6.2

13. Let H and K be separable Hilbert spaces. Show that if $T \in \mathcal{L}(H, K)$ is Hilbert–Schmidt, then it is unitarily equivalent to a matrix operator between ℓ^2-spaces where the matrix (a_{ij}) satisfies $\sum_{ij} |a_{ij}|^2 = \|T\|_{HS}^2$.

7. The Spectral theorem for compact self-adjoint operators on a Hilbert space

We now come to one of the great pinnacles of the theory of operators on Hilbert spaces, the spectral theorem for compact self-adjoint operators. Recall that $T \in \mathcal{L}(H)$ is **self-adjoint** if $T = T^*$.

Theorem 7.1. *Let H be a separable Hilbert space and suppose $T \in \mathcal{L}(H)$ is a compact self-adjoint operator. Then there exists an orthonormal basis of eigenvectors $\{e_n\}$ for T so that for all $x \in H$*

$$Tx = \sum_n \lambda_n \langle x, e_n \rangle e_n.$$

Moreover the eigenvalues λ_n are real and the sequence (λ_n) belongs to c_0. Finally the eigenspace of each non-zero eigenvalue is finite-dimensional.

Corollary 7.2. *If $T \in \mathcal{L}(H)$ is a compact self-adjoint operator, then $\|T\| = \max\{|\lambda| : \lambda$ is an eigenvalue for $T\}$.*

Recall that it is trivial that $\|T\| \geq \sup\{|\lambda| : \lambda$ is an eigenvalue for $T\}$; the deeper statement is that there is an eigenvalue λ such that $|\lambda| = \|T\|$. In fact we will prove this latter statement separately as an essential ingredient of the proof of the spectral theorem.

Exercise. Let $T \in \mathcal{L}(L^2([0,1]))$ be defined initially on $C([0,1])$ by $Tf(s) = sf(s)$. Then T is self-adjoint but has no eigenvectors.

Some comments are in order. Firstly, the converse statement – that if $\{e_n\}$ is an orthonormal basis for H, (λ_n) belongs to c_0, and T is given by $Tx = \sum_n \lambda_n x_n e_n$ for $x = \sum_n x_n e_n$, then T is compact – is also true and we saw this in the previous section as a consequence of Theorem 6.2. (There is no requirement for the eigenvalues to be real-valued here, but if they are, then T will be self-adjoint.) So the spectral theorem gives a complete structural description of the class of compact self-adjoint operators on a complex separable Hilbert space.

Secondly, we have also seen in the previous section that if $\{e_n\}$ is an orthonormal basis for H, if (λ_n) belongs to ℓ^∞, and if T given by $Tx = \sum_n \lambda_n \langle x, e_n \rangle e_n$ is compact, then $(\lambda_n) \in c_0$.

Finally, the statement that the eigenspace of each non-zero eigenvalue is finite-dimensional is a consequence of the fact that $(\lambda_n) \in c_0$: if there were a non-zero eigenvalue λ with an infinite-dimensional eigenspace, then λ would have to feature infinitely many times in the sequence (λ_n), preventing it from being in c_0.

So the main thing we have to prove is that if $T \in \mathcal{L}(H)$ a compact self-adjoint operator, then there exists an orthonormal basis of eigenvectors $\{e_n\}$ for T with *real* eigenvalues λ_n, and that we have the representation

$$Tx = \sum_n \lambda_n \langle x, e_n \rangle e_n.$$

We begin with two easy lemmas. The first says that eigenvalues of self-adjoint operators must be real.

Lemma 7.3. *Suppose $T \in \mathcal{L}(H)$ is a self-adjoint operator and that for some nonzero $x \in H$, $Tx = \lambda x$. Then $\lambda \in \mathbb{R}$.*

Proof. We have

$$\lambda\langle x, x\rangle = \langle \lambda x, x\rangle = \langle Tx, x\rangle = \langle x, T^* x\rangle = \langle x, Tx\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\langle x, x\rangle.$$

Since $x \neq 0$, this means that $\lambda \in \mathbb{R}$. □

The second says that eigenvectors corresponding to distinct eigenvalues are orthogonal.

Lemma 7.4. *Suppose $T \in \mathcal{L}(H)$ is a self-adjoint operator and that for some nonzero $x \in H$, $Tx = \lambda x$ and for some nonzero $y \in H$, $Ty = \mu y$ where $\lambda \neq \mu$. Then*

$$\langle x, y\rangle = 0.$$

Proof. We have, noting that λ and μ are both real,

$$\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle Tx, y\rangle = \langle x, T^* y\rangle = \langle x, Ty\rangle = \langle x, \mu y\rangle = \mu\langle x, y\rangle.$$

Since $\lambda \neq \mu$ this means that $\langle x, y\rangle = 0$. □

So ensuring orthonormality of the eigenvectors in Theorem 7.1 is not going to be the hardest task. Perhaps the deepest aspect of Theorem 7.1 is the *existence* of eigenvalues and eigenvectors. Recall that in the finite-dimensional case this is guaranteed because the characteristic polynomial $\det(T - \lambda I)$ must have a root $\lambda \in \mathbb{C}$ by the fundamental theorem of algebra. In the infinite-dimensional case we need a different argument.

Proposition 7.5. *Suppose $T \in \mathcal{L}(H)$ is a compact self-adjoint operator. Then either $\|T\|$ or $-\|T\|$ is an eigenvalue for T.*

This will be a consequence of the following lemma:

Lemma 7.6. *Suppose $T \in \mathcal{L}(H)$ is a self-adjoint operator. Then*

$$\|T\| = \sup_{\|x\| \leq 1} |\langle Tx, x\rangle|.$$

Proof. We certainly have that $|\langle Tx, x\rangle| \leq \|T\|\|x\|^2$, so $\sup_{\|x\|\leq 1} |\langle Tx, x\rangle| \leq \|T\|$. So we need to see the reverse inequality $\sup_{\|x\|\leq 1} |\langle Tx, x\rangle| \geq \|T\|$. To do this we use polarisation and the parallelogram law.

Let $K := \sup_{\|x\|\leq 1} |\langle Tx, x\rangle|$ so that $|\langle Tx, x\rangle| \leq K\|x\|^2$ for all $x \in H$.

Consider

$$\langle T(x \pm y), x \pm y\rangle = \langle Tx, x\rangle \pm 2\mathrm{Re}\langle Tx, y\rangle + \langle Ty, y\rangle$$

where we have used the fact that $T = T^*$.

So

$$\begin{aligned}4\mathrm{Re}\langle Tx, y\rangle &= \langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle \\ &\leq K\left(\|x+y\|^2 + \|x-y\|^2\right) = 2K\left(\|x\|^2 + \|y\|^2\right)\end{aligned}$$

using the parallelogram law. Now by multiplying x by a unimodular complex number chosen so that $\lambda\langle Tx, y\rangle$ is real and nonnegative we get

$$|\langle Tx, y\rangle| \leq \frac{K}{2}\left(\|x\|^2 + \|y\|^2\right)$$

for all x and y in H. Therefore

$$\sup_{\|x\|\leq 1, \|y\|\leq 1} |\langle Tx, y\rangle| \leq K.$$

If $Tx \neq 0$ take $y = Tx/\|Tx\|$ so obtain

$$\sup_{\|x\|\leq 1} \|Tx\| \leq K$$

which is the inequality we desired. □

Remark. If $T \in \mathcal{L}(H)$ is self-adjoint, then $\langle Tx, x\rangle$ is real for all $x \in H$. Indeed, $\langle Tx, x\rangle = \langle x, Tx\rangle = \overline{\langle Tx, x\rangle}$.

Proof. (of Proposition 7.5) We may assume that $\|T\| > 0$. By Lemma 7.6 there exists a sequence (x_n) satisfying $\|x_n\| \leq 1$ such that $|\langle Tx_n, x_n\rangle| \to \|T\|$ as $n \to \infty$. Since $\langle Tx_n, x_n\rangle$ is real (see the remark above) we can assume that $\langle Tx_n, x_n\rangle$ tends to either $+\|T\|$ or $-\|T\|$ as $n \to \infty$. We will assume that $\langle Tx_n, x_n\rangle \to +\|T\|$, the other case being similar.

Compactness of T tells us there is some $y \in H$ and a subsequence of (x_n), say x_{n_k}, such that $Tx_{n_k} \to y$ as $k \to \infty$. We claim that y is an eigenvector for T and that $Ty = \|T\|y$. Let us relabel the subsequence as (x_n). It is natural to examine $Tx_n - \|T\|x_n$, since if we apply T to this vector we get $T(Tx_n) - \|T\|(Tx_n)$ which tends to $Ty - \|T\|y$ since $Tx_n \to y$ and T is continuous

Indeed,

$$\|Tx_n - \|T\|x_n\|^2 = \|Tx_n\|^2 - 2\|T\|\langle Tx_n, x_n\rangle + \|T\|^2\|x_n\|^2 \leq 2\|T\|^2 - 2\|T\|\langle Tx_n, x_n\rangle$$

and the last term here tends to zero as $n \to \infty$. So $Tx_n - \|T\|x_n \to 0$ as $n \to \infty$. Continuity of T gives us that $T(Tx_n - \|T\|x_n) \to 0$ as $n \to \infty$. But we have seen above that the same sequence also tends to $Ty - \|T\|y$. Hence $Ty = \|T\|y$ as desired.

Finally, in order to see that $\|T\|$ is an eigenvalue it is necessary to check that $y \neq 0$. By the discussion above, $x_n \to y/\|T\|$ and if y were to equal zero we would have $x_n \to 0$. But then we would have $\langle Tx_n, x_n\rangle \to 0$, in contradiction to $\langle Tx_n, x_n\rangle \to \|T\| \neq 0$. □

Now we have everything we need in place to prove the spectral theorem.

Proof. (of Theorem 7.1.) Suppose $T := T_0 \in \mathcal{L}(H)$ is compact and self-adjoint. By Proposition 7.5, $\lambda_1 := \pm\|T\|$ is an eigenvalue for T. Let E_1 be the corresponding (finite-dimensional) eigenspace. Use the Gram–Schmidt process to create a (finite) orthonormal basis $\mathcal{F}_1$ for E_1. Let $P_1 : H \to H$ be orthogonal projection onto E_1. Then P_1 is of finite rank, is self-adjoint and so $T_1 := T - \lambda_1 P_1 \in \mathcal{L}(H)$ is a compact self-adjoint operator on H with norm $\|T|_{E_1^\perp}\|$ which is at most $|\lambda_1|$ and is strictly less than $|\lambda_1|$ unless $-\lambda_1$ is also an eigenvalue of T.

Repeating the argument for T_1 yields an eigenvalue λ_2 with $|\lambda_2| = \|T_1\| \le |\lambda_1|$ and a corresponding finite-dimensional eigenspace E_2 for T_1 and for T, which is perpendicular to E_1 by Lemma 7.4, together with a finite orthonormal basis $\mathcal{F}_2$ for E_2 such that, with P_2 denoting orthogonal projection onto E_2, $T_2 := T_1 - \lambda_2 P_2$ is a compact self-adjoint operator on H with $\|T_2\| = \|T_1|_{E_2^\perp}\| \le |\lambda_2| \le |\lambda_1|$ and $\|T_2\| < |\lambda_1|$.

Repeating the argument for T_2 yields an eigenvalue λ_3 with $|\lambda_3| = \|T_2\| \le |\lambda_2|$ and a corresponding finite-dimensional eigenspace E_3 for T_2 and for T, which is perpendicular to E_1 and E_2 by Lemma 7.4, together with a finite orthonormal basis $\mathcal{F}_3$ for E_3 such that, with P_3 denoting orthogonal projection onto E_3, $T_3 := T_2 - \lambda_3 P_3$ is a compact self-adjoint operator on H with $\|T_3\| = \|T_2|_{E_3^\perp}\| \le |\lambda_3| \le |\lambda_2|$ and $\|T_3\| < |\lambda_2|$.

At the n'th stage, we have inductively defined compact self adjoint operators $T_1, \dots, T_{n-1}$, eigenvalues $\lambda_1, \dots, \lambda_{n-1}$ for T with $|\lambda_j| = \|T_{j-1}\|$, $|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_{n-1}|$ (with no two consecutive equalities), mutually perpendicular finite-dimensional eigenspaces $E_1, \dots, E_{n-1}$ with corresponding finite orthonormal bases $\mathcal{F}_j$, projection operators $P_1, \dots, P_{n-1}$ onto $E_1, \dots, E_{n-1}$ respectively. Since T_{n-1} is a compact self-adjoint operator there is an eigenvalue λ_n with $|\lambda_n| = \|T_{n-1}\| \le |\lambda_{n-1}|$ and a corresponding finite-dimensional eigenspace E_n for T_{n-1} and for T, which is perpendicular to $E_1, \dots, E_{n-1}$ by Lemma 7.4, together with a finite orthonormal basis $\mathcal{F}_n$ for E_n such that, with P_n denoting orthogonal projection onto E_n, $T_n := T_{n-1} - \lambda_n P_n$ is a compact self-adjoint operator on H with $\|T_n\| = \|T_{n-1}|_{E_n^\perp}\| \le |\lambda_n| \le |\lambda_{n-1}|$ and $\|T_n\| < |\lambda_{n-1}|$.

This process will either halt once we reach an n such that $T_n = 0$, or else will continue *ad infinitum.* The first, finite-rank, case is easier, so we assume that we are in the second, non-terminating case. By our earlier observations the sequence $(|\lambda_n|)$ decreases to 0 as $n \to \infty$ (otherwise T would not have been compact).

We claim that for all $x \in H$ we have

$$Tx = \sum_{j=1}^{\infty} \lambda_j P_j x.$$

By construction, for each n we have $T = \sum_{j=1}^{n} \lambda_j P_j + T_n$ for all n, and $\|T_n\| = |\lambda_{n+1}| \to 0$ as $n \to \infty$. So $\|T - \sum_{j=1}^{n} \lambda_j P_j\| \to 0$ as $n \to \infty$, and this immediately implies $Tx = \sum_{j=1}^{\infty} \lambda_j P_j x$ for all $x \in H$.

Let L denote the linear span of $\bigcup_{j=1}^{\infty} E_j$. We claim that $\ker T = L^\perp$. First suppose that $Tx = 0$. Then, by our formula, $\sum_{j=1}^{\infty} \lambda_j P_j x = 0$ which means that $P_j x = 0$ for all j and hence x is perpendicular to each E_j and therefore belongs to $L^\perp$. Conversely, suppose $x \in L^\perp$. Then $P_j x = 0$ for all j and so by our formula above, $Tx = 0$ and so $x \in \ker T$.

Finally, let $\mathcal{F}_0$ be an orthonormal basis for $\ker T = L^\perp$. Then $\bigcup_{j=0}^{\infty} \mathcal{F}_j$ is an orthonormal basis for H consisting of eigenvectors of T since for $j \geq 1$, $e \in \mathcal{F}_j$ implies that $Te = \lambda_j e$, and if $e \in L^\perp \setminus \{0\}$, then $Te = 0$ and so e has eigenvalue 0.

Writing things out explicitly in terms of the orthonormal basis $\bigcup_{j=0}^{\infty} \mathcal{F}_j$ gives the formula in the statement of the theorem. $\square$

The spectral theorem for compact self-adjoint operators generalises to the class of compact **normal** operators, i.e. those for which $TT^* = T^*T$.

Theorem 7.7. *Let H be a separable (complex) Hilbert space and suppose $T \in \mathcal{L}(H)$ is a compact normal operator. Then there exists an orthonormal basis of eigenvectors $\{e_n\}$ for T so that for all $x \in H$*

$$Tx = \sum_n \lambda_n \langle x, e_n \rangle e_n.$$

Moreover the sequence (λ_n) belongs to c_0. Finally the eigenspace of each non-zero eigenvalue is finite-dimensional.

We discuss two posible approaches to this result in the exercises.

EXERCISES

1. Let $T \in \mathcal{L}(L^2([0,1]))$ be defined initially on $C([0,1])$ by $Tf(s) = sf(s)$. Show that T is self-adjoint but has no eigenvectors. Show directly that T is not compact.

2. Give an alternate proof of the end of the spectral theorem by arguing as follows. Let S be the closure of the linear span of all of the eigenvectors of T. We know that S is nontrivial by Proposition 7.5. Note that $\ker T \subseteq S$. We wish to show that $S = H$. If not, then $S^\perp$ is nontrivial. So if we can show that $S^\perp$ must contain an eigenvector, we have a contradiction. Note that $T(S) \subseteq S$ and $T(S^\perp) \subseteq S^\perp$ – the latter because if $x \in S^\perp$ and $y \in S$ we have $\langle Tx, y \rangle = \langle x, Ty \rangle = 0$. Let $\tilde{T} = T|_{S^\perp}$. Then $\tilde{T}$ is a compact self-adjoint operator on the Hilbert space $S^\perp$, and $\tilde{T} \neq 0$ (since otherwise $S^\perp \subseteq \ker T$.) Hence by Proposition 7.5 it has an eigenvector which is clearly also an eigenvector for T. Contradiction.

3. Suppose H is a separable Hilbert space and for some orthonormal basis $\{e_n\}$ of H we define T by $Tx = \sum_n \lambda_n \langle x, e_n \rangle e_n$ where $(\lambda_n) \in \ell^\infty$. Show that

(a) T is compact if and only if $(\lambda_n) \in c_0$,

(b) T is self-adjoint if and only if $\lambda_n \in \mathbb{R}$ for all n

(c) every operator of this form (i.e. with $(\lambda_n) \in \ell^\infty$ and not necessarily real) is **normal**, that is, it satisfies $T^*T = TT^*$.

4. Recall the spectral theorem for compact normal operators, Theorem 7.7 above.

(a) Discuss the need for the word "complex" in the statement of the theorem.

(b) Which parts of the spectral theorem for normal operators are straightforward and why?

(c) Is it true that eigenvectors of a normal T corresponding to distinct eigenvectors are orthogonal?

(d) Once we know that every compact normal operator T must have an eigenvalue λ with $|\lambda| = \|T\|$, can we conclude the proof as in the self-adjoint case?

(e)* Show, using the same ideas as in Propostion 7.5, that if T is a compact normal operator then it must have an eigenvalue λ with $|\lambda| = \|T\|$. (Do we need complexity of H here?)

5. Here we outline another approach to the spectral theorem for compact normal operators.

(a) Show that if $T \in \mathcal{L}(H)$ is any normal operator, then there exist self-adjoint operators $T_1, T_2 \in \mathcal{L}(H)$ such that $T = T_1 + iT_2$ and such that T_1 and T_2 commute. (**Hint:** Use the formula $z = (z + \overline{z})/2 + i(-i(z - \overline{z})/2)$ for complex numbers as inspiration.)

(b) Show that if T_1 and T_2 are *commuting* compact self-adjoint operators on H then there exists an orthonormal basis of H consisting of *common* eigenvectors for T_1 and T_2.

(c) Deduce that if T is compact and normal then H has an orthonormal basis of eigenvectors of T.

6. Suppose $T \in \mathcal{L}(H)$ is a compact self-adjoint operator which is also *positive-definite* in the sense that $\langle Tx, x\rangle \geq 0$ for all $x \in X$. Show that there is a compact self-adjoint operator $S \in \mathcal{L}(H)$ such that $S^2 = T$. Show that we can take S to be positive-definite (*and that there is a *unique* positive-definite compact self-adjoint $S \in \mathcal{L}(H)$ such that $S^2 = T$).

7. In this problem we work on the real Hilbert space $H = L^2([0, 1])$ and consider the operator $T \in \mathcal{L}(H)$ defined initially on $C([0, 1])$ by

$$Tf(s) = \int_0^1 K(s,t) f(t) dt$$

where

$$K(s,t) = (1-s)t \text{ for } 0 \leq t \leq s \leq 1 \text{ and } K(s,t) = (1-t)s \text{ for } 0 \leq s \leq t \leq 1.$$

(a) Show that T is a Hilbert–Schmidt operator on H and thus is compact.

(b) Show that T is self-adjoint. (You may change orders of integration without justification when the integrands are continuous.)

(c) Deduce using the spectral theorem that there is a countable orthonormal basis of eigenfunctions $\{\phi_n\}_{n\in\mathbb{N}}$ with $\phi_n \in L^2$, and real eigenvalues λ_n, with $\lambda_n \to 0$ as $n \to \infty$ such that

$$T\phi_n = \lambda_n \phi_n$$

for all $n \in \mathbb{N}$ and hence that

$$Tf = \sum_{n\in\mathbb{N}} \lambda_n \langle f, \phi_n\rangle \phi_n$$

for all $f \in L^2$.

(d) Assume that $\lambda \in \mathbb{R}$ is an eigenvalue of T and that $\phi \in L^2$ is an eigenfunction so that $T\phi = \lambda\phi$. Find formulas for ϕ' and ϕ'' by differentiating under the integral sign – which you may assume is valid – and deduce that

$$\lambda\phi'' + \phi = 0 \text{ and } \phi(0) = \phi(1) = 0.$$

(e) Show that if ϕ is a C^2 function on $[0,1]$ and $\lambda \in \mathbb{R}\setminus\{0\}$ is such that $\lambda\phi'' + \phi = 0$ and $\phi(0) = \phi(1) = 0$, then $\lambda = 1/(n\pi)^2$ for some $n \in \mathbb{N}$. Hence the eigenvalues of T are $1/(n\pi)^2$ for $n \in \mathbb{N}$. Show that the corresponding eigenspaces are one-dimensional and find the eigenfunctions.

Note: So we have "rederived" the completeness of the orthonormal system $\{\phi_n\}_{n\in\mathbb{N}}$ $= \{\frac{1}{2}\sin(n\pi\cdot)\}$ for $L^2([0,1])$ – a cornerstone of the theory of Fourier series – in the context of the spectral theory of the differential operator $f \mapsto f''$. This is the context in which Fourier series originated historically. Note that we obtained the smoothness of the eigenfuntions essentially for free. The arguments presented here can be applied to more general second-order ordinary differential operators of the form

$$f \mapsto Lf(s) = a(s)f''(s) + b(s)f'(s) + c(s)f(s)$$

(with certain natural boundary conditions at $s = 0$ and $s = 1$) where a, b and c are continuous functions on $[0,1]$. These appear in mathematical physics, both classical and quantum, and go by the name of Sturm–Liouville systems. In such cases we cannot necessarily solve the corresponding ODEs explicitly, but nevertheless the spectral theorem will tell us that we do have an orthonormal basis of smooth eigenfunctions ϕ_n for L, and it gives us information on the spectrum of such operators – i.e. the collection of eigenvalues – which is of great physical significance.

Maxwell Institute for Mathematical Sciences, The University of Edinburgh, JCMB, King's Buildings, Edinburgh EH9 3FD, United Kingdom

E-mail address: J.R.Wright@ed.ac.uk