Functional Analysis I

Michaelmas Term 2025

Lecturer: Stuart White

Version of January 3, 2026

0 Introduction

This is a work in progress for lectures notes to go alongside my lectures on B4.1: *Functional Analysis I*. I am updating and modifying previous versions by Melanie Rupflin, Luc Nguyen, which in turn are based on Hilary Priestley's notes for the course. I am greatful to them all for making these notes available to me. Naturally I am responsible for any errors.

Not having the strict time-limit imposed on a lecture course, the notes tend to go into various (interesting!) digressions and cover additional material which is meant to provide the reader with a "larger and clearer picture". Some parts of the material which are additional and are not covered in the lectures are clearly labeled (as *deep dives*). However, this is not always possible so to know the examinable material you should attend the lectures. I should stress the examinable material is summarised in the syllabus and covered in the lectures – nothing less or more is examinable.

Thanks too to Jan, for the style file for the deep dives, and to Austin for the ducks!

These notes are work in progress and are being constantly improved. I am very grateful to all who have helped, or will help me to improve them.

At the moment there is a discussion forum on the course page for comments / corrections to the notes. Once the term is over, I'd appreciate further comments and corrections by email to `stuart.white@maths.ox.ac.uk`.

0.1 Overview / Background

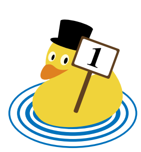
Will follow - it's always best to write these last!

0.2 Notation

We will write $\mathbb{F}$ for either $\mathbb{R}$ or $\mathbb{C}$ when it does not matter which of these is the underlying field of our vector space. This course it will rarely matter, but when we focus on operators, particularly from B4.2 onwards it will often be advantageous to work over $\mathbb{C}$, as this is algebraically closed. You'll be familiar with the fact that complex square matrices always have eigenvalues, as the characteristic equation must have a solution over $\mathbb{C}$, but real square matrices need not have any real eigenvalues. This phenomena persists into the infinite dimensional setting: the spectrum of a bounded linear operator on a complex Banach space is always non-empty, and for this reason I prefer to work over the complex field when studying operators. When one is studying Banach spaces in their own right, typically one takes $\mathbb{F}=\mathbb{R}$ (to avoid sometimes needing to do arguments involving taking real parts, see for example the Hahn–Banach theorem), but this is far less important than the advantages we get from taking $\mathbb{F}=\mathbb{C}$ when we study operators between Banach spaces.

Deep Dive

Anything marked as a *Deep Dive* covers material outside of the syllabus. It is only intended for those who are interested and eager to understand things in more depth. It is non-examinable and not necessary for the course. It goes above and beyond the material, often indicating links with other courses and parts of mathematics.



Even the eager readers should skip those parts on the first reading. More deep dives may appear as I revise the notes. The depth of deep dives may vary considerably from one dive to another.

0.3 Course Synopsis

- Brief recall of material from Part A Metric Spaces and Part A Linear Algebra on real and complex normed vector spaces, their geometry and topology and simple examples of completeness.
- The norm associated with an inner product and its properties.
- Banach spaces, exemplified by ℓ^p, L^p and $C(K)$, and spaces of differentiable functions.
- Finite-dimensional normed spaces, including equivalence of norms and completeness.
- Hilbert spaces as a class of Banach spaces having special properties (illustrations, but no proofs); examples (Euclidean spaces, ℓ^2 and L^2), projection theorem, Riesz Representation Theorem.
- Density. Approximation of functions, Stone-Weierstrass Theorem. Separable spaces; separability of subspaces.
- Bounded linear operators, examples (including integral operators).
- Continuous linear functionals. Dual spaces.
- Statement of the Hahn-Banach Theorem; applications, including density of subspaces and embedding of a normed space into its second dual.
- Adjoint operators.

Contents

1 Normed spaces and Banach spaces

In this section we introduce Banach spaces as complete normed spaces. Our main objective is to give a range of examples. The important special case of Hilbert spaces when the norm is induced from an *inner product* is the subject of Section 2.

1.1 Definitions and review from metric spaces

The key tool linking used to perform analysis on vector spaces is a *norm*, a suitable notion of distance compatible with the linear algebra structure.

Definition 1.1. Let X be a vector space (over either $\mathbb{F}=\mathbb{R}$ or $\mathbb{F}=\mathbb{C}$). A *norm* $\|\cdot\| : X \to \mathbb{R}$ is a function with the following properties:

(N1) for all $x \in X$, we have $\|x\| \geq 0$ with $\|x\| = 0 \Leftrightarrow x = 0$;

(N2) for all $x \in X$ and $\lambda \in \mathbb{F}$, we have $\|\lambda x\| = |\lambda|\|x\|$;

(N3) for all $x, y \in X$, the *triangle inequality* $\|x+y\| \leq \|x\| + \|y\|$ holds.

We call a pair $(X, \|\cdot\|)$ a *normed space*. We will often suppress explicit mention of the norm and say 'Let X be a normed space.'

Every norm $\|\cdot\|$ induces a metric

$$d : X \times X \to \mathbb{R}$$

via $d(x,y) := \|x-y\|$ and so all standard notions and properties of a metric space encountered in part A are applicable. Recall:

- Definition of convergence of a sequence (x_n) in X to $x \in X$:

$$x_n \to x \quad \Longleftrightarrow \quad d(x_n, x) \to 0 \quad \Longleftrightarrow \quad \|x_n - x\| \to 0.$$

- A sequence (x_n) in X is called *Cauchy* if:

$$\forall \varepsilon > 0,\ \exists N \text{ such that } \forall n, m \geq N,\ \|x_n - x_m\| < \varepsilon.$$

  This is written as $\|x_n - x_m\| \to 0$ as $m, n \to \infty$. Recall that a metric space X is called *complete* if and only if every Cauchy sequence in X converges to a point in X. For proving completeness it's worth reminding yourself of the following lemma. The proof, left as an exercise, is essentially the same as the last part of the proof of the that Cauchy sequences in $\mathbb{R}$ converge (using the Bolzano–Weierstrass theorem).

Lemma 1.2. *Let (x_n) be a Cauchy sequence in a normed space X. Then the following are equivalent:*

(i) (x_n) converges,

(ii) (x_n) has a convergent subsequence.

- A function $f : X \to Y$ is *continuous at $x \in X$* if

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ such that } \forall x_0 \in X, \|x - x_0\| < \delta \implies \|f(x) - f(x_0)\| < \varepsilon.$$

The function f is *continuous* if it is continuous at x for all $x \in X$. These definitions, and in fact all properties of metric spaces,[1] have sequential characterisations: continuous functions are those that preserve limits of sequences. Precisely f is continuous at $x \in X$ if and only if for all sequences (x_n) in X with $x_n \to x$, we have $f(x_n) \to f(x)$. for every $x \in X$ and every sequence (x_n) in X

$$x_n \to x \implies f(x_n) \to f(x).$$

Setting this out in the context of normed spaces, $f : X \to Y$ is continuous if and only if whenever (x_n) is a sequence in X with $\|x_n - x\|_X \to 0$, we have $\|f(x_n) - f(x)\|_Y \to 0$.

- A set $U \subset X$ is open if for every $x_0 \in U$ there exists $r > 0$ such that

$$B_r(x_0) := \{x \in X : \|x - x_0\| < r\} \subset U.$$

  Here $B_r(x_0)$ is the open ball centred at x_0 with radius r.

- By definition, a set $F \subset X$ is closed if its complement $X \setminus F$ is open. This can be characterised in terms of sequences: F is closed if and only if whenever (x_n) is a sequence of elements in F which converges to $x \in X$, then $x \in F$, i.e. F contains the limits of all convergent sequences.

The definition of the norm ensures that all the algebraic operations are automatically continuous:

- The scalar multiplication $\mathbb{F} \times X \to X$, $(\lambda, x) \mapsto \lambda x$ is continuous, where $\mathbb{F} \times X$ is given the product metric.
- The addition $X \times X \to X$, $(x, y) \mapsto x + y$ is continuous, where $X \times X$ is given the product metric (below we will give $X \times X$ a norm, but here we only need a metric).
- The norm $\|\cdot\| : X \to \mathbb{R}$ is continuous.

We also recall the appropriate notions of equivalence. First, the equivalence of different norms on the same space:

Definition 1.3. Let X be a vector space. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on X are *equivalent* if and only if there exist a constant $C > 0$ so that for all $x \in X$

$$C^{-1}\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1.$$

[1] The topology on a metric space is fully determined by knowledge of which sequences converge to which points. It is for this reason that any property which can be described using the metric can in some way be characterised by sequences. Therefore when I work with normed spaces, or more generally metric spaces, I have a tendancy towards giving sequence based arguments.

Deep Dive

Altthough this course, and Functional Analysis II do not rely on the Part A topology course, I can't resist pointing out that while one can't 'do everything with sequences' in a general topological space, there is a suitable generalisation of a sequence – known as a net – and one can do everything with nets. For example a function $f : X \to Y$ between topological spaces is continuous if and only if whenever $x_i \to x$ is a convergent net in X, then $f(x_i) \to f(x)$ is a convergent net in Y. I've found that the sort of functional analysis arguments I need to do in general topological spaces, work well with nets — sometimes all one needs to do is replace all the n's indexing the sequence with an i indexing a net! Of course you need to know about nets for this; the classic place which I learnt this from is Kelley's 'General topology', and another source is Willard's 'General Topology' (the section on nets is nice and short, but you'll need to dig around elsewher to learn that, for example, a subset is compact if and only if every net has a convergent subnet).

The normed spaces $(X, \|\cdot\|_1)$ and $(X, \|\cdot\|_2)$ are equivalent precisely when the metrics are strongly equivalent in the sense of the metric spaces course.[2] This ensures that all metric notions (open sets, convergent sequences, Cauchy sequences etc) are all the same in both norms.

Secondly, the notion of isomorphism between normed spaces. While there is only one possible notion of an isomorphism between vector spaces –

Definition 1.4. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. Say X and Y are *isometrically isomorphic* if there exists a surjective linear map $T\colon X \to Y$ with $\|T(x)\|_Y = \|x\|_X$ for all $x \in X$.[3] We call T an *isometric isomorphism*. Say X and Y are *isomorphic* if there exists an isomorphism $T\colon X \to Y$ of vector spaces which is also a homeomorphism of the underlying metric spaces, i.e T and T^{-1} are continuous.

We will have much more to say about the linear operators T appearing here in Section **??** and beyond. For now it is useful to know that $(X, \|\cdot\|_1)$ and $(X, \|\cdot\|_2)$ can be (isometrically) isomorphic, without being equivalent (Problem sheet 1 asks you to produce an example).

Deep Dive

You've met a number of theorems in your analysis courses so far that give conditions under which a continuous bijection is a homeomorphism, i.e. has continuous inverse. For example in prelims, a continuous bijection between two intervals has continuous inverse, or a continuous bijection from a compact metric (or topological) space into a metric space (or Hausdorff topological space) has continuous inverse. In functional analysis we have the following consequence of Banach's open mapping theorem, proved in B4.2 using the baire category theorem: if $T\colon X \to Y$ is a bijective continuous linear map between *Banach* spaces, then T^{-1} is continuous, and hence X and Y are isomorphic Banach spaces. This is sometimes called the Banach isomorphism theorem, and decreases the work needed to obtain an isomorphism between Banach spaces.

One of the key objects we study in this course are Banach spaces and linear maps between such spaces.

Definition 1.5. A normed space $(X, \|\cdot\|)$ is a Banach space if it is complete, i.e. if every Cauchy sequence in X converges.

A normed space is complete if and only if absolute convergence of series implies convergence of series:

Proposition 1.6. *Let X be a normed space. Then the following are equivalent*

(i) X is a Banach space,

(ii) Absolute convergence of series implies convergence, i.e. for sequences (x_n) in X and the corresponding partial sums $s_n := \sum_{k=1}^n x_k$ we have

$$\sum_{i=1}^{\infty} \|x_n\| < \infty \quad \Rightarrow \quad s_n \text{ converges to some } s \in X.$$

Proof. (i) $\Longrightarrow$ (ii): Suppose X is complete and $\sum_{n=1}^{\infty} \|x_n\| < \infty$. Then, for $s_n = \sum_{k=1}^n x_n$, the sequence (s_n) is Cauchy as for $m > n \geq N$

$$\|s_n - s_m\| = \|\sum_{k=n+1}^{m} x_k\| \leq \sum_{k=n+1}^{m} \|x_k\| \leq \sum_{k=N+1}^{\infty} \|x_k\| \to 0 \text{ as } N \to \infty.$$

[2] You might be suspicious of the terminolgy: normed spaces are called equivalent when their underlying metric structures are strongly equivalent. But while in the generality of metric spaces, equivalence of two metrics d_1 and d_2 on X – defined in Part A metric spaces to mean that the identity map Id_X is a homeomorphism – is not the same notion as strong equivalence, for normed spaces it is. This will all follow from equivalence between continuity and boundedness of linear maps in Section **??**.

[3] Such a map T is necessarily injective, so T is an isomorphism of vector spaces which is also an isometry.

As X is complete we thus obtain that s_n converges to some element $s \in X$.

(ii) $\Longrightarrow$ (i): Suppose (ii) holds and let (x_n) be a Cauchy sequence. Select a subsequence x_{n_j} so that

$$\|x_{n_j} - x_{n_{j+1}}\| \leq 2^{-j},$$

where the existence of such a subsequence is ensured by the fact that (x_n) is Cauchy. Then $\sum_{j=1}^{\infty} \|x_{n_{j+1}} - x_{n_j}\| \leq 1 < \infty$ so (ii) ensures that $\sum_{j=1}^{\infty}(x_{n_{j+1}} - x_{n_j})$ converges. Hence $x_{n_k} = x_{n_1} + \sum_{j=1}^{k-1}(x_{n_{j+1}} - x_{n_j})$ converges, so (x_n) has a convergent subsequence and must thus, by Lemma 1.2, itself converge. □

1.2 Examples

$(\mathbb{R}^n, \|\cdot\|_p)$ and $(\mathbb{C}^n, \|\cdot\|_p)$, $1 \leq p \leq \infty$. Consider $\mathbb{R}^n$, or $\mathbb{C}^n$, equipped with

$$\|x\|_p := \Big(\sum_i |x_i|^p\Big)^{1/p} \text{ for } 1 \leq p < \infty$$

respectively

$$\|x\|_\infty := \sup_{i \in \{1,\dots,n\}} |x_i|.$$

One can show that these are all norms, with the challenging bit being the proof of the triangle inequality

$$\|x+y\|_p = \Big(\sum_i |x_i + y_i|^p\Big)^{1/p} \leq \|x\|_p + \|y\|_p,$$

which is is the finite dimensional version of Minkowski's inequality.

Warning. This inequality does not hold if we were to extend the definition of $\|\cdot\|_p$ to $0 < p < 1$, and hence the above expression does not give a norm on $\mathbb{R}^n$ if $p < 1$.

We will write ℓ^p_n for the normed space $(\mathbb{R}^n, \|\cdot\|_p)$. A useful property to deal with the p norms $1 \leq p \leq \infty$ (and their generalisations to sequence and functions spaces) is Hölder's inequality (which we proved in much more generality in the Integration course – see Proposition 1.12).

Lemma 1.7 (Hölder's inequality in $\mathbb{R}^n$)**.** *For $1 \leq p, q \leq \infty$ with*

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{1}$$

we have that for any $x, y \in \mathbb{C}^n$

$$\Big|\sum_{i=1}^n x_i y_i\Big| \leq \|x\|_p \|y\|_q.$$

In (1) we use the convention that $\frac{1}{p} = 0$ for $p = \infty$, and one often calls numbers $p, q \in [1, \infty]$ satisfying (1) *Hölder conjugate exponents*.

Remark. As you will show on Problem sheet 1, we have that for all $1 \leq p < \infty$

$$\|x\|_\infty \leq \|x\|_p \leq n^{1/p}\|x\|_\infty.$$

Hence the ∞-norm is equivalent to every p-norm and thus, by transitivity, we have that $\|\cdot\|_p$ is equivalent to $\|\cdot\|_q$ for every $1 \leq p, q \leq \infty$. In fact as we will show in Section **??**, all norms on finite dimensional spaces are equivalent.

Deep Dive

With some small exceptions (which I neglected to mention in lectures), the spaces ℓ^p_n are not isometrically isomorphic as p varies. The first exception is if $n = 1$: then all the norms are the same! The only other exception is if $n = 2$, when ℓ^1_2 is isometrically isomorphic to ℓ^∞_2, via the map $(x,y) \to (x+y, x-y)$. In a deep dive below we discuss to how see this sort of thing for the sequence spaces ℓ^p, and come back to ℓ^1_n and ℓ^∞_n for $n \geq 3$.

Sequence spaces ℓ^p and c_0. An infinite dimensional analogue of $(\mathbb{R}^n, \|\cdot\|_p)$, respectively $(\mathbb{C}^n, \|\cdot\|_p)$ are the spaces of sequences ℓ^p, $1 \leq p \leq \infty$, where for $1 \leq p < \infty$

$$\ell^p := \Big\{ (x_j)_{j\in\mathbb{N}} : \sum_{j=1}^{\infty} |x_j|^p < \infty \Big\}$$

equipped with $\|\cdot\|_p$ where for $1 \leq p < \infty$

$$\|x\|_{\ell^p} = \Big(\sum_{j=1}^{\infty} |x_j|^p \Big)^{1/p},$$

while ℓ^∞ denotes the space of bounded sequences, equipped with

$$\|(x_j)\|_\infty := \sup_j |x_j|.$$

For any $1 \leq j \leq \infty$ we have that ℓ^p is a normed space (where we define addition and scalar-multiplication component-wise). Again the main difficulty is to obtain Minkowski's inequality, which is precisely the triangle inequality.

Example 1.8. ℓ^p is complete for $1 \leq p \leq \infty$.

This follows the standard procedure for showing completeness. Given a Cauchy sequence (x_n) in a normed space X:

1. identify a candidate x for $\lim x_n$;
2. show that $x \in X$;
3. show $\|x - x_n\|_X \to 0$ as $n \to \infty$.

Often, as in the proof below, which we recall from metric spaces, steps 2 and 3 can be performed simultaneously. The $p = \infty$ case is easier (and a special case of the Banach space $\mathscr{F}_b(\Omega)$ below, by taking $\Omega = \mathbb{N}$.)

Proof for $1 \leq p < \infty$. Let $(x^{(n)})$, be a Cauchy-sequence in ℓ^p and write $x^{(n)} = (x_j^{(n)})_{j=1^\infty}$. As for every $j \in \mathbb{N}$

$$|x_j^{(n)} - x_j^{(m)}| \leq \|x^{(n)} - x^{(m)}\|_2 \to 0$$

as $m, n \to \infty$, the sequence $(x_j^{(n)})$ is Cauchy in $\mathbb{F}$ so converges, say $x_j^{(n)} \to x_j$.

Fix $\varepsilon > 0$. Then there exists N so that for all $n, m \geq N$

$$\|x^{(n)} - x^{(m)}\|_p \leq \varepsilon.$$

Thus for every $K \in \mathbb{N}$ and for all $n \geq N$ we have that

$$\sum_{j=1}^{K} |x_j^{(n)} - x_j|^p = \lim_{m\to\infty} \sum_{j=1}^{K} |x_j^{(n)} - x_j^{(m)}|^p \leq \varepsilon^p.$$

Letting $K \to \infty$, it follows that for all $n \geq N$, we have $x^{(n)} - x \in \ell^p$, and so $x \in \ell^p$ (step 2), and $\|x^{(n)} - x\|_p \leq \varepsilon$ for $n \geq N$,[4], though for sums we were able to d so that $x^{(n)} \to x$ in ℓ^p. □

Again Hölder's inequality is valid. That is for every $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$ and any $(x_j) \in \ell^p$ and $(y_j) \in \ell^q$ we have that $\sum x_j y_j$ converges, i.e. the pointwise product $xy \in \ell^1$ and

$$|\sum_j x_j y_j| \leq \|(x_j)\|_p \|(y_j)\|_q.$$

Again this is a special case of Proposition 1.12 from Integration, taking the measure space to be $\mathbb{N}$ equipped with counting measure.

We will sometimes also consider the subspace

$$c_0 := \{(x_n) \in \ell^\infty : x_n \to 0\}$$

of ℓ^∞, which is closed[5] and hence, when equipped with the ℓ^∞- norm a Banach space (see Proposition 1.16 below)

Note that all of these sequence spaces contain a common subspace c_{00} – the collection of sequences which are eventually zero. You should check that c_{00} is dense in c_0, and in ℓ^p for $1 \leq p < \infty$. Accordingly it can not be a Banach space in any of the p-norms (by Proposition 1.16 below).

Deep Dive

In fact there is no norm on c_{00} under which it is a Banach space.

Theorem. *No infinite dimensional Banach space X can have a countable* Hamel basis.

A *Hamel basis* is what up to now we've just called a basis, i.e. a linearly independent spanning set. This is a purely algebraic notion so $\mathscr{S}$ is a Hamel basis for X when no non-trivial *finite* linear combination of elements of $\mathscr{S}$ can be zero, and every element of X can be written as a finite linear combination of elements of $\mathscr{S}$. While in a normed space we are allowed to consider infinite sums, these are not used to define Hamel bases. Certainly the polynomials have a countably infinite Hamel basis and so can not be a Banach space under any norm.

For background, by a Zorn's lemma argument every vector space has a Hamel basis. Zorn's lemma is a tool equivalent to the axiom of choice, which will appear in some other deep dives, but is definitely non-examinable. Zorn's lemma will be described in B1.2 (Set Theory) which will show that every vector space having a Hamel basis is another reformulation of the axiom of choice (that course will use the language of basis, rather than Hamel basis, as the vector spaces there do not come equipped with norms).

The normal, and in my view best way, to prove that no Banach space can have a countably infinite Hamel basis is through the Baire category theorem, which will be proved in B4.2 Functional Analysis 2. This states that a complete metric space is never a countable union of nowhere dense subsets.[a] We will see in Section 4 that finite dimensional subspaces of normed spaces are always closed, and it is an easy exercise to check that a proper closed subspace of a normed space is nowhere dense. Then the result follows. On problem sheet 2

[4] We have done a by-hand version of Fatou's lemma for infinite sums here. I make this remark only so you can compare it with the use of Fatou's lemma in one of the proofs that L^p is complete.

[5] This is very similar to Problem sheet 1, B.1(a).

we will see an alternative, slightly messier proof in Section C, using Riesz's lemma from Section **??**.

This result shows that Hamel bases, while they generally exist, are unlikely to be all that useful for working with Banach spaces. Even for spaces such as ℓ^1 or c_0, where we clearly only need countably many bits of data to specify elements, do not have countable bases. Instead, one needs the appropriate analytic notion of a basis – *Schauder bases* – to develop a satisfactory theory for Banach spaces. These appear in more detail at the end of the C4.1 course, but they are perfectly accessible to you now. A good book taking a basis first approach, and so developing interesting properties of the sequence spaces is Carothers 'A short course in Banach space theory'. Definitely the first 6 chapters of this book are very readable along side this course - but from chapter 3 onwards go in different direction.

[a] A nowhere dense set is a set whose closure has empty interior.

When I'm trying to build counter examples, my first thought is to check the finite dimensional case, and assuming that doesn't work then I tend to look for an example using sequences.

Deep Dive

While for $n \in \mathbb{N}$ fixed, all the n-dimensional normed spaces ℓ_n^p are equivalent, this fails in infinite dimensions. Also, just as in the finite dimensional case, by looking at the geometry of the unit balls one can see that no pair of these spaces is isometrically isomorphic. One thing you can look at here is the *modulus of convexity*, which quantifies the convexity of the unit ball. For a Banach space X, define the modulus of convexity $\delta_X : [0,2] \to [0,1]$ by

$$\delta_X(\varepsilon) = \inf\left\{1 - \left\|\frac{x+y}{2}\right\| : x, y \in X,\ \|x\| = \|y\| = 1,\ \|x-y\| \geq \varepsilon\right\}.$$

This measures how far the average of two points x and y of distance at least ε on the sphere must get pushed inside the unit ball. (There's some nice pictures to be drawn here; if someone draws them I'll be happy to include). You should be able to see that for ℓ^∞ and ℓ^1 we have $\delta = 0$. The modulus of convexity of ℓ^2, and indeed any Hilbert space, is $\delta(\varepsilon) = 1 - \sqrt{1 - \varepsilon^2/4} = \varepsilon^2/8 + O(\varepsilon^4)$ as $\varepsilon \to 0$.footnoteIt's small ε we care about, otherwise we wouldn't have used the notation ε. A two-dimensional geometric argument shows that this is the upper bound of the modulus of convexity for any normed space X; in this sense Hilbert space's have the "convexest" unit ball. For other p, we have

$$\delta_{\ell^p}(\varepsilon) = \begin{cases} (p-1)\frac{\varepsilon^2}{8} + o(\varepsilon^2), & 1 < p \leq 2; \\ \frac{\varepsilon^p}{p2^p}, & 2 \leq p < \infty. \end{cases}$$

While the modulus of convexity doesn't distinguish ℓ_n^∞ and ℓ_n^∞, for $n \geq 3$, you can see that nevertheless they are not isometrically isomorphic from the geometry of the unit balls: the ball of ℓ_n^1 has $2n$ extreme points, and the ball of ℓ_n^∞ has 2^n extreme points.[a]

Showing that in infinite dimensions none of the sequence spaces ℓ^p or c_0 are isomorphic is more challenging. By section **??** we will know that ℓ^∞ is not isomorphic to any of the other spaces — it is too big (not separable to be precise). By the end of the course, we will know that c_0, ℓ^1 are not isomorphic and also not isomorphic to any other ℓ^p space. In fact all these spaces are pairwise non-isomorphic. This is much harder, and we won't prove it in any of the functional analysis courses here. (The way this is normally done is through Pitt's theorem, that says that for $1 \leq p < q < \infty$, any bounded linear map (see Section 3) $T : \ell^q \to \ell^p$ is compact (a concept that will be defined in B4.2). Taking this fact for granted (a proof can be found in

Chapter 2 of Albaic and Kalton's "Topics in Banach Space Theory") you should be able to deduce that all these sequence spaces are pairwise non-isomorphic by the end of B4.2.

[a] I'll let you formalise or look up the definition of an extreme point. A picture ought to work in finite dimensions, but in infinite dimensions some unit balls need not have extreme points, for example c_0. I will make further deep dive remarks on this later in the notes.

Function spaces with supremum-norm

The supremum norm is often used to make spaces of bounded functions complete.

Example 1.9. Let Ω be any set. Then

$$\mathscr{F}^b(\Omega) := \{f : \Omega \to \mathbb{F} \text{ bounded}\}$$

is a Banach space under the norm $\|f\|_{\text{sup}} = \sup_{x\in\Omega} |f(x)|$. (We will often drop the subscript sup on these norms).

Proof of completeness of $\mathscr{F}^b(\Omega)$. This is another example of the three step process. If (f_n) is Cauchy in $\mathscr{F}^b(\Omega)$ then for each $x \in \Omega$, $(f_n(x))$ is Cauchy so converges to $f(x) \in \mathbb{F}$, say. Fixing $\varepsilon > 0$, there is $N \in \mathbb{N}$ with $|f^{(n)}(x) - f^{(m)}(x)| \leq \varepsilon$ for $m,n \geq N$ and all $x \in \Omega$. Taking limits as $m \to \infty$, $|f_n(x) - f(x)| \leq \varepsilon$ for all $x \in \Omega$ and $n \geq N$. Therefore f is bounded, and $\|f_n - f\| \leq \varepsilon$. Accordingly $\mathscr{F}^b(\Omega)$ is complete. $\square$

For a metric[6] space K, the vector space $C_b(K) := \{f : K \to \mathbb{F} \text{ continuous and bounded}\}$ is closed in $\mathscr{F}^b(K)$,[7] and hence $C_b(K)$ is a Banach space with the sup norm (see Proposition 1.16). If K is compact, then all continuous functions are bounded, and we write $C(\Omega) := \{f : \Omega \to \mathbb{F} \text{ continuous}\} = C_b(\Omega)$.

Example 1.10. Let K be a compact metric space. Then $C(K)$ is a Banach space with the sup-norm.

Similarly, on spaces of differentiable functions (with bounded derivatives) such as $C^1([0,1])$ – the space of functions on $[0,1]$ which have continuous derivatives (including at the end points), to get completeness we will need norms that are built using the sup norm of both the function and its derivative. Indeed, since $C^1([0,1])$ is dense in $C[0,1]$ (in fact the smaller set of polynomials is dense in $C[0,1]$ – see Section **??**), it is not closed in the sup-norm and so is not a Banach space with respect to $\|\cdot\|_{\text{sup}}$. Instead we use a standard trick for creating a norm on spaces like this; take the sum of two norms that we want to be able to control. Then a sequence which is Cauchy in the sum of the norms, will necessarily be Cauchy in each norm.

Example 1.11. $C^1([0,1])$ is a Banach space with $\|f\|_{C^1} := \|f\|_{\text{sup}} + \|f'\|_{\text{sup}}$

Proof. Suppose (f_n) is a Cauchy sequence in $C^{(}[0,1])$. Then (f_n) is $\|\cdot\|_{\text{sup}}$-Cauchy so converges to some $f \in C([0,1])$ and (f_n') is also $\|\cdot\|_{\text{sup}}$-Cauchy so converges to some $g \in C([0,1])$. Now prelims comes to the rescue: since (f_n) converges uniformly to f and each f_n' is differentiable and f_n' converges uniformly to g, it follows that f is differentiable and $f' = g$. Hence $f \in C^{(}[0,1])$, and then $\|f_n - f\|_{C^1} \to 0$. $\square$

Deep Dive

The map $f \mapsto (f', f(0))$ gives an isomorphism[a] from $C^1([0,1])$ to $C([0,1]) \times \mathbb{R}$ (we will discuss norms on the Cartesian product below. In fact, using a Schauder basis for $C([0,1])$ it is possible to show that $C([0,1]) \cong C([0,1]) \times \mathbb{R}$ though the isomorphism can not be isometric,[b] so that as Banach spaces $C^1([0,1])$ and $C([0,1])$ are isomorphic. But in applications we would most likely care about how our elements are

[6] this works fine for a topological space

[7] this is a consequence of the fact that a uniform limit of continuous functions is continuous — a 3ε or $\varepsilon/3$ argument depending on your taste.

realised as functions: so while these Banach spaces are abstractly isomorphic, it makes sense to understand them separately.

[a]This map is bounded and bijective between Banach spaces, so an isomorphism from Banach's isomorphism theorem; though you could show directly that the inverse map is bounded.

[b]We're implicitly working with reals here, so you can use the geometry of the unit ball: $C([0,1])$ has two extreme points, the constant functions ± 1, while the unit ball of $C([0,1]\times\mathbb{R}$ with the ℓ^2-norm has more.

Function spaces L^p, $1 \leq p \leq \infty$

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space (considering examples like Ω an interval in $\mathbb{R}$ or a measurable subset of $\mathbb{R}^n$ with Lebesgue measure will be more than sufficient for the course). Consider for $1 \leq p < \infty$ the space of functions

$$\mathscr{L}^p(\Omega) := \left\{ f : \Omega \to \mathbb{R} \text{ (or } \mathbb{C}) \text{ measurable so that } \int_\Omega |f|^p dx < \infty \right\}$$

respectively

$$\mathscr{L}^\infty := \left\{ f : \Omega \to \mathbb{R} \text{ (or } \mathbb{C}) \text{ measurable so that } \exists M \text{ with } |f| \leq M \text{ a.e.} \right\}.$$

This is not a course on Lebesgue integration, so typically in examples we will only work with measurable functions; you may assume that all functions you encounter are measurable when needed (though don't let that stop you briefly reminding yourself why). But of course not all measurable functions are integrable and that indeed for a general measurable function the integral might not even be defined, so justification is needed to consider integrals in general. However we also recall that the integral of a non-negative functions f is always defined though might be infinite.

We equip these spaces with

$$\|f\|_p := \left(\int_\Omega |f|^p dx \right)^{1/p} \text{ for } 1 \leq p < \infty$$

respectively

$$\|f\|_\infty := \text{ess sup}|f| := \inf\{M > 0 : |f| \leq M \text{ a.e. }\}.$$

We note that $\|\cdot\|_p$ is only a seminorm on $\mathscr{L}^p$ with $\|f-g\|_p = 0$ if and only if $f = g$ a.e. We can hence turn $(\mathscr{L}^p, \|\cdot\|)$ into a normed space by taking the quotient with respect to the equivalence relation

$$f \sim g \Longleftrightarrow f = g \text{ a.e.}.$$

The resulting quotient space

$$L^p(\Omega) := \mathscr{L}^p / \sim \text{ equipped with } \|\cdot\|_p$$

is one of the most important spaces of functions in the modern theory of PDE, and will be further developed in the course C4.3 Functional analytic methods for PDEs. Recall the following two key inequalities for L^p spaces; the first giving the triangle inequality, and the second will be crucial later in the course when we examine the dual spaces of $L^p(\Omega)$.

Proposition 1.12 (Minkowski and Hölder)**.**

- *The Minkowski-inequality, which is the triangle inequality for L^p holds: for $f, g \in L^p(\Omega)$, we have*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

- *Hölder's inequality holds: If $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$ where $\frac{1}{p} + \frac{1}{q} = 1$ then their product fg is integrable with*

$$\left| \int_\Omega fg dx \right| \leq \|f\|_p \|g\|_q.$$

None of the L^p norms are equivalent, though when Ω has positive and finite measure, we can estimate the L^p norm of functions by their L^q norm if $p < q$ and we have

$$L^\infty(\Omega) \subsetneq L^q(\Omega) \subsetneq L^p(\Omega) \subsetneq L^1(\Omega) \text{ for any } 1 < p < q < \infty. \tag{2}$$

As an example consider $\Omega = (0,2) \subset \mathbb{R}$ and $p = 2$, $q = 4$. Adding in a multiplication by the constant function $g = 1$ we can estimate, using Hölder's inequality,

$$\|f\|_{L^2}^2 = \int |f|^2 \cdot 1 dx \le \||f|^2\|_{L^2} \|1\|_{L^2} = \Big(\int_0^2 f^4\, dx\Big)^{1/2} \cdot \Big(\int_0^2 1\, dx\Big)^{1/2} = \sqrt{2}\|f\|_{L^4}^2,$$

so we get $\|f\|_{L^2} \le \sqrt{2}\|f\|_{L^4}$ and in particular that every $f \in L^4([0,2])$ is also an element of $L^2([0,2])$. The general case is discussed on the first problem sheet.

Warning.

- Note that the inclusions of the function spaces $L^p(\Omega)$ for sets Ω with bounded measure are the "other way around" compared with the inclusions of the sequence spaces ℓ^p.
- The inclusion (2) is wrong for unbounded domains, e.g. the constant function $f = 1$ is an element of $L^\infty(\mathbb{R})$ but isn't contained in any $L^p(\mathbb{R})$, $1 \le p < \infty$.

Remark. In practice it is can be useful to extend $\|\cdot\|_{L^p}$ to a function from the space of all (measurable) functions to $[0,\infty) \cup \{\infty\}$ by simply setting $\|f\|_{L^p} = \infty$ if $\int |f|^p = \infty$ (respectively for $p = \infty$ if $f \notin L^\infty$), and we note that also with this 'abuse of notation' the triangle and Hölder-inequality still hold (with the convention that $0 \cdot \infty = 0$ for Hölder's inequality). Similarly we can extend $\|\cdot\|_p$ to a function that maps all sequences to $[0,\infty) \cup \{\infty\}$ but we stress that while this notation/convention can be useful and used in the literature, these functions into $[0,\infty) \cup \{\infty\}$ are not norms as a norm is by definition a function into $[0,\infty)$.

Example 1.13. $(L^\infty(\Omega), \|\cdot\|_{L^\infty})$ is a Banach space.

The proof is more or less similar to the proof of completeness for $\mathscr{F}_b(\Omega)$, or a direct proof of completeness for $C(K)$ in the supremum norm, except that we have to take care of the almost everywhere nature of things.

Proof. Let (f_n) be a Cauchy sequence in $L^\infty(\Omega, \mathbb{R})$. For each $k \in \mathbb{N}$, there exists N_k such that

$$\|f_n - f_m\|_{L^\infty} \le \frac{1}{k} \text{ for all } n, m \ge N_k.$$

This means that, for each k and $m, n \ge N_k$ there is a null subset $Z_{k,m,n}$ of Ω such that

$$|f_n(x) - f_m(x)| \le \frac{1}{k} \text{ for } x \in \Omega \setminus Z_{k,m,n}.$$

Let $Z = \cup_k \cup_{n,m \ge N_k} Z_{k,n,m}$, which, as a countable union of null set, is null. Then,

$$|f_n(x) - f_m(x)| \le \frac{1}{k} \text{ for all } n, m \ge N, x \in \Omega \setminus Z. \tag{3}$$

So for almost all $x \in \Omega$, $(f_n(x))$ is Cauchy, and hence converges to some $f(x)$.

Being an almost everywhere limit of measurable functions, f is measurable. Sending $m \to \infty$ while keeping n fixed in (3) we get

$$|f_n(x) - f(x)| \le \frac{1}{k} \text{ for all } n \ge N_k, x \in \Omega \setminus Z.$$

This shows that $\|f_n - f\|_{L^\infty} \le \frac{1}{k}$ for all $n \ge N_k$. This implies on one hand that $f_n - f$ and hence f belong to $L^\infty(\Omega)$ and on the other hand that $f_n \to f$ in $L^\infty(\Omega)$. $\square$

Example 1.14. L^p is complete for $1 \leq p < \infty$.

We give two proofs. Firstly we consider the proof lectured in part A integration,[8] and show how this really fits into the abstract framework of proving completeness by showing that absolutely convergence series converge.

Proof. Let $(f_n)_{n=1}^{\infty}$ be a sequence in $L^p(\Omega)$ with $\sum_n \|f_n\|_p < \infty$. Define $g_n = \sum_{r=1}^{n} |f_r|$. This gives an increasing sequence of non-negative measurable functions which converges to $g = \sum_{n=1}^{\infty} |f_n|$ (g can of course take the value ∞ whenever this sum diverges). By Minkowski

$$\int g_n^p = \|g_n\|_p^p \leq \left(\sum_{r=1}^{n} \|f_n\|_p\right)^p \leq \left(\sum_{n=1}^{\infty} \|f_n\|_p\right)^p < \infty.$$

Therefore by the monotone convergence theorem g^p is integrable, i.e. $g \in L^p$, and hence g is finite almost everywhere. That is the sum $\sum_{n=1}^{\infty} f_n$ is absolutely convergent almost everywhere, and so converges almost everywhere, say to f.[9] Moreover, applying the triangle inequality pointwise $|f|^p \leq g^p$, so $f \in L^p$ by comparison.[10]

Finally, another application of the triangle inequality gives

$$|f - \sum_{r=1}^{n} f_r|^p \leq \left(\sum_{r=n+1}^{\infty} |f_r|\right)^p \leq g^p,$$

so the dominated convergence theorem gives $\|f - \sum_{r=1}^{n} f_r\|_p^p \to 0$, and $\sum_{n=1}^{\infty} f_n = f$ in L^p.[11] □

In the integration course proved the following facts (the first of which essentially came from in the middle of the proof given there that L^p was complete which essentially used the series argument above):

- If (f_n) is a Cauchy sequence in L^p (or a sequence converging in L^p to $f \in L^p$), then there exists a subsequence f_{n_k} which converges almost everywhere (to f). But we can not guarentee that f_n converges almost everwhere, only that a subsequence does.
- Given a sequence (f_n) in L^p with $f_n \to f$ almost everywhere, the *convergence theorems* (monotone convergence theorem, Fatou's lemma and the dominated convergence theorem) give tools you can try and use to prove that $f \in L^p$ and $f_n \to f$ in L^p.

While we only really used Fatou's lemma in the integration course as a tool for obtaining the dominated converge theorem, it can be very useful for obtaining convergence of f_n to f in the L^p norms. Let's see this in action in our second proof of completeness of L^p assuming the first fact above.

Second proof of completeness of L^p. Let $(f_n)_{n=1}^{\infty}$ be Cauchy in L^p, and let (f_{n_k}) be a subsequence which converges almost everywhere to a (necessarily measurable) f. Fix $\varepsilon > 0$. We know that

$$\|f_n - f_{n_j}\|_{L^p}^p = \int_{\Omega} |f_n - f_{n_j}|^p \, dx \leq \varepsilon^p \text{ for all } n, n_j \geq N.$$

As $j \to \infty$, the a.e. limit of the integrand is $|f_n - f|^p$. Moreover, the integrand is non-negative. By Fatou's lemma,[12] we have

$$\int_{\Omega} |f_n - f|^p \, dx \leq \liminf_{j \to \infty} \int_{\Omega} |f_n - f_{n_j}|^p \, dx \leq \varepsilon^p \text{ for all } n \geq N.$$

In other words, $\|f_n - f\|_{L^p} \leq \varepsilon$ for all $n \geq N$. This implies on one hand that $f_n - f$ and hence f belong to $L^p(\Omega)$ and on the other hand that $f_n \to f$ in $L^p(\Omega)$. □

[8] The case of L^p was lectured, but only the case of L^1 is in the lecture notes.

[9] This is Step 1 of the process for completeness (in this absolute convergence framework) by providing a candidate limit.

[10] Step 2: the limit is in the space it should be in.

[11] The expression in L^p relating to this sum means that we have justified convergence of the sum in the norm on L^p, as required for Step 3.

[12] Compare this with footnote [4].

Since the first fact above is fair game for the course – we proved it in integration – this is an approach I encourage you to keep in mind.

Deep Dive

Let's give another proof that every L^p Cauchy-sequence has a subsequence which is almost everywhere convergent by means of *convergence in measure*, and Borel-Cantelli arguments (see B8.1, or many past integration exams). I don't think we'll really need these tools in the course, but you are welcome to use them in L^p type examples if they help.

Say that a sequence f_n of measurable functions on $(\Omega, \mathscr{F}, \mu)$ *converges in measure* to a measurable function f when for all $\delta > 0$,

$$\mu(\{x \in \Omega : |f_n(x) - f(x)| > \delta\}) \to 0 \text{ as } n \to \infty.$$

Likewise, (f_n) is Cauchy in measure if for all $\delta > 0$,

$$\mu(\{x \in \Omega : |f_n(x) - f_m(x)| > \delta\}) \to 0 \text{ as } n, m \to \infty.$$

Convergence in measure, implies being Cauchy in measure. For the converse we find an almost everywhere convergent subsequence to obtain the proposed limit.

Lemma. Suppose (f_n) is Cauchy in measure. Then there exists a subsequence (f_{n_k}) which is convergent almost everywhere to f, and f_n converges to f in measure.

Proof. To follow. $\square$

Combining the above with the following lemma, every L^p Cauchy sequence has a subsequence which is convergent almost everywhere.

Lemma. Let (f_n) be Cauchy in $L^p(\Omega, \mathscr{F}, \mu)$. Then (f_n) is Cauchy in measure.

Proof. Fix $\varepsilon > 0$ and $\delta > 0$ and find N such that for $m, n \geq N$, we have $\|f_n - f_m\|_p < \varepsilon$. Therefore, for $n, m \geq N$,

$$\mu(\{x : |f_n(x) - f_m(x)|^{\geq} \delta^p\}) \leq \frac{\varepsilon^p}{\delta^p}.$$

Therefore (f_n) is Cauchy in measure. $\square$

Deep Dive

We should also discuss potential isomorphisms between all the L^p-spaces and all the ℓ^p spaces. Firstly none of the L^p spaces are isometrically isomorphic; the same modulus of convexity formula in one of the deep dives above works for L^p.

$L^2(\Omega, \mathscr{F}, \mu)$ is a Hilbert space, and as we will learn these are determined up to isometric isomorphism by the size of an orthonormal basis. In most examples of interest to us, $(\Omega, \mathscr{F}, \mu)$ has just the right 'size' to be separable (have a countable dense subset; see Section 6), and in this case it will be isometrically isomorphic to ℓ^2. This will always be the case when $\Omega \subset \mathbb{R}^n$ is a Lebesgue measurable subset (with non zero measure) equipped with Lebesgue measure.[a] As an example you might well be able to guess an isometric isomorphism between $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$.

A surprising theorem of Pełczyński shows that $L^\infty([0,1])$ and ℓ^∞ are isomorphic Banach spaces (though they are not isometrically isomorphic). This theorem proceeds through a back and forward principle known as Pełczyński's decomposition technique (see Albiac and Kalton, Theorem 2.3.3): if X and Y are Banach spaces such that X is isomorphic to a complemented subspace of Y[b], Y is isomorphic to a complemented subspace of X, X is isomorphic to $X \oplus X$, and Y is isomorphic to $Y \oplus Y$, then X and Y are isomorphic. Let $X = L^\infty([0,1])$ and $Y = \ell^\infty$. In fact both these spaces are what's known as *injective* (something we will find easier to define once we have bounded linear maps) which means that whenever either X or Y is isomorphic to a closed subspace of another Banach space, then this subspace is automatically complemented. We will be able to prove injectivity of ℓ^∞ as a consequence of Hahn–Banach by the end of the course. Injectivity of $L^\infty([0,1])$ is a bit harder; this can be found in Section 4.3 of Albiac and Kalton. Given these results to prove Pełczyński's theorem you need to see how to find ℓ^∞ as a closed subspace of L^∞ (have a go – it's not so bad), and L^∞ as a closed subspace of ℓ^∞ (this can be found as an extension exercise to Sheet 3). All of this is collected as Theorem 4.3.10 of Albiac and Kalton. To the best of my knowledge, Pełczyński's theorem is non-constructive and no explicit isomorphism is known.

Finally, ℓ^p and $L^p([0,1])$ are not isomorphic for other values of p. When $p = 1$ you'll be able to use something called Schur's property to distinguish ℓ^1 and $L^1([0,1])$ as an exercise in B4.2 and C4.1 (though I think the name 'Schur's property' will only be introduced in $C4.1$). To distinguish ℓ^p and $L^p([0,1])$ for $1 < p < 2$ and $2 < p < \infty$ one can show that ℓ^2 is a complemented subspace of $L^p([0,1])$ for any $1 < p < \infty$ (see Proposition 6.4.2 of Albiac and Kalton) but (by means of another Pełczyński decompositon technique) any complemented infinite dimensional subspace of ℓ^p is isomorphic to ℓ^p (see Theorem 2.2.4 of Albiac and Kalton).

[a] In general you need that Ω is σ-finite and $\mathscr{F}$ is the completion of a countably generated σ-algebra.
[b] we will discuss complemented subspaces in Section **??**

Some incomplete spaces

Example 1.15. We can construct many examples of non-complete spaces by equipping a well known space such as C_b, C^1, ℓ^p, L^p with the 'wrong' norm, or by choosing a subspace of a Banach space that is not closed. As an example we show that $C([0,1])$ equipped with $\|f\|_{L^1} = \int_0^1 |f|dx$ is not complete (which is actually an exercise from the metric spaces course and so I won't lecture it).

Proof. We give three proofs: one by direct argument, one via Corollary 1.6, and finally through density.

1. Let
$$g_n(x) = \begin{cases} (2x)^n & \text{for } x \in [0,1/2), \\ 1 & \text{for } x \in [1/2,1]. \end{cases}$$
For $n < m$, we have
$$\|g_n - g_m\|_{L^1} = \int_0^{1/2} [(2x)^n - (2x)^m]\,dx = \frac{1}{2(n+1)} - \frac{1}{2(m+1)},$$
so (g_n) is Cauchy. On the other hand, (g_n) is a decreasing sequence of non-negative functions which is bounded from above by 1. Its pointwise limit is the characteristic function of the interval $[1/2,1]$. By Lebesgue's dominated convergence theorem, g_n converges to $\chi_{[1/2,1]}$ in L^1 and there is no continuous function which is equal to $\chi_{[1/2,1]}$ almost everywhere,[13], and hence not in $C([0,1])$. In other words (g_n)

[13] $\chi_{[1/2,1]}$ is almost everywhere continuous, but not almost everywhere equal to a continuous function

has no limit in $C([0,1])$.[14]

2. For
$$f_n(x) := \begin{cases} 1 - n^2 x & \text{for } x \in [0, \frac{1}{n^2}] \\ 0 & \text{else} \end{cases}$$
we have that $\|f_n\|_1 = \frac{1}{2n^2}$ so $\sum \|f_n\|_{L^1}$ converges. However $\sum f_n$ cannot converge to an element of $C([0,1])$. Indeed suppose, seeking a contradiction, that $\sum f_n \to f$ converges in L^1 to a function $f \in C([0,1])$. Then, as continuous functions on compact sets are bounded, there exists some $M \in \mathbb{R}$ so that $f \leq M$ on $[0,1]$. Hence choosing $N \in \mathbb{N}$ so that $N \geq 2(M+1)$ we obtain that for any $n \geq N$ and any $x \in [0, \frac{1}{2N^2}]$
$$\sum_{j=1}^{n} f_j(x) - f(x) \geq \sum_{j=1}^{N} \frac{1}{2} - f(x) \geq N/2 - M \geq 1$$
and thus in particular $\|\sum_{j=1}^{n} f_j - f\|_1 \geq \frac{1}{2N^2} \not\to 0$.

3. We know from part A integration that $C([0,1])$ is a proper dense subspace of $L^1([0,1])$, so can not be complete (by Proposition 1.16).

$\square$

1.3 Constructions

We end this section with a brief collection of ways to construct new normed spaces from existing examples, and when this preserves completeness.

Subspaces We first note that for any given subspace Y of a normed space $(X, \|\cdot\|)$ we obtain a norm on Y simply by restricting the given norm to Y. For the resulting normed space $(Y, \|\cdot\|)$ we have

Proposition 1.16. *Let X be a Banach space, $Y \subset X$ a subspace. Then*
$$(Y, \|\cdot\|) \text{ is complete } \Leftrightarrow Y \subset X \text{ is closed }.$$

Proof. Suppose Y is complete, and (y_n) is a sequence in Y with $y_n \to x \in X$. As (y_n) converges in X, it is Cauchy. Therefore by completeness it converges to some $y \in Y$. Hence $x = y \in Y$ by uniqueness of limits and Y is closed.

Conversely, suppose Y is closed in X and let (y_n) be a Cauchy sequence in Y. By completeness of X, it follows that there exists $x \in X$ with $y_n \to x \in X$. But as Y is closed we must have that $x \in Y$ and hence that (y_n) converges in Y. Therefore Y is complete. $\square$

Direct sums Given two normed spaces X and Y we can define a norm on $X \times Y$ e.g. by
$$\|(x,y)\|_2 = (\|x\|^2 + \|y\|^2)^{1/2} \tag{4}$$
or more generally using any of the p-norms on $\mathbb{R}^2$ to define
$$\|(x,y)\|_p := \|(\|x\|, \|y\|)\|_p = (\|x\|^p + \|y\|^p)^{1/p} \text{ respectively } \|(x,y)\|_\infty := \max(\|x\|, \|y\|)$$

[14] Here we are using the fact that we know $C([0,1])$ is a subspace of $L^1([0,1])$ so limits are unique. When you did this exercise in the metric spaces course, this answer would not have been sufficient as we didn't have the space $L^1([0,1])$ to work with. In part A metric spaces you where supposed to deal with this by showing barehands that there is no continuous function which can arise as the L^1 limit.

where here and in the following we simply write $\|\cdot\|$ instead of $\|\cdot\|_X$ and $\|\cdot\|_Y$ if it is clear from the context what norm we are using. As all (the ℓ^p)-norms on $\mathbb{R}^2$ are equivalent, it follows that all the norms $\|(x,y)\|_p$ are equivalent on $X \times Y$. We tend to write $X \oplus_p Y$ for these spaces.

We note that for all of these norms on $X \times Y$ we obtain that $X \times Y$ is again a Banach space if both X and Y are Banach spaces. If X and Y are inner product spaces then one uses in general the norm (4) as for this choice of norm also the product $X \times Y$ will again be a inner product space with inner product $((x,y),(x',y')) = (x,x')_X + (y,y')_Y$, while none of the norms with $p \neq 2$ preserve the structure of an inner product space.

Deep Dive

We can also consider countable direct sums. Given normed spaces $(X_n)_{n=1}^{\infty}$, and $1 \leq p \leq \infty$, one can form the ℓ^p-direct sums. For $p = \infty$, let
$$X_\infty = \{(x_n) : x_n \in X_n,\ \sup \|x_n\| < \infty\}$$
with the norm $\|(x_n)\|_\infty = \sup \|x_n\|$. For $1 \leq p < \infty$ let
$$X_p = \{(x_n) : x_n \in X_n : \sum_{n=1}^{\infty} \|x_n\|^p < \infty\}$$
with the norm $\|(x_n)\|_p = (\sum_{n=1}^{\infty} \|x_n\|^p)^{1/p}$.

These spaces might be written as $(\oplus_{n=1}^{\infty} X_n)_p$. We could also consider a c_0-sum.

It's a good exercise in seeing if you understand the proofs that ℓ^p forms a Banach space to check these are norms, and that if each X_n is complete so too our the spaces X_p. This time of course all these norms will in general give rise to pairwise non-isomorphic spaces (as can be seen by taking each $X_n = \mathbb{F}$ when you get back the classical sequence spaces ℓ^p). Now you can start asking what sort of spaces you get if you take an infinite ℓ^p product say of a sequence of L^{q_n} spaces!

Sums of subspaces If $X_1, X_2 \subset X$ are subspaces of a normed space X then also
$$X_1 + X_2 := \{x_1 + x_2 : x_1 \in X_1, x_2 \in X_2\}$$
is again a subspace of X, but beware. Just because X_1 and X_2 are closed in X, it does not necessarily follow that $X_1 + X_2$ is closed; see example sheet 1 (question C1) for an example.

Quotients We saw the process of taking quotients to pass from the semi-norm on $\mathscr{L}^p$ to the normed space L^p above. This works generally. Given a vector space X and a semi-norm $|\cdot|$ on X, i.e. a function $|\cdot| : X \to [0,\infty)$ satisfying (N2) and (N3), we can consider the quotient space X/X_0 where $X_0 := \{x \in X : |x| = 0\}$. Then one can define a norm on X/X_0 by defining $\|x + X_0\| := |x|$, see problem sheet 1 for details.

Deep Dive

There are many reasons to be interested in quotient spaces more generally. Suppose X is a normed space, and Y is a subspace of X, when can we put a norm on X/Y? The solution is to define
$$\|x + Y\| = \inf\{\|x + y\| : y \in Y\} = \inf\{\|x - y\| : y \in Y\} = d(x, Y),$$
In general this is only a seminorm as if $\|x + Y\| = 0$, then there is a sequence $y_n \in Y$ with $\|x + y_n\| \to 0$. Noting that $-y_n \in Y$, it follows that $\|x + Y\| = 0$ if and only if x is in the closure $\overline{Y}$ of Y. In this way we get a norm

on X/Y precisely when Y is closed.[a].

This will be explored further in C4.1, but it is nice to know that the quotient of a Banach space by a closed subspace is again a Banach space (this is normally proved by showing absolute convergence implies convergence). Since the kernel of a continuous linear map is always closed, one can go on to work out what the right first isomorphism theorem should be in the setting of normed spaces (spoiler alert, there is a subtlety: it will not always be the case that the range space is isomorphic to the domain modulo the kernel, but it works for continuous linear maps between Banach spaces with closed range).

[a] When Y is not closed, you could follow the construction of taking a further quotient of X/Y by the null space of the seminorm. You can check that this gives the same thing as considering the quotient $X/\overline{Y}$, so in practise we only consider quotients by closed subspaces.

2 Inner product spaces and Hilbert spaces

In this section we turn to the important special case when the norm arises from an inner product, leading to the class of Hilbert spaces – one of the most central objects in mathematics. Just as with finite dimensional inner product spaces (and unlike Banach spaces), Hilbert spaces are completely classified upto isometric isomorphism by their *dimension*: the cardinality of an *orthonormal basis* (the appropriate notion of basis in the setting of Hilbert spaces).

2.1 Definitions and basic properties

Definition 2.1. An inner (scalar) product in a linear vector space X over $\mathbb{R}$ is a real-valued function on $X \times X$, denoted as $\langle x, y\rangle$, having the following properties:

(i) *Bilinearity*. For fixed y, $\langle x,y\rangle$ is a linear function of x, and for fixed x, $\langle x,y\rangle$ is a linear function of y.

(ii) *Symmetry*. $\langle x,y\rangle = \langle y,x\rangle$ for all $x,y \in X$.

(iii) *Positivity*. $\langle x,x\rangle > 0$ for $x \neq 0$.

When X is a vector space over $\mathbb{C}$, $\langle x,y\rangle$ is complex-valued and properties (i) and (ii) are replaced by

(i') *Sesquilinearity*. For fixed y, $\langle x,y\rangle$ is a linear function of x, and for fixed x, $\langle x,y\rangle$ is a skewlinear function of y, i.e.
$$\langle ax,y\rangle = a\langle x,y\rangle \text{ and } \langle x,ay\rangle = \bar{a}\langle x,y\rangle \text{ for all } a \in \mathbb{C}, x,y \in X.$$

(ii') *Skew symmetry*. $\langle x,y\rangle = \overline{\langle y,x\rangle}$ for all $x,y \in X$.

Warning. In some textbooks and courses, the sesquilinearity property is reversed: $\langle x,y\rangle$ is required instead to be skewlinear in x and linear in y. This particularly the case when one is coming from a quantum theory viewpoint, when the braket notion $\langle x|y\rangle$ is often used for the inner product.

An inner product $\langle\cdot,\cdot\rangle$ generates a norm, denoted by $\|\cdot\|$, as follows:
$$\|x\| = \langle x,x\rangle^{1/2}.$$

Then the positivity of the norm $\|\cdot\|$ follows from the positivity property (iii), and the homogeneity of $\|\cdot\|$ follows from the bi/sequi-linearity property (i)/(i'). The triangle inequality is a consequence of the Cauchy-Schwartz inequality below. The proof below is the same as in prelims.

Theorem 2.2 (Cauchy-Schwarz inequality)**.** *For $x, y \in X$,*

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Equality holds if and only if x and y are linearly dependent.

Proof. If $y = 0$, the conclusion is clear. Assume henceforth that $y \neq 0$. Replacing x by ax with $|a| = 1$ so that $a\langle x, y\rangle$ is real, we may assume without loss of generality that $\langle x, y\rangle$ is real.

For $t \in \mathbb{R}$, we compute using sesquilinearity and skew symmetry:

$$\|x+ty\|^2 = \langle x+ty, x+ty\rangle = \|x\|^2 + 2t\,\mathrm{Re}\,\langle x, y\rangle + t^2\|y\|^2. \tag{5}$$

By positivity, this quadratic polynomial in t is non-negative for all t. This implies that

$$(\mathrm{Re}\,\langle x, y\rangle)^2 - \|x\|^2\|y\|^2 \leq 0,$$

which gives the desired inequality. If equality holds, then there is some t_0 such that $x + t_0 y = 0$. The conclusion follows. $\square$

Note that the Cauchy-Schwartz identity ensures that the inner product $\langle\cdot,\cdot\rangle$ gives a continuous map $X \times X \to \mathbb{F}$.

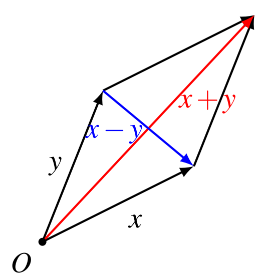
Proposition 2.3. *Let X be an inner product space. Then the* parallelogram law

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \text{ for all } x, y \in X, \tag{6}$$

holds.

Proof. Set $t = \pm 1$ in (5) and add the resulting identities. $\square$

This is illustrated[15] in $\mathbb{R}^2$ by



In fact the parallelogram law determines whether a norm comes from an inner product. See Sheet 2.A.3 for a proof.[16]

[15] Notice that the identity only involves the vectors x, y and so is verified in the 2-dimensional subspace $\mathrm{Span}(x, y)$ which we know is isometrically isomorphic to $\mathbb{R}^2$ with the usual inner product by means of the Gramm-Schmidt process. So if you've known the parallelogram law as a fact about parallelograms, then you've actually known the real case of the parallelogram law for inner product spaces!

[16] It's not hard to check that the expressions for the inner product in terms of the polarisation identities are the only things that can work: if you know the norm comes from an inner product simply multiply out the right hand sides . The difficulty is seeing that these identities do define an inner product. Note that the polarisation identity can be used in various other situations. For example, it shows that any isometric bijection between inner product spaces necessarily preserves the inner product.

Proposition 2.4. *Let $(X, \|\cdot\|)$ be a normed space satisfying the parallelogram law* (6). *Then the norm is induced from an inner product, which is given in terms of the norm by means of the* polarisation identity*:*

$$\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) \text{ when } \mathbb{F}=\mathbb{R},$$

and

$$\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) + \frac{1}{4}i(\|x+iy\|^2 - \|x-iy\|^2) \text{ when } \mathbb{F}=\mathbb{C}.$$

Definition 2.5. A linear vector space with an inner product is called an *inner product space*. If it is complete with the induced norm, it is called a *Hilbert space*.

Given an inner product space, one can complete it with respect to the induced norm.[17] Since the inner product is a continuous function on its factors, it can be extended to the completed space. The completed space is therefore a Hilbert space.

2.2 Examples

Example 2.6. 1. The space $\mathbb{C}^n$ or $\mathbb{R}^n$ is a Hilbert space with the standard inner product

$$\langle x,y\rangle = \sum_{k=1}^{n} x_k \bar{y}_k.$$

2. The space $\ell^2 = \{(x_1,x_2,\dots) = (x_n) : \sum_{n=1}^{\infty} |x_n|^2 < \infty\}$ is a Hilbert space with the inner product

$$\langle x,y\rangle = \sum_{n=1}^{\infty} x_k \bar{y}_k.$$

3. The space $C[0,1]$ of continuous functions on the interval $[0,1]$ is an incomplete inner product space with the inner product

$$\langle f,g\rangle = \int_0^1 f\,\bar{g}\,dx.$$

   You can see this as $C[0,1]$ is dense in $L^2([0,1])$, so can not be complete.

4. Let (Ω,μ) be a measure space, e.g. Ω is a subset of $\mathbb{R}^n$ and μ is the Lebesgue measure. The space $L^2(\Omega,\mu)$ of all complex-valued square integrable functions is a Hilbert space with the inner product

$$\langle f,g\rangle = \int_E f\,\bar{g}\,d\mu.$$

   The completeness of $L^2(E,\mu)$ is a special case of the Riesz-Fischer theorem on the completeness of the Lebesgue space $L^p(E,\mu)$.

5. A closed subspace of a Hilbert space is a Hilbert space.

6. Let $\mathbb{D}$ be the open unit disk in $\mathbb{C}$. The space $A^2(\mathbb{D})$ consists of all functions which are square integrable and holomorphic in $\mathbb{D}$ is a closed subspace of $L^2(\mathbb{D})$ and is thus a Hilbert space (known as Bergman space). You are asked to prove this on example sheet 2.[18]

[17] Right now the way we would complete a normed space is as per metric spaces: form the completion of the metric space and then extend both the addition, scalar multiplication and the norm by continuity to give the completion the structure of a Banach space. Fortunately there is a better way, which we might discuss at the end of the course.

[18] The Bergman space is an example of a *reproducing kernel Hilbert space*. Unlike the L^2 spaces, whose elements are equivalence classes of functions on a space, the elements of $A^2(\mathbb{D})$ are functions on $\mathbb{D}$ – elements are equal if and only if they agree exactly. Moreover you can recover the value of $f(z)$ from taking a suitable inner product; see Sheet 2.

7. The space $H^2(\mathbb{T})$ of all functions $f \in L^2(-\pi,\pi)$ whose Fourier series are of the form $\sum_{n\geq 0} a_n e^{inx}$ is a closed subspace of $L^2(-\pi,\pi)$ and is thus a Hilbert space. You will be able to see this by noting that the n-th Fourier coefficient of f is given by $\langle f, e_n\rangle$, where $e_n(x) = \frac{1}{2\pi i}e^{inx}$. In this way $H^2(\mathbb{T})$ is a countable intersection of closed sets, so closed. This space is known as Hardy space, and appears in applications to harmonic analysis.

Examples 2, 6 and 7, and 4 (provided $(\Omega,\mathscr{F},\mu)$ is small enough for $L^2(\Omega)$ to be separable – see a footnote to a deep dive in the previous section – and big enough so that $L^2(\Omega)$ is not finite dimensional) are all isometrically isomorphic. Indeed, as we will see in the next subsection there is a unique infinite dimensional separable Hilbert space. But nevertheless, the different presentations of these Hilbert spaces

2.3 Orthogonality

Definition 2.7. Two vectors x and y in an inner product space X are said to be orthogonal if $\langle x,y\rangle = 0$. For $Y \subset X$, define $Y^\perp$ as the space of all vectors $v \in X$ which are orthogonal to Y, i.e. $\langle v,y\rangle = 0$ for all $y \in Y$. When Y is a subspace of X, $Y^\perp$ is called the *orthogonal complement* of Y in X.

We shall see that a Hilbert space always decomposes as the direct sum of a closed subspace and its orthogonal complement, just as you are familiar with in finite dimensions. First we collect the properties of orthogonal complements that don't require completeness.

Proposition 2.8. *Let Y be a subset of an inner product space X. Then*

(i) $Y^\perp$ *is a closed subspace of* X.

(ii) $Y \subset Y^{\perp\perp}$.

(iii) If $Y \subset Z \subset X$, *then* $Z^\perp \subset Y^\perp$.

(iv) $(\overline{\operatorname{span} Y})^\perp = Y^\perp$.

(v) If Y and Z are subspaces of X such that $X = Y + Z$ and $Z \subset Y^\perp$, then $Y^\perp = Z$.

Proof. Most of this is left as an exercise / to be recalled from linear algebra. In (i), to see Y is closed suppose $x_n \in Y^\perp$ has $x_n \to x \in X$. Then for $y \in Y$, we have

$$0 = \langle x_n, y\rangle \to \langle x,y\rangle,$$

so $\langle x,y\rangle = 0$, and hence $x \in Y^\perp$. For (v), take $x \in Y^\perp$ and by hypothesis write $x = y+z$ with $y \in Y$ and $z \in Z$. Then, as $x \in Y^\perp$,

$$0 = \langle x,y\rangle = \langle y,y\rangle + \langle z,y\rangle = \|y\|^2,$$

since $z \in Y^\perp$. Therefore $y = 0$ and $x = z \in Z$, i.e. $Y^\perp \subseteq Z$. □

Our main goal in this section is the following theorem, which we will prove at a bit later:

Theorem 2.9 (Projection theorem). *If Y is a closed subspace of a Hilbert space $\mathscr{H}$, then Y and $Y^\perp$ are complementary subspaces: $\mathscr{H} = Y \oplus Y^\perp$, i.e. every $x \in \mathscr{H}$ can be decomposed uniquely as a sum of a vector in Y and in $Y^\perp$.*

In an inner product space context, we will reserve the $\oplus$ symbol for this orthogonal complementation, i.e. write $X = Y \oplus Z$ when Y, Z are subspaces with $Z = Y^\perp$ and $X + Y = X$.[19].

[19]This is compatible with our use of $\oplus$ for products in the previous section. If we take inner product spaces X, Y and equip $X \times Y$ with the inner product $\langle (x_1,y_1),(x_2,y_2)\rangle = \langle x_1,x_2\rangle + \langle x_2,y_2\rangle$ and identify X with the subspace $\{(x,0) : x \in X\}$ and Y with $\{(0,y) : y \in Y\}$ of $X \times Y$, then $X \times Y$ can be written as $X \oplus Y$.

Deep Dive

More generally, as per linear algebra, we subspaces Y, Z of a vector space are *complemented* if $Y \cap Z = \{0\}$ and $Y + Z = X$. Using the axiom of choice, every subspace of a vector space has a complementary subspace: Take a basis A_Y for Y,[a] and extend it to a basis A_X or X.[b] Then take $Z = \mathrm{Span}(A_X \setminus A_Y)$. Note how this proof is the same as the finite dimensional proof of taking a basis for Y and extending it to a basis for X. However, in a normed space context, we don't learn anything about Z. For example, if X is Banach, and Y is closed, when can one take a complementary subspace Z to be closed (so also a Banach space)?

For this reason in Banach space we say subspaces Y, Z are complemented when they are *closed subspaces* with $Y + Z = X$ and $Y \cap Z = \{0\}$. (In C4.1 this is called topologically complemented, to compare with the notion of algebraic complementation of the previous paragraph).

In a Hilbert space, the projection theorem shows that all closed subspaces are complemented. Strikingly this characterises Hilbert space.

Theorem (Lindenstrass and Tzafriri, 1971)**.** *Let X be a Banach space such that every closed subspace has a closed complement. Then there exists an equivalent norm on X under which it is a Hilbert space.*

[a] using Zorn's lemma to obtain a maximal linearly independent set, which is a basis; see B1.2

[b] using Zorn's lemma again to obtain a maximal linearly independent set containing B_Y.

Before proving the projection theorem, let us collect some consequences.

Corollary 2.10. *If Y is a closed subspace of a Hilbert space $\mathscr{H}$, then $Y = Y^{\perp\perp}$ (which is short hand for $(Y^\perp)^\perp$).*

Proof. We have $\mathscr{H} = Y \oplus Y^\perp = Y^\perp \oplus Y^{\perp\perp}$ from the projection theorem. So $Y \subseteq Y^{\perp\perp}$ with $\mathscr{H} = Y^\perp + Y$. The result follows from Proposition 2.8(v). $\square$

Definition 2.11. The *closed linear* span of a set S in a normed space X is the smallest closed linear subspace of X containing S, i.e. the intersection of all such subspaces. We write $\overline{\mathrm{Span}}(S)$ for this subspace, which is the closure of the span of S.[20]

Proposition 2.12. *Let S be a set in a Hilbert space $\mathscr{H}$. Then $\overline{\mathrm{Span}}(S) = S^{\perp\perp}$.*

Proof. Exercise. $\square$

To prove the projection theorem, we use the following geometrical result.

Theorem 2.13 (Closest point in a closed convex subset)**.** *Let K be a non-empty closed convex[21] subset of a Hilbert space $\mathscr{H}$. Then, for every $x \in X$, there is a unique point $k \in K$ which is closer to x than any other points of K, i.e. a unique $k \in K$ with*

$$\|x - k\| = \inf_{y \in K} \|x - y\|.$$

By translating (replace x by 0 and K by $\{k - x : k \in K\}$, the closest point theorem is equivalent to the statement that every non-empty closed convex subset of a Hilbert space has a unique element of minimal norm.

Proof. Let

$$d = \inf_{z \in K} \|x - z\| \geq 0$$

[20] Check that the closure of a subspace is a subspace, so that the closure of the span of S is a closed subspace containing S. Since any subspace containing S contains the span of S, any closed subspace containing S must contain the closure of the span of S.

[21] i.e. if $x, y \in K$ and $0 < \lambda < 1$, then $\lambda x + (1 - \lambda y) \in K$

and $y_n \in K$ be a minimizing sequence, i.e.

$$\lim_{n\to\infty} d_n = d, \qquad d_n = \|x - y_n\|.$$

Applying the parallelogram law (6) to $\frac{1}{2}(x-y_n)$ and $\frac{1}{2}(x-y_m)$ yields

$$\left\|x - \frac{1}{2}(y_n + y_m)\right\|^2 + \frac{1}{4}\|y_n - y_m\|^2 = \frac{1}{2}(d_n^2 + d_m^2).$$

Since K is convex, $\frac{1}{2}(y_n+y_m) \in K$ and so $\left\|x - \frac{1}{2}(y_n+y_m)\right\| \geq d$. This and the above implies that (y_n) is a Cauchy sequence. Let y be the limit of this sequence, which belongs to K as K is closed. We then have by the continuity of the norm that $\|x-y\| = \lim \|x-y_n\| = d$, i.e. y minimises the distance from x.

That y is the unique minimiser follows from the same reasoning above. If y' is also a minimiser, we apply the parallelogram law to $\frac{1}{2}(x-y)$ and $\frac{1}{2}(x-y')$ to obtain

$$d^2 + \frac{1}{4}\|y-y'\|^2 \leq \left\|x - \frac{1}{2}(y+y')\right\|^2 + \frac{1}{4}\|y-y'\|^2 = \frac{1}{2}(\|x-y\|^2 + \|x-y'\|^2) = d^2.$$

This implies that $y = y'$. □

Deep Dive

The closest point theorem also holds for some, but not all, other Banach spaces. As you can have a go at on Sheet 2 (C.2), if the unit ball of a Banach space is *uniformly convex* (think of as 'round enough') then the closest point theorem holds. In particular it is valid for ℓ^p and L^p for $1 < p < \infty$. But the uniqueness portion of the closest point theorem fails for ℓ^1 and ℓ^∞ even in two dimensions.[a] We will see an example of a closed convex subset of a Banach (in fact an affine subspace, i.e. a translation of a subspace) which does not have an element of minimal norm on a problem sheet.

[a] Existence of a closest point in to a closed convex set in finite dimensions is a consequence of compactness; see Section 4.

Proof of the Projection Theorem. Certainly $Y \cap Y^\perp = \{0\}$. It remains to show that $X = Y + Y^\perp$.

Take any $x \in X$ and, since Y is a non-empty closed convex subset of X, there is a point $y_0 \in Y$ which is closer to x than any other points of Y by Theorem 2.13. To conclude, we show that $x - y_0 \in Y^\perp$.[22] Indeed, for all $y \in Y$ and $t \in \mathbb{R}$, we have

$$\|x-y_0\|^2 \leq \|x - \underbrace{(y_0 - ty)}_{\in Y}\|^2 = \|x-y_0\|^2 + 2t\,\mathrm{Re}\,\langle x-y_0, y\rangle + t^2\,\|y\|^2.$$

It follows that $2t\,\mathrm{Re}\,\langle x-y_0,y\rangle + t^2\,\|y\|^2 \geq 0$ for all $t \in \mathbb{R}$. This implies $\mathrm{Re}\,\langle x-y_0,y\rangle = 0$. This concludes the proof if the scalar field is real.

If the scalar field is complex, we proceed as before with t replaced by it to show that $\mathrm{Im}\,\langle x-y_0,y\rangle = 0$. □

[22] We can see this geometrically for $\mathbb{F} = \mathbb{R}$. Fix some $y \in Y$, and consider the plane $\mathrm{Span}(y, y_0)$. The closest point of x to $\mathrm{Span}(y_0, y)$ is still y_0; but we know from 3-dimensional geometry that the closest point of x to this plane is giving by dropping the perpendicular of x to the plane: hence $x - y_0$ is orthogonal to y.

2.4 Orthonormal bases

Definition 2.14. A subset S of a Hilbert space X is called an orthonormal set if $\|x\| = 1$ for all $x \in S$ and $\langle x, y\rangle = 0$ for all $x \neq y \in S$.

S is called an orthonormal basis (or a complete orthonormal set) for X if S is an orthonormal set and its closed linear span is X.

Theorem 2.15. *Every Hilbert space contains an orthonormal basis.*

Proof. The proof is only examinable when the Hilbert space $\mathscr{H}$ is *separable*, i.e. contains a countable dense subset S. In this case label the elements of S as $y_1, y_2, \dots$. Applying the Gram-Schmidt process[23] we obtain an orthonormal set $B = \{e_1, e_2, \dots\}$ (which might terminate after some finite stage) such that, for every n, the span of $\{e_1, \dots, e_n\}$ contains $y_1, \dots, y_n$. As $\bar{S} = X$, this implies that $X = \overline{\operatorname{span} B}$, and so X is the closed linear span of B.

Deep Dive

In general we need the axiom of choice – in fact the statement that every Hilbert space has an orthonormal basis is equivalent to the axiom of choice – in the equivalent form of *Zorn's Lemma*. Zorn's Lemma will be described in B1.2 (Set Theory), and shown to be equivalent to the axiom of choice there. It allows us to produce sets which are maximal with respect to certain properties.[a] Let S be a maximal orthonormal set[b] in $\mathscr{H}$. If $\overline{\mathrm{Span}}(S) \neq \mathscr{H}$, then this is a proper closed subset of $\mathscr{H}$, so by the projection theorem, there exists $x \in \mathscr{H}$ orthogonal to $\overline{\mathrm{Span}}(S)$, which we can normalise to have $\|x\| = 1$. Then $S \cup \{x\}$ is orthonormal, contradicting maximality of S. Hence $\overline{\mathrm{Span}}(S) = \mathscr{H}$ and S is an orthonormal basis for $\mathscr{H}$.

[a] Precisely: Given a non-empty partially ordered set $\mathscr{P}$ with the property that every chain $\mathscr{C}$ (i.e. a collection $\mathscr{C} \subset \mathscr{P}$ with the property that for all $x, y \in \mathscr{C}$ either $x \leq y$ or $y \leq x$) has an upper bound (i.e. there exists $z \in \mathscr{P}$ with $x \leq z$ for all $x \in \mathscr{C}$. Then Zorn's Lemma ensures that $\mathscr{P}$ has a maximal element, i.e. some $z \in \mathscr{P}$ with $z \geq x$ for all $x \in \mathscr{P}$. (In B1.2 this will be set out when $\mathscr{P}$ is a collection of sets ordered by inclusion satisfying this property, rather than using the language or partially ordered sets.)

[b] If you do B1.2 it's a good exercise in using Zorn to show this exists

$\square$

Given a finite orthonoromal set $e_1, \dots, e_n$ in an inner product space X, we can always decompose $x \in X$ as

$$x = \sum_{r=1}^{n} \langle x, e_r\rangle e_r + (x - \sum_{r=1}^{n} \langle x, e_r\rangle e_r),$$

where the first term lies in $\mathrm{Span}(e_1, \dots, e_n)$ and the second lies in $\mathrm{Span}(e_1, \dots, e_n)^{\perp}$. In this way $X = \mathrm{Span}(e_1, \dots, e_n) \oplus \mathrm{Span}(e_1, \dots, e_n)^{\perp}$.[24] The element $\sum_{r=1}^{n} \langle x, e_r\rangle e_r$ is the unique closest point in $\mathrm{Span}(e_1, \dots, e_n)$ to x (we don't need completeness of X for this as $\mathrm{Span}(e_1, \dots, e_n)$ is finite dimensional. The following is a consequence of Pythagoras:

Proposition 2.16 (Pythagorean theorem)**.** *Let X be an inner product space and $S = \{x_1, x_2, \dots, x_m\}$ be a finite orthonormal set in X. For every $x \in X$, there holds*

$$\|x\|^2 = \sum_{n=1}^{m} |\langle x, x_n\rangle|^2 + \Big\|x - \sum_{n=1}^{m} \langle x, x_n\rangle x_n\Big\|^2.$$

[23] The Gram-Schmidt process is usually applied to a set of finitely many linearly independent vectors yielding an orthogonal basis of the same cardinality. In our setting, we will lose the latter property as the vectors y_i's are not necessarily linearly independent.

[24] We will later see that more generally, all finite dimensional subspaces of normed spaces have closed complements as a consequence of the Hahn–Banach theorem

Corollary 2.17 (Bessel's inequality)**.** *Let X be a Hilbert space and $S = \{x_1, x_2, \dots\}$ be an orthonormal sequence in X. Then, for every $x \in X$, there holds*

$$\sum_{n=1}^{\infty} |\langle x, x_n\rangle|^2 \leq \|x\|^2.$$

We can characterise when an orthonormal sequence forms a basis in terms of always having equality in Bessel's inequality (this is known as Parseval's identity). The proof is strictly speaking only examinable in B4.2, but we've done all the work, so let's give it here.

Theorem 2.18 (Characterising bases)**.** *Let $\mathscr{H}$ be a Hilbert space and $S = \{e_1, e_2, \dots\}$ be an orthonormal sequence in $\mathscr{H}$. Then the following are equivalent:*

1. *S is an orthonormal basis for $\mathscr{H}$;*
2. *$\|x\|^2 = \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2$ for all $x \in \mathscr{H}$ (i.e. Parseval's identity holds)*
3. *$x = \sum_{n=1}^{\infty} \langle x, e_n\rangle e_n$ for all $x \in \mathscr{H}$;*
4. *$\langle x, y\rangle = \sum_{n=1}^{\infty} \langle x, e_n\rangle \overline{\langle y, e_n\rangle}$, for all $x, y \in \mathscr{H}$.*

In this case the map $\mathscr{H} \to \ell^2$ given by $x \mapsto (\langle x, e_n\rangle)_{n=1}^{\infty}$ is an isometric isomorphism.

Proof. 1$\Rightarrow$3: Note that $\sum_{r=1}^{n} \langle x, e_r\rangle e_r$ is the closest point in $\mathrm{Span}(e_1, \dots, e_n)$ to x (as $x - \sum_{r=1}^{n} \langle x, e_r\rangle$ is orthogonal to $e_1, \dots, e_n$). Since $x \in \mathrm{Span}(e_1, e_2, \dots)$, it follows that

$$\|x - \sum_{r=1}^{n} \langle x, e_r\rangle e_r = d(x, \mathrm{Span}(e_1, \dots, e_n)) \to 0,$$

proving 3. 3 $\Longrightarrow$ 4 is obtained from computing the inner product $\left\langle \sum_{r=1}^{n} \langle x, e_r\rangle e_r, \sum_{s=1}^{n} \langle y, e_s\rangle e_s \right\rangle$ and using continuity of the inner product. 4 $\Longrightarrow$ 2 folows from the definition of the norm in terms of the inner product. Finally for 2 $\Longrightarrow$ 1, if 2 holds, then $\|x - \sum_{r=1}^{n} \langle x, e_r\rangle e_r\| \to 0$ as $n \to \infty$ (by Proposition 2.16, giving 1.

For the last part, condition 2 ensures we have defined an isometric linear map. For surjectivity, given $(\alpha_n) \in \ell^2$, the series $\sum_{n=1}^{\infty} \alpha_n e_n$ is absolutely convergent so converges to x in $\mathscr{H}$, which is then mapped onto (α_n). □

Deep Dive

More generally you can check that if S is an orthonormal basis for a Hilbert space $\mathscr{H}$, then $\mathscr{H}$ is isometrically isomorphic to

$$\ell^2(S) := \{f : S \to \mathbb{F} : \sum_{s \in S} |f(s)|^2 < \infty\},$$

(which is given the inner product you would expect). Here the sum of positive elements over this (potentially uncountable) set is given by

$$\sum_{s \in S} |f(s)|^2 = \sup\{\sum_{s \in F} |f(s)|^2 : F \subset S \text{ is finite}\}$$

(which is exactly the definition you would get from taking the Lebesgue integral on S with counting measure).

Deep Dive

While in a Hilbert space the characterisation above allows us to define a basis as an orthonormal set which has dense linear span, and we learn that every x can be written (uniquely) as $\sum_{r=1}^{n} \alpha_n e_n$ (with convergence in $\mathscr{H}$) this is a feature of Hilbert space. We have to be more careful with the definition of a basis in a Banach space. For $X = C([0,1])$, the sequence $1, x, x^2, \dots$ is linearly independent, and has dense linear span (as we will see in Section 5). But it is not true that every $f \in C([0,1])$ can be written as a convergent series $\sum_{n=1}^{\infty} \alpha_n x^n$ (with convergence in $C([0,1])$, with it's canonical sup-norm). Functions which can be written in this way are infinitely differentiable. Therefore $1, x, x^2, \dots$ does not form a *Schauder basis* for $C([0,1])$. This space does have a Schauder basis: a sequence $f_1, \dots$ such that every $f \in C([0,1])$ can be written uniquely as $\sum_{n=1}^{\infty} \alpha_n f_n$ for some unique scalars α_n, and further this basis can be taken to consist of polynomial functions. But the firs thing that might come to mind doesn't work. This will be explored further in C4.1; see also the books by Corothers and by Albiac and Kalton.

3 Bounded linear operators between normed vector spaces

Whenever we introduce a class of mathematical objects it is always important to understand the appropriate maps between these objects. In the setting of vector spaces, we study linear maps. In the setting of metric spaces we look at continuous maps, or perhaps contractive, or even isometric maps. For our normed spaces the right maps to consider are the continuous linear maps (as well as contractive and isometric linear maps).

3.1 Boundedness and continuity

Recall that a map $T : V \to W$ between vector spaces is linear if $T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$ for all $x, y \in X$ and scalars $\lambda, \mu \in \mathbb{F}$. Continuity of a map $T : X \to Y$ is a local property: T is continuous if and only if it is continuous at x for all $x \in X$. But for a linear map, we can use linearity to translate continuity at one point to continuity at all other points, so we only need to check continuity at 0. This leads to the following important proposition.

Proposition 3.1. *Let $T : X \to Y$ be a linear map betwen normed spaces. The following are continuous:*

(i) *T is Lipschitz continuous,*

(ii) *T is continuous,*

(iii) *T is continuous at* 0*,*

(iv) *there exists $K > 0$ such that $\|T(x)\| \leq K\|x\|$ for all $x \in X$.*

Proof. $(i) \Rightarrow (ii) \Rightarrow (iii)$ are immediate. Suppose that T is continuous at 0. Then there is some $\delta > 0$ such that

$$\|Tx\| = \|Tx - T0\| \leq 1 \text{ for } \|x\| = \delta.$$

It follows that, for any $x \neq 0$,

$$\|Tx\| = \frac{\|x\|}{\delta} T\left(\frac{\delta x}{\|x\|}\right) \leq \frac{\|x\|}{\delta}.$$

Clearly, this continues to holds for $x = 0$ and we can take $K = \frac{1}{\delta}$ in condition (iv).

Finally assume (iv) holds, so let $K > 0$ have $\|T(x)\| \leq K\|x\|$ for all $x \in X$¿ Now we use linearity, to get

$$\|T(x) - T(y)\| = \|T(x - y)\| \leq K\|x - y\|,$$

for all $x, y \in X$. That is T is Lipschitz continuous (with Lipschitz constant at most K). $\square$

The last condition is often the most useful, both for establishing continuity

Definition 3.2. Let X and Y be normed spaces (aways assumed to be over the same field $\mathbb{F}$). Then we say that $T : X \to Y$ is a *bounded linear operator* if T is linear and there exists $K > 0$ so that

$$\|T(x)\|_Y \leq K\|x\|_X \text{ for all } x \in X. \tag{7}$$

Define the *operator norm* of a bounded linear operator $T : X \to Y$ by

$$\|T\| = \inf\{K > 0 : \|T(x)\| \leq K\|x\| \text{ for all } x \in X\}.$$

Write $\mathscr{B}(X,Y)$ for the collection of all bounded linear operators from X to Y.

Warning. T being a bounded linear operator does not mean that $T(X) \subset Y$ is bounded. Indeed, the only linear operator with a bounded image is the trivial operator that maps each $x \in X$ to $T(x) = 0$.

We will often abbreviate the space $\mathscr{B}(X,X)$ of bounded linear operators from a normed space X to itself by $\mathscr{B}(X)$.[25] An important special case is the 'bounded linear functionals', i.e. bounded linear functions from a normed vector space to the corresponding field $\mathbb{F} = \mathbb{R}$ (respectively $\mathbb{F} = \mathbb{C}$ for complex vector spaces) and this so called dual space $X^* := \mathscr{B}(X,\mathbb{F})$ will be discussed in far more detail in chapters 7 and **??**.

Note that the infiumum in the definition of the operator norm is attained, i.e. for a bounded linear operator $T : X \to Y$, we have[26]

$$\boxed{\|T(x)\| \leq \|T\|\|x\| \text{ for all } x \in X.}$$

The set of continuous linear maps between normed spaces is a vector space (by AOL). We check that the operator norm gives $\mathscr{B}(X,Y)$ the structure of a normed space. Needless to say, we shall later be interested in when this is complete. Spoiler alert: $\mathscr{B}(X,Y)$ is complete if and only if Y is complete.

Proposition 3.3. *Let X and Y be normed spaces. Then $\|\cdot\|$ is a norm on $\mathscr{B}(X,Y)$. Also, for $T \in \mathscr{B}(X,Y)$, we have (except in the case when $X = \{0\}$)*

$$\|T\|_{\mathscr{B}(X,Y)} = \sup_{x \in X, x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{x \in X, \|x\|=1} \|Tx\| = \sup_{x \in X, \|x\| \leq 1} \|Tx\|.$$

Proof. Note that $\|T\| \geq 0$, and if $\|T\| = 0$, then we have $T(x) = 0$ for all x (by positivity of the norm on Y). Hence $T = 0$.

Let $K = \sup\{\|T(x)\| : \|x\| \leq 1,\ x \in X\}$. Then for $x \neq 0$,

$$\|T(x)\| = \left\| T\left(\frac{x}{\|x\|}\right) \right\| \|x\| \leq K\|x\|,$$

and so $\|T\| \leq K$. But for $x \in X$ with $\|x\| \leq 1$, we have $\|T(x)\| \leq \|T\|$. Taking the supremum over all such x we get $K \leq \|T\|$.

Using this characterisation of the norm, we get

$$\|(\lambda T)\| = \sup\{\|(\lambda T)(x)\| : \|x\| \leq 1\} = |\lambda| \sup\{\|T(x)\| : \|x\| \leq 1\} = |\lambda|\|T\|,$$

[25] In some texts, $\mathscr{B}(X,Y)$ is also denoted as $\mathscr{L}(X,Y)$.

[26] Take a sequence (K_n) satisfying (7) with $K_n \to \|T\|$ and use limits preserve weak inequalities.

for $T \in \mathscr{B}(X,Y)$ and $\lambda \in \mathbb{F}$.

Finally for $S,T \in \mathscr{B}(X,Y)$, and $x \in X$, the triangle inequality (in Y) gives

$$\|(S+T)(x)\| \leq \|S(x)\| + \|T(x)\| \leq (\|S\| + \|T\|)\|x\|,$$

so that $\|S+T\| \leq \|S\| + \|T\|$. $\square$

Remark. Note that if Y is an inner product space, then we have

$$\|T\|_{\mathscr{B}(X,Y)} = \sup\{|\langle Tx, y\rangle| : x \in X, y \in Y, \|x\|_X = \|y\|_Y = 1\}.$$

This is a consequence of (i) and the fact that $\|Tx\|_Y = \sup_{y \in Y, \|y\|_Y = 1} |\langle Tx, y\rangle|$.

Warning. For general bounded linear operators, one cannot expect that there exists $x \in X$ so that $\|Tx\| = M\|x\|$, i.e. the supremum $\sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}$ is in general not achieved. Some examples can be found on the problem sheets.

Deep Dive

In the case of bounded linear functionals $f : X \to \mathbb{F}$, whether the supreumum $\|f\| = \sup\{|f(x)| : \|x\| \leq 1\}$ is attained is related to the geometry of the unit ball – a theme you may be getting used to. You'll quickly be able to see using the Riesz representation theorem in Section **??** that this supremum is attained whenever X is a Hilbert space, but actually it works for any uniformly convex Banach space. If you solved Exercise 2.C.2 then you should also be able to prove that if X is a uniformly convex Banach space and $f \in X^*$, then there exists a unique $x \in X$ with $\|x\| = 1$ satisfying $f(x) = \|x\|$. In particular bounded linear functionals on L^p attain their norms for $1 < p < \infty$. (This is a result that we'll also be able to see directly later in the course when we determine the general form of a bounded linear functional on L^p).

The question of which Banach spaces X have the property that all bounded linear functionals attain their norms has a very interesting answer: a theorem of James shows that this characterises reflexivity of X. More on reflexivity in Section **??**, and James' theorem will return in further deep dives.

We note that for any $T \in \mathscr{B}(X,Y)$ both the kernel $\ker(T) := \{x \in X : T(x) = 0\}$ of T and its image $TX =: \{Tx : x \in X\}$ are subspaces (of X respectively Y), but that while $\ker(T)$ is always closed, as it can be viewed as the preimage of the closed set $\{0\}$ under a continuous operator, the image TX is in general not closed.

3.2 Examples

In order to prove that a map $T : X \to Y$ is a bounded linear operator we need to:

(1) Potentially check that T does map into Y, i.e. $Tx \in Y$ for all $x \in X$;[27]

(2) Verify that T is linear;

(3) Find some M so that for all $x \in X$
$$\|Tx\|_Y \leq M\|x\|_X.$$

[27] A well posed question should be clear whether or not you can assume that the map specified does take values in Y, or whether you are expected to prove this. But in your own work if you write down a map, do make sure you check that it does take values where you say it does!

Whether (1) is needed will depend on context - is there is a discussion to be had about whether $T(x) \in Y$. In cases where the codomain Y is a space like ℓ^p or L^p, it may be necessary to bound some sum or integral to do this. In that case, most likely the bound you get will directly feed into the proof of (3), and it is worth doing these at the same time. See for example the multiplication by functions on $L^2([0,1])$ example below. In most examples we encounter, linearity will be routine; it will typically be enough just to note briefly why the map is linear[28] but a proof in the spirit of Prelims Linear Algebra does not need to be given unless there is good reason to.

Now let us give some examples.

Shift operators For $1 \le p \le \infty$, define the shift operators $L, R : \ell^p \to \ell^p$ by

$$R((x_1, x_2, x_3, \dots)) := (0, x_1, x_2, x_3, \dots) \text{ and } L((x_1, x_2, x_3, \dots)) := (x_2, x_3, x_4, \dots)$$

Here L and R are certainly linear and map ℓ^p into ℓ^p. The map R is isometric, i.e. $\|R(x)\| = \|x\|$ for all $x \in X$, and so certainly bounded with $\|R\| = 1$. The map L is not isometric, as it has kernel $\{(x_n) \in \ell^p : x_1 = 0\}$. But L is bounded as

$$\|L(x)\| = \sum_{n=2}^{\infty} |x_n| \le \sum_{n=1}^{\infty} |x_n| = \|x\|, \quad x = (x_n) \in \ell^p$$

which shows $\|L\| = 1$. Taking $x = (0,1,0,0,\dots)$, we have $\|x\| = 1$ and $\|L(x)\| = 1$ so $\|L\| = 1$.

As we will see when we look at dual operators in the last section of the course, for $1 \le p < \infty$ the left shift on ℓ^p is related to the right shift on ℓ^q, where q is the Hölder conjugate of p. You'll be able to exploit this in B4.2, when you compute specta of operators.

You can equally look at shift operators on c_0 with analogous results. Expect a nice relationship between the left and right shift operators on c_0 and the right and left shift operators on ℓ^1.

Co-ordinate projections On ℓ^p, define the co-ordinate projection $\mathrm{ev}_n : \ell^p \to \mathbb{F}$ by $\mathrm{ev}_n(x) = x_n$, i.e. the map which evaluates the sequence in the n-th position. Then ev_n is linear and bounded with $\|\mathrm{ev}_n\| = 1$.[29]

Multiplication by functions on $C[0,1]$ Let $X = C([0,1])$, as always equipped with the supremum norm and let $g \in C^0([0,1])$. Then define $M_g : C([0,1]) \to C([0,1])$ by $M_g(f) = fg$. This does map into $C([0,1])$ as the pointwise product of continuous functions is continuous. Certainly M_g is linear, and for $f \in C([0,1])$, we have

$$\|M_g(f)\|_\infty = \sup_{t \in [0,1]} |f(t)g(t)| \le \|f\|_\infty \|g\|_\infty,$$

so that M_g is bounded and $\|M_g\| \le 1$. Taking $f \in C([0,1])$ to be $f(t) = 1$ for all t, we have $\|f\| = 1$ and $M_g(f) = g$ so $\|M_g\| \ge \|g\|_\infty$ and hence $\|M_g\| = \|g\|_\infty$.[30]

Note that there was nothing special about $[0,1]$ here. The same works for $C(K)$ where K is any compact metric space (or compact Hausdorff topological space).

[28] For example a sentence like 'T is linear as integration is linear'

[29] As $|x_n| \le \|x\|_p$, while for the standard element $e_n \in \ell^p$ we have $\|e_n\| = 1$ and $|\mathrm{ev}_n(e_n)| = 1$.

[30] This f is the constant function 1, so I would normally write it as $1 \in C([0,1])$, the function with $1(t) = 1$ for all t (where the 1 on the right hand side lies in $\mathbb{F}$). This notation is useful as $C([0,1])$ is not just a Banach space; it is also an *algebra* with the additional multiplication given by the pointwise multiplication. The constant function 1 is the identity for this multiplication: $1g = g$ for all $g \in C([0,1])$. This is what we used in the calculation above. More on *Banach algebras* in some deep dives.

Multiplication by functions on $L^2([0,1])$) Consider instead $g \in L^\infty([0,1])$ and let $X = L^2([0,1])$ (equipped of course with the L^2 norm). Then we can define a map $M_g : X \to X$ by $M_g(f) = fg$. This time we should note that it is the case that $fg \in L^2([0,1])$ when $f \in L^2([0,1])$. Recalling that for $g \in L^\infty([0,1])$, we have $|g(t)| \leq \|g\|_\infty$ almost everywhere, we get the estimate

$$\int_0^1 |(M_g f)(t)|^2 dt = \int_0^1 |f(t)|^2 |g(t)|^2 dt \leq \|g\|_\infty^2 \int |f(t)|^2 dt$$

which both shows $fg \in L^2([0,1])$ and gives the estimate

$$\|M_g f\|_{L^2} \leq \|g\|_{L^\infty} \|f\|_{L^2} \text{ for all } f \in X.$$

By linearity of the integral[31], M_g is linear. Putting all this together, M_g is a bounded linear map from $L^2([0,1])$ to $L^2([0,1])$ and $\|M_g\| \leq \|g\|_{L^\infty}$.

Again we have $\|M_g\| = \|g\|_\infty$. To see this, fix $C < \|g\|_\infty$ (if $\|g\|_\infty = 0$, then $g = 0$ a.e. and hence $M_g(f) = 0$ a.e., and $M_g = 0$). By definition of $\|g\|_\infty$ the set $\Omega_C = \{t \in [0,1] : |g(t)| > C\}$ (which is measurable) has positive measure. Let χ_{Ω_C} denote its indicator function, which lies in $L^2([0,1])$. Then

$$\|M_g \chi_{\Omega_C}\|^2 = \int_{\Omega_C} |g(x)|^2 \geq C^2 \int_{\Omega_C} = C^2 \|\chi_{\Omega_C}\|^2$$

Accordingly $\|M_g\| \geq C$. Since $C < \|g\|_\infty$ was arbitrary $\|M_g\| \geq \|g\|_\infty$.

At the same time one can show that for $g(t) = t$, and any $f \in L^2([0,1])$

$$\|Tf\|_{L^2} < \|f\|_{L^2}$$

(this proof is a nice exercise related to the part A course in integration) so this gives an example of an operator for which the supremum $\sup_{f \neq 0} \frac{\|Tf\|}{\|f\|}$ is not attained for any element of the Banach space $X = L^2([0,1])$.

Deep Dive

Planting seeds for the spectrum of an operator in B4.2, this multiplication operator M_g (for $g(t) = t$) is a, or perhaps the, classic example of a bounded operator on $L^2([0,1])$ with no eigenvalues; yet the spectrum of M_g — those $\lambda \in \mathbb{F}$ for which $M_g - \lambda I$ is not invertible is non-empty. In fact the spectrum in this case is $[0,1]$.

There is a converse to the previous result: If a measurable g is such that $fg \in L^2([0,1])$ for all $f \in L^2([0,1])$, then g is an element of $L^\infty([0,1])$. This is a consequence of the Closed graph theorem, which will be treated in B4.2 Functional Analysis 2. As a consequence of this fact you can show that if $T \in \mathscr{B}(L^2([0,1])$ has $TM_g = M_g T$ for all $g \in L^\infty([0,1])$, then there exists $h \in L^\infty([0,1])$ such that $T = M_h$.

Linear maps between Euclidean Spaces We know that any linear map $T : \mathbb{C}^n \to \mathbb{C}^m$ can be written as

$$Tx = Ax \text{ for some } A \in M_{m \times n}(\mathbb{C}).$$

For the purpose of discussing the operator norm of T, we will equip $\mathbb{C}^n$ with the Euclidean ℓ^2-norm in this section. Certainly from the formula giving matrix multiplication, T is continuous, so bounded.

There are several different norms on the space of matrices, including the analogues of the p-norms on $\mathbb{R}^n$. One that can be useful is the analogue of the Euclidean norm (i.e. of the case $p = 2$) given by

$$\|A\|_2 := \Big(\sum_{i,j} |a_{ij}|^2\Big)^{1/2}$$

[31] Which takes work in the integration course, but we now just quote

which is also called the Frobenius norm or the Hilbert-Schmidt norm and is widely used in Numerical Analysis. A useful property of this norm is that it gives a simple way of obtaining an upper bound on the operator norm of the corresponding map $T:\mathbb{C}^n\to\mathbb{C}^m$ as follows: for $x\in\mathbb{C}^n$ the Cauchy-Schwartz inequality gives

$$\|Tx\|^2=\sum_{i=1}^m (Ax)_i^2=\sum_{i=1}^m\Big(\sum_{j=1}^n a_{ij}x_j\Big)^2\leq\sum_{i=1}^m\Big(\sum_{j=1}^n a_{ij}^2\Big)\cdot\Big(\sum_{j=1}^n x_j^2\Big)=\|A\|_2^2\|x\|^2.$$

Therefore $\|T\|\leq\|A\|_2$. However, for most matrices we have $\|T\|<\|A\|$. For example the identity operator on $\mathbb{C}^2$ certainly has $\|I\|=1$, but the Hilbert-Schmidt norm of the associated matrix is 2. In the case when $m=n$, we can make some progress by diagonalisation. If $A=A^*$, i.e. A is hermitian (equal to its own conjugate transpose),[32] then you can diagonalise A, by finding an orthonormal basis of eigenvectors for A (and hence T). It is then straightforward to see that

$$\|T\|=\max\{|\lambda_1|,\dots.|\lambda_n|\},\quad \lambda_i \text{ the eigenvalues of } A.$$

In general A need not be hermitian, but A^*A always will be. We have that

$$\|T\|=\max\{|\lambda_1|^{1/2},\dots.|\lambda_n|^{1/2}\},\quad \lambda_i \text{ the eigenvalues of } A^*A.$$

We could do this now – have a go – but we will see it right at the end of the course as a consequence of the *C^*-identity* for bounded operators on a Hilbert space.

Integral operator on $C([0,1])$: Let $X=C([0,1])$ as always be equipped with the sup-norm. Given any $k\in C([0,1]\times[0,1])$ we map each $x\in X$ to the function $Tx:[0,1]\to\mathbb{F}$ that is given by

$$Tx(s):=\int_0^1 k(s,t)x(t)\mathrm{d}t$$

where the integral is well defined as the integrand is bounded (by $\|k\|_\infty\|x\|_\infty$), and $Tx\in C[0,1]$ by, for example, the continuous parameter DCT.[33] The function k is often called an *integral kernel* or a kernel (which is unfortunate as it has nothing to do with the meaning of the word kernel in the context of the kernel of a linear map or homomorphism). Think of Tx as being given by a continuous version of matrix multiplication over the interval.

Then T is linear (as integration is linear) and for any $s\in[0,1]$ we can bound

$$|Tx(s)|\leq\int_0^1|k(s,t)x(s)|\mathrm{d}s\leq\|k\|_\infty\|x\|_\infty.$$

Therefore T is a bounded linear operator on $C([0,1])$ with $\|T\|\leq\|k\|_\infty$.

[32] we will have much more to say about the adjoint operation in the last section of the course, both for operators on Hilbert spaces and the dual of an operator between Banach spaces.

[33] Here's the proof. Given $s_0\in[0,1]$ and any sequence $s_n\to s_0$, we need to show $Tx(s_n)\to Tx(s_0)$. To this end we set $f_n(t):=k(s_n,t)x(t)$ and $f(t):=k(s_0,t)x(t)$ and observe that

- $f_n(t)\to f(t)$ for every $t\in[0,1]$, so in particular $f_n\to f$ a.e.
- $|f_n|\leq g$ on $[0,1]$ for the constant function $g:=\|k\|_\infty\|x\|_\infty$ which is of course integrable over the interval $[0,3]$.

Hence, by the dominated convergence theorem of Lebesgue, we have that

$$\lim_{n\to\infty}(Tx)(s_n)=\lim_{n\to\infty}\int_0^1 f_n(t)\mathrm{d}t=\int_0^1\lim_{n\to\infty}f_n(t)\mathrm{d}t=\int_0^1 f(t)dt=(Tx)(s_0)$$

as claimed. In this case we could get away with a Riemann integral argument as everything in sight is continuous on closed and bounded sets, so you can deduce continuity of Tx usnig uniform continuity of the continuous function k on the compact space $[0,1]\times[0,1]$.

An unbounded operator To show that a proposed linear operator $T : X \to Y$ is unbounded, you'll want to find a bounded sequence (x_n) (typically all of norm 1) such that $(T(x_n))$ is unbounded. Here is an example of an unbounded linear functional. Let X be the set of polynomial functions on $[0,1]$ equipped with the sup norm, and let $T : X \to \mathbb{C}$ be given by $T(p) = p'(1)$. Then T is unbounded. Indeed, the polynomial $p_n(t) = t^n$ has $\|p_n\|_\infty = 1$ for all n (as we work over $[0,1]$) but $(Tp_n) = n \to \infty$.

Deep Dive

Perhaps some of you are complaining that the space of polynomials X above is not a Banach space? What is an example of an unbounded linear functional $X \to \mathbb{F}$ when X is Banach? Or indeed an unbounded linear map $X \to Y$ where X is Banach and Y is a normed space?

Firstly, it follows from the axiom of choice that every infinite dimensional normed space X admits an unbounded linear functional. The idea, which will be given in C4.1, is to take an infinite linearly independent set of vectors (x_n) each of norm 1, and extend this arbitrarily to a Hamel basis (using Zorn's lemma). Then you can define a functional $T : X \to \mathbb{F}$ by sending each x_n to n and sending other basis elements to 0. This is a linear map, as linear maps are uniquely determined by their behaviour on a Hamel basis, in just the same way as in prelims linear algebra, and by construction $\|T(x_n)\| \geq n \to \infty$ with $\|x_n\| = 1$ so T is unbounded.

It is possible to find models of ZF without AC for which every linear map from a Banach space to a normed space is bounded.[a] However my take is that the axiom of choice is a true statement when we're studying functional analysis! So what this means is that you won't be writting down any everywhere defined unbounded linear maps on a Banach space any time soon: every linear map $T : X \to Y$ you explicitly construct on a Banach space X is going to be bounded. But beware, that means that you have to define your operator on all elements of the domain, and it must map into a normed space Y, i.e. $T(x) \in Y$ for all $x \in X$. There are many interesting examples of 'densely defined' unbounded operators (and a very interesting theory crucial to formalising quantum mechanics, which we can start to build once we have the closed graph theorem for bounded operators).

Finally, be in no doubt that you still need to prove your operators are bounded directly. While it's useful to know that without the axiom of choice, it's possible for all everwhere defined linear operators to be bounded, appealing to this deep dive isn't a valid way to proceed in an exam!

[a] Garnir's paper 'Solovay's axiom and Functional Analysis, Springer Lecture Notes in Mathematics, 399, 189-204, 1974' shows that this holds for a model with dependent choice and the hypothesis that every set of reals is Lebesgue measurable.

Projections onto complemented subspaces

Deep Dive

Given a Banach space X, recall that a closed subspace $Y \subset X$ is called complemented when there exists a *closed* subspace Z such that $Y + Z = X$ and $Y \cap Z = \{0\}$ (so that, using the Banach isomorphism theorem from B4.2, X is isomorphic as a Banach space to the product $Y \times Z$). This can be characterised using operators: a closed subspace Y is complemented in X if and only if there exists $P \in \mathscr{B}(X)$ with $P^2 = P$ and $P(X) = Y$, i.e. Y is the range of a bounded idempotent (also called a projection). Given such a Z the map P is given by $P(y+z) = y$ (which is well defined). The point is that this is bounded (which we get as a consequence of the Banach isomorphism theorem). In the reverse direction, given such a P, one can take $Z = (I-P)(X)$ (and check this is a closed subspace which complements Y). More on this in C4.1.

In the Hilbert space setting when we use the term *projection* we typically mean the orthogonal projection onto a closed subspace: given a closed subspace $Y \subset \mathscr{H}$, the orthogonal projection onto Y is the projection

$P \in \mathscr{B}(\mathscr{H})$ onto Y corresponding to the decomposition $\mathscr{H} = Y \oplus Y^\perp$, i.e. $P(x)$ is the unique closest point to Y for all $x \in \mathscr{H}$ (by the proof of the projection theorem). Once we have the notion of the Hilbert space adjoint (defined in Section **??**) you can check that in addition to $P = P^2$, we also have $P = P^*$. In fact the properties $P = P^2 = P^*$ for a bounded operator$P \in \mathscr{B}(\mathscr{H})$ characterise being the orthogonal projection onto $P(\mathscr{H})$ (which is necessarily closed). This can be deduced from the exercises in B4.2.

3.3 Properties of (the space of) bounded linear operators

The space of bounded linear operators is a normed space, so we want to know when it is complete. This happens when the target space is complete.[34] Note how the proof follows the standard 'completeness strategy' discussed in section 1.

Theorem 3.4. *Let X be any normed space and let Y be a Banach space. Then $\mathscr{B}(X,Y)$ (equipped with the operator norm) is complete and therefore is a Banach space.*

Proof. Let (T_n) be a Cauchy-sequence in $\mathscr{B}(X,Y)$. Then for every $x \in X$ we have that

$$\|T_n x - T_m x\| \leq \|T_n - T_m\| \, \|x\| \to 0$$

as $m,n \to \infty$, so $(T_n x)$ is a Cauchy sequence in Y and, as Y is complete, thus converges to some element in Y which we call Tx.

We now show that the resulting map $x \mapsto Tx$ is an element of $\mathscr{B}(X,Y)$ and $T_n \to T$ in $\mathscr{B}(X,Y)$, i.e. $\|T - T_n\| \to 0$.

We first note that the linearity of T_n (and (AOL)) implies that also T is linear. Given any $\varepsilon > 0$ we now let N be so that for $m,n \geq N$ we have $\|T_n - T_m\| \leq \varepsilon$. Given any $x \in X$, continuity of the norm gives

$$\|Tx - T_n x\| = \|\lim_{m\to\infty} T_m x - T_n x\| = \lim_{m\to\infty} \|T_m x - T_n x\| \leq \varepsilon \|x\|.$$

Hence T is bounded (as $\|Tx\| \leq (\|T_n\| + \varepsilon)\|x\|$ for all x) and so an element of $\mathscr{B}(X,Y)$ with $\|T - T_n\| \leq \varepsilon$ for all $n \geq N$, so as $\varepsilon > 0$ was arbitrary we obtain that $T_n \to T$ in the sense of $\mathscr{B}(X,Y)$. $\square$

We note in particular that if X is a Banach-space then the space $\mathscr{B}(X) := \mathscr{B}(X,X)$ of bounded linear operators from X to itself is a Banach space and that for any normed space X the dual space $X^* = \mathscr{B}(X,\mathbb{R})$ (respectively $X^* = \mathscr{B}(X,\mathbb{C})$ if X is a complex vector space) is complete as both $\mathbb{R}$ and $\mathbb{C}$ are complete.

Given any normed spaces X, Y and Z and any linear operators $T \in \mathscr{B}(X,Y)$ and $S \in \mathscr{B}(Y,Z)$ we can consider the composition $ST = S \circ T : X \to Z$ and observe that:

Proposition 3.5. *The composition ST of two bounded linear operators $S \in \mathscr{B}(Y,Z)$ and $T \in \mathscr{B}(X,Y)$ between normed spaces X,Y,Z is again a bounded linear operator and we have*

$$\|ST\|_{\mathscr{B}(X,Z)} \leq \|S\|_{\mathscr{B}(Y,Z)} \, \|T\|_{\mathscr{B}(X,Y)}.$$

Proof. The only thing we should prove is the estimate,[35] which follows as for $x \in X$, we have

$$\|STx\| = \|S(Tx)\| \leq \|S\| \, \|Tx\| \leq \|S\| \, \|T\| \, \|x\|.$$

$\square$

[34] In fact $\mathscr{B}(X,Y)$ is complete if and only if Y is complete; the converse direction will follow from the Hahn-Banach theorem; see sheet 4.

[35] we are already very familar with the fact that the composition of linear operators is linear, and the composition of continuous maps is continuous

Remark. The proposition implies in particular that for sequences $T_n \to T$ in $\mathscr{B}(X,Y)$ and $S_n \to S$ in $\mathscr{B}(Y,Z)$ also

$$S_nT_n \to ST \text{ in } \mathscr{B}(X,Z)$$

since

$$\|S_nT_n - ST\| \leq \|(S_n - S)T_n\| + \|S(T_n - T)\| \leq \|S_n - S\|\|T_n\| + \|S\|\|T_n - T\| \to 0$$

where we use in the last step that $\|T_n\|$ is bounded since T_n converges. That is multiplication (i.e. composition) of operators

Deep Dive

Let X be a Banach space. Then the space of bounded linear operators $\mathscr{B}(X)$ is an example of a *unital Banach algebra*. A unital Banach algebra is a Banach space A together with an associative multiplication $A \times A \to A$ which has an identity element 1 with $1x = x1 = x$ for all $x \in A$ (for $\mathscr{B}(X)$, the identity is I_X) such that the multiplication interacts with the Banach space addition and scalar mutliplication in the way you would expect, [a] and satisfying

$$\|ab\| \leq \|a\|\|b\|, \qquad \text{for all } a, b \in A.$$

The last condition, which is Proposition 3.5 for $\mathscr{B}(X)$, shows that the multiplication is jointly continuous. Note that the multiplication need not be commutative the example of composition of operators in $\mathscr{B}(X)$ is not generally commutative.

We have seen some other Banach algebras already: $C(K)$ with pointwise multiplication, and $L^\infty([0,1])$ with pointwise multiplication (defined almost everywhere), are both Banach algebras with the supremum and essential supremum norms. In fact the map $M_\bullet$ sending $g \in C([0,1])$ to the mutliplication operator $M_g \in \mathscr{B}(C([0,1]))$ discussed in the previous section is a Banach algebra *homomorphism*: $M_\bullet$ is linear in g, and preserves the multiplication $M_{gh} = M_gM_h$ (which in this case is M_hM_g). In our example we found that $\|M_g\| = \|g\|_\infty$ so $M_\bullet$ is isometric, so certainly bounded. In general just as linear maps need not be automaically bounded, so too Banach algebras homomorphisms are not always bounded (though there are many nice situations where they are)! Here's another unital Banach algebra:

$$\ell^1(\mathbb{Z}) = \{(x_n)_{n\in\mathbb{Z}} : \sum_{n\in\mathbb{Z}} |x_n| < \infty\}$$

with the ℓ^1 norm and convolution multiplication

$$(xy)_n = \sum_r x_r y_{n-r}.$$

What is the identity?

When X is a Banach space, any closed subalgebra[b] $A \subset \mathscr{B}(X)$ containing the identity is a unital Banach algebra. Conversely, if A is a unital Banach algebra with identity 1, consider the homomorphism $M_\bullet : A \to \mathscr{B}(A)$ given by $M_a(b) = ab$. [This generalises the multiplication map on $C([0,1])$.] This is a homomorphism as $(M_aM_b)(c) = M_a(M_b(c)) = a(bc) = (ab)(c) = M_{ab}(c)$ and $\|M_a(b)\| \leq \|a\|\|b\|$ so $M_a \in \mathscr{B}(A)$ with $\|M_a\| \leq \|a\|$ and from taking $b = 1$, we get $\|M_a\| \geq \|a\|/\|1\|$, so M_a is bounded below, and hence the image $\{M_a : a \in \mathscr{B}(A)\}$ is closed in $\mathscr{B}(A)$.

Unital Banach algebras provide the right abstract framework for spectral theory, which we will develop in B4.2 for operators in $\mathscr{B}(X)$. As you do that it's worth going through and seeing that it all works fine in a unital Banach algebra with no real changes to the arguments.

[a]Have a go at axiomising this motivated by the relations you find in $\mathscr{B}(X)$
[b]i.e. a closed subspace also closed under the multiplication

We also note that for operators $T \in \mathscr{B}(X)$ from a normed space X to itself we can consider the composition of T with itself, and more generally powers $T^n = T \circ T \circ \ldots \circ T \in \mathscr{B}(X)$ which, by the above proposition have norm
$$\|T^n\| \leq \|T\|^n.$$
We can use this to define suitable power series of operators.[36]

Example 3.6. Let X be a Banach space and let $A \in \mathscr{B}(X)$. Then[37]
$$\exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$
converges in $\mathscr{B}(X)$ and hence $\exp(A)$ is a well defined element of $\mathscr{B}(X)$.

Proof. We know that
$$\sum_{k=0}^{\infty} \|\frac{1}{k!} A^k\| \leq \sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} = \exp(\|A\|) < \infty,$$
i.e. that the series converges absolutely. As X is complete and thus, by Theorem 3.4, also $\mathscr{B}(X)$ is complete we hence obtain from Corollary 1.6 that the series converges. □

Deep Dive

This is the starting point of a fundamental tool in studying operators: *functional calculi*. A functional calculus gives a consistent way of defining $f(T)$ for a suitable bounded operators T, and suitable classes of functions $f : D \to \mathbb{C}$, for suitable $D \subset \mathbb{C}$. At the moment you can use Taylor's theorem to extend the example above and define $f(T)$ whenever f is a holomorphic function $\mathbb{C} \to \mathbb{C}$, and also define $f(T)$ when f is given by a power series with radius of convergence exceeding $\|T\|$. The sort of thing you might like is given two functions f and g, to have that $(f \circ g)(T) = f(g(T))$. You'll get a chance to do something like this with power series on exercise sheet 3. This gives you a first functional calculus, but once we've defined the spectrum it's possible to build more sophisticated functional calculus, such as the *holomorphic functional calculus* which allows you to define $f(T)$ whenever f is a holomorphic function on the spectrum of a bounded operator T (or more generally an element in a Banach algebra), or later the continuous and Borel functional calculi, which work for self-adjoint (and more generally normal) operators on a Hilbert space.

3.4 Invertibility

Just as in finite dimensions we shall be interested in when bounded operators are invertible in $\mathscr{B}(X)$, i.e. when a bounded linear operator is bijective and the inverse map is bounded.

Definition 3.7. An element $T \in \mathscr{B}(X)$ is called invertible (short for *invertible in* $\mathscr{B}(X)$) if there exists $S \in \mathscr{B}(X)$ so that $ST = TS = I_X$.[38] When it exists, S is called the inverse of T written T^{-1}.

[36]The following works equally well for an element of a Banach algebra
[37]Here $A^0 = I$, the identity operator on X.
[38]It is necessary that S is a two sided inverse. Going back to our left and right shift operators we have $LR = I$ on ℓ^p but $RL \neq I$.

If we only talk about $T : X \to X$ being 'invertible as a function between sets', we sometimes say that T is algebraically invertible and that a function $S : X \to X$ is an algebraic inverse of T if $ST = TS = I$ (but not necessarily $S \in \mathscr{B}(X)$).

Deep Dive

The Banach isomorphism theorem (a consequence of the Banach open mapping theorem) will tell you that if $T \in \mathscr{B}(X)$ is algebraically invertible and X is a Banach space, then T is invertible. This will be proved in B4.2.

In many applications, including spectral theory which will be discussed in B4.2 Functional Analysis II, the following lemma turns out to be useful to prove that an operator is invertible. The statement should be reminisant of the convergent geometric series from prelims. In fact the proof (when set out in the right way) is also the same telescoping sum argument as prelims.

Lemma 3.8 (Convergence of Neumann-series)**.** *Let X be a Banach space and let $T \in \mathscr{B}(X)$ be so that $\|T\| < 1$. Then the operator $I - T$ is invertible with*

$$(I-T)^{-1} = \sum_{j=0}^{\infty} T^j \in \mathscr{B}(X).$$

Proof of Lemma 3.8. As $\|T\| < 1$ we know that $\sum \|T^k\| \le \sum \|T\|^k < \infty$ so, by Corollary 1.6, the series converges

$$S_n := \sum_{k=0}^{n} T^k \to S = \sum_{k=0}^{\infty} T^k \text{ in } \mathscr{B}(X).$$

As

$$(I-T)S_n = I - T + T - T^2 + T^2 - \ldots - T^n + T^n - T^{n+1} = I - T^{n+1}$$

and $\|T^{n+1}\| \le \|T\|^{n+1} \to 0$, we can pass to the limit $n \to \infty$ in the above expression to obtain that $(I-T)S = I$ and similarly $S(I-T) = I$ so $S = (I-T)^{-1}$. $\square$

Corollary 3.9. *Let X be a Banach space. Then the invertible operators on X are open. Precisely, if $T \in \mathscr{B}(X)$ be invertible, then for any $S \in \mathscr{B}(X)$ with $\|S\| < \|T^{-1}\|^{-1}$ we have that $T - S$ is invertible.*

Proof. Fix invertible $T \in \mathscr{B}(X)$, and let $S \in \mathscr{B}(X)$ have $\|S\| < \|T^{-1}\|^{-1}$. As T is invertible (which by definition means that also $T^{-1} \in \mathscr{B}(X)$) we obtain can write $T - S = T(I - T^{-1}S)$ and note that $T^{-1}S \in \mathscr{B}(X)$ with $\|T^{-1}S\|_{\mathscr{B}(X)} \le \|T^{-1}\|\|S\| < 1$. By Lemma 3.8 we thus find that $(I - T^{-1}S)$ is invertible with $(I - T^{-1}S)^{-1} = \sum_{j=0}^{\infty}(T^{-1}S)^j \in \mathscr{B}(X)$ and hence $T - S$ is the composition of two invertible operators and thus invertible, compare also A.1 on Problem Sheet 2.

Since $T^{-1} \neq 0$, $\|T^{-1}\| \neq 0$, and it follows that the invertible operators are open. $\square$

Notice that if $T \in \mathscr{B}(X)$ is invertible then for $x \in X$,

$$\|x\| = \|T^{-1}T(x)\| \le \|T^{-1}\|\|T(x)\|.$$

This suggests:

Definition 3.10. Let X, Y be normed spaces and $T \in \mathscr{B}(X,Y)$. Say that T is *bounded below* if there exists $C > 0$ such that

$$\|T(x)\| \ge C\|x\| \quad \text{for all } x \in X.$$

We have seen that invertible operators are bounded below (by the norm of the inverse). Hence being bounded below is necessary for invertibility, and for an algebraically invertible $T \in \mathscr{B}(X)$, we have that T is invertible if and only if it is bounded below. An operator which is bounded below is certainly injective. When the domain is complete, operators which are bounded below also have closed range.

Proposition 3.11. *Let X be a Banach space, Y be a normed space and $T \in \mathscr{B}(X,Y)$ be bounded below. Then $T(X)$ is closed in Y.*

Proof. Suppose that (x_n) is a sequence with $T(x_n) \to y \in Y$. Let $C > 0$ be such that $\|Tx\| \geq C\|x\|$ for all x, so that $\|x_n - x_m\| \leq C^{-1}\|Tx_n - Tx_m\| \to 0$, as (Tx_n) is Cauchy. By completeness of X, there exists $x \in X$ with $x_n \to x$. By continuity of T, $Tx_n \to Tx$, so $y = Tx \in T(X)$. □

4 Finite dimensional normed spaces

In this section we will explain why for finite dimensional spaces most of the questions raised in the previous chapters do not arise, and hence why you never had to discuss issues of continuity, and completeness in your early courses on finite dimensional normed spaces. We shall see in particular that

- all norms on a finite dimensional space are equivalent,
- all linear maps defined on a finite dimensional space are bounded,
- all finite dimensional spaces are complete.

We shall furthermore see that the Heine-Borel Theorem from part A and Prelims for $\mathbb{R}$ and $\mathbb{R}^n$, that assures that bounded and closed sets in $\mathbb{R}^n$ are compact, remains valid in general finite dimensional normed spaces. Moreover, the Heine-Borel property characterises finite dimensional normed spaces: if the unit ball of a normed space is compact, then the space must be finite dimensional.

Deep Dive

For this reason we shall need to work with 'weaker forms' of compactness in infinite dimensions, such as the 'weak sequential compactness' of the unit ball of a reflexive space (such as L^p for $1 < p < \infty$) which you will see in B4.2, and is a crucial tool in PDE, or the Banach-Alaglou theorem that the unit ball of a dual space X^* is compact in the weak*-topology from C4.1 (which underpins the weak sequential compactness of unit balls in reflexive spaces).

4.1 Equivalence of norms and its consequences

We start out by proving that all finite dimensional normed spaces are equivalent by comparing them to the Eucidlian spaces ℓ_n^2. The proof works identically for real and complex scalars.

Proposition 4.1. *Let V be a normed space with basis $e_1, \dots, e_n$. Then the linear map $T : \ell_n^2 \to V$ given by $T(\lambda_1, \dots, \lambda_n) = \sum_{i=1}^n \lambda_i e_i$ gives an isomorphism of normed spaces.*

Proof. From Cauchy Schwartz we have

$$\|T(\lambda)\| = \|\sum_{i=1}^n \lambda_i e_i\| \leq \sum_{i=1}^n |\lambda_i|\|e_i\| \leq (\sum_{i=1}^n |\lambda_i|^2)^{1/2}(\sum_{i=1}^n \|e_i\|^2)^{1/2}.$$

That is $\|T(\lambda)\| \leq C\|\lambda\|_2$, for $C = (\sum_{i=1}^n \|e_i\|^2)^{1/2}$. Therefore T is bounded.

The unit sphere $S_{\ell_n^2} = \{\lambda \in \ell_n^2 : \|\lambda\|_2 = 1\}$ is closed and bounded so compact, and the map $\lambda \to \|T\lambda\|$ is continuous, so attains its minimum value m on $S_{\ell_n^2}$ at $\lambda \in S_{\ell_n^2}$. Since $\|\lambda\|_2 = 1$, we have $T(\lambda) \neq 0$, and hence $m \neq 0$. Therefore, for $\lambda \in S_{\ell_n^2}$, we have $\|T(\lambda)\| \geq m$, and so by homogenity, $\|T(\lambda)\| \geq m\|\lambda\|_2$ for all $\lambda \in \ell_n^2$. That is T is bounded below, so an isomorphism. $\square$

Note that we could obtain control on $\|T\|$, for example $\|T\| \leq n^{1/2}$ by taking each $\|e_i\| = 1$, the proof above does not give a way of controlling $\|T^{-1}\|$ explciitly; it only produces the constant m showing that T^{-1} is bounded. We will have more to say about this on the exercise sheet and in a deep dive.

It follows that two normed spaces of the same dimension (over the same field) are isomorphic and all norms on a finite dimensional space are equivalent.

Corollary 4.2. *Any two finite dimensional normed spaces with the same dimension (over the same field $\mathbb{R}$ or $\mathbb{C}$) are isomorphic. Any two norms on a finite dimensional space are equivalent.*

Proof. The first statement follows as given n-dimensional normed spaces V and W with bases $e_1, \dots, e_n$ for V and $f_1, \dots, f_n$ for W respectively, the unique linear bijection $T : V \to W$ which has $T(e_i) = f_i$ is an isomorphism (by applying Propositon 4.1 twice). For the second statement given two different norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on V, applying the argument of the previous sentence, the identity map $(V, \|\cdot\|_1) \to (V, \|\cdot\|_2)$ is an isomorphism, i.e the two norms are equivalent. $\square$

Also all linear maps whose domain is a finite dimensional normed space are automatically continuous.

Corollary 4.3. *Let X be a finite dimensional normed space and let Y be any normed space (not necessarily finite dimensional). Then any linear map $T : X \to Y$ is an element of $\mathscr{B}(X,Y)$, i.e. a bounded linear operator.*

Proof. Given any such T we set for every $x \in X$

$$\|x\|_T := \|x\|_X + \|Tx\|_Y.$$

We can easily check that this defines a norm on the finite dimensional space X which, by the previous corollary, must hence be equivalent to $\|\cdot\|_X$. In particular, there exists a constant $C \in \mathbb{R}$ so that

$$\|Tx\|_Y \leq \|x\|_T \leq C\|x\|_X$$

which ensures that T is bounded and hence an element of $\mathscr{B}(X,Y)$. $\square$

Since ℓ_n^2 is complete, and completeness is a property preserved by (strong) equivalence of metric spaces (and hence isomorphism of normed spaces), it follows that all finite dimensional normed spaces are complete. This gives the following important corollary

Corollary 4.4. *Every finite dimensional subspace of a normed vector space X is complete and hence closed.*

Proof. Let Y be a finite dimensional subspace of X. Then $Y \cong \ell_n^2$, where $n = \dim Y$, so Y is complete. Then suppose $y_n \in Y$ has $y_n \to x \in X$. Then (y_n) is convergent in X, so Cauchy in X, and hence Cauchy in Y. Since Y is complete, there exists $y \in Y$ with $y_n \to y$. By uniqueness of limits in a metric space, $y = x$ and hence Y is closed in X. $\square$

Warning. Not every subspace of a normed vector space X is closed.

Example 4.5. Consider $C([0,2])$ as a subspace of $(L^1([0,2]), \|\cdot\|_{L^1})$. Then the sequence $(f_n)_{n\in\mathbb{N}} \subset C([0,2])$ defined by

$$f_n(t) = \begin{cases} t^n, & 0 \le t \le 1 \\ 1, & t > 1 \end{cases}$$

is a Cauchy sequence in $L^1([0,2])$ with limit $f(t) = \begin{cases} 0 & 0 \le t \le 1 \\ 1 & t > 1 \end{cases}$ however $f \notin C([0,2])$.

At a more abstract level we could also argue as follows: $C([0,2])$ is a proper subspace of $L^1([0,1])$ however, as we shall see later, $C([0,2])$ is dense in $L^1([0,1])$, so the closure of $C([0,2])$ in $L^1([0,1])$ is $\overline{C([0,2])}^{L^1} = L^1([0,1]) \neq C([0,2])$.

4.2 Compactness

We now turn to the Heine-Borel property, and show that norm-compactness of unit balls characterises finite dimensional normed spaces. Recall that a subset K of a metric (or a topological space, though formally this course only considers metric spaces) is compact if ever open cover of K has a finite subcover. For metric spaces (and so in particular for normed spaces), compactness is equivalent to sequential compactness, i.e. every sequence in K has a subsequence which converges to an element of K. A further useful equivalent characterisation of compactness in metric spaces is that K is compact if and only K is complete and totally bounded (which means that for every $\varepsilon > 0$ there exists a finite ε-net, i.e. a finite set of points $x_1, \dots, x_m \in K$ so that $K \subset \bigcup_{i=1}^m B_\varepsilon(x_i)$).

Theorem 4.6. *Let X a normed space. Then the following are equivalent*

(1) $\dim(X) < \infty$.

(2) Every subset $Y \subset X$ that is bounded and closed is compact.

(3) The unit sphere $S := \{x \in X : \|x\| = 1\}$ is compact.

The implication (1)⇒(2) follows from the Heine–Borel theorem, as if X is finite dimensional, then it is isomorphic to a Euclidean space ℓ_n^2 (where $n = \dim X$) and for this space the Heine–Borel theorem shows that all closed and bounded subsets are compact.[39] (2) $\Longrightarrow$ (3) follows as the unit sphere is closed (by continuity of the norm) and bounded (by definition).

For the remaining implication (3)⇒(1) it is useful to first understand how this works in the setting of an inner product space.[40] Suppose that X is an infinite dimensional inner product space, and using Gramm-Schmidt produce an orthonormal sequence $(e_n)_{n=1}^\infty$ in X. By Pythagoras, we have $\|e_n - e_m\| = \sqrt{2}$ for $n \neq m$, so the bounded sequence (e_n) has no Cauchy subsequence, and hence the sphere of X can not be sequentially compact. This proves (3)⇒(1) for inner product spaces.

In general, we can not rely on Pythagoras, so we use the following useful lemma of Reisz to be able to inductively construct a sequence of points which

Proposition 4.7 (Riesz's Lemma)**.** *Let X be a normed vector space and $Y \subsetneq X$ a closed subspace. Then to any $\varepsilon > 0$ there exists an element $x \in S \subset X$ in the unit sphere so that*

$$\operatorname{dist}(x, Y) := \inf\{\|x - y\| : y \in Y\} \ge 1 - \varepsilon.$$

[39]Note that the properties of a subset being closed, bounded, or compact are all preserved by homeomorphisms, so preserved by isomorphisms of normed spaces: if $T: X \to Y$ is an isomorphism and K is closed (or bounded, or compacT), then $T(K)$ is closed (or bounded or compact).

[40]It is often useful to understand how to prove general Banach space results in the Hilbert space case first.

Proof. We can assume without loss of generality that $\varepsilon \in (0,1)$.

As $Y \neq X$ is closed we know that the set $X \setminus Y$ is open and non-empty, so we can choose some $x^* \in X \setminus Y$ and use that $d := \mathrm{dist}(x^*, Y) > 0$, as $X \setminus Y$ must contain some ball $B_\delta(x)$ which ensures that $d \geq \delta > 0$.

By the definition of the infimum, we can now select $y^* \in Y$ so that $d \leq \|x^* - y^*\| < \frac{d}{1-\varepsilon}$ and claim that $x := \frac{x^*-y^*}{\|x^*-y^*\|}$ has the desired properties. Clearly $\|x\| = 1$, i.e. $x \in S$ as desired, and we furthermore have that

$$\begin{aligned}\mathrm{dist}(x,Y) &= \inf_{y\in Y}\|x-y\| = \inf_{y\in Y}\|\frac{x^*}{\|x^*-y^*\|} - \frac{y^*}{\|x^*-y^*\|} - y\| = \inf_{\tilde{y}\in Y}\|\frac{x^*}{\|x^*-y^*\|} - \tilde{y}\| \\ &= \inf_{\hat{y}\in Y}\|\frac{x^*-\hat{y}}{\|x^*-y^*\|}\| = \frac{\mathrm{dist}(x^*,Y)}{\|x^*-y^*\|} \geq 1-\varepsilon\end{aligned}$$

where we used twice that Y is a subspace, to replace the infimum over $y \in Y$ first by an infimum over $\tilde{y} = \frac{y^*}{\|x^*-y^*\|} + y$ and then an infimum over $\hat{y}$ which is related to $\tilde{y}$ by $\tilde{y} = \frac{\hat{y}}{\|x^*-y^*\|}$. □

Deep Dive

It is natural to ask whether you can take $\varepsilon = 0$ in Reisz's lemma. For example if X is a Hilbert space and Y is a proper subspace then you can take any x in the unit sphere of $Y^\perp$ and for $y \in Y$, we have $\|x-y\|^2 = \|x\|^2 + \|y\|^2 \geq 1$, so $\mathrm{dist}(x,Y) = 1$.

In fact for a given Banach space X, being able to take $\varepsilon = 0$ in Reisz's lemma for all proper closed subspaces Y is equivalent to the question of whether every bounded linear functional on X attains its norm (which we noted in an earlier deep dive is equivalent to reflexivity of X by James' Theorem). One direction proceeds by considering kernels of functionals. Given $f \in X^*$ with $\|f\| = 1$, let $Y = \ker f$ a proper closed subspace of X. If we can take $\varepsilon = 0$ in Reisz's lemma then we get some $x \in X$ with $\|x\| = 1$ and $d(x,Y) = 1$. But, for $y \in Y$, $|f(x)| = |f(x-y)| \leq \|x-y\|$, so $d(x,Y) \geq |f(x)|$. On the other hand we can find a sequence (z_n) with $f(z_n) = 1$ and $\|z_n\| \to 1$. Then $y_n = x - f(x)z_n \in Y$ and $\|x - y_n\| = \|f(x)z_n\| \to |f(x)|$, and hence $d(x,Y) = |f(x)|$.

For the reverse direction we really need quotient spaces. Suppose that bounded linear functionals on X attain their norms, and let Y be a proper closed subspace of X. Fix a norm 1 functional f on the quotient space X/Y (which we didn't define), and consider the norm 1 bounded linear functional $g = f \circ q : X \to \mathbb{F}$ where $q : X \to X/Y$ is the quotient map. Let $x \in X$ have $\|x\| = 1$ and $|g(x)| = 1$ (by hypothesis). For $y \in Y$, as by construction $g(y) = 0$, we have $\|x-y\| \geq |g(x)| = 1$, hence $d(x,Y) \geq 1$. But as $\|x\| = 1$, we have $d(x,Y) = 1$.

For a non-reflexive space X, one can also take $\varepsilon = 0$ in Reisz's lemma when Y is finite dimensional.

Proof of Theorem 4.6 (3)⇒(1). Suppose $\dim(X) = \infty$ and there We may thus choose a sequence of linearly independent elements $y_k \in X$, $k \in \mathbb{N}$. Then the subspace $Y_k := \mathrm{span}\{y_1, \ldots, y_k\} \subsetneq Y_{k+1}$ is finite dimensional, so by Corollary 4.4, a closed proper subspace of Y_{k+1}. Applying Proposition 4.7 with $\varepsilon = \frac{1}{2}$ (viewing Y_k as a subspace of Y_{k+1} instead of X) thus gives us a sequence of elements $y_k \in Y_{k+1} \cap S$ with $\mathrm{dist}(y_k, Y_k) \geq \frac{1}{2}$. In particular for every $k > l$ we have $\|y_k - y_l\| \geq \mathrm{dist}(y_k, Y_{l+1}) \geq \mathrm{dist}(y_k, Y_k) \geq \frac{1}{2}$ so no subsequence of (y_k) can be a Cauchy-sequence. Therefore S is not sequentially compact (and as a metric space) S is not compact. □

4.3 The Banach–Mazur compactum

It is tempting to think with the results of the previous two subsections that there is not much left to say about finite dimensional normed spaces; and this is true if we only care about their normed space structure up to *isomorphism*. But if we are interested in the metric properties of finite dimensional Banach spaces, there are many interesting directions.

Deep Dive

Definition. Let X,Y be n-dimensional normed spaces. The *Banach–Mazur* distance between X and Y is

$$\rho(X,Y) = \inf : \{\|T\|\|T^{-1}\| : T \text{ is an isomorphism between } X \text{ and } Y\}.$$

Note that this is not a metric in the usual sense on a collection of isomorphic normed spaces: $\rho(X,Y) \geq 1$, with equality if and only if X and Y are isometrically isomorphic.[a] Also, we have a multiplicative triangle inequality: $\rho(X,Z) \leq \rho(X,Y)\rho(Y,Z)$ when X,Y and Z are all isomorphic. From this it's not hard to see that $\log\rho(\cdot,\cdot)$ gives a metric on the collection of isometric isomorphism classes of n-dimensional normed spaces.

Theorem. *Let $n \in \mathbb{N}$. The collection $Q(n)$ of isometric isomorphism classes of n-dimensional normed spaces is compact in with the Mazur distance.*

Once we know that $Q(n)$ is compact, it follows that it is bounded; there is a constant $C(n)$ such that for all n-dimensional spaces X,Y we have $\rho(X,Y) \leq C(n)$. In the exercises we shall give some methods for estimating $C(n)$ directly:

- firstly by comparing with ℓ^1_n we can show $\rho(X,\ell^1_n) \leq n$ for all n-dimensional spaces X, and hence $C(n) \leq n^2$.
- secondly we get a better (but more tricky) estimate $C(n) \leq n$ by comparing with ℓ^2_n and obtaining $\rho(X,\ell^2_n) \leq n^{1/2}$ for all n dimensional spaces.

Given there is some constant $C(n)$ showing that $Q(n)$ is bounded, here is an outline of a proof of compactness of $Q(n)$; this is likely to become an extensional exercise in a future year!

Proof. Let $X = \mathbb{F}^n$ be a fixed n-dimensional space, which we equip with the Euclidian norm denoted in this proof by $\|\cdot\|_{\mathrm{euc}}$. A sequence in $Q(n)$ can be realised by a sequence of norms $\|\cdot\|_m$ on X satisfying

$$C(n)^{-1/2}\|x\|_{\mathrm{euc}} \leq \|x\|_m \leq C(n)^{1/2}\|x\|_{\mathrm{euc}}, \quad x \in X.$$[b]

Then performing a diagonal argument we can find a subsequence of these norms, say $(\|\cdot\|_{k_m})$ such that $(\|x\|_{k_m})$ converges for all x in the countable dense set $\mathscr{Q}(i)^n$ of X. From the bounds above, it follows that $(\|x\|_{k_m})$ converges for all $x \in X$, and the resulting function $\|x\|$ inherits the properties $\|x+y\| \leq \|x\|+\|y\|$ and $\|\lambda x\| \leq |\lambda|\|x\|$ from the norm properties of each $\|\cdot\|_{k_m}$. Also

$$C(n)^{-1/2}\|x\|_{\mathrm{euc}} \leq \|x\| \leq C(n)^{1/2}\|x\|_{\mathrm{euc}}, \quad x \in X,$$

so that $\|x\| = 0$ only if $x = 0$, and accordingly $\|\cdot\|$ is a norm.

Finally, note that the subsequence $(X,\|\cdot\|_{k_m})$ converges to $(X,\|\cdot\|)$ in the Banach–Mazur sense. The point is that we have to upgrade from a pointwise convergence result to obtain a uniform estimate using compactness. Indeed if $S_m : (X,\|\cdot\|_{k_m}) \to (X,\|\cdot\|)$ is the identity map, then we have $\|S_m x\| \to \|x\|$ for all $x \in X$ and we need to show $\|S_m\|$ and $\|S_m^{-1}\| \to 1$. Fix $\varepsilon > 0$ and by compactness fix a finite $\varepsilon C(n)^{-1/2}$ net N for the ball of radius $C(n)^{1/2}$ in $\|\cdot\|_{\mathrm{euc}}$,[c] Fix $\varepsilon > 0$ and find m_0 large enough so that for $m \geq m_0$, and $x \in N$, we have $|\|x\|_{k_m} - \|x\|| \leq \varepsilon$. Then for y with $\|y\|_{k_m} = 1$, there exists $x \in N$ with $\|x-y\|_{\mathrm{euc}} \leq \varepsilon C(n)^{-1/2}$ so $\|x-y\|_{k_m} \leq \varepsilon$ and $\|x-y\| \leq \varepsilon$. Therefore

$$\|y\| \leq \varepsilon + \|x\| \leq \|x\|_{k_m} + 2\varepsilon \leq \|y\|_{k_m} + 3\varepsilon = (1+3\varepsilon)\|y\|_{k_m},$$

and hence $\|S_m\| \leq (1+3\varepsilon)$. Similarly $\|S_m^{-1}\| \leq (1+3\varepsilon)$. This shows that $(X, \|\cdot\|_{k_m})$ converges to $(X, \|\cdot\|)$ in the Banach–Mazur sense. □

[a] For this, if $(T_n) : X \to Y$ is a sequence of isomorphisms with $\|T_n\|\|T_n^{-1}\| \to 1$, we can scale so that both $\|T_n\|, \|T_n^{-1}\| \to 1$. Then for large enough n, both T_n and T_n^{-1} lie in the ball of radius 2 in $\mathscr{B}(X,Y)$ which is a finite dimensional normed space so compact. Therefore, we can pass to convergent subsequences to that $T_{k_n} \to S : X \to Y$ and $T_{k_n}^{-1} \to R : Y \to X$. We then check that S is invertible with inverse R and that each $\|S\| = \|R\| = 1$, so S is an isometric isomorphism between X and Y. In general this shows that the infiumum in the definition of the Banach–Mazur distance is attained (since we work with finite dimensional normed spaces).

[b] Given any finite dimensional normed space Y we can find an isomorphism $T : Y \to X$ with $\|T\|, \|T^{-1}\| \leq C(n)^{1}2/$ (arguing in the same way as the previous footnote). Then X equipped with the norm $x \mapsto \|T^{-1}x\|$ is an isometric copy of Y and this norm has $C(n)^{-1/2}\|x\|_{\mathrm{euc}} \leq \|T^{-1}x\| \leq C(n)^{1/2}\|x\|_{\mathrm{euc}}$ as needed.

[c] i.e. a finite set N in the unit ball of $(X, \|\cdot\|)$ such that for all $y \in X$ with $\|y\|_{\mathrm{euc}} \leq C(n)^{1/2}$, there exists $x \in N$ with $\|x - y\| \leq \varepsilon C(n)^{-1/2}$. The point of these constants $C(n)^{1/2}$ is that N gives an ε net for all the balls $(X, \|\cdot\|_{k_m})$ and $(X, \|\cdot\|)$.

Deep Dive

A major direction of study, known as the *Ribe programme*, originates in Ribe's rigidity theorem which (loosely speaking) says that two Banach spaces X and Y are uniformly equivalent as metric spaces (i.e. there is a bijection $f : X \to Y$ which is uniformly continuous and f^{-1} is uniformly continuous, but note no assumption about how f interacts with the vector space structure is made) if and only if they have the same finite dimensional subspaces (precisely there is $K > 0$ such that for every finite dimensional subspace F of X, there exists a bounded linear map $T : F \to Y$ with $\|x\|_X \leq \|Tx\|_Y \leq K\|x\|_X$ for all $x \in X$ and vice versa exchanging X and Y). Therefore isomorphism invariant properties of Banach spaces that depend only on finite dimensional subspaces can be described entirely in terms of the metric geometry (and do not need the linear structure). The Ribe programme studies this phenomena explicitly aiming to uncover the hidden properties of metric spaces that correspond to finite dimensional Banach space properties. This has led to insights from Banach spaces giving rise to completely unexpected applications in other fields. A nice thing to read is the introduction to Assaf Naor's survey: "An Introduction to the Ribe programme".

5 Density of subspaces and the Stone–Weierstrass Theorem

Just as in prelims analysis we sometimes approximated real numbers by sequences of rational numbers, or in integration we approximated integrable functions by simple functions, we often want to be able to approximate elements of our normed space by elements with some nicer property. The relevant definitions really belong to the land of metric spaces (and apply to normed spaces with the metric space coming from the norm).[41]

Definition. Let X be a metric space. Recall that a subset $A \subset X$ is *dense* if the closure $\overline{A}$ of A is all of X, i.e. $\overline{A} = X$. The metric space X is *separable* if it has a countable dense subset.

It follows that A is dense in X if and only if for all $x \in X$ and $\varepsilon > 0$ there exists $a \in A$ with $d(x,a) < \varepsilon$ which happens if and only if for all $x \in X$, there is a sequence (a_n) in A with $a_n \to x$.

We will come back to separability in Section 6. Here we discuss how to extend operators from dense subspaces and give the Stone–Weierstrass theorem, which gives a fundamental tool for obtaining dense subspaces of $C(K)$. We will look at more examples in Section 6.2.

[41] The same definitions apply equally to topological spaces.

5.1 Density of subspaces and extensions of bounded linear operators by density

An important feature of dense subsets D of normed spaces is that a bounded linear operator on X is fully determined by its values on D. This is particularly useful if we are working on a space that contains a subspace of "well-understood" objects, e.g. the space of polynomials in the space of real valued continuous functions or the space of real valued smooth functions on $[0,1]$ in $(L^2([0,1]),\|\cdot\|_{L^2})$. The fact that the behaviour on a dense subspace fully describes a bounded linear operator is encapsulated in the next lemma.[42]

Lemma 5.1. *Let X be a normed space, $D \subset X$ a dense subset and let Z be a normed space. Then for operators $T, S \in \mathscr{B}(X,Z)$ we have*

$$T|_D = S|_D \iff T = S.$$

In particular, the only element $T \in \mathscr{B}(X,Z)$ with $T|_D = 0$ is $T = 0$.

Proof of Lemma 5.1. Suppose $S|_D = T|_D$. For any $x \in X$ we can choose a sequence $d_n \to x$ with $d_n \in D$ to conclude that since both T and S are continuous

$$Tx = \lim_{n\to\infty} Td_n = \lim_{n\to\infty} Sd_n = Sx.$$

The reverse direction is immediate. □

Theorem 5.2. *Let X be a normed space, let Y be a dense subspace of X (which we equip with the norm of X) and let Z be a Banach space. Then any $T \in \mathscr{B}(Y,Z)$ has a unique extension $\tilde{T} \in \mathscr{B}(X,Z)$, i.e. there exists a unique bounded linear operator $\tilde{T} : X \to Z$ so that $\tilde{T}y = Ty$ for every $y \in Y$ and we furthermore have that*

$$\|\tilde{T}\|_{\mathscr{B}(X,Z)} = \|T\|_{\mathscr{B}(Y,Z)}.$$

Proof. The uniqueness is covered by Lemma 5.1.

Let $x \in X$ be any element. Then as Y is dense there exists a sequence y_n of elements of Y so that $y_n \to x$. The key observation is that the sequence $(T_0(y_n))_{n=1}^{\infty}$ converges. Indeed, $(y_n)_{n=1}^{\infty}$ is Cauchy in Y, so $\|T_0(y_n - y_m)\| \leq \|T_0\|\|y_n - y_m\| \to 0$ as $n, m \to \infty$. Therefore $(T_0(y_n))_{n=1}^{\infty}$ is Cauchy so converges in Z, proving the observation.

We can now define $Tx = \lim T_0(y_n)$. To see this is well defined suppose that both (y_n) and (z_n) are sequences in Y with $y_n \to x$ and $z_n \to x$. Then form the alternating sequence $y_1, z_1, y_2, z_2, \ldots$, which converges to x. Therefore

$$T_0(y_1), T_0(z_1), T_0(y_2), T_0(z_2), T_0(y_3), \ldots$$

converges by the observation in the previous parapgraph. Passing to subsequences it follows that $\lim T_0(y_n) = \lim T_0(z_n)$ and T is well defined.

Then $T : X \to Z$ is linear (using continuity of addition and scalar multiplication),[43] and extends T_0 (as for $y \in Y$ we can use the constant sequence $y, y, \ldots$ to see $T(y) = T_0(y)$). Finally for $x \in X$, and (y_n) in Y with $y_n \to x$, we have

$$\|Tx\| = \lim \|T_0(y_n)\| \leq \lim \|T_0\|\|y_n\| = \|T_0\|\|x\|,$$

so $\|T\| \leq \|T_0\|$. For the other direction

$$\|T\| = \sup_{x\in X,\ \|x\|\leq 1} \|T(x)\| \geq \sup_{y\in Y,\ \|y\|\leq 1} \|T_0(y)\| = \|T_0\|.$$

□

[42] Note that this lemma has nothing to do with normed spaces and linear operators. It is equally valid for continuous maps on a metric space.

[43] Given sequences $y_n \to x_1$ and $z_n \to x_2$ from Y and $\lambda \in \mathbb{F}$, the point is that $\lambda y_n + z_n \to \lambda x_1 + x_2$. Then

$$T(\lambda x_1 + x_2) = \lim T_0(\lambda y_n + z_n) = \lambda \lim T_0(y_n) + \lim T_0(z_n) = \lambda T(x_1) + T(x_2).$$

Deep Dive

It's worth comparing this result with the traditional prelims analysis exercise that asks you to take a continuous function $f : [0,1) \to \mathbb{R}$ and show that f has a (necessarily unique) continuous extension to $[0,1]$ if and only if f is uniformly continuous. (The only if is a consequence of compactness of $[0,1]$ so that every continuous function on $[0,1]$ is uniformly continuous). The extension part of this works more generally: Let $A \subset X$ be a dense subset of a metric space and let Z be a complete metric space. If $f : A \to Z$ is uniformly continuous then it has a (necessarily unique) continuous extension to X. I think of this as 'extension by uniform continuity' (though you will often see it described as extending by continuity). Of course when X,Y,Z are normed spaces and $T_0 : Y \to Z$ are all as in Theorem 5.2, the map T_0 is certainly not uniformly continuous (unless it is the zero map). But boundedness of T_0 ensures that it is uniformly continuous on any bounded subset of Y, and hence extends (uniquely) on any bounded subset of Y to its closure in X.

5.2 The Theorem of Stone-Weierstrass and Density of Polynomials in the space of continuous functions

The goal of this section is to identify suitable dense subspaces of the space $C_{\mathbb{R}}(K)$ of *real-valued* continuous functions on a compact metric space K.[44] As always we equip $C_{\mathbb{R}}(K)$ with the sup-norm and recall that since continuous functions on compact sets are bounded this is well defined.

The first result in this direction is Weierstrass' approximation theorem from 1885: the polynomial functions are uniformly dense in $C_{\mathbb{R}}([a,b])$ for any $a < b$. Taking $[a,b] = [0,1]$, for any $f \in C_{\mathbb{R}}([0,1])$ and $\varepsilon > 0$ we can find a polynomial $p(x) = \sum_{r=1}^{n} a_r x^r$ for some $a_0, \dots, a_n \in \mathbb{C}$, so that $\|f - p\|_\infty < \varepsilon$. But note that while the polynomials have dense linear span in $C_{\mathbb{R}}([0,1])$, it is not true that every continuous function $[0,1]$ can be written as an infinite combination $\sum_{n=0}^{\infty} a_n x^n$ with convergence in $C_{\mathbb{R}}([0,1])$: power series are always infinitely differentiable. An explicit proof of Weierstrass' approximation theorem using Bernstein polynomials is given as a bonus question on the example sheets.

We aim for Marshall Stone's vast generalisation of Weierstrass's approximation theorem in 1937. First note that if K is a metric space and $x \neq y$ in K, then there exists $g \in C_{\mathbb{R}}(K)$ with $g(x) \neq g(y)$: we can take $g(z) = d(z,x)$.[45]

Definition 5.3. Let K be a compact metric (or compact Hausdorff) space. We say that a subset $D \subset C_{\mathbb{R}}(K)$ *separates points* if for all $p,q \in K$ with $p \neq q$ there exists a function $g \in D$ so that $g(p) \neq g(q)$.

[44] The theorem we're aiming for works for compact Hausdorff topological spaces, and we shall prove it in that generality, but since Part A topology is not a prerequisite for this course I will state the results both for compact metric spaces and compact Hausdorff topological spaces! You can ignore all references to compact Hausdorff spaces if you prefer and, and there is absolutely no harm in imagining that the compact space K is contained in $\mathbb{R}^n$ if you prefer. But for some of the abstract applications of Stone–Weierstrass that I use regularly it's necessary to have compact Hausdorff spaces. And, when we give the proof, the only thing that will matter is that a closed subset of a compact metric space or compact Hausdorff topological space is again compact.

That said one fundamental difference between working with metric spaces and general topological spaces is that for a metric space K it is always easy to produce continuous functions $f : K \to \mathbb{R}$: for each $x \in K$, the function $d(\cdot, x)$ is a continuous function on K. For a general topological space X there might be no non-constant continuous functions: a crude example is give by taking X to be any set equipped with the indiscrete topology, so that all points are dense. In a compact Hausdorff topological space K one has Urysohn's lemma which shows that for any two disjoint closed subsets A,B in K there exists a continuous function $f : K \to [0,1]$ with $f(a) = 0$ for $a \in A$ and $f(b) = 1$ for $b \in B$. More generally Urysohn's lemma works for a *normal space*, one for which any two disjoint closed subsets are contained in disjoint open sets: showing that compact Hausdorff spaces are normal is a standard exercise in using compactness. I'll say a bit more about Urysohn's lemma in some footnotes / deep dives below, but we do not use it in the proof of Stone–Weierstrass: the Stone–Weierstrass theorem *assumes* that the subspace contains enough continuous functions to separate points (and the proof contains an argument which upgrades this to separate disjoint closed subsets as in Urysohn's lemma).

[45] The same is true for a compact Hausdorff space K, using Urysohn's lemma.

Let K be a compact metric (or compact Hausdorff) space. We are interested in dense subspaces $D \subset C_{\mathbb{R}}(K)$. By Lemma 5.1 any dense subspace $D \subset C_{\mathbb{R}}(K)$ must separate points. We will be interested in those subspaces $D \subset C_{\mathbb{R}}(K)$ which *contain constant functions* i.e. for each $\lambda \in \mathbb{R}$, the function $f(k) = \lambda$ for all $k \in K$ lies in D. In this case we can use the vector space operations to rescale separation of points as in the following lemma (which is left as an exercise).

Lemma 5.4. *Let K be a compact metric (or compact Hausdorff) space and let $D \subset C_{\mathbb{R}}(K)$ be a subspace containing constant functions. The following are equivalent:*

1. *D separates points,*
2. *for any $p \neq q \in K$, $\exists g \in D$ with $g(p) = 0$ and $g(q) = 1$,*
3. *for any $p \neq q \in K$, $a, b \in \mathbb{R}$, $\exists g \in D$ with $g(p) = a$ and $g(q) = b$.*

Definition 5.5. Let K be a compact metric or (compact Hausdorff) space. Say that $A \subset C_{\mathbb{R}}(K)$ is a *subalgebra* of $C_{\mathbb{R}}(K)$ if it is a subspace with the property that A is closed under pointwise multiplication, i.e. if $f, g \in A$, then $fg \in A$.

We now reach the Stone–Weierstrass theorem for real coefficients. We come back to the case of complex coefficients in the next subsection.

Theorem 5.6 (Stone–Weierstrass theorem: real version)**.** *Let K be a compact metric (or compact Hausdorff) space, and let $A \subset C_{\mathbb{R}}(K)$ be a subalgebra which contains constant functions and separates points. Then A is dense in $C_{\mathbb{R}}(K)$.*

Warning. Traditionally the Stone–Weierstrass theorem above is proved by means of first proving a Stone–Weierstrass theorem for lattices, and then deducing Theorem 5.6 from the lattice theorem. In this way Theorem 5.6 is known as the 'subalgebra version' of Stone–Weierstrass in the 2023-24 lecture notes. There's no formal syllabus change here: the syllabus speaks of 'the Stone–Weierstrass theorem', but you will certainly find the lattice (and subalgebra) version on previous exams. We did not lecture the lattice version this year, and you can be confident that if relevant the exam will be clear that we are using a subalgebra version of the theorem. Similarly the previous years courses did not consider the complex scalar version of Stone–Weierstrass we give in Subsection 5.3 (and so that is unlikely to be found on many past papers). I will describe the lattice version in a deep dive at the end of the subsection.

We will follow a proof from the 1980's, the key step of which is contained in the next lemma which uses compactness to upgrade the separation of points to separation of disjoint closed sets. Recall that a closed subset of a compact metric (or a compact Hausdorff) space is compact.

Lemma 5.7. *Let K be a compact metric (or compact Hausforff space), and let $A \subset C_{\mathbb{R}}(K)$ be a subalgebra which contains constant functions and separates points. Then for E, F disjoint closed subsets of K there exists $f \in A$ with $-1 \leq f(x) \leq 1$ for all $f \in K$, and $f(x) \leq -1/2$ for $x \in E$ and $f(x) \geq 1/2$ for $x \in F$.*

Proof. Fix $x \in E$. We claim that there exists $g_x \in A$ with $g_x(x) = 0$, $0 \leq g_x \geq 1$, and $g_x(y) > 0$ for all $y \in F$. To prove this, for each $y \in F$, there exists $h_y \in A$ with $h_y(x) = 0$ and $h_y(y) > 0$ and $h_y \geq 0$ (using Lemma 5.4 to get the first and second condtiion; then replace h_y by h_y^2 to ensure that $h_y \geq 0$ everywhere, using that A is a subalgebra). Then $U_y = \{z \in K : h_y(z) > 0\}$ is an open set containing y, so by compactness of F, there exists $y_1, \ldots, y_r \in F$ such that $U_{y_1} \cup \cdots \cup U_{y_r} \supseteq F$. Then

$$g_x = \frac{\sum_{j=1}^{r} h_{y_j}}{\| \sum_{j=1}^{r} h_{y_j} \|_\infty} \in A$$

and has $g_x(x) = 0$, $0 \leq g_x \leq 1$, and $g_x(y) > 0$ for $y \in F$ as claimed.

Note that g_x is a continuous function so attains its minimum on the compact set F, so we can fix $m_x \in \mathbb{N}$ such that $g_x(z) \geq 2/m_x$ for all $z \in F$.

We now run a second compactness argument. Let $V_x = \{x \in K : g_x(z) < 1/(2m_x)\}$ which is an open subset of K containing x. By compactness of E, there exists $x_1, \dots, x_s \in E$ such that $V_{x_1} \cup \cdots \cup V_{x_s} \supseteq E$. For $i = 1, \dots, s$ and $n \in \mathbb{N}$, Bernouli's inequality[46] gives

$$(1 - g_{x_i}^n)^{m_{x_i}^n} \geq 1 - (m_{x_i} g_{x_i})^n.$$

Therefore on the set V_{x_i} (where $m_{x_i} g_{x_i} < 1/2$) we have

$$(1 - g_{x_i}(z)^n)^{m_{x_i}^n} \geq 1 - (m_{x_i} g_{x_i}(z))^n \geq 1 - \frac{1}{2^n} \to 1, \quad z \in V_{x_i}$$

as $n \to \infty$. Now we consider the same function restricted to F, where $m_{x_i} g_{x_i} \geq 2$. Using the difference between squares at the first inequality (as $0 \leq g_{x_i} \leq 1$), and estimating crudely at the second (again using $g_{x_i} > 0$), we have

$$(1 - g_{x_i}(z)^n)^{m_{x_i}^n} \leq \frac{1}{(1 + g_{x_i}(z)^n)^{m_{x_i}^n}} \leq \frac{1}{(m_{x_i} g_{x_i}(z))^n} \leq \frac{1}{2^n} \to 0, \quad z \in F.$$

Then

$$1 - (1 - g_{x_i}(z)^n)^{m_{x_i}^n} \leq \frac{1}{2^n} \to 0, \quad z \in V_{x_i} \text{ and } 1 - (1 - g_{x_i}(z)^n)^{m_{x_i}^n} \geq 1 - \frac{1}{2^n} \to 1, \quad z \in F.$$

For each $i = 1, \dots, s$ we can find $n_i \in \mathbb{N}$ such that

$$h_i = 1 - (1 - g_{x_i}(z)^{n_i})^{m_{x_i}^{n_i}}$$

has $0 \leq h_i \leq 1$ and $h_i(z) \leq 1/4$ for $z \in V_i$, while $h_i(z) \geq (3/4)^{1/s}$ for $z \in F$. Note that as A is an algebra, $h_i \in A$, and also so too is $h = h_1 h_2 \dots h_s$. Then $0 \leq h \leq 1$ and $h(z) \leq 1/4$ for $z \in \bigcup_{i=1}^s V_i \supseteq E$ and $h(z) \geq 3/4$ for $z \in F$. Taking $f = 2h - 1$ (which is also in A) proves the lemma. $\square$

We can now prove the real Stone–Weierstrass theorem.

Proof of Theorem 5.6. Let A be a subalgebra of $C_{\mathbb{R}}(K)$ which contains constant functions and separates points. Suppose that $\overline{A} \neq C_{\mathbb{R}}(K)$. Then by Reisz's lemma (Proposition 4.7), there exists $f \in C_{\mathbb{R}}(X)$ with $\|f\| = 1$ and $d(f, A) > 3/4$. So $-1 \leq f \leq 1$. Define disjoint closed subsets of K by

$$E = \{x \in K : f(x) \leq -1/4\} \text{ and } F = \{x \in K : f(x) \geq 1/4\}.$$

By (a scaled version of) Lemma 5.7, there exists $g \in A$ with $-1/2 \leq g \leq 1/2$ and $g \leq -/4$ on E and $g \geq 1/4$ on F. Then $\|f - g\| \leq 3/4$ by checking all cases.[47]This is a contradiction proving the theorem. $\square$

Corollary 5.8 (Weierstrass' approximation theorem)**.** *Let $a < b$. The real polynomial functions on $[a, b]$ are $\|\cdot\|_\infty$ dense in $C_{\mathbb{R}}([a, b])$.*

Proof. The polynomial functions are a subalgebra which contains constant functions and separates points. Indeed the polynomial $p(x) = x$ separates points! $\square$

[46]Bernouli's inequality is $(1 + z)^n \geq 1 + nz$ for $z \geq -1$ and $n \in \mathbb{N}$.

[47]If $z \in E$, then $-1 \leq f(z), g(z) \leq -1/4$ so $|f(z) - g(z)| \leq 3/4$ and similarly if $z \in F$, then $14 \leq f(z), g(z) \leq 1$. While if $z \notin E \cup F$, then $-1/4 \leq f(z) \leq 1/4$ while $-1/2 \leq g(z) \leq 1/2$, so $|f(z) - g(z)| \leq 3/4$.

Corollary 5.9. *Let K be a compact subset of $\mathbb{R}^n$. Then polynomials in n-variables restricted to K are dense in $C_{\mathbb{R}}(K)$.*

Example 5.10 (An application of Weierstrass's theorem)**.** We claim that the only continuous real valued function $f \in C_{\mathbb{R}}([0,1])$ for which

$$\int_0^1 f(t)t^n dt = 0 \text{ for every } n \in \mathbb{N}$$

is the zero function. To see this, we let $X = C_{\mathbb{R}}([0,1])$ (as always equipped with the sup norm) and we note that any function $f \in C_{\mathbb{R}}([0,1])$ induces a bounded linear functional $F \in X^* = \mathscr{B}(X,\mathbb{R})$ defined by $F(h) = \int_0^1 f(t)h(t)dt$, where we note that F is bounded since $|F(h)| \leq \|f\|_\infty \|h\|_\infty$, so $\|F\|_{X^*} \leq \|f\|_\infty$. If f satisfies the displayed equation then, by linearity, $F(p) = 0$ for every polynomial. Since the polynomials are dense in X we can thus apply Lemma 5.1 to obtain that $F = 0$, in particular $F(f) = \int f^2(t)dt = 0$. But as $f^2 \geq 0$ this implies that $f^2 = 0$ a.e. and so as f is continuous indeed $f = 0$.

Deep Dive

Let's come back and discuss the lattice versions of Stone–Weierstrass. Fix a compact metric (or compact Hausdorff) space K. A *linear sublattice* $L \subset C_{\mathbb{R}}(K)$ is a subspace L which has

$$f,g \in L \Rightarrow \max(f,g) \in L \text{ and } \min(f,g) \in L.$$

Theorem (Real Stone-Weierstrass-Theorem, lattice form)**.** *Let K be a compact metric (or compact Hausdorff). Let L be a linear sublattice of $C_{\mathbb{R}}(K)$ which contains constant functions and separates points. Then L is dense in $C_{\mathbb{R}}(K)$.*

The strategy to prove (which can be found in full in the 2023-24 lecture notes) is as follows. Fix some $f \in C_{\mathbb{R}}(K)$ and $\varepsilon > 0$. Then:

- show that for each $x \in K$, there is some $g_x \in L$ with $g_x(x) = f(x)$ and $g_x(y) < f(y) + \varepsilon$ for all $y \in K$. (For each y there exists $h_y \in L$ with $h_y(x) = f(x)$ and $h_x(y) = f(y)$. Then $h_y < f$ on an open neighbourhood of y. Taking the minimum of a number of these h_y given by compactness yields the claim).
- on a suitable small neighbourhood of x we have $f - \varepsilon < g_x$, while $g_x < f + \varepsilon$ globally. Running a second compactness argument (and now taking a finite maxima) gives a single $g \in L$ with $f - \varepsilon < g < f + \varepsilon$.

Note how the double use of compactness to control g first from above and then from below resonates with the proof of Lemma 5.7.

One then normally deduces the real subalgebra version of Stone–Weierstrass from the lattice version by means of showing that:

Lemma. Let K be a compact metric (or compact Hausdorff) space, and let $A \subset C_{\mathbb{R}}(K)$ be a subalgebra. Then the closure of A is a linear sublattice.

For this we need to show that if $f \in \overline{A}$, then $|f| \in \overline{A}$. Since $|f| = \sqrt{f^2}$, this is done by showing that if $g \in A$ with $g \geq 0$, then $g^{1/2} \in \overline{A}$. The 2023-24 notes give details of how to use the contraction mapping theorem to do this; proofs can also be found in the recommended books, as the lattice approach is the standard way to obtain Stone–Weierstrass. Since we've already proved the algebra version of Stone–Weierstrass we can cheat and deduce the lemma by approximating the function $g(t) = t^{1/2}$ on the compact set $[-\|f\|_\infty, \|f\|_\infty]$

by polynomial function. Using such approximations (which we could get from the classical Weierstrass approximation theorem) gives the lemma.

It is also possible to obtain the lattice version of Stone–Weierstrass from the subalgebra version by showing that if L is a linear sublattice of $C_{\mathbb{R}}(K)$ which contains constants, then the closure $\overline{L}$ is a subalgebra, for which it suffices to check that $f \in A \implies f^2 \in A$ (by means of the usual trick $(f+g)^2 = f^2 + 2fg + g^2$).[a] Note that if L is a linear lattice containing constants, then for $f \in L$ and $\lambda \in \mathbb{R}$, the element $\max(f-\lambda, 0) \in L$. On the interval $[-\|f\|_\infty, \|f\|_\infty]$ we can uniformly approximate the function $g(t) = t^2$ by a suitable finite linear combination of functions of the form $t \mapsto \max(t-\lambda, 0)$.[b]. Using such approximations we can approximate f^2 by elements of L. That said, I think this isn't a sensible thing to do: if we want the lattice version of Stone–Weierstrass best is to follow the double compactness argument above and just prove it directly.

[a] Note the cute symmetry: to get from the lattice to the algebra version of Stone–Weierstrass we approximate the square root function; to get from the algebra to the lattice version we approximate the square function.

[b] We don't need a Weierstrass approximation theorem for this, it can be done by hand.

Deep Dive

Recall that in integration we showed that the continuous functions are dense in L^p (for $p < \infty$). Although off syllabus we can improve this to obtain density of compactly supported smooth functions inside L^p-spaces for $p < \infty$; this is often very useful in applications to PDES. Let's first do so via a combination of Stone–Weierstrass plus the existence of smooth bump functions (which appears as a bonus prelims exercise in analysis 2).[a]

Theorem. *Let $1 \le p < \infty$. Then $C_c^\infty(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} | f \text{ is smooth and has compact support}\}$ is dense in $L^p(\mathbb{R})$.*[b].

Proof. Let $f \in L^p(\mathbb{R})$ and $\varepsilon > 0$. We already know from integration that there is a continuous function g of compact support on $\mathbb{R}$ such that $\|f-g\|_p < \varepsilon/2$. Let $I_1 = [-M, M]$ be such that $g(x) = 0$ for all $x \notin I_1$ and $I_2 = [-M-1, M+1]$. Then $A = \{h \in C_{\mathbb{R}}(I_2) : f \text{ is smooth}\}$ is a subalgebra of $C_{\mathbb{R}}(I_2)$, which is dense by Stone–Weierstrass. For $\delta = \varepsilon/2(2M+2)^{1/p}$ we can find $h \in A$ with $\sup_{x \in I_2} |h(x) - g(x)| \le \delta$. If we let $b : \mathbb{R} \to [0,1]$ be a bump function which is 1 on I_1 and 0 outside I_2 then we can view hb as defined on all of $\mathbb{R}$ where it has compact support and $\sup_{x \in \mathbb{R}} |h(x)b(x) - g(x)| \le \delta$. Then $\|hb - g\|_p^p \le (2M+2)\delta^p = (\varepsilon/2)^p$. □

We could produce a version of this for $\mathbb{R}^n$ but let's do so in a more explicit way, which illustrates the idea of *mollifying* an integrable function to obtain a good smooth approximation to the original function (which need not be continuous). This technique is widely applicable in the theory of and applications to PDES; for more see the C4.3 course on Functional Analytic Methods for PDEs.

Theorem. *For any $1 \le p < \infty$ and $K \subset \mathbb{R}^n$ the space $C^\infty(K)$ of smooth real valued functions is dense in $L^p(K)$.*

Proof. We let $\phi : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$\phi(x) := c \begin{cases} \exp(-\frac{1}{1-|x|^2}), & |x| < 1 \\ 0 & \textit{else} \end{cases}$$

where $c > 0$ is chosen so that $\int_{\mathbb{R}^n} \phi(x)dx = 1$ and set $\phi_\varepsilon(x) := \frac{1}{\varepsilon^n}\phi(\frac{x}{\varepsilon})$. These smooth functions ϕ_ε (which are often called 'mollification kernels' or a family of 'standard mollifiers') have $\int_{\mathbb{R}^n} \phi_\varepsilon = 1$ and are zero outside of $B_\varepsilon(0)$. One can get a sequence f_ε of smooth functions that approximates a given $f \in L^p(K)$ as follows: We extend f by zero outside of K to get a function that is defined on all of $\mathbb{R}^n$ and then set

$$f_\varepsilon := \phi_\varepsilon * f, \text{ i.e. define } f_\varepsilon(x) := \int_{\mathbb{R}} \phi(x-y)f(y)dy.$$

Then one can check that $f_\varepsilon \in C^\infty(\mathbb{R}^n)$ with derivatives $D^\alpha f_\varepsilon = (D^\alpha \phi_\varepsilon) * f$ (follows from the differentiation theorem from Part A Integration) and one can indeed prove with more care that $f_\varepsilon \to f$ in L^p. $\square$

All these statements are false for $p = \infty$ as you can easily see when trying to approximate step functions by continuous functions. We give two quick sketches.

[a] Precisely, given a closed and bounded interval I contained in an open interval U, there exists a smooth function $f : \mathbb{R} \to [0,1]$ with $f(x) = 1$ for $x \in I$ and $f(x) = 0$ for $x \in \mathbb{R} \setminus U$.

[b] Compact support means that the closure of the set $\{x : f(x) \neq 0\}$ is compact, i.e. f has compact suppose if f is zero outside a compact subset of $\mathbb{R}$.

5.3 Complex Stone–Weierstrass

We now turn to complex coefficients. In keeping with my preferences for scalars I will write $C(K)$ for $C_{\mathbb{C}}(K)$, the complex valued continuous functions on a compact metric (or compact Hausdorff) space K.

Example 5.11. Let $\mathbb{D}$ be the open unit disc in $\mathbb{C}$ and let $K = \overline{\mathbb{D}}$ be the closed unit disc in $\mathbb{C}$. Define $A = \{f \in C(\overline{D}) : f|_{\mathbb{D}} \text{ is holomorphic}\}$. This is a closed[48] subalgebra which separates points but it is not all of $C(\overline{\mathbb{D}})$.

What goes wrong in Example 5.11 is that $f(z) = \bar{z}$ can not be uniformly approximated by holomorphic functions. But essentially this is the only problem: the algebra A is not closed under the additional operation we have with complex scalars of complex conjugation. We define a subalgebra of $C(K)$ in just the same way as for real scalars: a subspace closed under pointwise multiplication, and likewise the notion of separating points.

Theorem 5.12 (Complex version of Stone–Weierstrass)**.** *Let K be a compact metric (or compact Hausdorff) space. Let $A \subset C(K)$ be a subalgebra which:*

- *is closed under complex conjugation*
- *separates points*
- *contains constant functions.*

Then A is dense in $C(K)$.

The proof amounts to taking real and imaginary parts.

Proof. Let $A_{\mathbb{R}} = \{\mathrm{Re}\, f : f \in A\} \subset C_{\mathbb{R}}(K)$. This is subalgebra of $C_{\mathbb{R}}(K)$ which is contained in A (as $\mathrm{Re}\, f = (f + \overline{f})/2)$) which contains constants and separates points so is dense in $C_{\mathbb{R}}(K)$. Now given $f \in C(K)$, and $\varepsilon > 0$, find $g, h \in A_{\mathbb{R}}$ such that $\|\mathrm{Re}\, f - g\|, \|\mathrm{Im}\, f - h\| < \varepsilon/2$. Then $g + ih \in A$ and $\|f - (g + ih)\| < \varepsilon$. $\square$

Example 5.13. Let $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. A trigonometric polynomial is a function $\mathbb{T} \to \mathbb{C}$ of the form $p(z) = \sum_{r=-n}^{n} c_r z^r$ for some n and scalars $c_{-n}, \cdots, c_n \in \mathbb{C}$. These are the finite combinations of the terms appearing in Fourier series. By Stone–Weierstrass the trigonometric polynomials are uniformly dense in $C(\mathbb{T})$

[48] If $f_n \in A$ has $f_n \to f \in C(\overline{\mathbb{D}})$ uniformly, then f is holomorphic as the uniform limit of holomorphic functions (as a consequence of Morera's theorem.

Deep Dive

Just as we learnt that the polynomial functions are uniformly dense in $C_{\mathbb{R}}[0,1]$ but not every continuous function is given as a uniformly convergent power series, so too there are continuous functions on $\mathbb{T}$ which can not be written as a uniformly convergent infinite series $\sum_{-\infty}^{\infty} c_n z^n$. We will see a way of obtaining this from the uniform boundedness principle in B4.2.

Let $1 \leq p < \infty$. As $C(\mathbb{T})$ is dense in $L^p(\mathbb{T})$ and the trigonometric polys are dense in $C(\mathbb{T})$, and the L^p-norm is controlled in terms of the supremum norm (as $\mathbb{T}$ has finite measure), it follows that the trigonometric polynomials are dense in $L^p(\mathbb{T})$. Indeed, given $\varepsilon > 0$ and $f \in L^p(\mathbb{T})$, there exists $g \in C(\mathbb{T})$ with $\|f - g_{-}p| < \varepsilon/2$ and then a trigonometric polynomial h with $\|g-h\|_\infty < \varepsilon/2(2\pi)^{1/p}$ so that $\|g-h\|_p \leq \varepsilon/2$. In this way $\|f-h\|_p < \varepsilon$. Taking $p = 2$ it follows that the trigonometric polynomials have dense linear span in $L^2(\mathbb{T})$. Therefore the elements $e_n(z) = z^n/(\sqrt{2\pi})$ form an orthonormal basis for $L^2(\mathbb{T})$ and so by Theorem 2.18 every element $f \in L^2(\mathbb{T})$ is equal to its Fourier series with convergence in $L^2(\mathbb{T})$. This can also be proved much more directly (as discussed in B4.2).

Deep Dive

The operation of complex conjugation on $C(K)$ is something that we can abstract to a general (Banach algebra): * is an involution on the algebra A if $(x^*)^* = x$, $(xy)^* = y^*x^*$ and * is conjugate linear $(\lambda x + \mu y)^* = \overline{\lambda} x^* + \overline{\mu} y^*$ for all $x, y \in A$ and $\lambda, \mu \in \mathbb{C}$. In the setting of a Banach algebra, we would call the involution *isometric* when $\|x^*\| = \|x\|$.

The example of complex conjugation in $C(K)$ as an involution is a bit misleading as this algebra is abelian so the reason for requiring that involutions reverse the order of multiplication is not apparent. Perhaps a more familiar example is the operation of conjugate transpose on complex matrices. We will see a very important example of these involutions in the last section of the course when we look at the adjoint of an operator on a Hilbert space.

6 Separability

With tools for obtaining interesting examples of dense subspaces at hand, we now come back to separability.

6.1 Definition and basic properties

Recall:

Definition 6.1. A normed space X is *separable* if it has a countable dense subset and *inseparable* otherwise.

Separability is a topological property invariant under isomorphism (and hence under equivalence of norms). Of course it is possible for the same space to be separable in one norm, and not separable in a different non-equivalent norm.[49]

Lemma 6.2. *(i) Let X be a vector space and let $\|\cdot\|$ and $|||\cdot|||$ be two norms on X that are equivalent. Then $(X, \|\cdot\|)$ and $(X, |||\cdot|||)$ are either both separable or both inseparable.*

(ii) Let X and Y be isomorphic normed spaces. Then X and Y are either both separable or both inseparable.

[49] For example $L^\infty[0,1]$ is separable in the L^1-norm but not in the L^∞-norm.

Just as when we start out with sets we tend to view countable as reasonably small, and uncountable as large, separability is the right notion of *smallness* for metric and normed spaces. As we will see, separability often allows us to perform arguments powered by countability. If we want to view separability as being small, then we need to have the following proposition.

Proposition 6.3. *Let (X,d) be a separable metric space, and $Y \subset X$. Then Y is separable (with the subspace metric). In particular any subspace of a separable normed space is separable.*

Proof. Let $(x_n)_{n=1}^{\infty}$ be a countable dense sequence in X. For each $n,m \in \mathbb{N}$, if $B_{1/m}(x_n) \cap Y \neq \emptyset$ fix some $y_{m,n}$ in this intersection. Then the countable collection of those $y_{m,n}$ which have been chosen is dense in Y. □

Deep Dive

This is not true for topological spaces; there exists a topological space X with a countable dense subset, together with a subspace Y which does not have a countable dense subset (when equipped with the subspace topology). The notion for topological spaces which does pass to subspaces is being second countable; every point having a countable basis of neighbourhoods. But this is far from the same as being separable: every metric space is second countable as a topological space.

Proposition 6.4. *Let X be a normed space. Then X is separable if and only if there is a sequence $(x_n)_{n=1}^{\infty}$ with dense linear span.*

Proof. If X is separable, then it has a countable dense set, so this set certainly has dense linear span. Conversely if $(x_n)_{n=1}^{\infty}$ has dense span, then the countable set

$$\{\sum_{i=1}^{n} n \in \mathbb{N},\ \lambda_i \in \mathbb{Q}(i)\}$$

is dense (use coefficients from $\mathbb{Q}$ in the case that $\mathbb{F} = \mathbb{R}$).[50] □

6.2 Examples and non-examples

We have the following examples of separable normed spaces.

- Finite dimensional normed spaces are separable: they have a finite basis, so certainly a sequence with dense linear span.
- Both c_0 and the sequence spaces ℓ^p for $1 \leq p < \infty$ as the subspace c_{00} of finitely supported sequences (spanned by the canonical elements $e_n = (0,\dots,0,1,0,\dots)$ withe 1 in the n-th position) is dense in all these spaces. Indeed for ℓ^p, give any element $x = (x_1,\dots) \in \ell^p$ we can now use that since $\sum_{j=1}^{\infty} |x_j|^p$ converges, the cut-off sequences $x^{(k)} := (x_1,\dots,x_k,0,0,\dots)$ approximate x in the sense of ℓ^p, namely $\|x - x^{(k)}\|_{\ell^p} = \left(\sum_{j\geq k+1} |x_j|^p\right)^{1/p} \to 0$ as $k \to \infty$. We thus conclude that c_{00} is dense in ℓ^p. Likewise for $(x_n) \in c_0$, the condition that $x_n \to 0$, shows that the cut off sequences $\|x^{(k)} - x\|_{\infty} \to 0$.
- For $K \subset \mathbb{R}^n$ compact, $C(K)$ is separable, as by Stone–Weierstrass the monomial functions have dense linear span.

[50] I leave the details as an exercise.

- For $K \subset \mathbb{R}^n$ compact, and $1 \leq p < \infty$, $L^p(K)$ is separable. One way to do this is to argue in the same way we showed that trigonometric polynomials are dense in $L^p(\mathbb{T})$ by combining:
  - $C(K)$ is dense in $L^p(K)$
  - $C(K)$ is norm separable
  - $\|f\|_p \leq \mu(K)^{1/p}\|f\|_\infty$ for $f \in C(K)$.

  Alternatively, and more directly for $K = [a,b]$, the space of step functions, that is finite linear combinations of characteristic functions of intervals, is dense in L^p. We then note that given any interval $[c,d] \subset [a,b]$ with real endpoints, we can choose $c_n, d_n \in \mathbb{Q}$ so that $c_n \to c$ and $d_n \to d$ and that this guarantees that $\chi_{[c_i,d_i]} \to \chi_{[c,d]}$ in L^p as $\|\chi_{[c,d]} - \chi_{[c_i,d_i]}\| \leq \big(|c-c_i| + |d-d_i|\big)^{1/p} \to 0$. Hence also the span of all characteristic functions $\chi_{[c,d]}$ of intervals with rational endpoints is dense in L^p and as the set of such functions $\{\chi_{[c,d]}, c < d, c, d \in \mathbb{Q}\}$ is countable, $L^p([a,b])$ is separable.

- More generally, the function spaces $L^p(E)$ are separable for any measurable subset $E \subset \mathbb{R}^n$ and any $1 \leq p < \infty$. It suffices to prove this for $E = \mathbb{R}^n$ (as in general $L^p(E)$ is a subspace of $L^p(\mathbb{R}^n)$. Since $\mathbb{R}^n = \bigcup_{r=1}^\infty \overline{B}_r$, where $\overline{B_r}$ is the closed ball of radius r, the dominated convergence theorem version of the baby MCT shows that $\bigcup_{r=1}^\infty L^p(\overline{B}_r)$ is dense in $L^p(\mathbb{R}^n)$. But each of these spaces is separable, so $L^p(\mathbb{R}^n)$ is separable. Alternatively, and more directly, one could use step functions corresponding to rectangles in $\mathbb{R}^n$ with rational end points.

Deep Dive

Stone–Weierstrass shows that for a compact Hausdorff space K, $C(K)$ will be separable if and only if there exists a countable collection (f_n) of continuous functions which separate points in K. Certainly if $C(K)$ has a countable dense set, then this set must separate points. Conversely, if $(f_n)_n$ is a countable collection of continuous functions which separate points, then the countable collection of the constant function 1 together with all finite products $f_{i_1} \dots f_{i_n}$ over all $n \in \mathbb{N}$ has dense span by Stone–Weierstrass.[a]

Therefore given any separable compact metric space K, the space $C(K)$ is separable, as if (x_n) is countable dense in K, then the functions $f_n(y) = d(y, x_n)$ are a countable collection of functions which separate points in K. In fact for a compact Hausdorff space K, separability of $C(K)$ is equivalent to *metrisability* of K, i.e. the existence of a metric d inducing the topology on K. We can use the techniques in C4.1 to prove the reverse direction.[b] We can be more direct (though essentially it's the same proof): if (f_n) is a countable dense subset of the unit ball of= $C(K)$, then we get a map $\theta : K \to \prod_{n=1}^\infty \overline{D}$ given by $k \mapsto (f_1(k), f_2(k), \dots,)$. Give $\prod_{n=1}^\infty \overline{\mathbb{D}}$ the product metric from one of the exercises in the metric spaces course,[c] and check that θ is a continuous bijection. Since K is compact and $\prod_{n=1}^\infty \overline{\mathbb{D}}$ is a metric space (so Hausdorff), θ is a homeomorphism onto $\theta(K)$, which is a subspace of the metric space $\prod_{n=1}^\infty \overline{\mathbb{D}}$.

[a] Note that the span of this set is closed under multiplication, so is an algebra. It is the algebra generated by f_i and 1, i.e., the smallest subalgebra of $C(K)$ containing these elements.

[b] The unit ball of $C(K)^*$ will be metrisable in the weak*-topology when you know what this means, and the map sending $k \in K$ to $\mathrm{ev}_k \in C(K)^*$, defined by $\mathrm{ev}_k(f) = f(k)$ will give a homeomorphism of K onto $\{\mathrm{ev}_k : k \in K\} \subset \mathrm{Ball}(C(K)^*)$.

[c] As $\overline{\mathbb{D}}$ is already bounded this can be given by $d((x_n),(y_n)) = \sum_{n=1}^\infty 2^{-n}|x_n - y_n|$, with the point being that $x^{(k)} = (x_n^{(k)})$ has $x^{(k)} \to y$ as $k \to \infty$ if and only if $x_n^{(k)} \to y_n$ as $k \to \infty$ for all n.

Deep Dive

Let $1 \leq p < \infty$ and $(\Omega, \mathscr{F}, \mu)$ a measure space.. We hinted in one of the earlier deep dives at when $L^p(\Omega, \mathscr{F}, \mu)$ is separable. This happens when$\otimes$ is σ-finite, and $\mathscr{F}$ is the completion of a countably generated σ-algebra. The latter condition rules out stuff like $L^p(\Omega, \mathscr{F}, \mu)$, where μ is the counting measure on an uncountable set.[a] So the result will follow in just the same way as for $L^p(\mathbb{R}^n)$ above once we show that $L^p(\Omega, \mathscr{F}, \mu)$ is separable when Ω is finite, and $\mathscr{F}$ is the completion of a countably generated σ-algebra. For this, show that the linear combinations of characteristic functions of a countable collection of subsets which generate a σ-algebra which completes to $\mathscr{F}$ is dense.

[a] You should be able to show that such a space is not separable.

Example 6.5 (ℓ^∞ and L^∞ are inseparable)**.** The sequence space $(\ell^\infty, \|\cdot\|_\infty)$ and the function spaces $L^\infty(\Omega)$, $\Omega \subset \mathbb{R}^n$ any non-empty open set, are inseparable.

We provide the proof of this result for the sequence space ℓ^∞ and note that a very similar proof, using characteristic functions of sets, shows that also L^∞ is inseparable. The standard idea for proving non-separability directly is to find an uncountable set of elements which are all a fixed distance apart.

Proof. For $I \subseteq \mathbb{N}$, let

$$x_n^{(I)} = \begin{cases} 1 & n \in I; \\ 0 & n \notin I. \end{cases}$$

Note that $A = \{x^{(I)} : I \subseteq \mathbb{N}\} \subset \ell^\infty$ and for $I \neq J$, $\|x^{(I)} - x^{(J)}\|_\infty = 1$. So no countable subset of A can be dense in A, and hence the metric space A (with the metric induced from ℓ^∞ is not separable. Therefore ℓ^∞ is not separable by Proposition 6.3. □

Corollary 6.6. *$\ell^p \not\cong \ell^\infty$ for $1 \leq p < \infty$ and $\ell^\infty \not\cong c_0$.*

Warning. $L^\infty([0,1])$ is contained in $L^1([0,1])$ but $L^1([0,1])$ is separable but $L^\infty([0,1])$ is not. This is not a contraction to Proposition 6.3, $L^\infty([0,1])$ is a vector subspace of $L^1([0,1])$ but not a subspace as a normed space (since the norm is not the same).

6.3 A small outlook

Finally we want to give a brief outlook on the use of separability and density of subspaces.

- **Bases**. We have already seen the notion of an orthonormal basis for a Hilbert space, and used these to see that all separable infinite dimensional Hilbert spaces are isomorphic (see Theorems 2.15 and 2.18). While for Hilbert spaces the notion of orthonormal basis works outside the separable setting, the appropriate notion for Banach spaces of a *Schauder basis* very much requires separability.

  Since a basis is supposed to provide a useful co-ordinate system, the usual linear algebra notion of a basis (a linearly independent spanning set, often called a Hamel basis in functional analysis), is not that useful for separable infinite dimensional Banach spaces, as every such basis is necessarily uncountable, but one of the points of separability is that the spaces is small enough that we should be able to use a countable co-ordinate system. To do so we must allow convergent infinite series: A *Schauder basis* is a sequence $(s_n)_{n=1}^\infty$ in a Banach space X such that every element $x \in X$ has a unique norm-convergent representation

$x = \sum_{n=1}^{\infty} \lambda_n s_n$ for some scalars $\lambda_n \in \mathbb{F}$.[51] A lot of work examining Banach spaces, particularly the classical sequence spaces, went through a careful analysis of bases (see the book of Albiac and Kalton), leading to nice characterisations of reflexivity in the presence of a Schauder basis in terms of properties of the basis.

One thing a basis enables you to do is obtain an infinite matrix representation for operators. We've already had a couple of examples of operators defined by infinite matrices on the problem sheet, and the spectral theorem for compact self-adjoint operators on a Hilbert space (found across B4.2 / C4.1) will give you a nice generalisation to a class of operators on a Hilbert space that can be diagonalised.

For a long time it was an open problem dating back to the foundations of the subject of Banach spaces whether every separable Banach space admits a basis. This was resolved negatively by Per Enflo in 1973 (but as you can imagine by the time this took, all the naturally occurring separable Banach spaces you're familiar with have bases).

- **Approximating by a nice set**. In applications, it is often possible to reduce the proof of a property or inequality to first proving the claim for a dense subset of "nice" elements of the space, such as smooth functions in case of L^p and then a second step that uses the density of such functions to prove that this property extends to the whole space. Similarly, as a bounded linear operator $T \in \mathscr{B}(Y,Z)$, Z a Banach space, that is defined on a dense subspace $Y \subset X$ has a unique extension to an element $T \in \mathscr{B}(X,Z)$, in many instances one defines operators first on a dense subset of "nice" elements (e.g. continuous functions) and then extends this operator to the whole space.

  Many instances of such arguments can be seen in the Part C course on Functional analytic methods for PDEs.

- **Finite dimensional approximations** For separable spaces there exists a sequence of finite dimensional subspaces $Y_1 \subset Y_2 \subset \ldots$ of X so that $\bigcup Y_i$ is dense in X. This property is used in many instances (be it to try to prove the existence of a solution of a problem, like a PDE, or more practically in numerics to obtain an approximate solution) when considering problems on separable Banach spaces (e.g. subspace of L^p, $1 \leq p < \infty$). The idea of this method (also called Galerkin's method) is to first determine solutions $x_n \in Y_n$ of approximate problems defined on the finite dimensional spaces Y_n, where results from Linear Algebra such as the rank-nullity theorem apply (and e.g. ensure that an operator $T : Y_j \to Y_j$ is invertible if and only if it is injective) and then hope to obtain that x_n converges to a solution x of the original problem (in some sense, usually one only obtains so called "weak convergence", see Part C courses on Functional Analysis and Fixed Point Methods for Nonlinear PDEs), respectively in applications in numerical analysis that x_n provides a good approximation of the solution.

  Finite dimensional approximations are also crucial in more abstract situations. A lot of my own work has focused on the classification of C^*-algebras – certain subalgebras of $\mathscr{B}(H)$ (which will appear in a later deep dive).

- Some proofs intrinsically use separability (either through back and forward arguments, or other approaches that really rely on building maps on a countable dense set through means of an inductive construction).[52]. Again in the area I work, for some results (but certainly not all) there can be a stark difference between the separable and non-separable. In particular all the classification results for C^*-algebras (and the earlier work for von Neumann algebras, going back to von Neumann!) on which they rely, use separability in an essential way and there are examples showing the results fail without it. In other places separability may be a convenience,

[51] It is necessary to keep track of the order of a basis as there are Schauder bases (s_n) with the property that some rearrangements $(s_{\sigma(n)})_{n=1}^{\infty}$ (corresponding bijections $\sigma : \mathbb{N} \to \mathbb{N}$) are not Schauder bases. Of course those bases we can 'reorder' are better (these are called unconditional Schauder bases).

[52] An example of a sort of argument like this can hopefully be found in your solution to 3.C.5!

- As a personal piece of propaganda, I believe most of modern functional analysis and its applications are to separable objects, or things built from separable objects. For example if $\mathscr{H}$ is the infinite dimensional separable Hilbert space, then $\mathscr{B}(\mathscr{H})$ is in separable,[53] but it is the bounded operators on a separable space. Moreover, it turns out we're

7 The Hahn-Banach Theorem and applications

Recall the dual space X^* of a normed space X is given by

$$X^* := \mathscr{B}(X, \mathbb{F})$$

and we know this space is complete. In all our natural examples it is not difficult to give elements of the dual space; for example for $X = \ell^p$, and $y = (y_n) \in \ell^q$ for q the Hölder conjugate of p, the function $f_y(x) = \sum x_n y_n$ gives an element $f_y \in (\ell^p)^*$ (using Hölder's inequality to show f_y is bounded). Indeed, we will come back to this in the next section to show that for $1 \leq p < \infty$, the dual of ℓ^p is given by ℓ^q. However in general it is not immediate that there are *any* non-zero continuous linear functionals on an arbitrary normed space X, i.e. why is $X^* \neq \{0\}$.[54]

Our aim is to prove the Hahn Banach theorem[55] and show that bounded linear functionals always exist, and moreover there are enough continuous functionals to see the norms of elements (the precise statement behind this slogan is Corollary 7.2).

7.1 Statement of the Hahn-Banach extension theorem

We note that if $f \in X^*$ and Y is a subspace of X (as always equipped with the same norm to turn it into a normed space), then we can restrict any $f \in X^*$ to obtain an element $f|_Y$ of Y^*, where we of course set $f|_Y(y) := f(y)$. We note that the definition of the operator norm immediately implies that $\|f|_Y\|_{Y^*} \leq \|f\|_{X^*}$.

Conversely we may ask whether we can extend a functional $g \in Y^*$ to a bounded linear operator $G \in X^*$, where we call such a G an extension of g provided $G|_Y = g$. We know how to do this if Y is dense in X (and there the extension is unique; see Theorem 5.2). Hahn–Banach shows that we can always extend bounded linear functionals from Y to X and retain control of the norm.

Theorem 7.1 (Hahn-Banach existence of a bounded extension)**.** *Let X be a (real or complex) normed space, $Y \subset X$ a subspace and let $f \in Y^*$ be any given element of the dual space of Y. Then there exists an extension $F \in X^*$ of f, i.e. an element F of X^* so that $F|_Y = f$, so that*

$$\|F\|_{X^*} = \|f\|_{Y^*}.$$

We will give the proof (in the separable case, and modulo a Zorn's lemma argument in general) in the next section. As an immediate consequence we can always find bounded linear functionals that witness the norms of elements of normed spaces, and so bounded linear functionals separate points in normed spaces.

Corollary 7.2. *Let X be a normed space. Then for any $x \in X \setminus \{0\}$ there exists an element $f \in X^*$ with $\|f\| = 1$ so that $f(x) = \|x\|$.*

[53] can you see this, for example by embedding ℓ^∞ into $\mathscr{B}(\mathscr{H})$ as a family of diagonal operators

[54] As mentioned in a deep dive in Section 3.2, we can use the axiom of choice to produce unbounded linear functionals on X, but this argument can not be adjusted to directly produce a bounded linear functional.

[55] As a warning there are many different variations and consequences of Hahn–Banach, and in papers these all tend to be referred to as 'by Hahn–Banach' leaving the reader to work out which version/consequence is intended. We'll see several more in C4.1, with my personal favourite – i.e. the one I've used the most – appearing on one of the C4.1 exercises

Proof. Let $Y = \mathrm{Span}(x)$ and define $g(\lambda x) = \lambda\|x\|$ for $\lambda \in \mathbb{F}$. Then $g \in Y^*$ with $\|g\| = 1$ and hence g has an extension $f \in X^*$ with $\|f\| = 1$ and $f(x) = g(x) = \|x\|$. □

Corollary 7.3 (Bounded linear functionals separate points)**.** *Let X be a normed space and suppose $x \neq y$ in X. Then there exists $f \in X^*$ with $f(x) \neq f(y)$.*

Proof. Apply the previous corollary to $x - y \neq 0$. □

Comparing X^* and X we get the following.

- By definition of the operator norm we have:

  For every $f \in X^*$ we have $\|f\|_{X^*} = \sup_{x\in X, \|x\|_X = 1} |f(x)|$

  but the supremum might not be attained.

- By Hahn Banach we have:

  For every $x \in X$ we have $\|x\|_X = \sup_{f\in X^*, \|f\|_{X^*}=1} |f(x)|$

  and the supremum is attained.

It's really useful to keep in mind how one of these is the definition and the other a consequence of Hahn–Banach. Notice too that by Hahn–Banach

For every $f \in X^*$ we have $\|f\|_{X^*} = \sup_{\Phi\in(X^*)^*, \|\Phi\|_{(X^*)^*}=1} |\Phi(f)|$

and here the supremum is attained. That is we might not attain $\|f\| = \sup_{\|x\|\leq 1} |f(x)| = 1$, but we will always attain the supremum when we work over elements of the unit ball of $(X^*)^*$ (a space we will come back to in the last section of the course).

Warning. The Theorem of Hahn-Banach is specific to functionals, that is maps from a vector space to the corresponding field $\mathbb{F}$, and does not hold true for linear operators between two normed spaces.

One can e.g. show that there is no continuous linear extension of the identity map $\mathrm{Id} : c_0 \to c_0$ to a map $f : \ell^\infty \to c_0$ where $c_0 \subset \ell^\infty$ denotes the closed subspace of all sequences that tend to zero.

Deep Dive

A Banach space X with the property that whenever Y is a Banach space and F is a closed subspace of Y, and $T : F \to X$ is a bounded linear operator, then there exists a bounded linear extension $\tilde{T} : F \to X$ is called *injective*. X is called *isometrically injective* when we can arrange for $\|\tilde{T}\| = \|T\|$. The Hahn Banach theorem tells is that $\mathbb{F}$ is isometrically injective, and you can use this to show (see exercise 4.C.1) that ℓ^∞ is also isometrically injective. The space c_0 is not injective. In exercise 4.C.3 we show that c_0 is not a complemented subspace of its bidual $\ell^\infty \cong (c_0)^{**}$, i.e. there is no bounded linear operator $P \in \mathscr{B}(\ell^\infty)$ with $P^2 = P$ and $P(\ell^\infty) = c_0$. From this c_0 can not be injective, as otherwise we could extend the identity operator on c_0 to such a P.

More generally, there are no infinite dimensional separable injective Banach spaces (see Chapters 5 and 6 of Albaic and Kalton).

7.2 Extending functionals controlled by sublinear functionals

We now discuss the proof of Hahn–Banach. First in the real case, and we will come back to the complex case afterwards. In the statement of Theorem 7.1 we extended bounded linear functionals; those controlled by a multiple of the norm. In fact we can extend functionals satisfying a weaker notion of control: those controlled by a sublinear functional.

Definition 7.4. Let X be a real vector space. Then $p : X \to \mathbb{R}$ is called sublinear if for every $x, y \in X$ and every $\lambda \geq 0$ we have that

$$p(x+y) \leq p(x) + p(y) \text{ and } p(\lambda x) = \lambda p(x).$$

We also note that every norm, and indeed every seminorm, on X is a sublinear functional. There are also many other constructions that yield sublinear functions that are important in applications (as discussed e.g. in Part C Further Functional Analysis), such as the so called Minkowski functional associated to each convex set C that contains the origin.

To get a simple example of a sublinear functional that is not induced by a semi-norm, we can consider any linear function $p : X \to \mathbb{R}$, or to a get a more geometric example consider $p : \mathbb{R}^n \to \mathbb{R}$ that is defined by $p(x) = \max(x_n, 0)$, i.e. that is given by the distance of a point x to the halfspace $\{x : x_n \leq 0\}$.

The general version of the Hahn-Banach extension theorem (for real vector spaces) is

Theorem 7.5 (Real Hahn-Banach extension theorem)**.** *Let X be a real vector space, $Y \subset X$ a subspace and $p : X \to \mathbb{R}$ sublinear. Suppose that $f : Y \to \mathbb{R}$ is a linear functional with the property that*

$$f(y) \leq p(y) \textit{ for all } y \in Y.$$

Then there exists a linear extension $F : X \to \mathbb{R}$ so that

$$F(x) \leq p(x) \textit{ for all } x \in X.$$

Since a norm is a sublinear functional, Theorem 7.1 (for real normed spaces) is a special case of Theorem 7.5.

Proof that Theorem 7.1 for real vector spaces follows from Theorem 7.5. Given $f \in Y^*$, let $p(x) = \|f\|_{Y^*}\|x\|$ so that $f(y) \leq p(y)$ for all $y \in Y$. Then Theorem 7.5 gives a linear extension $F : X \to \mathbb{R}$ with $F|_Y = f$ and $F(x) \leq p(x)$ for all $x \in X$. Note that we also have $-F(x) = F(-x) \leq p(-x) = p(x)$ for all x, so that $|F(x)| \leq \|f\|\|x\|$ for all $x \in X$. Therefore F is bounded and has $\|F\|_{X^*} \leq \|f\|_{Y^*}$. Since F extends f we have $\|F\|_{X^*} = \|f\|_{Y^*}$. ☐

The strategy behind the proof of Theorem 7.5 is to extend one-dimension at a time by means of the following:

Lemma 7.6 (1-step extension lemma)**.** *Let X be a real vector space, $p : X \to \mathbb{R}$ sublinear and let Y be a subspace of X and $x_0 \in X \setminus Y$. Then for any linear $f : Y \to \mathbb{R}$ for which $f(y) \leq p(y)$ for all $y \in Y$ there exists a linear extension $\tilde{f} : \mathrm{Span}(Y \cup \{x_0\}) \to \mathbb{R}$ so that*

$$\tilde{f}(x) \leq p(x) \textit{ for all } x \in \mathrm{Span}(Y \cup \{x_0\}). \tag{8}$$

Proof of Lemma 7.6. Write $\tilde{Y} = \mathrm{Span}(Y \cup \{x_0\}$ and note that every $\tilde{y} \in \tilde{Y}$ can be uniquely written as

$$\tilde{y} = y + \lambda x_0 \text{ for some } \lambda in \mathbb{R}$$

so given any number $r \in \mathbb{R}$ we obtain a well defined linear map $\tilde{f}_r : \tilde{Y} \to \mathbb{R}$ if we set

$$\tilde{f}_r(y + \lambda x_0) := f(y) + \lambda r \text{ for every } y \in Y \text{ and } \lambda \in \mathbb{R}$$

and note that $\tilde{f}_r|_Y = f$ no matter how r is chosen. We now need to show that we can choose r so that this function $\tilde{f}$ has the required property that $\tilde{f}_r(\tilde{y}) \le p(\tilde{y})$ for all $\tilde{y} \in \tilde{Y}$, which is equivalent to

$$\lambda r \le p(y + \lambda x_0) - f(y) \text{ for all } y \in Y, \lambda \in \mathbb{R}. \tag{9}$$

We first note that for $\lambda = 0$ this is trivially true no matter how r is chosen as by assumption $f \le p$ on Y.
For $\lambda > 0$ the above inequality (9) holds true if and only if

$$r \le \tfrac{1}{\lambda}\left[p(y + \lambda x_0) - f(y)\right] = p(\tfrac{1}{\lambda}y + x_0) - f(\tfrac{1}{\lambda}y)$$

for all $y \in Y$ or equivalently, setting $v = \frac{1}{\lambda}y$ and using that Y is a vector space, if and only if

$$r \le \inf_{v \in Y}\big(p(v + x_0) - f(v)\big). \tag{10}$$

For $\lambda < 0$ we write $\lambda = -|\lambda|$ to rewrite (9) as $-|\lambda| r \le p(y - |\lambda| x_0) - f(y)$. We hence obtain that (9) is satisfied for all $\lambda < 0$ and $y \in Y$ if and only if

$$r \ge -|\lambda|^{-1}(p(y - |\lambda| x_0) - f(y)) = f(|\lambda|^{-1}y) - p(|\lambda|^{-1}y - x_0),$$

i.e. if and only if r is chosen so that

$$r \ge \sup_{w \in Y}\big(f(w) - p(w - x_0)\big). \tag{11}$$

For $\tilde{f}_r$ to be the required extension we thus need to choose r so that both (10) and (11) hold, which is possible provided

$$\inf_{v \in Y}\big(p(v + x_0) - f(v)\big) \ge \sup_{w \in Y}\big(f(w) - p(w - x_0)\big).$$

However this easily follows since for any $v, w \in Y$ we have that

$$\begin{aligned}\big(p(v + x_0) - f(v)\big) - \big(f(w) - p(w - x_0)\big) &= p(v + x_0) + p(w - x_0) - f(v + w)\\ &\ge p(v + w) - f(v + w) \ge 0\end{aligned}$$

where we use the sublinearity of p in the second and the assumption that $f \le p$ on Y in the last step. $\square$

We illustrate how to use the one-step extension lemma to prove Theorem 7.1 for a separable space,

Proof of Theorem 7.1 for a separable real normed space X. Write $Y_0 = Y$, and using separability of X we can find $y_1, y_2, \dots \in X$ such that $\mathrm{Span}(Y \cup \{y_n : n \in \mathbb{N}\})$ is dense in X. Write $Y_n = \mathrm{Span}(Y \cup \{y_1, \dots, y_n\})$, and let g_0 be the original $f \in Y^*$.

Suppose inductively we have found a linear $g_n : Y_n \to \mathbb{R}$ extending g_r for $r < n$ (there is no extension requirement in the base case when $n = 0$) with $|g_n(z)| \le \|g\|\|z\|$ for all $z \in Y_n$. Then use the one step extension lemma to obtain an extension $g_{n+1} : Y_{n+1} \to \mathbb{R}$ with $|g_{n+1}(z)| \le \|g\|\|z\|$ for $z \in Y_{n+1}$. Then $X_0 = \bigcup_{n=0}^{\infty} Y_n$ is a dense subspace of X and we get a well defined element $f \in X_0^*$ given by $f(x) = g_n(x)$ for any n for which $x \in Y_n$ with $\|f\| \le \|g\|$. Extending this to X by density (using Theorem 5.2) gives the result. $\square$

Deep Dive

The (non-examinable) proof of the general version of Hahn-Banach extension theorem (Theorem 7.5) uses the one-step extension lemma together with a Zorn's lemma powered maximality argument. If you've done the set theory course I recommend having a go at doing this. If not the details will be in C4.1 (modulo Zorn's lemma, which I'm happy to view as a black box – it's equivalent to an axiom after all – I claim this is easier than the arguing using separability and density, though note that such arguments only use countable choice). An *controlled* extension of $g : Y \to \mathbb{R}$ is a pair (Y_1, g_1) with Y_1 a subspace of X containing Y and g_1 a linear extension of g to Y_1 satisfying $g_1(x) \leq p(x)$ for all $x \in Y_1$. These are ordered by saying $(Y_1, g_1) \leq (Y_2, g_2)$ if and only if $Y_1 \subset Y_2$ and $g_2|_{Y_1} = g_1$.[a] Zorn's lemma guarantees the existence of a maximal controlled extension $(\tilde{Y}, f)$ of g. Now if $\tilde{Y} \neq X$, then we can use the one-step extension lemma to extend this by one further dimension, contradicting maximality.

[a] This condition precisely says that the graph of g_1 is a subset of the graph of g_2.

For complex spaces we shouldn't use sublinear functions; these are designed to work with real scalars, but we can use seminorms:

Theorem 7.7 (Complex Hahn-Banach extension theorem)**.** *Let X be a complex normed space and $s : X \to [0, \infty)$ a seminorm on X. Given a subspace $Y \subseteq X$ and $f : Y \to \mathbb{C}$ linear with $|f(y)| \leq s(y)$ for all $y \in Y$, there exists $g : X \to \mathbb{C}$ linear with $g|_Y = f$ and $|g(x)| \leq s(x)$ for all $x \in X$.*

The proof of this theorem is to forget that X is a complex space, and view it as a real normed space and look at Re g as a real linear functional. Historically it took 10 or so years for the complex version of Hahn Banach to be deduced from the real scalar case, but I'd encourage you to have a go – it won't take you that long! It'll be discussed in C4.1.

Deep Dive

It is natural to ask how unique is a Hahn–Banach extension? The answer is given by the proof of Lemma 7.6. The one step extension is unique if and only if

$$\inf_{v \in Y} \big(p(v + x_0) - f(v)\big) = \sup_{w \in Y} \big(f(w) - p(w - x_0)\big).$$

In the case of one important application of Hahn–Banach: for each $x \neq 0$ in a normed space X, there exists $f \in X^*$ with $f(x) = \|x\|$ and $\|f\| = 1$, one can use these ideas and show:

Theorem. *Let X be a real normed space and $x_0 \in X$. There is a unique $f \in X^*$ with $\|f\| = 1$ and $f(x_0) = \|x_0\|$ if and only if the norm of X is* smooth *at x_0 in the sense that for all $z \in X$,*

$$\lim_{h \to 0} \frac{\|x_0 + hz\| - \|x_0\|}{h}$$

exists.

Again we see how geometric properties of the unit ball interact with other properties of normed spaces: here the uniqueness of 'norming functionals'. Note that no norm can be smooth at 0 (for the same reason that the modulus function is not differentiable. It's worth satisfying yourself that the norm of ℓ^p is smooth at all non-zero points if and only if $1 < p < \infty$ (you can see this with a nice picture in 2- dimensions for ℓ^1 and ℓ^∞).

There are interesting connections between the norm being smooth at all non-zero points and notions of convexity: if X^* is *strictly convex* ($\|\lambda x + (1-\lambda)y\| < 1$ whenever $0 < \lambda < 1$ and $\|x\| = \|y\| = 1$ with $x \neq y$)[a] then the norm of X is smooth at all non-zero points. Likewise if X^* has a smooth norm at all non-zero points, then X is strictly convex. You can find results of this nature set out in the book by Fabian et al.

[a] draw a picture to understand what this is saying. Note that uniformly convex spaces are strictly convex.

7.3 Further applications of the Hahn-Banach Theorem

We have already seen that the Hahn–Banach theorem shows that bounded functionals separate points (Corollary 7.3). There are many further 'Hahn–Banach separation theorems', which have a strong connection to convexity (see C4.1) and powerful applications both to PDES but also to other advanced topics in functional analysis an its applications across pure mathematics. Here's a separation theorem which works in the same way as Corollary 7.2 which separates points from closed subspaces.

Proposition 7.8. *Let X be a normed space, Y a proper closed subspace of X. Then for any $x_0 \in X \setminus Y$ there exists an element $f \in X^*$ with $\|f\| = 1$ so that*

$$f|_Y = 0 \textit{ while } f(x_0) = \mathrm{dist}(x_0, Y).$$

Note that since Y is closed we necessarily have $\mathrm{dist}(x_0, Y) > 0$.

Proof of Proposition 7.8. We define a suitable linear map g on the subspace $U = \mathrm{Span}\,(Y \cup \{x_0\})$ and then use Hahn-Banach to extend g to f. To this end we note that every $u \in U$ can be written uniquely as $u = y + \lambda x_0$ for some $\lambda \in \mathbb{R}$ and $y \in Y$ so that defining

$$g(y + \lambda x_0) := \lambda d, \text{ where } d := \mathrm{dist}(x_0, Y) > 0$$

gives a well defined linear map on Y which has the property that $g(x_0) = d$ and $g|_Y = 0$.

For any $u = y + \lambda x_0 \in U$, we have

$$\|y + \lambda x_0\| = |\lambda|\,\|x_0 - (-\lambda^{-1} y)\| \geq |\lambda| \inf_{\tilde{y} \in Y} \|x_0 - \tilde{y}\| = |\lambda|\, d = |g(y + \lambda x_0)|,$$

so that $\|g\| \leq 1$. For the reverse inequality, given $0 < \varepsilon < 1$, choose $y \in Y$ with $\|x - y\| < \frac{1}{1-\varepsilon} d$. Then

$$\frac{|g(x_0 - y)|}{\|x_0 - y\|} = \frac{d}{\|x_0 - y\|} > 1 - \varepsilon.$$

Since ε was arbitrary between 0 and 1, it follows that $\|g\| = 1$.

Now apply Hahn–Banach to obtain $f \in X^*$ with $f|_Y = 0$, $f(x_0) = d$ and $\|f\| = \|g\| = 1$. □

From this we get another proof of Reisz's lemma (Lemma 4.7).[56]

Proof that Proposition 7.8 implies Reisz's lemma. Given $0 < \varepsilon < 1$ and a proper closed subspace Y of X, let f be as in Proposition 7.8 and find $x \in X$ with $\|x\| = 1$ and $|f(x)| > 1 - \varepsilon$. For $y \in Y$, we have

$$\|x - y\| \geq |f(x - y)| = |f(x)| > 1 - \varepsilon,$$

so $\mathrm{dist}(x, Y) \geq 1 - \varepsilon$. □

Next we revisit the concepts of annihilators we saw in part A linear algebra in the context of algebraic dual spaces, and examine these analytically. Let's start with the definitions.

[56] But note that the original proof did not use the axiom of choice, whereas going via Hahn–Banach as in this proof does does. On the up side the proof makes it completely transparent why being able to find vectors which attain the norm of functionals gives rise to being able to take $\varepsilon = 0$ in Reisz's lemma from an earlier deep dive.

Definition 7.9. Given any subset $A \subset X$, we define the annihilator of A to be

$$A^\circ := \{f \in X^* : f|_A = 0\}.$$

Furthermore, for subsets $B \subset X^*$ we define

$$B_\circ := \{x \in X : f(x) = 0 \text{ for all } f \in B\} = \bigcap_{f \in B} \ker(f).$$

This is called the *pre-anhilator* of B

In the purely algebraic setting, we know[57] that if we start with a subset S of a vector space X, then form the algebriac pre-annhilator of the algebraic annhilator[58] of S we recover the span of S. This is not going to work analytically as both annihilators and pre-annhilators are closed.[59] The necessary modification is that the pre-annhilator or the annhilator gives the closed linear span of S (Theorem **??**), and we can use annhilators to test for having dense span.

Proposition 7.10. *Let X be a normed space. Then the following hold true:*

(i) Let $A \subset X$. Then $\mathrm{Span}(A)$ *is dense if and only if the annihilator of A is trivial, i.e. $A^\circ = \{0\} \subset X^*$*

(ii) If $B \subset X^*$ is so that $\mathrm{Span}(B)$ *is dense in X^* then $B_\circ = \{0\} \subset X$.*

Proof. (i) Suppose first that $\mathrm{Span}(A)$ is dense. Then for any $f \in A^\circ$, we have by linearity that also $f|_Y = 0$ where we set $Y = \mathrm{Span}(A)$. As Y is dense in X we thus get that $f = 0$ by Lemma 5.1.

Conversely, suppose that $\mathrm{Span}(A)$ is not dense. Then $Y = \overline{\mathrm{Span}}(A)$ is a closed proper subspace of X so we can choose $x_0 \in X \setminus Y$ and apply Proposition 7.8 to obtain an $f \in X^*$ with $f|_Y = 0$ and $f(x_0) = \|x_0\| \neq 0$ so have found an element $f \neq 0$ of A°.

(ii) Suppose $\mathrm{Span}(B)$ is dense in X^*. Then for any $f \in X^*$ there exists a sequence $f_n \in \mathrm{Span}(B)$ with $f_n \to f$. Therefore for $x \in B_\circ$, we have $f(x) = \lim f_n(x) = 0$. Since $f \in X^*$ is arbitrary, $x = 0$, by Hahn–Banach (in the form of Corollary 7.2). Therefore $B_\circ = \{0\}$.

□

The converse to part (ii) above is false; though it'll be easiest to wait until we have some explicit computations of dual spaces to give an example.

Deep Dive

The proof of (ii) of shows why the converse is false, as we didn't need that $f_n \to f$ in the norm of X^* just that $f_n(x) \to f(x)$ for all $x \in X$, i.e. if $\mathrm{Span}(B)$ has the property that it is dense in the weak*-topology (which will be introduced in C4.1)[a] then $B_\circ = \{0\}$. This now does characterise when a pre-anhilator vannishes: $B_\circ = \{0\}$ if and only if B has weak* dense span in X^*.

To give examples we should look for subspaces B which are weak*-dense but not norm dense. For reasons that will not be clear now (but follow from my favourite version of Hahn-Banach to be given in C4.1) it is necessary to use a non-reflexive space[b] to obtain an example, so – looking ahead – we should work with one of c_0, ℓ^1 or ℓ^∞ as these are easy to work with. Take $X = \ell^1$ when X^* is canonically identified with ℓ^∞, and let $B = c_{00} \subset \ell^\infty$. This is separable, so can not be norm dense in ℓ^∞, but identifying ℓ^∞ with $(\ell^1)^*$ we have

[57] Most likely this was only proved in finite dimensions; but the same result works using the axiom of choice in the right place to produce suitable functionals

[58] i.e. the subset $\{f \in X' : f(x) = 0 \text{ for all } x \in S\}$ of the algebraic dual space X'.

[59] The description of $B_\circ$ in terms of an intersection of kernels shows that $B_\circ$ is closed. While if $f_n \in A^\circ$ has $f_n \to f \in X^*$, then for any $x \in A$, we have $f_n(x) \to f(x)$ so $f(x) = 0$, i.e. $f \in A^circ$, and A° is closed.

$B_\circ = \{0\}$. What was going on? c_{00} is weak* dense in ℓ^∞. In fact if we start with any non-reflexive Banach space X, and consider the canonical map $j : X \to X^{**}$ from Section **??**, you can check that $j(X) \subset X^{**}$ is not norm dense, but does have $j(X)_\circ = \{0\}$ in X^*.

[a] this topology is not metrisable so we will have to avoid sequence arguments, but that doesn't cause a difficulty here.
[b] See Section **??**

Theorem 7.11. *Let A be any subset of a normed space X. Then*

$$\overline{\mathrm{Span}}(A) = (A^\circ)_\circ.$$

Proof. Note that for any $B \subseteq X^*$, we have $B_\circ$ is a closed subspace in X. Further, for $a \in A$, and $f \in A^\circ$, we have $f(a) = 0$, so that $a \in (A^\circ)_\circ$. Since $\overline{\mathrm{Span}}(A)$ is the smallest closed subspace containing A, and $(A^\circ)_\circ$ is a closed subspace containing A, it follows that $\overline{\mathrm{Span}}(A) \subseteq (A^\circ)_\circ$.

For the reverse inclusion if $x \notin \overline{\mathrm{Span}}(A)$, then by Proposition 7.8, there exists $f \in \mathrm{Span}(A)_\circ \subseteq A_\circ$ with $f(x) \neq 0$. Then $x \notin (A^\circ)_\circ$. □