Functionalanalysis 1

Melanie Rupflin

November 13, 2020

Acknowledgement: These lecture notes and the problem sheets are based on material from the lecture notes and exercises of Hilary Priestley from MT 2017 as well as some further sources, such as the lecture notes of M. Struwe and the book of D. Werner on Funktionalanalysis.

Contents

Chapter 1

Banach spaces

1.1 Definitions and basic properties

Recall:

Definition 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 so that $\forall x, y \in X$, $\forall \lambda \in \mathbb{F}$

(N1) $\|x\| \geq 0$ with $\|x\| = 0 \Leftrightarrow x = 0$

(N2) $\|\lambda x\| = |\lambda| \|x\|$

(N3) $\|x + y\| \leq \|x\| + \|y\|$ (Triangle inequality)

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

Recall that every norm $\|\cdot\|$ induces a metric

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

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

We recall in particular:

- Definition of convergence of a sequence (x_n):

$$x_n \underset{n\to\infty}{\longrightarrow} x \quad \Longleftrightarrow \quad \|x_n - x\| \underset{n\to\infty}{\longrightarrow} 0$$

- (x_n) is a Cauchy-sequence if $\forall \varepsilon > 0 \quad \exists N$ so that $\forall n, m \geq N$

$$\|x_n - x_m\| < \varepsilon$$

- A function $f : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ is continuous if and only if for every $x \in X$ and every sequence (x_n) in X

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

i.e.

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

- A set $\Omega \subset X$ is open if for every $x_0 \in \Omega$ there exists a $r > 0$ so that
$$B_r(x_0) := \{x \in X : \|x - x_0\| < r\} \subset \Omega.$$
- By definition, a set $F \subset X$ is closed if F^c is open and we have the following equivalent characterisations of closed sets:
  - F is closed if and only if F contains all its limit points
  - F is closed if and only if for every sequence (x_n) that consists of elements $x_n \in F$ and that converges $x_n \to x$ in X we have that the limit x is again an element of F.
- We also recall that $x \mapsto \|x\|$ is a continuous map and hence that if $x_n \to x$ then of course also $\|x_n\| \to \|x\|$.

Notation: We will use the convention that $A \subset B$ simply means that A is a subset of B, not necessarily a proper subset, i.e. allowing for $A = B$. If our assumption is that A is proper subset of B then we will either explicitly say so or write $A \subsetneqq B$.

We also recall that two norms $\|\cdot\|$ and $\|\cdot\|'$ are equivalent if and only if there exist a constant $C > 0$ so that for all $x \in X$
$$C^{-1}\|x\| \leq \|x\|' \leq C\|x\|,$$
or equivalently if there exist two constants $C_{1,2} \in \mathbb{R}$ so that for all $x \in X$
$$\|x\| \leq C_1\|x\|' \text{ and } \|x\|' \leq C_2\|x\|,$$
and that equivalent norms lead to equivalent definitions of convergence, Cauchy sequences, open and closed sets,....

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

Definition 2. A normed space $(X, \|\cdot\|)$ is a Banach space if it is complete, i.e. if every Cauchy sequence in X converges.

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

Proof.
"$\Rightarrow$":
Let (y_n) be so that $y_n \in Y$, $y_n \to x \in X$. Then (y_n) is a Cauchy sequence in Y so converges in $(Y, \|\cdot\|)$ to some $y \in Y$. Hence $x = y \in Y$ by uniqueness of limits. Hence Y is closed.

"$\Leftarrow$":
If (y_n) is a Cauchy sequence in $(Y, \|\cdot\|)$, it is also a Cauchy sequence in $(X, \|\cdot\|)$ and must hence converge in X, say $y_n \to x \in X$. But as Y is closed we must have that $x \in Y$ and hence that (y_n) converges in $(Y, \|\cdot\|)$. Thus Y is complete. □

WARNING. Many properties of finite dimensional normed spaces are NOT true for general infinite dimensional spaces, or maps between such spaces. A few examples of this are:

- Linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ (or indeed, as we shall see later, linear maps from any finite dimensional space to any normed space Y) are always continuous
  BUT
  not all linear maps $L : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ from a Banach space $(X, \|\cdot\|_X)$ are continuous.
- Bounded, closed sets in $\mathbb{R}^n$ are compact (Heine-Borel-Theorem)
  BUT
  while compact sets are always bounded and closed, the converse is WRONG for infinite dimensional spaces
- Every subspace of $\mathbb{R}^n$ is a closed set
  BUT
  not all subspaces of infinite dimensional spaces are closed.

Our intuition can further be wrong as we are used to thinking about Euclidean spaces $\mathbb{R}^n$ whose norm is introduced by an inner product via $\|x\| = (x, x)^{1/2}$.

We recall that an inner product $(\cdot, \cdot) : X \times X \to \mathbb{R}$ is a map that is symmetric $(x, y) = (y, x)$ if $\mathbb{F} = \mathbb{R}$, respectively hermitian $(x, y) = \overline{(y, x)}$ if $\mathbb{F} = \mathbb{C}$, that is linear in the first variable and positive definite and call a vector space X together with an inner product $(\cdot, \cdot)$ an inner product space.

An important special case of Banach spaces are spaces whose norm is induced by an inner product and these spaces will play a key role in the HT course B4.2 Functional Analysis 2.

Definition 3. A Hilbert space is an inner product space $(X, (\cdot, \cdot))$ which is complete (wrt the induced norm $\|x\| = (x, x)^{1/2}$).

WARNING. There are several important properties that hold true in $\mathbb{R}^n$, and more generally in Hilbert spaces, but that do not hold for general Banach spaces. Examples of this include

- In $\mathbb{R}^n$ (and indeed any Hilbert space, cf. B4.2 Functional Analysis 2) minimal distances to closed subspaces are attained, i.e. given any closed subspace $S \subset X$ of a Hilbert space X and any $p \in X$ there exists a unique element $s_0 \in S$ so that
  $$\|p - s_0\| = \inf_{s \in S} \|p - s\|.$$
  In Banach spaces this does not hold true in general.
- If $\mathbb{R}^n = W \oplus V$ for two *orthogonal* subspaces W and V then the projection $P_V : v + w \mapsto v$ is so that
  $$\|P_V(z)\| \leq \|z\|.$$
  This is not true for general direct sums $\mathbb{R}^n = W \oplus V$ of subspaces that are not orthogonal (a picture illustrates that nicely) and is in particular not true for general Banach spaces $X = W \oplus W$ where there is not even a notion of "orthogonal".

1.2 Examples

$(\mathbb{R}^n, \|\cdot\|_p)$, $1 \leq p \leq \infty$

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 Δ-inequality

$$\|x+y\|_p = \big(\sum_i (x_i+y_i)^p\big)^{1/p} \le \|x\|_p + \|y\|_p.$$

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 $p<1$.

A useful property to deal with the p norms $1\le p\le\infty$ (and their generalisations to sequence and functions spaces) is Hölder's inequality

Lemma 1.2 (Hölder's inequality in $\mathbb{R}^n$)**.** *For $1\le p,q\le\infty$ with*

$$(\star) \qquad \frac{1}{p}+\frac{1}{q}=1$$

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

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

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

The proof of this inequality (both for $\mathbb{R}^n$ as well as the analogues for the sequence and function spaces ℓ^p and L^p) can be found in most textbooks. Here we will simply use Hölder's inequality without proof.

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

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

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 $1\le p,q\le\infty$.

Sequence spaces $(\ell^p,\|\cdot\|_p)$

An infinite dimensional analogue of $(\mathbb{R}^n,\|\cdot\|_p)$, respectively $(\mathbb{C}^n,\|\cdot\|_p)$ are the spaces of sequences $(\ell^p,\|\cdot\|_p)$, $1\le p\le\infty$, where for $1\le p<\infty$

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

while ℓ^∞ denotes the space of bounded sequences, equipped with $\|\cdot\|_p$ where for $1\le p<\infty$

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

while for $p=\infty$

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

For any $1\le j\le\infty$ we have that $(\ell^p,\|\cdot\|_p)$ is a normed space (where we define addition and scalar-multiplication component-wise) and one can furthermore prove:

- the spaces $(\ell^p, \|\cdot\|_p)$ are all complete and hence Banach spaces, we carry out the proof of this for $p = 2$ in the next section.
- the Hölder inequality holds true, i.e. for every $1 \le p, q \le \infty$ 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 and

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

- In particular, for $(x_j), (y_j) \in \ell^2$

$$\big((x_j), (y_j)\big) := \sum x_j \overline{y}_j$$

is well defined and one can easily check that this is an inner product that induces the ℓ^2-norm, hence making $(\ell^2, \|\cdot\|_2)$ into a Hilbert-space.

We will sometimes also consider the subspace

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

of ℓ^∞, which is closed and hence, when equipped with the ℓ^∞- norm a Banach space.

Function spaces $(L^p(\Omega), \|\cdot\|_{L^p})$, $1 \le p \le \infty$

Let $\Omega \subset \mathbb{R}$ be an interval, or more generally any measurable subset of $\mathbb{R}^n$. Consider for $1 \le p < \infty$ the space of functions

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

respectively

$$\mathcal{L}^\infty := \{f : \Omega \to \mathbb{R} \text{ measurable so that } \exists M \text{ with } |f| \le M \text{ a.e. }\}.$$

Here and in the following all integrals are computed with respect to the Lebesgue measure and we shall only ever consider functions that are measurable so you may assume in any application that the functions you encounter are measurable without having to provide a proof for this. Conversely, we recall that 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

$$\|f\|_{L^p} := \Big(\int_\Omega |f|^p dx\Big)^{1/p} \text{ for } 1 \le p < \infty$$

respectively

$$\|f\|_{L^\infty} := \text{ess sup}|f| := \inf\{M : |f| \le M \text{ a.e. }\}.$$

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

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

The resulting quotient space

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

is one of the most important spaces of functions in the modern theory of PDE (as developed e.g. in the course C4.3 Functional analytic methods for PDEs) and has the following properties: For any (measurable) set $\Omega \subset \mathbb{R}^n$

- $L^p(\Omega)$, $1 \leq p \leq \infty$ is a Banach space (completeness of L^1 was proven in A.4 Integration)
- The so called Minkowski-inequality (=triangle inequality for $\|\cdot\|_{L^p}$) holds true
$$\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$$
- 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
$$\Big|\int_\Omega fg dx\Big| \leq \|f\|_{L^p}\|g\|_{L^q}.$$
- In particular, for $f, g \in L^2(\Omega)$
$$(f,g)_{L^2} := \int f\overline{g}\, dx$$
is a well defined inner product that induces the L^2 norm, so L^2 is a Hilbert space.

None of the L^p norms are equivalent, though for bounded domains (and sets with finite measure), we can estimate the L^p norm of functions by their L^q norm if $p < q$ and have that for any $1 < p < q < \infty$

$$L^\infty(\Omega) \subsetneqq L^q(\Omega) \subsetneqq L^p(\Omega) \subsetneqq L^1(\Omega). \tag{1.1}$$

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,

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

so we get $\|f\|_{L^2} \leq \sqrt{2}\|f\|_{L^4}$ 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. The inclusion (1.1) 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 $L^p(\mathbb{R})$, $1 \leq p < \infty$.

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 $[0,\infty)$.

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.

Function spaces with supremum-norm

If we consider vector spaces of bounded functions $f : \Omega \to \mathbb{F}$, Ω some given subset of $\mathbb{R}$ or $\mathbb{R}^n$, such as

- $\mathcal{F}^b(\Omega) := \{f : \Omega \to \mathbb{F} \text{ bounded }\}$
- $C_b(\Omega) := \{f : \Omega \to \mathbb{F} \text{ continuous and bounded }\}$

or on compact sets simply $C(\Omega) := \{f : \Omega \to \mathbb{F} \text{ continuous}\} = C_b(\Omega)$, we can consider the supremum norm, denoted either by $\|\cdot\|_\infty$ or $\|\cdot\|_{sup}$ or (in case of C_b) often also by $\|\cdot\|_{C^0}$ that is simply defined by

$$\|f\|_{sup} := \sup\{|f(x)| : x \in \Omega\}.$$

Similarly, on spaces of differentiable functions (with bounded derivatives) such as $C^1([0,1])$ we will generally use norms that are built using the sup norm of both the function and its derivative such as $\|f\|_{C^1} := \|f\|_{sup} + \|f'\|_{sup}$.

It is important to note that convergence with respect to the supremum norm is the same as uniform convergence of functions, so as seen in Prelims and Part A analysis lectures, one often proves convergence of a given sequence f_n in three steps: First we prove that the sequence converges pointwise to some function f which is then the only candidate for the limit of f_n as uniform convergence implies pointwise convergence. We then need to check that f is in the corresponding space and finally to establish uniform convergence of f_n to f.

Product of normed spaces

Given two normed spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ we can define a norm on $X \times Y$ e.g. by

$$\|(x,y)\| = (\|x\|^2 + \|y\|^2)^{1/2} \tag{1.2}$$

or more generally using any of the p-norms on $\mathbb{R}^2$ to define

$$\|(x,y)\| := \|(\|x\|, \|y\|)\|_p = (\|x\|^p + \|y\|^p)^{1/p} \text{ respectively } \|(x,y)\| := \max(\|x\|, \|y\|)$$

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.

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 (1.2) 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.

Sums of subspaces

If $X_{1,2} \subset X$ are subspaces of a normed space $(X, \|\cdot\|_X)$ 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

WARNING.

$$X_{1,2} \subset X \text{closed} \not\Rightarrow X_1 + X_2 \text{ closed}.$$

Quotientspaces

Given a vector space X and a seminorm $|\cdot|$ 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.

This is the process whereby L^p spaces are obtained from the corresponding $\mathcal{L}^p$ spaces by identifying functions which are equal a.e.

1.3 Completeness

The spaces discussed above are all complete. The proof of completeness often follows the following rough pattern: Given a Cauchy sequence (x_n) in a normed space $(X, \|\cdot\|_X)$

1. Identify a candidate x for $\lim x_n$
2. Show that $x \in X$ and $\|x - x_n\|_X \to 0$ as $n \to \infty$.

We illustrate this by proving the completeness of some of the spaces introduced in the previous section:

Completeness of $(C_b(\Omega, \mathbb{R}), \|\cdot\|_{sup}), \quad \Omega \subset \mathbb{R}^n$

Given a Cauchy sequence (f_n) in $(C_b(\Omega), \|\cdot\|_{sup})$ we have that for every $x \in \Omega$

$$|f_n(x) - f_m(x)| \to 0 \text{ as } n, m \to \infty$$

i.e. $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$ so, as $\mathbb{R}$ is complete, converges to some limit. We define as candidate for the limit of the sequence of functions f_n the function $f(x) := \lim_{n\to\infty} f_n(x)$ obtained by this pointwise convergence and now show that

Claim: $f \in C_b(\Omega)$ and $\|f_n - f\|_{sup} \to 0$ (i.e. $f_n \to f$ uniformly).

Proof. Let $\varepsilon > 0$. As (f_n) is a Cauchy sequence, there exists some N so that for every $n, m \geq N$

$$\|f_n - f_m\|_{sup} \leq \varepsilon.$$

Thus for every $x \in \Omega$, $n \geq N$ we have

$$|f_n(x) - f(x)| = |f_n(x) - \lim_{m\to\infty} f_m(x)| \leq \varepsilon.$$

This implies in particular that f is bounded, namely that $\sup_{x\in\Omega} |f(x)| \leq \|f_N\|_{sup} + \varepsilon$, and that for every $n \geq N$, $\|f - f_n\|_{sup} < \varepsilon$. As $\varepsilon > 0$ was arbitrary this proves that f_n converges to f with respect to the supremum norm. Finally we obtain that $f \in C_b(\Omega)$ as f is uniform limit of a sequence of continuous functions and hence continuous (c.f. Analysis II and Part A Metric spaces, is proved using $\varepsilon/3$ argument).

Completeness of $(\ell^2(\mathbb{R}), \|\cdot\|_2)$

Let $(x^{(n)})$, $x^{(n)} = (x_j^{(n)})_{j\in\mathbb{N}}$, be a Cauchy-sequence in $(\ell^2, \|\cdot\|_2)$. As for every $j \in \mathbb{N}$

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

the sequence $(x_j^{(n)}) \subset \mathbb{R}$ is Cauchy so converges, say $x_j^{(n)} \underset{n\to\infty}{\longrightarrow} x_j$.
Claim: $x = (x_j) \in \ell^2$ and $\|x - x^{(n)}\|_2 \underset{n\to\infty}{\longrightarrow} 0$.
Proof: Let $\varepsilon > 0$. Then as $(x^{(n)})$ is Cauchy there exists N so that for all $n, m \geq N$

$$\|x^{(n)} - x^{(m)}\|_2 \leq \varepsilon.$$

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

$$\sum_{j=1}^{K} \left|x_j^{(n)} - x_j\right|^2 = \lim_{m\to\infty} \sum_{j=1}^{K} \left|x_j^{(n)} - x_j^{(m)}\right|^2 \leq \varepsilon^2.$$

As this holds for every K we can take $K \to \infty$ to get that $\|x^{(n)} - x\|_2^2 \leq \varepsilon^2$ for every $n \geq N$. As $\varepsilon > 0$ was arbitrary, we thus obtain that $\|x^{(n)} - x\|_2 \underset{n\to\infty}{\longrightarrow} 0$. As above we also get that $x \in \ell^2$ as

$$\|x\|_2 \overset{\Delta}{\leq} \|x^{(n)} - x\|_2 + \|x^{(n)}\|_2 < \infty.$$

(Note that here we use the above mentioned "abuse of notation" of defining $\|\cdot\|_2$ for arbitrary sequence by setting $\|x\|_2 = \infty$ if $x \notin \ell^2$ to be able to already talk of $\|x\|_2$ when we do not yet know that $x \in \ell^2$.) $\square$

Useful results to prove completeness

For the proof of completeness it is often useful to note:

Lemma 1.3. *Let (x_n) be a Cauchy sequence in a normed space $(X, \|\cdot\|)$. Then the following are equivalent:*

(i) (x_n) converges

(ii) (x_n) has a convergent subsequence

Proof. <u>(i) ⇒ (ii)</u> is trivial

<u>(ii) ⇒ (i)</u>
Suppose $x_{n_k} \to x$. Given any $\varepsilon > 0$, we can choose N so that for all $n, m \geq N$

$$\|x_n - x_m\| < \varepsilon/2$$

and furthermore choose K so that for $k \geq K$

$$\|x_{n_k} - x\| < \varepsilon/2.$$

Then for $n \geq N$ we have, choosing some $k \geq K$ so that $n_k \geq N$,

$$\|x - x_n\| \overset{\Delta}{\leq} \|x - x_{n_k}\| + \|x_{n_k} - x_n\| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

$\square$

As a consequence we obtain that a normed space is complete if and only if absolute convergence of series implies convergence of series:

Corollary 1.4. *Let $(X, \|\cdot\|)$ be a normed space. Then the following are equivalent*

(i) $(X, \|\cdot\|)$ *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. $\underline{(i) \Rightarrow (ii)}$
If $\sum_{n=1}^{\infty} \|x_n\| < \infty$ then s_n is a Cauchy sequence in $(X, \|\cdot\|)$ since for $m > n \geq N$

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

As $(X, \|\cdot\|)$ is complete we thus obtain that s_n converges to some element $s \in X$.

$\underline{(ii) \Rightarrow (i)}$
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 1.3, itself converge. □

Example (Examples of non-complete spaces)**.** 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 $C^0([0,1])$ equipped with $\|f\|_{L^1} = \int_0^1 |f| dx$ is not complete.

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 $\|f_n\|_{L^1} = \frac{1}{2n^2}$ 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 $\|\sum_{j=1}^n f_j - f\|_{L^1} \geq \frac{1}{2N^2} \nrightarrow 0$.

Chapter 2

Bounded linear operators between normed vector spaces

The most important class of maps between normed spaces are:

Definition 4. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ 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, i.e. $T(x + \alpha\tilde{x}) = Tx + \alpha T\tilde{x}$ for all $x, \tilde{x} \in X$ and $\alpha \in \mathbb{F}$, and T has the property that there exists some number $M \in \mathbb{R}$ so that

$$\|Tx\|_Y \leq M\|x\|_X \text{ for all } x \in X. \tag{2.1}$$

We let

$$L(X,Y) := \{T : X \to Y \text{ bounded linear operator }\}$$

which we always equip with the so called *operator norm*, which is defined by

$$\|T\|_{L(X,Y)} := \inf\{M : (2.1) \text{ holds true}\}, \quad T \in L(X,Y).$$

We will often abbreviate the space $L(X,X)$ of bounded linear operators from a normed space $(X, \|\cdot\|)$ to itself by $L(X)$.

We will later see that an important special case is the space of '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 $X^* := L(X, \mathbb{F})$ will be discussed in far more detail in chapters 6 and 7.

One can easily check that $\|\cdot\|_{L(X,Y)}$ is a norm on $L(X,Y)$ and as this is the only norm on $L(X,Y)$ that we shall use, we will often write for short $\|T\|$ for the norm of an operator $T \in L(X,Y)$ (provided it is clear from the context what X and Y are and with respect to which norms on X and Y the operator norm has to be computed). In applications the following equivalent expressions for the norm of an operator are often more useful than the above definition

Remark. For $T \in L(X,Y)$, $X \neq \{0\}$, we have

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

and we have in particular that for any $x \in X$

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

i.e. the infimum in the definition of the norm of a bounded linear operator is actually a minimum. Conversely, the supremum in the above expressions for the norm of an operator is in general not achieved, and we shall see examples of this later.

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 $T(x) = 0$.

One of the main reasons why $L(X, Y)$ gives a very natural class of operators between normed spaces is that it can be equivalently characterised as the space of *continuous* linear maps:

Proposition 2.1. *Let $(X, \|\cdot\|)$ and $(Y, \|\cdot\|)$ be normed spaces and let $T : X \to Y$ be linear. Then the following are equivalent:*

(i) T is Lipschitz continuous

(ii) T is continuous

(iii) T is continuous at $x_0 = 0$

(iv) $T \in L(X, Y)$.

Proof. *(i) ⇒ (ii) ⇒ (iii)* are trivial

(iii) ⇒ (iv)
Suppose that T is continuous at $x_0 = 0$ but that $T \notin L(X, Y)$, i.e. that there exists no $M \in \mathbb{R}$ so that the required inequality $\|Tx\| \leq M\|x\|$ holds for all $x \in X$. Then there exists a sequence x_n so that

$$\|Tx_n\| > n\|x_n\|.$$

Then $\tilde{x}_n := \frac{x_n}{\|Tx_n\|}$ (which is well defined as $Tx_n \neq 0$) satisfies $\|\tilde{x}_n\| \leq \frac{1}{n} \to 0$, i.e. converges to $x_0 = 0$. By continuity of T at 0 we must thus have that also $T\tilde{x}_n \to T(0) = 0 \in Y$ and hence $\|T\tilde{x}_n\| \to 0$ which contradicts the fact that by construction

$$\|T\tilde{x}_n\| = \frac{1}{\|Tx_n\|}\|Tx_n\| = 1.$$

(iv) ⇒ (i)
Let $M \in \mathbb{R}$ be so that (2.1) holds. Then as T is linear we obtain that for any $x, \tilde{x} \in X$

$$\|Tx - T\tilde{x}\| = \|T(x - \tilde{x})\| \leq M\|x - \tilde{x}\|,$$

i.e. T is Lipschitz continuous. □

In order to prove that a map $T : (X, \|\cdot\|) \to (Y, \|\cdot\|)$ is a bounded linear operator we need to

(1) Check that T is well defined, in particular that $Tx \in Y$ for all $x \in X$

(2) Check that T is linear (which is usually routine and in such situations does not need a long explanation or proof)

(3) Find some M so that for all $x \in X$

$$\|Tx\|_Y \leq M\|x\|_X.$$

As (1) and (3) often require similar arguments, in particular when working with spaces like ℓ^p or L^p where the key step is to be able to bound a sum/integral respectively to prove that it is finite, one often discusses these two steps at the same time.

We remark that to show that a linear map $T : X \to Y$ is an element of $L(X, Y)$ we just require some (possibly far from optimal) number M for which (3) holds and that any such M will be an upper bound on the operator norm. If we need to additionally determine the norm of T then we usually proceed as follows:

(i) Determine a candidate M for $\|T\|$ and show that

$$\|Tx\| \leq M\|x\| \text{ for every } x \in X.$$

This proves that $\|T\| \leq M$

(ii) Prove that there exists a sequence (x_n) in X so that

$$\frac{\|Tx_n\|}{\|x_n\|} \to M.$$

This establishes that $\|T\| \geq M$.

Instead of (ii) one might be tempted to try to find some element $x \in X$ so that $\|Tx\| = M\|x\|$, but

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.

We note that for any $T \in L(X, Y)$ 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.

2.1 Examples

Shift operators and projections on ℓ^p, $1 \leq p \leq \infty$

Define the shift operators $L, R : \ell^p \to \ell^p$ by

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

and for $k \in \mathbb{N}$ the projections $\pi_k : \ell^p \to \mathbb{F}$ by $\pi_k((x_1, x_2, x_3, \ldots)) = x_k$.

Claim: $L, R \in L(\ell^p) = L(\ell^p, \ell^p)$ with $\|L\| = \|R\| = 1$ while $\pi \in L(\ell^p, \mathbb{F}) = (\ell^p)^*$ also with $\|\pi_k\| = 1$.

Proof: Clearly all three operators are linear and well defined and for every $x \in \ell^p$ we have $\|Rx\|_p = \|x\|_p$ and hence of course $R \in L(\ell^p, \ell^p)$ with $\|R\| = 1$ (indeed R preserves norms, i.e. is so called isometric which is a much stronger property than merely having $\|R\| = 1$). For L and π_k we immediately see from the definition of the ℓ^p norm that

$$\|Lx\|_p \leq \|x\|_p \text{ as well as } |\pi_k(x)| \leq \|x\|_p$$

so that both are bounded linear operators (namely $L \in L(\ell^p, \ell^p)$ and $\pi_k \in (\ell^p)^*$) and the corresponding operator norms are bounded from above by $\|L\| \leq 1$ and $\|\pi_k\| \leq 1$. To see that also $\|L\| \geq 1$ we may use that $\|L(0,1,0,\ldots)\|_p = \|(1,0,\ldots)\|_p = 1 = \|(0,1,0,\ldots)\|_p$, while choosing $x = e^{(k)}$, the sequence that is defined by $e^{(k)} = (\delta_{kj})_{j\in\mathbb{N}}$, we also get that $1 = |\pi_k(x)| = \|x\|_p$ and hence that $\|\pi_k\| \geq 1$.

Definition 5. We call a linear function $T : X \to Y$ isometric if for every $x \in X$ we have $\|Tx\| = \|x\|$.

We note that if $T \in L(X, Y)$ is both isometric and bijective, then we have that also T^{-1} is linear and isometric (so in particular a bounded linear operator) as for every $x \in X$

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

Such a map is called an *isometric isomorphism* and the spaces X and Y are called isometrically isomorphic, written for short as $X \cong Y$.

Multiplication by functions (i)

Let $X = C^0([0,1])$, as always equipped with the supremum norm and let $g \in C^0([0,1])$. Then

$$T : X \to X \text{ defined by } (Tf)(x) := f(x)g(x)$$

is linear, well defined (as the product of continuous functions is continuous) and bounded as

$$\|Tf\|_{sup} \leq \|g\|_{sup} \|f\|_{sup}.$$

In particular $\|T\| \leq \|g\|_{sup}$ and choosing $f \equiv 1$ we get $Tf = g$ so as $\|f\|_{sup} = 1$ also

$$\|T\| = \sup_{h\in X, \|h\|_{sup}=1} \|Th\|_{sup} \geq \|Tf\| = \|g\|_{sup},$$

so indeed $\|T\| = \|g\|_{sup}$.

Multiplication by functions (ii)

Consider instead $g \in L^\infty([0,1])$ and let $X = L^2([0,1])$ (equipped of course with the L^2 norm). Then the map $T : X \to X$ defined as above is well defined as

$$\int_0^1 |(Tf)(t)|^2 dt = \int_0^1 f^2(t)g^2(t)dt \leq \|g\|_{L^\infty}^2 \int |f(t)|^2 dt$$

so

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

and thus $\|T\| \leq \|g\|_{L^\infty}$. Indeed one can show that $\|T\| = \|g\|_{L^\infty}$, though to prove this for general functions requires a careful argument using some techniques from Part A integration, that are not used elsewhere in the course. We hence only consider as an example $g(t) = t$. Then $\|g\|_{L^\infty} = 1$, so the above calculation implies that $\|T\| \leq 1$ while choosing $f_n := \chi_{[1-\frac{1}{n},1]}$ gives

$$\|Tf_n\|_{L^2}^2 = \int_{1-\frac{1}{n}}^1 t^2 dt \geq \tfrac{1}{n}(1 - \tfrac{1}{n})^2,$$

so as $\|f_n\|_{L^2}^2 = \frac{1}{n}$ we have $\frac{\|Tf_n\|_{L^2}}{\|f_n\|_{L^2}} \geq 1 - \frac{1}{n} \to 1$ so also $\|T\| \geq 1$ and hence $\|T\| = 1 = \|g\|_{L^\infty}$.

At the same time one can show that for 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 $X = L^2([0,1])$.

Linear maps between Euclidean Spaces

We know that any linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ can be written as

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

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

$$\|A\| := \big(\sum_{i,j} |a_{ij}|^2\big)^{\frac{1}{2}}$$

which is also called 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 $T : \mathbb{R}^n \to \mathbb{R}^m$

Lemma 2.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $Tx = Ax$ for some $A \in M_{m\times n}(\mathbb{R})$ where we equip $\mathbb{R}^n$ and $\mathbb{R}^m$ with the Euclidean norm. Then $T \in L(\mathbb{R}^n, \mathbb{R}^m)$ and its operator norm is bounded by the Hilbert-Schmidt norm of A*

$$\|T\| \leq \|A\|.$$

Remark. For most matrices we have

$$\|T\| < \|A\|$$

and computing $\|T\|$ can be difficult. For *symmetric* $n \times n$ matrices however we can easily show (using material from Prelims Linear Algebra) that

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

Proof of Lemma 2.2.

$$\|Tx\|^2 = \sum_{i=1}^m (Ax)_i^2 = \sum_{i=1}^m \big(\sum_{j=1}^n a_{ij}x_j\big)^2 \overset{\text{C.S.}}{\leq} \sum_{i=1}^m \big(\sum_{j=1}^n a_{ij}^2\big) \cdot \big(\sum_{j=1}^n x_j^2\big) = \|A\|^2\|x\|^2$$

□

Integral operator on $C([0,3],\mathbb{R})$:

Let $X = C([0,3])$ as always equipped with the sup-norm. Given any $k \in C([0,3]\times[0,3])$ we map each $x \in X$ to the function $Tx : [0,3] \to \mathbb{R}$ that is given by

$$Tx(t) := \int_0^3 k(s,t)x(s)ds$$

where the integral is well defined as the integrand is bounded,

$$|k(s,t)x(s)| \leq \|k\|_{sup} \cdot \|x\|_{sup},$$

and thus (Lebesgue) integrable over the bounded interval $[0,3]$. Here the supremum norms of k and x are computed over the corresponding domains, i.e. $[0,3]\times[0,3]$ respectively $[0,3]$.

Claim: $T \in L(X)$

Proof. T is obviously linear and for any $t \in [0,3]$ we can bound

$$|Tx(t)| \leq \int_0^3 |k(s,t)x(s)|ds \leq 3\|k\|_{sup}\|x\|_{sup}.$$

Provided we show that $T : X \to X$ is actually well defined, we will thus obtain that $T \in L(X)$ with $\|T\| \leq 3\|k\|_{sup}$. To prove that T is well defined we have to show that for any function $x \in C([0,3])$ also Tx is continuous on $[0,3]$, i.e. that for any $t_0 \in [0,3]$ and any sequence $t_n \to t_0$ $Tx(t_n) \to Tx(t_0)$. To this end we set $f_n(s) := k(s,t_n)x(s)$ and $f(s) := k(s,t_0)x(s)$ and observe that

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

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

$$\lim_{n\to\infty}(Tx)(t_n) = \lim_{n\to\infty}\int_0^3 f_n(s)ds \overset{DCT}{=} \int_0^3 \lim_{n\to\infty} f_n ds = \int_0^3 f(s)ds = (Tx)(t_0)$$

as claimed.

□

2.2 Properties of (the space of) bounded linear operators

2.2.1 Completeness of the space of bounded linear operators

An important property of the space of bounded linear operators is that it "inherits" the completeness of the target space.

Theorem 2.3. *Let $(X, \|\cdot\|)$ be any normed space and let $(Y, \|\cdot\|)$ be a Banach space. Then $L(X,Y)$ (equipped with the operator norm) is complete and thus a Banach space.*

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

$$\|T_n x - T_m x\| \leq \|T_n - T_m\|\,\|x\| \underset{n,m\to\infty}{\longrightarrow} 0$$

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 $L(X,Y)$ and $T_n \to T$ in $L(X,Y)$, 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 $x \in X$ we thus have

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

Hence T is bounded (as $\|Tx\| \leq (\|T_n\| + \varepsilon)\|x\|$ for all x) and so an element of $L(X,Y)$ 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 $L(X,Y)$. □

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

2.2.2 Composition and invertibility of bounded linear operators

Given any normed spaces $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$ and any linear operators $T \in L(X,Y)$ and $S \in L(Y,Z)$ we can consider the composition $ST = S \circ T : X \to Z$ and observe that

Proposition 2.4. *The composition ST of two bounded linear operators $S \in L(Y,Z)$ and $T \in L(X,Y)$ between normed spaces X, Y, Z is again a bounded linear operator and we have*

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

Proof. Clearly ST is linear. Given any $x \in X$ we can furthermore bound

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

which implies the claim. □

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

$$S_n T_n \to ST \text{ in } L(X,Z)$$

since

$$\|S_n T_n - ST\| \overset{\Delta}{\leq} \|(S_n - S)T_n\| + \|S(T_n - T)\| \leq \|S_n - S\|\|T_n\| + \|S\|\|T_n - T\| \to 0$$

where we use in the last step that $\|T_n\|$ is bounded since T_n converges.

We also note that for operators $T \in L(X)$ from a normed space $(X, \|\cdot\|)$ to itself we can consider the composition of T with itself, and more generally powers $T^n = T \circ T \circ \ldots \circ T \in L(X)$ which, by the above proposition have norm

$$\|T^n\| \leq \|T\|^n.$$

We conclude in particular

Remark. Let X be a Banach space and let $A \in L(X)$. Then

$$\exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

converges in $L(X)$ and hence $\exp(A)$ is a well defined element of $L(X)$.

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 2.3, also $L(X)$ is complete we hence obtain from Corollary 1.4 that the series converges. □

In many applications, including spectral theory as discussed in chapter 8, the following lemma turns out to be useful to prove that an operator is invertible:

Lemma 2.5 (Convergence of Neumann-series). *Let X be a Banach space and let $T \in L(X)$ be so that $\|T\| < 1$. Then the operator $Id - T$ is invertible with*

$$(Id - T)^{-1} = \sum_{j=0}^{\infty} T^j \in L(X).$$

Here and in the following we use the following definition.

Definition 6. An element $T \in L(X)$ is called invertible (short for *invertible in* $L(X)$) if there exists $S \in L(X)$ so that $ST = TS = \mathrm{Id}$.

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 $ST = TS = \mathrm{Id}$ (but not necessarily $S \in L(X)$).

Corollary 2.6. *Let $T \in L(X)$ be invertible. Then for any $S \in L(X)$ with $\|S\| < \|T^{-1}\|^{-1}$ we have that $T - S$ is invertible*

We will discuss the topic of invertibility of linear operators in more detail later in the course (see chapter 8)

Proof of Lemma 2.5. As $\|T\| < 1$ we know that $\sum \|A^k\| \le \sum \|A\|^k < \infty$ so, by Corollary 1.4, the series converges

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

As

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

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

Proof of Corollary 2.6. As T is invertible (which by definition means that also $T^{-1} \in L(X)$) we obtain can write $T - S = T(\mathrm{Id} - T^{-1}S)$ and note that $T^{-1}S \in L(X)$ with $\|T^{-1}S\|_{L(X)} \le \|T^{-1}\|\|S\| < 1$. By Lemma 2.5 we thus find that $(\mathrm{Id} - T^{-1}S)$ is invertible with $(\mathrm{Id} - T^{-1}S)^{-1} = \sum_{j=0}^{\infty}(T^{-1}S)^j \in L(X)$ and hence $T - S$ is the composition of two invertible operators and thus invertible, compare also Q.1 on Problem Sheet 2. □

Remark. We obtain in particular that if $T \in L(X)$ is so that $\|\mathrm{Id} - T\| < 1$ then T is invertible. Denoting by

$$GL(X) := \{T \in L(X) : T \text{ is invertible }\}$$

we thus know that the open unit ball $B_1(\mathrm{Id}) := \{T \in L(X) : \|T - \mathrm{Id}\| < 1\}$ around the identity is fully contained in $GL(X)$ and more generally that for any $T \in GL(X)$ so that $B_\delta(T) \subset GL(X)$, for $\delta = \frac{1}{\|T^{-1}\|} > 0$, so $GL(X)$ is an open subset of $L(X)$.

Remark. As you will show on Problem sheet 2, for $S \in L(X)$ algebraically invertible we have that $S^{-1} \in L(X)$ if and only if

$$(\star) \qquad \exists \delta > 0 \text{ so that } \forall x \in X \text{ we have } \|S(x)\| \ge \delta\|x\|.$$

We will furthermore see that for any $S \in L(X,Y)$ satisfying $(\star)$ we have that the image SX is closed, compare Proposition 8.1.

Chapter 3

Finite dimensional normed spaces

In this chapter 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, completeness,... in your prelims/part A courses on Linear Algebra. 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 Theorem of Heine-Borel seen in 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 and that indeed a normed space is finite dimensional if and only if the assertion of this theorem holds.

To begin with, we prove the following important special case of the equivalence of norms, upon which we shall later base the proof of this result for general finite dimensional spaces:

Proposition 3.1. *Any norm $\|\cdot\|$ on $\mathbb{R}^m$, $m \in \mathbb{N}$, is equivalent to the euclidean norm $\|x\|_2 := \big(\sum_{i=1}^m x_i^2\big)^{1/2}$ and hence all norms on $\mathbb{R}^m$ are equivalent.*

Proof. We first remark that the last part of the proposition simply follows from the transitivity of the relation of norms being equivalent, so it remains to show that for any norm $\|\cdot\|$ there exist constants $C_{1,2} \in \mathbb{R}$ so that for every $x \in \mathbb{R}^m$

$$\|x\| \leq C_1\|x\|_2 \text{ and } \|x\|_2 \leq C_2\|x\|.$$

To get the first inequality we note that for any $x = (x_1, \ldots, x_m) = \sum_{i=1}^m x_i e_i \in \mathbb{R}^m$

$$\|x\| \overset{\Delta}{\leq} \sum_{i=1}^m |x_i|\, \|e_i\| \overset{\text{C.S.}}{\leq} \big(\sum_{i=1}^m |x_i|^2\big)^{1/2}\big(\sum_{i=1}^m \|e_i\|^2\big) = C_1\|x\| \tag{3.1}$$

where we set $C_1 := \big(\sum_{i=1}^m \|e_i\|^2\big)^{1/2}$.

For the proof of the reverse inequality we give two slightly different variants, which are however based on the same core idea and use in particular the Theorem of Heine-Borel in Euclidean space $(\mathbb{R}^m, \|\cdot\|_2)$.

Variant 1 (Using that continuous functions on compact sets achieve their minimum:)
We note that the function $f(x) := \|x\|$ is a Lipschitz-continuous function from $(\mathbb{R}^m, \|\cdot\|_2)$ to $\mathbb{R}$ (though of course not an element of $L(\mathbb{R}^n, \mathbb{R})$ as not linear) as the reverse triangle inequality combined with (3.1) allows us to bound

$$|f(x) - f(y)| = \big|\|x\| - \|y\|\big| \le \|x - y\| \le C_1\|x - y\|_2.$$

As the (euclidean) unit sphere $S = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is a closed and bounded subset of $(\mathbb{R}^n, \|\cdot\|_2)$ and hence by Heine-Borel compact, we know that $f|_S$ achieves its minimum in some point $x^* \in S$. As $\|\cdot\|$ is a norm, we know that $f(x^*) > 0$ and set $C_2 := \frac{1}{f(x^*)}$. With this choice of C_2 we then get that for every $x \in X$

$$C_2\|x\| = C_2\Big\|\|x\|_2 \cdot \frac{x}{\|x\|_2}\Big\| = C_2\|x\|_2 f\Big(\underbrace{\frac{x}{\|x\|_2}}_{\in S}\Big) \ge C_2\|x\|_2 f(x^*) = \|x\|_2$$

which gives the reverse inequality claimed above.

Variant 2 (Proof by contradiction) Suppose that there exists no constant C_2 so that the inequality $\|x\|_2 \le C_2\|x\|$ holds true for every $x \in X$. Then we can choose a sequence of elements $x^{(n)} \in \mathbb{R}^m \setminus \{0\}$ so that $\|x^{(n)}\|_2 \ge n\|x^{(n)}\|$. The renormalised sequence $\tilde{x}^{(n)} = \frac{x^{(n)}}{\|x^{(n)}\|_2}$ then consists of elements of the euclidean unit sphere S which as observed above is compact and thus has a subsequence that converges $\tilde{x}^{(n_j)} \to x \in S$ with respect to the euclidean norm $\|\cdot\|_2$. As $x \in S$, we know that $x \neq 0$ and thus $\|x\| \neq 0$ which contradicts the fact that

$$\|x\| \overset{\Delta}{\le} \|x - \tilde{x}^{(n_j)}\| + \|\tilde{x}^{(n_j)}\| \le C_1\|x - \tilde{x}^{(n_j)}\|_2 + \frac{1}{n_j} \to 0.$$

□

We note that the exact same proof (replacing all $\mathbb{R}$ with $\mathbb{C}$) applies also if the field is $\mathbb{F} = \mathbb{C}$ and hence yields that all norms on $\mathbb{C}^m$ are equivalent. More generally we obtain

Theorem 3.2. *Let X be any finite dimensional space. Then any two norm $\|\cdot\|$ and $\|\cdot\|'$ on X are equivalent.*

To simplify the notation we again carry out the proof just for real vector spaces and note that the exact same proof (with all $\mathbb{R}$ replaced by $\mathbb{C}$) applies for complex vector spaces.

Proof. Let $m = \dim(X)$. Choosing a basis $f_1, \ldots, f_m$ of X we know from Prelims Linear Algebra that the map

$$Q : \mathbb{R}^m \ni (\mu_1, \ldots, \mu_m) \mapsto \sum_{i=1}^m \mu_i f_i \in X$$

is a linear bijection. Given any two norms $\|\cdot\|_X$ and $\|\cdot\|'_X$ on X we obtain two norms $\|\cdot\|_{\mathbb{R}^m}$ and $\|\cdot\|'_{\mathbb{R}^m}$ on $\mathbb{R}^m$ by defining for every $x \in \mathbb{R}^n$

$$\|x\|_{\mathbb{R}^m} := \|Q(x)\|_X \text{ respectively } \|x\|'_{\mathbb{R}^m} := \|Q(x)\|'_X.$$

We note that these norms are chosen so that the maps $Q : (\mathbb{R}^m, \|\cdot\|_{\mathbb{R}^m}) \to (X, \|\cdot\|_X)$ and $Q : (\mathbb{R}^m, \|\cdot\|'_{\mathbb{R}^m}) \to (X, \|\cdot\|'_X)$ are isometric and hence, as they are bijections, so are their inverses (i.e. Q and Q' are isometric isomorphisms). Using that, by Proposition 3.1, all norms on $\mathbb{R}^m$ are equivalent and hence that there exist constants $C_{1,2}$ so that

$$\|x\|_{\mathbb{R}^m} \le C_1\|x\|'_{\mathbb{R}^m} \text{ and } \|x\|'_{\mathbb{R}^m} \le C_2\|x\|_{\mathbb{R}^m}$$

we now conclude that for any $y \in X$

$$\|y\|_X = \|Q^{-1}(y)\|_{\mathbb{R}^m} \leq C_1 \|Q^{-1}(y)\|'_{\mathbb{R}^m} = C_1 \|y\|'_X$$

and similarly $\|y\|'_X \leq C_2 \|y\|_X$, establishing the equivalence of norms. □

Based on this result it is now easy to prove

Theorem 3.3. *Let $(X, \|\cdot\|_X)$ be a finite dimensional normed space and let $(Y, \|\cdot\|_Y)$ be any normed space (not necessarily finite dimensional). Then any linear map $T : X \to Y$ is an element of $L(X, Y)$, 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 theorem, 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 $L(X, Y)$. □

An important conclusion of this result is

Corollary 3.4. *Let $(X, \|\cdot\|)$ be a finite dimensional normed space. Then $(X, \|\cdot\|)$ is homeomorphic to $\mathbb{F}^m$, $m = \dim(X)$ and $\mathbb{F} = \mathbb{R}$ respectively $\mathbb{F} = \mathbb{C}$, and a linear homeomorphism from $\mathbb{F}^m$ to $(X, \|\cdot\|)$ can be obtained by choosing any basis $f_1, \ldots, f_m$ of X and defining*

$$Q : \mathbb{F}^m \ni (\mu_1, \ldots, \mu_m) \to \sum_{i=1}^{m} \mu_i f_i \in X. \tag{3.2}$$

We recall that a map $f : M \to \tilde{M}$ between two metric spaces is called a homeomorphism if f is invertible and both f and f^{-1} are continuous. We also recall that the image $f(C)$ of a closed set C under a homeomorphism f is again closed as it can be viewed as the preimage of C under the continuous function f^{-1}.

As we already know from Linear Algebra that Q is a bijection, this corollary immediately follows from Theorem 3.3 which implies that the linear maps Q, Q^{-1} are continuous.

Combing the equivalence of norms with the completeness of $\mathbb{R}$ and $\mathbb{C}$ furthermore allows us to prove

Theorem 3.5. *Every finite dimensional normed space $(X, \|\cdot\|)$ is complete, i.e. a Banach space.*

Proof. We first recall from Prelims Analysis and Part A metric space that $\mathbb{F}^n$, $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ equipped with the euclidean norm $\|\cdot\|_2$ is complete and remark that this can be easily proved by showing that a sequence in $\mathbb{R}^n$ converges/is a Cauchy-sequence if and only if all of its components converge/are Cauchy-sequences in $\mathbb{R}$. (We stress that this statement is wrong in infinite dimensional spaces such as the sequence spaces ℓ^p).

Let now Q be as in (3.2). Given a Cauchy-sequence (x_n) in X we conclude that since Q^{-1} is a bounded linear operator from $(X, \|\cdot\|_X)$ to $(\mathbb{F}^n, \|\cdot\|_2)$ we have

$$\|Q^{-1}(x_n) - Q^{-1}(x_m)\|_2 = \|Q^{-1}(x_n - x_m)\|_2 \leq \|Q^{-1}\| \, \|x_n - x_m\|_X \underset{n,m\to\infty}{\longrightarrow} 0,$$

i.e. that $Q^{-1}(x_n)$ is a Cauchy sequence in $(\mathbb{R}^n, \|\cdot\|_2)$ and therefore converges to some y. Setting $x = Q(y)$ we hence obtain that

$$\|x_n - x\|_X = \|Q(Q^{-1}(x_n) - y)\|_X \leq \|Q\| \, \|Q^{-1}(x_n) - y\|_2 \to 0,$$

i.e. that the original Cauchy sequence (x_n) in X converges. □

As an immediate conclusion of the above result we obtain

Corollary 3.6. *Every finite dimensional subspace of a normed vector space $(X, \|\cdot\|)$ is complete and hence closed.*

WARNING. Not every subspace of a normed vector space $(X, \|\cdot\|)$ is closed.

Example. 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

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

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

We recall that the Theorem of Heine Borel ensures that every subset of $\mathbb{R}^n$ respectively of $\mathbb{C}^n$ that is bounded and closed is automatically compact. While the reverse implication, i.e. that a compact set is always bounded and closed, is valid in every normed space (and indeed more generally in every metric space), for general normed spaces closedness and boundedness does not imply compactness. Indeed, the analogue of the Heine-Borel Theorem holds true in a normed space if and only if the space has finite dimension:

Theorem 3.7. *Let $(X, \|\cdot\|)$ 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.

Remark. We recall that by definition a set K is compact if every open cover of K has a finite subcover. We also recall that for metric spaces (and hence in particular for normed space) compactness is equivalent to sequential compactness, i.e. to the property that every sequence in K has a subsequence which converges in 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 $x_1, \ldots, x_m \in K$ so that $Y \subset \bigcup_{i=1}^m B_\varepsilon(x_i)$).

For the difficult implication in the proof of 3.7, i.e. (3) $\Rightarrow$ (1) we shall use the following useful property of closed subspaces of normed vector spaces.

Proposition 3.8 (Riesz-Lemma)**.** *Let $(X, \|\cdot\|)$ 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*

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

Proof of Proposition 3.8. 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 $d := \text{dist}(x^*, Y) > 0$, 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

$$\begin{aligned}\text{dist}(x, Y) &= \inf_{y \in Y} \|x - y\| = \inf_{y \in Y} \|\frac{x^*}{\|x^* - y^*\|} - \frac{y^*}{\|x^* - y^*\|} - y\| = \inf_{\tilde{y} \in Y} \|\frac{x^*}{\|x^* - y^*\|} - \tilde{y}\| \\ &= \inf_{\hat{y} \in Y} \|\frac{x^* - \hat{y}}{\|x^* - y^*\|}\| = \frac{\text{dist}(x^*, Y)}{\|x^* - y^*\|} \geq 1 - \varepsilon\end{aligned} \tag{3.3}$$

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 $\tilde{y} = \frac{\hat{y}}{\|x^*-y^*\|}$. □

Proof of Theorem 3.7.
(1) ⇒ (2)
Let Y be a closed and bounded set. Then the image of Y under the homeomorphism $Q^{-1} : X \to \mathbb{F}^m$ obtained in Corollary 3.4 is also closed and, as Q^{-1} is a bounded linear operator, also bounded and hence by the Theorem of Heine-Borel a compact subset of $\mathbb{R}^m$. Hence $Y = Q(Q^{-1})$ is the image of a compact set under the continuous function Q and hence itself compact.

(2) ⇒ (3):
Is trivial as S is clearly closed and bounded.

(3) ⇒ (1):
We argue by contradiction and assume that S is compact but $\dim(X) = \infty$. We may thus choose a sequence of linearly independent elements $y_k \in X$, $k \in \mathbb{N}$. Then the subspace $Y_k := \text{span}\{y_1, \ldots, y_k\} \subsetneqq Y_{k+1}$ is finite dimensional, so by Corollary 3.6, a closed proper subspace of Y_{k+1}. Applying Proposition 3.8 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 $\text{dist}(y_k, Y_k) \geq \frac{1}{2}$. In particular for every $k > l$ we have $\|y_k - y_l\| \geq \text{dist}(y_k, Y_{l+1}) \geq \text{dist}(y_k, Y_k) \geq \frac{1}{2}$ so no subsequence of (y_k) can be a Cauchy-sequence. Having thus constructed a sequence (y_k) in $S \subset X$ that does not contain a convergent subsequence we conclude that S is not sequentially compact and hence not compact leading to a contradiction. □

Remark. In the special case that X is an inner product space, rather than a general normed space, then the proof that (3) ⇒ (1) can be simplified significantly and does not require the use of Proposition 3.8: Given any sequence y_k of linearly independent elements of X, we can apply the Gram-Schmidt method from Prelims Linear Algebra to obtain a sequence x_k of orthonormal elements of X which hence have the property that $\|x_k - x_l\|^2 = \|x_k\|^2 - 2(x_k, x_l) + \|x_l\|^2 = 2$ which ensures that no subsequence of (s_n) can be Cauchy and hence that S is not sequentially compact.

Chapter 4

Density of subspaces and the Theorem of Stone-Weierstrass

4.1 Density of subspaces and extensions of bounded linear operators

We recall

Definition 7. Let $(X, \|\cdot\|)$ be a normed space. Then a subset $D \subset X$ is dense if its closure $\overline{D}$ is given by the whole space X, i.e. $\overline{D} = X$.

Remark. A useful equivalent characterisation is that a subset $D \subset X$ is dense in X if and only if for every $x \in X$ there exists a sequence of elements $y_n \in D$ so that $y_n \underset{n\to\infty}{\longrightarrow} x$, i.e. $\|x - y_n\| \underset{n\to\infty}{\longrightarrow} 0$ or equivalently if and only if for every $x \in X$ and every $\varepsilon > 0$ there exists $y \in D$ so that $\|x - y\| < \varepsilon$.

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 $(L^2([0, 1]), \|\cdot\|_{L^2})$.

Theorem 4.1. *Let $(X, \|\cdot\|_X)$ be a normed space, let Y be a dense subspace of X (which we equip with the norm of X) and let $(Z, \|\cdot\|_Z)$ be a Banach space. Then any $T \in L(Y, Z)$ has a unique extension $\tilde{T} \in L(X, Z)$, 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*

$$\|\tilde{T}\|_{L(X,Z)} = \|T\|_{L(Y,Z)}$$

We first prove the following simpler result which can be useful in applications.

Lemma 4.2. *Let $(X, \|\cdot\|_X)$ be a normed space, $D \subset X$ a dense subset and let $(Z, \|\cdot\|_Z)$ be a normed space. Then for operators $T, S \in L(X, Z)$ we have*

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

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

Proof of Lemma 4.2. We can prove the non-trivial direction "⇒" as follows: 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

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

□

Proof of Theorem 4.1. Let $x \in X$ be any element. Then as Y is dense there exists a sequence y_n of elements of Y so that $y_n \to x$.

Claim: Ty_n converges and the limit $z = \lim_{n\to\infty} Ty_n$ depends only on x and not on the chosen sequence y_n.

Once proven, this claim allows us define $\tilde{T}x := \lim_{n\to\infty} Ty_n$ to obtain a well defined map $\tilde{T} : X \to Z$. This map will be linear as T is linear, as we can interchange limits and addition/scalar multiplication and know that the obtained limit is independent of the chosen approximating sequence. Furthermore

$$\|\tilde{T}x\|_Z = \lim_{n\to\infty} \|Ty_n\|_Z \leq \|T\| \lim_{n\to\infty} \|y_n\|_X = \|T\|\|x\|_X$$

so that $\tilde{T} \in L(X, Z)$ with $\|T\| \geq \|\tilde{T}\|$. The reverse inequality follows from the definition of the operator norm, as

$$\|\tilde{T}\|_{L(X,Z)} = \sup_{x\in X, \|x\|_X=1} \|\tilde{T}x\|_Z \geq \sup_{y\in Y, \|y\|_X=1} \|\tilde{T}y\|_Z = \|T\|_{L(Y,Z)}.$$

Hence, once the claim is proven, we obtain the desired extension which, by Lemma 4.2, is furthermore unique.

It thus remains to prove the claim. To this end we remark that if $y_n \to x$ then (y_n) is a Cauchy sequence and hence also

$$\|Ty_n - Ty_m\|_Z \leq \|T\|\|y_n - y_m\| \underset{n,m\to\infty}{\longrightarrow} 0.$$

So (Ty_n) is a Cauchy sequence in the Banach space Z and must thus converge to some limit z. To prove that the limit does not depend on the choice of the sequence of elements of Y that approximate x, let $\tilde{y}_n$ be any alternative sequence in Y that converges to x. Then the argument above implies that $T\tilde{y}_n$ converges to some limit $\tilde{z}$ and one way to see that $z = \tilde{z}$ is to consider a third sequence $\hat{y}_n \to x$ chosen as $\hat{y}_n = y_n$ for n odd and $\hat{y}_n = \tilde{y}_n$ for n even. Then also $T\hat{y}_n$ must converge to a limit $\hat{z}$ which must agree with the limit of both of the subsequences $T\hat{y}_{2n}$ and $T\hat{y}_{2n+1}$, i.e. we must have that $z = \hat{z} = \tilde{z}$ which establishes the claim and thus completes the proof of the theorem. □

4.2 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 $C(K) = C(K, \mathbb{R})$ of *real-valued* continuous functions on a compact subset $K \subset \mathbb{R}^n$. As always we equip $C(K)$ with the sup-norm and recall that since continuous functions on compact sets are bounded this is well defined.

We begin by exploring what properties are necessary for a subspace $L \subset C(K)$ to be dense. To this end we first note that given any two points $p, q \in K$ with $p \neq q$ we can choose a continuous function $g \in C(K)$ so that $g(p) \neq g(q)$, e.g. by letting $g(x) = 1 - \frac{\|x-p\|}{\delta}$ in $B_\delta(p) \cap K$ and $g \equiv 0$ outside of this ball for some number $0 < \delta < \|p - q\|$. As an aside we note that with a bit more care we could also construct such a function g which is smooth, compare also section 4.3.

We now observe that since $C(K)$ contains a function g with $g(p) \neq g(q)$, $p \neq q$ any given points, also L must have this property: Indeed, if $L \subset C(K)$ is dense, then there must be a sequence $f_n \in L$ so that $\|f_n - g\|_{sup} \to 0$ and hence in particular

$$\begin{aligned} |f_n(p) - f_n(q)| &\geq |g(p) - g(q)| - |f_n(p) - g(p)| - |f_n(q) - g(q)| \\ &\geq |g(p) - g(q)| - 2\|f_n - g\|_{sup} \to |g(p) - g(q)| > 0. \end{aligned} \tag{4.1}$$

A necessary condition for a subspace $L \subset C(K)$ to be dense is hence that it *separates points*

Definition 8. We say that a subset $D \subset C(K)$ separates points if for all $p, q \in K$ with $p \neq q$ there exists a function $g \in D$ so that $g(p) \neq g(q)$.

Remark. It can be useful to note that for a subspace $L \subset C(K)$ that contains the constant functions, the following two properties are equivalent

$$L \text{ separates points} \iff \text{for any } p \neq q \quad \exists g \in L \text{ with } g(p) = 0 \text{ and } g(q) = 1.$$

The direction "$\Leftarrow$" is trivial, while the implication "$\Rightarrow$" can be seen as follows: Given $p \neq q$ we let $\tilde{g}$ be any function in L for which $\tilde{g}(p) \neq \tilde{g}(q)$. Then defining $g(x) := \frac{\tilde{g}(x) - \tilde{g}(p)}{\tilde{g}(q) - \tilde{g}(p)}$, which is again an element of L as L is a subspace and contains the constant functions, we obtain the desired function with $g(p) = 0$ and $g(q) = 1$.

For our first density result for $C(K)$ we furthermore want our subspace to be closed under the operation of taking the (pointwise) maximum or minimum of two elements of L, i.e. to be a so called linear sublattice:

Definition 9. A subspace $L \subset C(K)$ is called a linear sublattice if

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

We note that if L is a linear sublattice then also the minimum and maximum of any finite number of elements $f_1, \ldots, f_m$ is contained in L since we can iteratively write $\max(f_1, \ldots, f_m) = \max(f_1, \max(f_2, \ldots, f_m)) = ...$ and furthermore remark that L is a sublattice if and only if

$$f \in L \Rightarrow |f| \in L$$

as one can easily check using e.g. that $|f| = \max(f, -f)$ and $\max(f, g) = \frac{1}{2}(f + g) + \frac{1}{2}|f - g|$.

We stress again that here we consider real valued functions f (and note that this definition would make no sense for complex valued functions f).

We now prove our first main result of this section, which gives a density result for general sublattices:

Theorem 4.3 (Stone-Weiserstrass-Theorem, lattice form). *Let $K \subset \mathbb{R}^n$ be a compact set and let $C(K)$ be the space of continuous real-valued functions on K which is equipped with the sup-norm. Let L be a subspace of $C(K)$ which is such that*

(i) L is a linear sublattice

(ii) L contains the constant functions

(iii) L separates points in K.

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

Remark. To see that just (i) and (iii) are not sufficient we can e.g. consider $\{f \in C([0,1]) : f(1) = 0\}$ which is a linear sublattice that separates points but is of course not dense in $C([0,1])$ as we cannot approximate any function $f \in C([0,1])$ with $f(1) \neq 0$ by elements of this space.

As a first step towards the proof of the theorem, we show:

Lemma 4.4. *Let $L \subset C(K)$ be a subspace that contains the constant functions and that separates points. Then for any function $f \in C(K)$ and any two points $p, q \in K$ there exists a function $f_{p,q} \in L$ so that*

$$f_{p,q}(p) = f(p) \text{ and } f_{p,q}(q) = f(q).$$

Furthermore, for any $\varepsilon > 0$ there exists an open neighbourhood $U_{p,q}^{\varepsilon} \subset K$ of $\{p, q\}$ in K so that

$$|f - f_{p,q}| < \varepsilon \text{ on } U_{p,q}^{\varepsilon}.$$

Proof of Lemma 4.4. If $p = q$ then we simply choose $f_{p,q}$ to be the constant function $f_{p,p} \equiv f(p)$, which by assumption is an element of L. So let $p \neq q$. As L separates points and contains the constant functions, we can choose $g \in L$ as in the above remark so that $g(p) = 0$ and $g(q) = 1$.

Then $f_{p,q}$ defined by $f_{p,q}(x) := f(p) + (f(q) - f(p)) \cdot g(x)$ is a linear combination of elements of L and hence also an element of L and has the desired properties that $f_{p,q}(p) = f(p) + 0 = f(p)$ and $f_{p,q}(q) = f(p) + (f(q) - f(p)) \cdot 1 = f(q)$.

The final claim of the lemma now simply follows from the fact that $f - f_{p,q}$ is continuous and hence $U_{p,q}^{\varepsilon} := (f - f_{p,q})^{-1}\big((-\varepsilon, \varepsilon)\big)$ is an open subset of K that contains both p and q. □

Based on this lemma we can now prove the lattice version of the Theorem of Stone-Weierstrass as follows:

Proof. Given $f \in C(K)$ and $\varepsilon > 0$ we let $f_{p,q}$ and $U_{p,q}^{\varepsilon}$ be the functions and sets obtained in the above lemma. Let $p \in K$ be any given point, which we consider to be fixed for the first step of the proof. Then $\{U_{p,q}^{\varepsilon}\}_{q \in K}$ is an open cover of K so we can find finitely many points $q_1, \ldots, q_m$ (allowed to depend on the fixed point $p \in K$) so that $K = \bigcup_{i=1}^{m} U_{p,q_i}^{\varepsilon}$.

We recall that on the sets U_{p,q_i}^{ε} we have $|f - f_{p,q_i}| < \varepsilon$ and hence in particular $f_{p,q_i} < f + \varepsilon$. Defining $g_p := \min(f_{p,q_1}, \ldots f_{p,q_m})$ we hence obtain a function $g_p \in L$ which satisfies $g_p < f + \varepsilon$ on all of K and is furthermore so that $g_p(p) = f(p)$ as all functions $f_{p,q}$ have the property that $f_{p,q}(p) = f(p)$.

To turn these functions g_p, for which we have a good upper bound on $g_p - f$, into a function g for which we have both a good upper and a good lower bound on $g - f$, we now want to take a suitable maximum of functions g_{p_i}. To this end we note that since each g_p is continuous and $g_p(p) = f(p)$ we can choose an open neighbourhood V_p of p in V (e.g. by setting $V_p := (f - g_p)^{-1}(-\varepsilon, \varepsilon))$ so that $g_p > f - \varepsilon$ on V_p. As above we can now use the compactness of K to conclude that the open cover $K = \bigcup_{p \in K} V_p$ has a finite subcover $K = \bigcup_{i=1}^{k} V_{p_i}$ and finally set

$$g := \max\{g_{p_1}, \ldots, g_{p_k}\},$$

which is in L as L is a sublattice. As g is the maximum of functions that satisfy $g_{p_i} < f + \varepsilon$ on all of K we have of course still $g < f + \varepsilon$ on K, but now know additionally that for every $x \in K$ there is some i so that $x \in V_{p_i}$ and hence $g(x) \geq g_{p_i}(x) > f(x) - \varepsilon$. Combined we thus obtain that the element $g \in L$ that we constructed satisfies $\|f - g\|_{sup} < \varepsilon$. As $\varepsilon > 0$ and $f \in C(K)$ were arbitrary this completes the proof that L is dense in $C(K)$. □

As a major application of this result we now prove:

Theorem 4.5 (Theorem of Weierstrass (respectively Stone-Weierstrass) on approximation of continuous functions by polynomials)**.** *Let $K \subset \mathbb{R}^n$ be compact. Then the space of polynomials is dense in $C(K)$, i.e. for every $f \in C(K)$ there exists a sequence of polynomials p_n on K so that $p_n \to f$ in the sense of $(C(K), \|\cdot\|_{sup})$, i.e. uniformly.*

This theorem was first proven by Weierstrass in the case of K a compact interval, while the proof of the more general form of the theorem given above is due to Stone. Hence one generally talks of the Theorem of Weierstrass if K is a compact interval and of the Theorem of Stone-Weierstrass otherwise.

We note that the space of polynomials trivially contains the constant functions and also separates points (for this already the linear functions would be sufficient). It has furthermore the extra structure of being a subalgebra of the algebra of continuous functions

Definition 10. A subspace $A \subset C(K)$ is a subalgebra if A contains the constant functions and

$$f, g \in A \Rightarrow fg \in A$$

where $(fg)(x) = f(x)g(x)$ is obtained by pointwise multiplication.

The Theorem of Weierstrass on the density of polynomials in $C(K)$ is hence a special case of the following more general result:

Theorem 4.6 (Stone-Weierstrass Theorem, subalgebra form)**.** *Let $A \subset C(K)$ be a subalgebra of $C(K)$ which separates points. Then A is dense in $C(K)$.*

We derive this theorem from the lattice form of the Stone-Weierstrass theorem by using

Proposition 4.7. *If $A \subset C(K)$ is a subalgebra of $C(K)$ that is closed then A is a linear sublattice.*

Based on this proposition which is proven below and the lattice form of the Stone-Weierstrass theorem we can now immediately prove the subalgebra form of the theorem:

Proof of Theorem 4.6. Given a subalgebra A as in the theorem we can easily check that $\bar{A}$ is also a subalgebra and hence, by Proposition 4.7, $\bar{A}$ is a linear sublattice. As A contains the constant functions and separates points the same holds true also for $\bar{A}$ so we may apply the sublattice version of the Theorem of Stone Weierstrass to conclude that $\bar{A}$ is dense in $C(K)$, so $\bar{A} = C(K)$. Hence that A is dense in $C(K)$. □

Proof of Proposition 4.7. We need to prove that if $f \in A$ then also $|f| \in A$. As A is closed, this follows provided we can construct a sequence $f_n \in A$ which converges $f_n \to |f|$ in $C(K)$, i.e. uniformly. We note that it suffices to prove this claim for elements of A with $\|f\| \leq 1$ as we may then obtain an approximating sequence for $f \in C(K)$ with $\|f\| > 1$ by setting $\tilde{f} = \frac{f}{\|f\|}$, choosing an approximating sequence $\tilde{f}_n \in A$ for $|\tilde{f}|$ and then deducing that $f_n := \|f\|\tilde{f}_n \in A$ converges to $|f|$.

So let $f \in A$ be so that $\|f\| \leq 1$. We claim that a sequence $f_k \in A$ that converges to $|f|$ can be obtained by setting $f_0 = 0$ and defining for $k = 0, 1, 2, \ldots$

$$f_{k+1} = f_k + \frac{1}{2}(f^2 - f_k^2).$$

As the f_k are obtained by taking linear combinations and products of elements of the subalgebra A we know that $f_k \in A$. We now show that for every $k \geq 0$ and every $x \in K$

$$(*) \qquad 0 \leq f_k(x) \leq |f(x)| \text{ and } f_{k+1}(x) \geq f_k(x).$$

This clearly holds true if $k = 0$ since $f_0 = 0$ and $f_1 = \frac{1}{2}f^2 \geq f_0$. Furthermore, if $(*)$ holds true for k then also

$$0 \leq f_k \leq f_{k+1} = f_k + \frac{1}{2}(|f| - f_k)(|f| + f_k) \leq f_k + |f| - f_k = |f|,$$

and

$$f_{k+2} - f_{k+1} = f_{k+1} - f_k - \frac{1}{2}(f_{k+1}^2 - f_k^2) = (f_{k+1} - f_k)(1 - \frac{1}{2}(f_{k+1} + f_k)) \geq 0$$

so, by induction, $(*)$ holds true for every k. We hence conclude that for every $x \in K$ the sequence $f_k(x)$ is bounded above and monotone increasing and thus converges to some limit $g(x) \geq 0$ which, by (AOL), must satisfy $g(x) = g(x) + \frac{1}{2}(f^2(x) - g^2(x))$, i.e. must be given by $g(x) = |f(x)|$. This establishes that $f_n \to |f|$ pointwise. Finally, to prove that f_n converges indeed in the sense of $C(K)$, i.e. uniformly, we can apply the following lemma, which is a generalisation of Dini's Theorem encountered in Prelims, to the decreasing sequence $|f| - f_n$. □

Lemma 4.8. *Let K be a compact subset of some metric space (M, d) and let $g_n : K \to \mathbb{R}$ be a sequence of continuous functions which is decreasing, i.e. so that $g_n(x) \geq g_{n+1}(x)$ for every x, and which converges pointwise to zero. Then $g_n \to 0$ uniformly.*

Proof. Let $\varepsilon > 0$ be any given number. We set $G_n := \{x \in K : g_n(x) < \varepsilon\}$ and note that since g_n is monotone the sets G_n are increasing, $G_1 \subset G_2 \supset \ldots$. As the g_n are continuous these sets are open in K and as $g_n \to 0$ pointwise, we furthermore know that $\bigcup_{n\geq 1} G_n = K$. So as K is compact, the open cover $\{G_j\}$ has a finite subcover and hence there exists a number N so that $G_N = K$. Hence for all $x \in K$ and $n \geq N$ we have $0 \leq g_n(x) \leq g_N(x) < \varepsilon$ which proves that g_n converges uniformly. □

Remark (Non-examinable)**.** There are various direct ways of proving Weierstrass's theorem on the density of polynomials e.g. in $C([0, 1])$. One can prove e.g. that given any $f \in C([0, 1])$ the so called Bernsteinpolynomials

$$p_n(t) := \sum_{i=1}^{n} \binom{n}{k} t^k (1 - t)^k f(\frac{k}{n})$$

converge to f uniformly, or follow the original proof of Weierstrass using the Weierstrass transform.

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

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

is the zero function. To see this, we let $X = C[0, 1]$ (as always equipped with the sup norm) and we note that any function $f \in C[0, 1]$ induces a bounded linear functional $F \in X^* = L(X, \mathbb{R})$ defined by $F(x) = \int_0^1 f(t)x(t)dt$, where we note that F is bounded since $|Fx| \leq \|f\|_{sup}\|x\|_{sup}$, so $\|F\|_{X^*} \leq \|f\|_{sup}$. If f satisfies $(\star)$ then, by linearity, $F(p) = 0$ for every polynomial. Since the polynomials are dense in X we can thus apply Theorem 4.1 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 $f = 0$.

4.3 Approximation of functions in Lp

In many applications where one works with L^p spaces, the following result is very useful

Theorem 4.9. *For any $1 \leq p < \infty$ and any compact set $K \subset \mathbb{R}^n$ the space $C^\infty(K)$ of smooth functions is dense in $L^p(K)$.*

WARNING. This result is wrong for $p = \infty$ as you can easily see when trying to approximate step functions by continuous functions.

The proof of this result is non-examinable (though the result and its applications are examinable), and we only sketch the proof to introduce the important concept of *mollifying* an integrable function to obtain a smooth function which is a good approximation to the given (in general not even continuous) function. This concept is widely used in theory of and applications to PDEs and more on this topic can be found in particular in the courses B4.3 Distribution Theory and C4.1 Functional analytic methods for PDEs.

Sketch of proof of Theorem 4.9 (non-examinable).
We let $\phi : \mathbb{R}^n \to \mathbb{R}$ be defined by

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

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 easily 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 that $f_\varepsilon \to f$ in L^p (though this proof requires more care and uses properties of L^p functions that we do not require elsewhere in the course). $\square$

Chapter 5

Separability

Many but not all spaces we have encountered so far have the following useful property

Definition 11. A normed space $(X, \|\cdot\|)$ is called separable if there exists a countable set $D \subset X$ which is dense. A space which is not separable is called inseparable.

To prove that a space is separable/inseparable it can be useful to note

Lemma 5.1. *(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 $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces which are isometrically isomorphic, i.e. so that there exists a linear bijection $i : X \to Y$ so that $\|i(x)\|_Y = \|x\|_x$ for all $x \in X$. Then $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are either both separable or both inseparable.

Proof. As equivalent norms lead to the same notion of convergent sequences we obtain that a set $D \subset X$ is dense in $(X, \|\cdot\|)$ if and only if it is dense in $(X, \|\cdot\|')$. Hence (i) follows. Similarly, to obtain (ii) we note that if $D \subset X$ is dense and if $i : X \to Y$ is an isometric isomorphism then $\tilde{D} := i(D) \subset Y$ is dense as for any $y \in Y$ we can choose $d_n \in D$ so that $d_n \to x := i^{-1}(y)$ and thus get a sequence $\tilde{d}_n = i(d_n) \in \tilde{D}$ that converges to y since

$$\|y - \tilde{d}_n\|_Y = \|i(x) - i(d_n)\|_Y = \|i(x - d_n)\|_Y = \|x - d_n\|_X \underset{n\to\infty}{\longrightarrow} 0.$$

□

We also recall from prelims that

- $\mathbb{Q} \subset \mathbb{R}$ and $\mathbb{Q} + i\mathbb{Q} \subset \mathbb{C}$ are dense, so both $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$ are separable
- Finite products $A_1 \times \ldots \times A_n$ and countable unions $\bigcup_{j\in\mathbb{N}} A_j$ of countable sets are countable.

This allows us to show in particular

Proposition 5.2. *Every finite dimensional normed space $(X, \|\cdot\|_X)$ is separable.*

For simplicity of notation we will carry out this proof just for real normed spaces and remark that the exact same proof, with $\mathbb{R}$ replaced by $\mathbb{C}$ and $\mathbb{Q}$ replaced by $\mathbb{Q} + i\mathbb{Q}$ applies in the complex case.

Proof. We first show that $\mathbb{R}^n$ equipped with the 1 norm is separable. Indeed, given any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and any $\varepsilon > 0$ we can use that $\mathbb{Q}$ is dense in $\mathbb{R}$ to choose $q_i \in Q$ so that $|x_i - q_i| < \frac{\varepsilon}{n}$ and hence $\|x - (q_1, \ldots, q_n)\| < \varepsilon$. As $\mathbb{Q}^n$ is countable we thus get that $(\mathbb{R}^n, \|\cdot\|_1)$ is separable. As every other norm $\|\cdot\|$ on $\mathbb{R}^n$ is equivalent to $\|\cdot\|_1$ we thus get that also $(\mathbb{R}^n, \|\cdot\|)$ is separable thanks to Lemma 5.1.

Given any other real finite dimensional vector space $(X, \|\cdot\|_X)$ we let $Q : \mathbb{R}^n \to X$ be the isomorphims introduced in (3.2) and note that Q is an isometric isomorphism if we equip $\mathbb{R}^n$ with the norm $\|x\| := \|Qx\|_X$. The separability of $(X, \|\cdot\|)$ thus follows from the separability of $(\mathbb{R}^n, \|\cdot\|)$ and Lemma 5.1. $\square$

While many of the spaces we have seen so far are separable, not all of them are and the most prominent examples of non-separable spaces are

Proposition 5.3 (ℓ^∞ and L^∞ are inseparable)**.** *The sequence space $(\ell^\infty(\mathbb{F}), \|\cdot\|_\infty)$ 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 incomplete.

Proof. We recall that the set $A := \{a = (a_1, a_2, \ldots) : a_i \in \{0, 1\}\}$ is uncountable and note that for this subset of ℓ^∞ the distance of any two elements $a \neq \tilde{a}$ is $\|a - \tilde{a}\|_\infty = 1$.

Let now D be any dense subset of ℓ^∞. Then given any $a \in A$ there must be an element $d_a \in D$ so that $\|d_a - a\| < \frac{1}{2}$ and we define a function $f : A \to D$ by assigning to each a such an element d_a. We note that $d_a = d_{\tilde{a}}$ implies that $\|a - \tilde{a}\| = \|a - d_a + d_{\tilde{a}} - \tilde{a}\| \overset{\Delta}{\leq} \|a - d_a\| + \|\tilde{a} - d_{\tilde{a}}\| < 1$ and hence that $a = \tilde{a}$ so this map is injective. Since A is uncountable, we thus obtain that any dense subset of ℓ^∞ is uncountable. Hence ℓ^∞ is inseparable. $\square$

To prove separability of spaces we can use the following two results

Lemma 5.4. *Let $(X, \|\cdot\|_X)$ be a normed space, $Y \subset X$ a subspace (which we equip as always with the same norm $\|\cdot\|_X$). Suppose that $D \subset (Y, \|\cdot\|_X)$ is dense and that $Y \subset (X, \|\cdot\|_X)$ is dense. Then also $D \subset X$ is dense.*

WARNING. Here we crucially use that the sets $D \subset Y$ and $Y \subset X$ are dense with respect to the same norm.

This lemma follows from a simple $\varepsilon/2$ argument: Given $x \in X$ and $\varepsilon > 0$ we use that Y is dense in X to choose $y \in Y$ so that $\|x - y\|_X < \frac{\varepsilon}{2}$ and then use that $D \subset Y$ is dense to choose $d \in D$ with $\|d - y\|_X < \frac{\varepsilon}{2}$.

Proposition 5.5. *Let $(X, \|\cdot\|)$ be a normed space and suppose that there exists a countable set S so that span($\bar{S}$) is dense in X. Then X is separable.*

Here we recall that the span of a subset $A \subset X$ is the set of all **finite** linear combinations, i.e.

$$\operatorname{span}(S) := \{\sum_{j=1}^{N} \lambda_j s_j : \lambda_j \in \mathbb{F},\, s_j \in S,\, N \in \mathbb{N}\}$$

and note that since $\operatorname{span}(S) \subset \operatorname{span}(\bar{S})$ the above proposition implies in particular that if there is a countable set S whose span is dense in X then X is separable

As before, we carry out the proof for real normed spaces and remark that the exact same proof, with $\mathbb{R}$ and $\mathbb{Q}$ replaced by $\mathbb{C}$ and $\mathbb{Q} + i\mathbb{Q}$, apply in the complex case.

Proof. We prove that if $\mathrm{span}(\bar{S})$ is dense, then also the set

$$Y := \{\sum_{i=1}^{N} a_i s_i : a_i \in \mathbb{Q},\, s_i \in S,\, N \in \mathbb{N}\}$$

of rational linear combinations of elements of the set S is dense in X.

Indeed given any $x \in X$ and any $\varepsilon > 0$ we can first use that $\mathrm{span}(\bar{S})$ is dense in X to determine $\bar{s}_j \in \bar{S}$ and $a_j \in \mathbb{R}$, $j = 1, \ldots, N$, so that $\|x - \sum_{j=1}^N a_j \bar{s}_j\| < \varepsilon/3$. In a second step we can now use that every element in the closure $\bar{S}$ of a set can be approximated by elements of the set S to determine $s_j \in S$ so that for every j we have $|a_j|\|s_j - \bar{s}_j\| < \frac{\varepsilon}{3N}$. Finally, we use that $\mathbb{Q}$ is dense in $\mathbb{R}$ to determine rational numbers b_i so that for every j also $|a_j - b_j| \cdot \|s_j\| < \frac{\varepsilon}{3N}$. All in all we hence obtain an element $y = \sum_{j=1}^N b_j s_j$ of Y for which

$$\begin{aligned} \|x - y\| &\overset{\Delta}{\leq} \|x - \sum_{j=1}^N a_j \bar{s}_j\| + \|\sum_{j=1}^N (a_j \bar{s}_j - b_j s_j)\| < \varepsilon/3 + \sum_{j=1}^N \|a_j \bar{s}_j - a_j s_j + a_j s_j - b_j s_j\| \\ &\overset{\Delta}{\leq} \varepsilon/3 + \sum_{j=1}^N |a_j|\|s_j - \bar{s}_j\| + \sum_{j=1}^N |a_j - b_j|\|s_j\| < \varepsilon. \end{aligned} \tag{5.1}$$

To conclude the proof of the proposition it is thus enough to show that Y is countable. Writing $S = \{s_1, s_2, \ldots\}$ we obtain a surjective map $f : A \to S$ from the set

$$A := \bigcup_{N \in \mathbb{N}} \{(a_1, \ldots) : a_k \in \mathbb{Q} \text{ and } a_k = 0 \text{ for every } k \geq N+1\}$$

of finite rational sequences to Y by defining $s(a) := \sum_j a_j s_j$ for $a \in A$ (we note that this sum is well defined as only finitely many terms are non-zero). As A is the countable union of sets that are bijective to $\mathbb{Q}^N$, we know that A is countable and hence that also Y is countable. $\square$

We are now in the position to prove that the following important Banach spaces are separable.

Proposition 5.6 (Separablility of ℓ^p and L^p for $1 \leq p < \infty$ and of $C(K)$)**.**

- *$C(K)$ is separable for any compact set $K \subset \mathbb{R}^n$.*
- *The sequence space $\ell^p(\mathbb{F})$, $F = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, is separable for $1 \leq p < \infty$.*
- *The function spaces $L^p(K)$ is separable for any compact subset $K \subset \mathbb{R}^n$ and any $1 \leq p < \infty$.*

We remark that more generally $L^p(\Omega)$ is separable for arbitrary (measurable) domains $\Omega \subset \mathbb{R}^n$ and $1 \leq p < \infty$.

Proof of Proposition 5.6.
Proof of (i):
The set of monomials $\{x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}, \alpha \in \mathbb{N}_0^n\}$ is countable and by Weierstrass's Theorem we know that its span, i.e. the space of polynomials, is dense in $C(K)$. By Proposition 5.5 we thus get that $C(K)$ is separable.

Proof of (ii): We let $Y := \mathrm{span}(S)$ where the countable set $S = \{e^{(k)}, k \in \mathbb{N}\}$ consists of all sequences $e^{(k)}$ for which the $e_j^{(k)} = \delta_j^k$.

Given any element $x = (x_1, \ldots) \in \ell^p$ we can now use that since $\sum_{j=1}^\infty |x_j|^p$ converges, we obtain that the cut-off sequences $x^{(k)} := (x_1, \ldots, x_k, 0, 0, \ldots)$ approximate x in the sense of ℓ^p, namely

$\|x - x^{(k)}\|_{\ell^p} = \big(\sum_{j\geq k+1} |x_j|^p\big)^{1/p} \to 0$ as $k \to \infty$. We thus conclude that Y is dense in ℓ^p and thus obtain from Proposition 5.5 that Y is separable.

Proof of (iii), Variant 1 (using Theorem 4.9):
From Theorem 4.9 we know that $C^\infty(K)$ is dense in $L^p(K)$ and hence in particular that $C(K)$ is dense in $L^p(K)$. We now claim that $C(K)$ is separable also when equipped with the L^p norm instead of the usual sup-norm. Indeed as $\|f\|_{L^p} \leq \big(\int_K \|f\|_{sup}^p\big)^{1/p} \leq \mathcal{L}^n(K)^{1/p}\|f\|_{sup}$ we know that if a sequence f_n converges to some element $f \in C(K)$ in the usual sense of the sup-norm, then also

$$\|f_n - f\|_{L^p} \leq \mathcal{L}^n(K)^{1/p}\|f_n - f\|_{sup} \to 0.$$

Hence any set $D \subset C(K)$ that is dense with respect to the sup-norm, will also be dense with respect to the L^p norm. In particular the set of polynomials is dense also in $(C(K), \|\cdot\|_{L^p})$. Combined with the density of $C(K) \subset L^p(K)$ and Lemma 5.4 we thus conclude that the space of polynomials is dense in $L^p(K)$. The claim that L^p is separable hence again follows from Proposition 5.5 and the fact that the space of polynomials is spanned by the countable set $\{x^\alpha, \alpha \in \mathbb{N}_0^n\}$ of monomials.

Proof of (iii), Variant 2 (using density of step functions) for $K = [a, b]$:
We use without proof the fact that 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 we obtain from Proposition 5.5 that $L^p([a, b])$ is separable. □

Once we have established that a space e.g. $C(K)$ is separable, we get for free that also any subspace (equipped with the same norm) is separable:

Proposition 5.7. *Let $(X, \|\cdot\|_X)$ be a separable normed space and let Y be a subspace of X. Then also $(Y, \|\cdot\|_X)$ is separable.*

Here it is very important that the subspace is equipped with the norm of X, not any other norm. E.g. we can see $L^\infty([0, 1])$ as a subspace of $L^1([0, 1])$ and the above proposition implies that if we were to equip $L^\infty([0, 1])$ with the L^1 norm (which is not often done in practice as we would end up with a space that is not complete) then this would give us a separable normed space, while $L^\infty([0, 1])$ equipped with the 'correct norm', i.e. the L^∞ norm, is not separable (however it is complete which in practice is more important).

Proof of Proposition 5.7. As X is separable there exists a countable dense subset $D_X := \{x_k, k \in \mathbb{N}\} \subset X$. To prove that Y is separable, we now need to determine a subset of Y that is dense in Y. To this end, we use that (by the definition of the infimum) we can choose for any $k, n \in \mathbb{N}$ an element $y_{k,n} \in Y$ so that

$$\|x_k - y_{k,n}\| \leq \operatorname{dist}(x_k, Y) + \frac{1}{n} = \inf_{y\in Y} \|x_k - y\| + \frac{1}{n}$$

note that $D := \{y_{k,n}, k, n \in \mathbb{N}\}$ is countable. We claim that $D \subset Y$ is dense. Indeed, given any $y \in Y$ and any $\varepsilon > 0$ we can first use that D_X is dense in X to find some x_k with $\|y - x_k\| < \frac{\varepsilon}{3}$, which we note implies in particular that $\operatorname{dist}(x_k, Y) < \frac{\varepsilon}{3}$. Choosing n large enough so that $\frac{1}{n} < \frac{\varepsilon}{3}$ we hence know that $\|x_k - y_{k,n}\| < \operatorname{dist}(x_k, Y) + \frac{\varepsilon}{3} < \frac{2\varepsilon}{3}$ and hence get that

$$\|y - y_{k,n}\| \leq \|y - x_k\| + \|x_k - y_{k,n}\| < \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} = \varepsilon.$$

□

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

- Existence of a basis? One might ask whether for separable space $(X, \|\cdot\|)$ there is a 'useful notion' of basis of a space, and whether in a separable space one can expect such a basis to have good properties (e.g. be countable). There are several notions of basis for normed spaces that will be discussed in part C Functional Analysis, and while every space admits a so called Hamel basis S (a set S so that every element of X can be written as *finite* linear combination of elements in S and that is so that that every finite subset of S is linearly independent), such Hamel bases are not much use in practice as one can show that a Hamel basis of a Banach space is either finite (if X is finite dimensional) or else uncountable (c.f. Part C Functional Analysis). A more useful notion of basis is that of a Schauder basis (a set $\{s_1, s_2, ...\}$ so that every $x \in X$ has a unique norm-convergent representation $x = \sum_{j=1}^{\infty} \lambda_j s_j$) and such a Schauder basis exists only for separable spaces (though as you will see in Part C Functional Analysis not for every separable space).

  In the special case of Hilbert-spaces one can show the following stronger and very useful result which you will prove in B4.3 Functional Analysis 2: Every separable Hilbert space has a countable orthonormal basis $\{e_n\}$, where a set S is called an orthogonal basis of a Hilbert space X if its elements are orthogonal, have $\|s\| = 1$ and $\mathrm{span}(S)$ is dense in X.

- Simplifications of proofs: The main application of separability in the present course will be that it will allow us to give a proof of some of our main results in case of separable spaces, most notably the Theorem of Hahn-Banach that we discuss in the next section, that avoid the use of Zorn's lemma.

- 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 $T \in L(Y, Z)$, Z a Banach space, that is defined on a dense subspace $Y \subset X$ has a unique extension to an element $T \in L(X, Z)$, 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.

- Approximating problems on infinite dimensional spaces by finite dimensional problems:

  For separable spaces there exists a sequence of finite dimensional subspaces $Y_1 \subset Y_2 \subset ...$ 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.

Chapter 6

The Theorem of Hahn-Banach and applications

The most important special case of linear operators between Banach spaces is the space of bounded linear *functionals*, i.e. bounded linear maps into $\mathbb{R}$ (respectively $\mathbb{C}$).

Definition 12. Let $(X, \|\cdot\|)$ be a normed space. Then the dual space of X is defined as

$X^* := L(X, \mathbb{F})$ equipped with the operator norm $\|f\|_{X^*} := \inf\{M : |f(x)| \leq M\|x\| \text{ for every } x \in X\}$,

where as always $\mathbb{F} = \mathbb{R}$ if $(X, \|\cdot\|)$ is a real normed space, respectively $\mathbb{F} = \mathbb{C}$ for complex spaces.

We remark that since $\mathbb{R}$ (and $\mathbb{C}$) are complete we know from Theorem 2.3 that the dual space of any normed space is complete.

We also recall that for any $f \in X^*$

$$|f(x)| \leq \|f\|\,\|x\| \text{ for all } x \in X.$$

We note that if $f \in X^*$ and Y is a subspace of X (as always equiped 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 have already seen that if Y is dense in X such an extension not only exists, but is furthermore unique and indeed the extension operator $E : Y^* \to X^*$ is an isometric isomorphims, compare Theorem 4.1. While this result holds true for linear opererators into a general Banach space, the results that we will prove in this chapter are valid only for elements of the dual space, i.e. functions that map into the corresponding field $\mathbb{F} = \mathbb{R}$ respectively $\mathbb{F} = \mathbb{C}$.

The main result that we prove in this chapter is the Theorem of Hahn-Banach, that assures in particular that we can indeed extend any element $f \in Y^*$, Y an arbitrary subspace of X, to an operator $F \in X^*$ without increasing its operator norm, compare Theorem 6.1 below.

6.1 The Theorem of Hahn-Banach for real normed spaces

In this section we assume that X is a real vector space (which we can assume wlog to be non-trivial, i.e. $X \neq \{0\}$) and will discuss complex spaces in the next section 6.2.

We discuss several versions and consequences of the Theorem of Hahn-Banach whose main purpose is to establish the existence of linear functionals with some prescribed properties, such as being an extension of a given functional from a subspace Y to the whole space X.

The results we state in this section are all valid for general (real) normed spaces $(X, \|\cdot\|)$, but we shall only give their proofs in the case that X is separable to avoid the need for an argument using Zorn's lemma in favour of proofs that use the usual induction. The general version of Hahn-Banach and further applications that go beyond the confines of this course will be discussed in the Part C course on Functional Analysis.

The first version of the Theorem of Hahn-Banach shows that we can not only extend bounded linear functionals from a subspace to the whole space, but we can do this in a way that does not increase their operator norm, namely

Theorem 6.1 (Theorem of Hahn-Banach on the existence of a bounded extension)**.** *Let X be a 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 note that for any extension $F \in X^*$ we trivially have the inequality

$$\|F\|_{X^*} = \sup_{x\in X, \|x\|=1} |F(x)| \geq \sup_{y\in Y, \|y\|=1} |F(y)| = \sup_{y\in Y, \|y\|=1} |f(y)| = \|f\|_{Y^*}$$

so to prove the above result it is enough to prove that there exists a linear extension of f so that $|F(x)| \leq p(x)$ for all $x \in X$ where we set $p(x) := \|f\| \, \|x\|$. We note that as F is linear, this condition is equivalent to having

$$F(x) \leq p(x) = \|f\| \, \|x\| \text{ for all } x \in X$$

as this then also implies that $-F(x) = F(-x) \leq \|f\|\|x\|$. We also recall that we are dealing with real vector spaces, and hence functionals with values in $\mathbb{R}$, so the above inequality is well defined.

Indeed the general version of the Hahn-Banach Theorem assures that such an extension exists for a much larger class of functions p than just the $p(x) = \|f\|\|x\|$ that we obtain in the context of Theorem 6.1, namely for all $p : X \to \mathbb{R}$ that are so called sublinear:

Definition 13. 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 note that we do not require that p is non-negative. 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 Functional Analysis), such as the so called Minkowski functional associated to each convex set C that contains the orgin, compare section 6.4

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 Theorem of Hahn-Banach (for real vector spaces) is

Theorem 6.2 (Theorem of Hahn-Banach (general sublinear version)). *Let X be a real vectorspace, $Y \subset X$ a subspace and $p : X \to \mathbb{R}$ sublinear. Suppose that $f : Y \to \mathbb{R}$ is a linear functional with the property that*

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

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

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

In this course we shall only give the full proof of the more specific version of the Theorem of Hahn-Banach stated in Theorem 6.1 and this only in the case that X is separable.

As the first step of the proof of this more restricted result and of the proof of the general version of Hahn-Banach are however identical, we will carry out this first step in the general setting. This first step is to prove that we can obtain the required extension to subspaces that are obtained by "adding one dimension" and relies crucially on scalars being real, rather than complex:

Lemma 6.3 (1-step extension lemma). *Let X be a real vector space, $p : X \to \mathbb{R}$ sublinear and let $Y, \tilde{Y}$ be subspaces of X which are so that there exists some $x_0 \in X$ so that*

$$\tilde{Y} = span(Y \cup \{x_0\}).$$

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 $\tilde{f} : \tilde{Y} \to \mathbb{R}$ so that

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

Proof of Lemma 6.3. If $x_0 \in Y$ then the claim is trivial as $\tilde{Y} = Y$. So suppose instead that $x_0 \notin Y$. Then we can write every $\tilde{y} \in \tilde{Y}$ uniquely 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 $\tilde{f}_r(\tilde{y}) \leq p(\tilde{y})$ for all $\tilde{y} \in \tilde{Y}$, which is equivalent to

$$\lambda r \leq p(y + \lambda x_0) - f(y) \text{ for all } y \in Y, \lambda \in \mathbb{R}. \tag{6.1}$$

We first note that for $\lambda = 0$ this is trivially true no matter how r is chosen as by assumption $f \leq p$ on Y.

For $\lambda > 0$ the above inequality (6.1) holds true if and only if

$$r \leq \tfrac{1}{\lambda}\big[p(y + \lambda x_0) - f(y)\big] = p(\tfrac{1}{\lambda}y + x_0) - f(\tfrac{1}{\lambda}y)$$

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

$$r \leq \inf_{v \in Y} \big(p(v + x_0) - f(v)\big). \tag{6.2}$$

For $\lambda < 0$ we write $\lambda = -|\lambda|$ to rewrite (6.1) as $-|\lambda| r \leq p(y - |\lambda| x_0) - f(y)$. We hence obtain that (6.1) is satisfied for all $\lambda < 0$ and $y \in Y$ if and only if

$$r \geq -|\lambda|^{-1}(p(y - |\lambda| x_0) - f(y)) = f(|\lambda|^{-1} y) - p(|\lambda|^{-1} y - x_0),$$

i.e. if and only if r is chosen so that

$$r \geq \sup_{w \in Y} \big(f(w) - p(w - x_0)\big). \tag{6.3}$$

For $\tilde{f}_r$ to be the required extension we thus need to choose r so that both (6.2) and (6.3) hold, which is possible provided

$$\inf_{v \in Y} \big(p(v + x_0) - f(v)\big) \geq \sup_{w \in Y} \big(f(w) - p(w - x_0)\big).$$

However this easily follows since for any $v, w \in Y$ we have that

$$\begin{aligned}\big(p(v + x_0) - f(v)\big) - \big(f(w) - p(w - x_0)\big) &= p(v + x_0) + p(w - x_0) - f(v + w) \\ &\geq p(v + w) - f(v + w) \geq 0\end{aligned}$$

where we use the sublinearity of p in the second and the assumption that $f \leq p$ on Y in the last step. □

The proof of the general version of Hahn-Banach now uses this lemma together with an argument based on Zorn's lemma to obtain the desired extension as a maximal element of a partially ordered set of pairs $(\tilde{Y}, \tilde{f})$ of subspaces $\tilde{Y}$ of X that contain Y and extensions $\tilde{f}$ of f with $\tilde{f} \leq p$. As mentioned, this will be carried out in detail in C4.1 Functional Analysis.

Here we instead use the above lemma to give

Proof of Theorem 6.1 for separable X. Let $(X, \|\cdot\|)$ be a separable space and let $D = \{x_1, x_2, \ldots\}$ be a dense subset of X. Given a subspace Y of X we now define an increasing sequence of subspaces Y_i of X that contain Y by setting $Y_0 = Y$ and then defining iteratively

$$Y_i := \operatorname{span}(Y_{i-1} \cup \{x_i\}).$$

Given $f \in Y^*$ we let $p : X \to \mathbb{R}$ be the sublinear function that is defined by $p(x) := \|f\| \, \|x\|$, set $f_0 = f$ and now iteratively obtain linear extensions $f_{i+1} : Y_{i+1} \to \mathbb{R}$ of $f_i : Y_i \to \mathbb{R}$, i.e. linear functions with $f_{i+1}|_{Y_i} = f_i$, satisfying $f_{i+1} \leq p$ on Y_{i+1} from Lemma 6.3. On $Y_\infty := \bigcup_{i=1}^\infty Y_i$ we hence obtain a linear function $\tilde{f} : Y_\infty \to \mathbb{R}$ by defining $f(x) := f_i(x)$ for $x \in Y_i$, which is well defined as by construction $f_i = f_j$ on $Y_i \cap Y_j = Y_{\min(i,j)}$. By construction $\tilde{f}|_Y = f_0|_{Y_0} = f$ and $\tilde{f} \leq p$ which, as observed previously (c.f. the remark made after the statement of the theorem) implies that $\tilde{f} \in (Y_\infty)^*$ with

$$\|\tilde{f}\|_{(Y_\infty)^*} = \|f\|_{Y^*}.$$

We finally note that since $D \subset X$ is dense and $D \subset Y_\infty$, we have that Y_∞ is a dense subspace of X so we can extend $\tilde{f}$ to the desired element $F \in X^*$ with $\|F\|_{X^*} = \|\tilde{f}\|_{(Y_\infty)^*} = \|f\|_{Y^*}$ using Theorem 4.1. □

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.

6.2 The Theorem for Hahn-Banach for complex normed spaces

We remark that both the bounded extension version of the Theorem of Hahn-Banach, i.e. Theorem 6.1, as well as the general version stated in Theorem 6.2 using sublinear functions, have an extension to complex vector space. For simplicity here we only state and prove the former:

Theorem 6.4. *Let $(X, \|\cdot\|)$ be a complex normed space, Y a subspace of X and $f \in Y^*$. Then there exists an extension $F \in X^*$ of f so that*

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

Proof. Let $f : Y \to \mathbb{C}$ be a $\mathbb{C}$ linear map. Writing $f = f_1 + \mathrm{i} f_2$ we note that both $f_1, f_2 : Y \to \mathbb{R}$ are $\mathbb{R}$-linear and as $f(\mathrm{i}y) = \mathrm{i}f(y) = -f_2(y) + \mathrm{i}f_1(y)$ we have $f_2(y) = -f_1(\mathrm{i}y)$, i.e. $f(y) = f_1(y) - \mathrm{i}f_1(\mathrm{i}y)$. We aim to obtain the desired extension by extending the real valued function f_1 to a function F_1 and then setting $F(x) = F_1(x) - \mathrm{i}F_1(\mathrm{i}x)$.

To this end we first note that

$$|f_1(y)| \le |f(y)| \le \|f\|\|y\| \text{ for all } y \in Y.$$

As we we can view $(X, \|\cdot\|)$ and its subspace Y also as a real vector-spaces (simply ignoring the additional structure obtained from multiplying by i) we hence know that $f_1 : Y \to \mathbb{R}$ is a bounded $\mathbb{R}$-linear functional. Using the version of Hahn-Banach for real normed spaces that we have already proven, we can thus extend f_1 to a bounded $\mathbb{R}$-linear functional $F_1 : Y \to \mathbb{R}$ that also satisfies

$$\|F_1(x)\| \le \|f\|\|x\| \text{ for every } x \in X. \tag{6.4}$$

We now set $F(x) := F_1(x) - \mathrm{i}F_1(\mathrm{i}x)$ and claim that this is the desired $\mathbb{C}$-linear extension of f with $\|F\| = \|f\|$. By construction $F|_Y = f$ and $F(\mathrm{i}x) = \mathrm{i}F(x)$ so we get that F is not only $\mathbb{R}$-linear but indeed $\mathbb{C}$-linear, so it remains to check that $\|F\| \le \|f\|$, i.e. that for any $x \in X$ we have $|F(x)| \le \|f\|\|x\|$.

Given any $x \in X$ we set $\theta = -\mathrm{Arg}(F(x)) \in \mathbb{R}$, which implies that $F(e^{\mathrm{i}\theta}x) = e^{\mathrm{i}\theta}F(x) = e^{\mathrm{i}\theta}|F(x)|e^{\mathrm{i}\mathrm{Arg}(F(x))} = |F(X)|$ is real and hence $F(e^{i\theta}x) = \mathrm{Re}(F(e^{i\theta}x)) = F_1(e^{i\theta}x)$. We can thus use (6.4) and the $\mathbb{C}$ linearity of F to obtain the desired estimate of

$$|F(x)| = F(e^{i\theta}x) = F_1(e^{i\theta}x) \le \|f\|\,\|e^{i\theta}x\| = \|f\|\,\|x\|.$$

□

6.3 Some applications of the Theorem of Hahn-Banach

As a first application of the Theorem of Hahn-Banach we obtain the following useful result

Proposition 6.5. *Let $(X, \|\cdot\|)$ be a normed space. Then for any $x \in X \setminus \{0\}$ there exists an element $f \in X^*$ with $\|f\| = 1$ so that $f(x) = \|x\|$.*

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 $f(x) = g(x) = \|x\|$. □

This result has several useful consequences, including the following 'dual characterisations' of the norms on X and its dual space X^*

Corollary 6.6. *Let $(X, \|\cdot\|)$, $X \neq \{0\}$, be a normed space. Then*

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

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

Proof. We already observed that the second statement is an easy consequence of the definition of the operator norm on X. For the proof of (i) we observe that Proposition 6.5 implies that $\|x\|_X \leq \sup_{f \in X^*, \|f\|_{X^*}=1} |f(x)|$ while the reverse inequality is trivially true since $|f(x)| \leq \|f\|\|x\| = \|x\|$ for every $f \in X^*$ with $\|f\| = 1$. $\square$

We note that while the supremum in (ii) is in general not achieved, Proposition 6.5 implies that the supremum in (i) is always achieved.

A further important consequence of Proposition 6.5 is that it allows us to separate points

Corollary 6.7. *Let $(X, \|\cdot\|)$ be a normed space. Then for any two elements $x \neq y$ of X there exists an element $f \in X^*$ so that*

$$f(x) \neq f(y).$$

This corollary follows as Proposition 6.5 allows us to choose $f \in X^*$ so that $f(x-y) = \|x-y\| \neq 0$.

6.4 Geometric interpretation and further applications

We first note that the kernel of an element $f \in X^* \setminus \{0\}$ has codimension 1 , namely

Lemma 6.8. *Let X be a normed space and let $f : X \to \mathbb{F}$, $\mathbb{F} = \mathbb{R}$ respectively $\mathbb{F} = \mathbb{C}$ be linear so that $f \neq 0$. Then for any $x_0 \in X$ for which $f(x_0) \neq 0$ we have that*

$$\mathit{span}(\mathit{ker}(f) + \{x_0\}) = X$$

Proof. Let $x_0 \in X$ be so that $f(x_0) \neq 0$. Given any $x \in X$ we set $\lambda := \frac{f(x)}{f(x_0)}$ and note that $f(x - \lambda x_0) = 0$. Hence $x - \lambda x_0 \in \ker(f)$ and thus $x \in \operatorname{span}(\ker(f) + \{x_0\})$. As $x \in X$ was arbitrary, this establishes the claim. $\square$

Geometrically we can think of Corollary 6.7 as follows: As Lemma 6.8 implies that the kernel of f has codimension 1 we can think of the sets $\{x : f(x) = \lambda\}$ as hyperplanes in X (that is shifts of a subspace with codimension 1) that divides our space X into two parts, namely $\{x : f(x) < \lambda\}$ and $\{x : f(x) > \lambda\}$. The above corollary hence ensures that we can separate any two points by a hyperplane, with x and y on either side of it.

A slightly more general form of this result that we can prove is that we can separate points from closed subspaces

Proposition 6.9. *Let $(X, \|\cdot\|)$ be a Banach 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*

$$f|_Y = 0 \text{ while } f(x_0) = \operatorname{dist}(x_0, Y).$$

We note that since Y is closed we necessarily have $\text{dist}(x_0, Y) > 0$.

Proof of Proposition 6.9. We define a suitable linear map g on the subspace $U = \text{span}(Y \cup \{x_0\})$ 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

$$g(y + \lambda x_0) := \lambda d, \text{ where } d := \text{dist}(x_0, Y) > 0$$

gives a well defined linear map on Y which has the property that $g(x_0) = d$ and $g|_Y = 0$.

To see that $\|g\|_{U^*} \leq 1$ we note that for any $u = y + \lambda x_0 \in U$

$$\|y + \lambda x_0\| = |\lambda| \, \|x_0 - (-\lambda^{-1} y)\| \geq |\lambda| \inf_{\tilde{y} \in Y} \|x_0 - \tilde{y}\| = |\lambda| \, d = |g(y + \lambda x_0)|.$$

To prove that $\|g\| = 1$ it hence remains to prove that that also $\|g\| \geq 1$ or equivalently that for any $\varepsilon > 0$ there exists $x \in X \setminus \{0\}$ so that $\frac{|g(x)|}{\|x\|} \geq 1 - \varepsilon$, where it is of course enough to consider $\varepsilon \in (0, 1)$. To obtain such an x we note that by the definition of $d = \text{dist}(x_0, Y)$, we can find for any $c > d$, an element $y \in Y$ so that $\|x_0 - y\| < c$. We chose $c = \frac{1}{1-\varepsilon} d$, which is strictly larger than d since $d > 0$, and hence obtain an element $x_0 - y \in X$ for which $\frac{|g(x_0-y)|}{\|x_0-y\|} = \frac{d}{\|x_0-y\|} \geq \frac{d}{c} = 1 - \varepsilon$ as required. Having thus shown that $\|g\| = 1$ we now obtain the required $f \in X^*$ with $f|_Y = 0$, $f(x_0) = d$ and $\|f\| = \|g\| = 1$ by applying the Theorem of Hahn-Banach. ☐

There are far stronger versions of such 'geometric forms of Hahn-Banach' which will be discussed in the part C course on Functional Analysis. As already simple examples in $\mathbb{R}^2$ show, we cannot expect to separate sets by straight lines without imposing some constraints on their geometry. Unsurprisingly, a key role is played by the convexity of sets and one of the results of Part C Functional Analysis will be to prove that if A is closed and B is compact and if both sets are convex, then these sets can be strictly separated by a hyperplane in the sense that there exists an element $f \in X^*$ and a number $\lambda \in \mathbb{R}$ so that

$$\sup_{a \in A} f(a) < \lambda < \inf_{b \in B} f(b).$$

The proof of this result uses the sublinear version of the Theorem of Hahn-Banach and the fact that for an open convex set C containing the origin, one can define a sublinear function by $p(x) := \inf\{\lambda > 0 : x \in \lambda C\}$ (called the Minkowski functional).

Such general results play an important role also in applications to PDEs and in other advanced topics in functional analysis but go beyond the remit of this course.

To formulate another useful consequence of Proposition 6.9 we introduce the following notation

Definition 14. 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 $T \subset X^*$ we define

$$T_\circ := \{x \in X : f(x) = 0 \text{ for all } f \in T\} = \bigcap_{f \in T} \ker(T).$$

We may now prove

Proposition 6.10. *Let $(X, \|\cdot\|)$ be a normed space. Then the following hold true:*

(i) Let $S \subset X$. Then span(S) is dense if and only if the annihilator of S is trivial, i.e. $S^0 = \{0\} \subset X^*$

(ii) If $T \subset X^*$ is so that span(T) is dense in X^* then $T_\circ = \{0\} \subset X$.

Proof. (i) Suppose first that span(S) is dense. Then for any $f \in S^\circ$, we have by linearity that also $f|_Y = 0$ where we set $Y = \text{span}(S)$. As Y is dense in X we thus get that $f = 0$ by Lemma 4.2.

Conversely, suppose that span(S) is not dense. Then $Y = \overline{\text{span}(S)}$ is a closed proper subspace of X so we can choose $x_0 \in X \setminus Y$ and apply Proposition 6.9 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 S°.

(ii) Suppose that there exists $x \in T_\circ$ with $x \neq 0$. Then by Corollary 6.7 there exists $f \in X^*$ so that $f(x) \neq f(0) = 0$. If span(T) is dense in X^* we can however find a sequence (f_n) of elements of span(T) that converges $f_n \to f$ in the sense of X^*. Note that since $x \in T_\circ$ we have that $f_n(x) = 0$ which leads to a contradiction as $0 \neq f(x) = \lim_{n\to\infty} f_n(x)$.

□

We note that as the kernel of any element $f \in X^*$ is closed, we know that $T_\circ$ is an intersection of closed subspaces and hence itself a closed subspace of X. Also one can easily check from the definition that the annihilator of any set $A \subset X$ is closed subspace of X^*. Furthermore we have

Lemma 6.11. *Let A be any subspace of a normed space X. Then*

$$\bar{A} = (A^\circ)_\circ.$$

Proof. <u>"⊂"</u> As $(A^\circ)_\circ$ is closed, it suffices to prove that $A \subset (A^\circ)_\circ$. Let $a \in A$. Then by definition of the annihilator of A we know that $f(a) = 0$ for any $f \in A^\circ$ and hence that $a \in (A^\circ)_\circ$.

<u>"⊃"</u> Suppose that there exists an element $x \in (A^\circ)_\circ$ so that $x \notin \bar{A}$. As $\bar{A}$ is a closed subspace of X we know from Proposition 6.9 that there exists some $f \in X^*$ so that $f(x) \neq 0$ but with $f|_{\bar{A}} = 0$, i.e. with $f \in (\bar{A})^\circ$ and hence $f \in A^\circ$. But this contradicts the assumption that $x \in (A^\circ)_\circ$.

□

Chapter 7

Dual-spaces, second duals and completion

In this chapter we further discuss the special properties of functionals, describe the dual spaces of some important spaces encountered earlier, take a first look at the second dual X^{**} of a normed space, that is the dual space of the dual space X^* of X and explain how a space always embedds into its second dual and how this can be used to view a non-complete normed space as a subspace of a complete space.

7.1 Another basic property of functionals

To begin with, we note that for linear functionals, we have the following characterisation of continuity

Lemma 7.1. *Let X be a normed space and let $f : X \to \mathbb{F}$, $\mathbb{F} = \mathbb{R}$ respectively $\mathbb{F} = \mathbb{C}$ be linear. Then the following are equivalent*

$$\textit{ker}(f) \textit{ is closed} \iff f \in X^*.$$

Proof. "$\Leftarrow$" As f is continuous and $\{0\}$ is closed we get that the preimage $\ker(f) = f^{-1}(\{0\})$ is also closed.
"$\Rightarrow$" The claim is trivial if $f = 0$ so suppose that $f \neq 0$ and let x_0 be so that $f(x_0) \neq 0$, where (after replacing x_0 by a multiple of x_0) we can assume without loss of generality that $f(x_0) = 1$.

We first note that since $\ker(f)$ is closed, we know that $\text{dist}(x_0, \ker(f)) > 0$, compare problem sheet 0. We now claim that for every $x \in X$

$$|f(x)| \leq \delta^{-1}\|x\| \text{ where } \delta := \text{dist}(x_0, \ker(f)) > 0$$

which will of course imply that $f \in X^*$.

This claim is trivial for $x \in \ker(f)$ so suppose instead that $f(x) \neq 0$. We note that since $f(x_0) = 1$ and since f is linear we have that $x - f(x)x_0 \in \ker(f)$. Hence also $-\frac{1}{f(x)}(x - f(x)x_0) \in \ker(f)$ and must thus have distance of at least δ from x_0 which implies that

$$\delta \leq \|x_0 + \frac{1}{f(x)}(x - f(x)x_0)\| = \frac{\|x\|}{|f(x)|}$$

and thus that indeed $|f(x)| \leq \delta^{-1}\|x\|$. $\square$

7.2 Dual spaces of particular spaces

We recall from Linear Algebra that if X is a finite dimensional space then we can associate to each basis $e_1, \ldots, e_n$ of X a dual basis $f_1, \ldots, f_n$ of X^* by defining $f_i(e_j) = \delta_{ij}$. In particular X and its dual X^* are isomorphic.

WARNING. As so often in this course, the finite dimensional case leads to the wrong intuition for general normed spaces. In general, the dual space can have very different properties than a space itself, e.g. we can have that X is separable while X^* is inseparable, we will have that the dual space of any normed space is complete, even if the space X itself is not complete....

The one exception to this warning are Hilbert-spaces, for which you will prove the following beautiful theorem in B4.2 Functional analysis 2

Theorem 7.2 (Riesz-Representation Theorem (contents of B4.2))**.** *Let $(X, (\cdot,\cdot))$ be a Hilbert-space. Then the map $\iota : X \to X^*$ defined by $\iota(x)(y) := (x, y)$ is an isometric isomorphism from X to X^*.*

To describe the dual space of a given space X, we would like to find another normed space which is isometrically isomorphic to X, written for short as $X \cong Y$, i.e. for which there exists a bijective linear map $L : Y \to X$ so that

$$\|Ly\|_X = \|y\|_Y \text{ for all } y.$$

Often it is not too difficult to find a space Y and a map L so that $L : Y \to X^*$ is isometric, i.e. so that $\|Ly\|_X = \|y\|_Y$ for all y, and hence also injective, but it can be difficult to find a space Y that is large enough so that it represents all elements of X^*, i.e. so that the map L is surjective, respectively to prove that a candidate Y for the dual space has this property.

In general, determining the dual space of a given normed space $(X, \|\cdot\|)$ can be difficult and already the dual spaces of some very familiar spaces such as $C([a, b])$ or $L^\infty([a, b])$ can be complicated and their description is beyond the scope of this course, though we remark that in both cases the dual spaces can be identified with a suitable space of (signed) measures. Examples of elements of $(C[0, 1])^*$ are e.g. the map $T : f \mapsto \int_{[0,1]} fgdx$ for any $g \in L^1([0, 1])$ but also maps like $T : f \mapsto f(\frac{1}{2})$ which one can interpret as the integral of f with respect to a δ-measure that is concentrated at $x = \frac{1}{2}$.

On the other hand, for the sequence spaces ℓ^p and the function spaces L^p we have the following characterisations if $1 \leq p < \infty$, which we stress do not apply if $p = \infty$. For simplicity we consider real valued functions, though the results and their proofs also apply in the complex case (with some extra complex conjugates).

Theorem 7.3 (Dual space of L^p)**.** *Let $\Omega \subset \mathbb{R}^n$ be measurable, $1 \leq p < \infty$ and let $1 < q \leq \infty$ be so that $\frac{1}{p} + \frac{1}{q} = 1$. Then $(L^p(\Omega))^* \cong L^q(\Omega)$, and an isometric isomorphism is given by the map $L : L^q(\Omega) \to (L^p(\Omega))^*$ where for $f \in L^q(\Omega)$ the linear map $L(f) \in (L^p(\Omega))^*$ is defined by*

$$L(f) : L^p(\Omega) \ni g \mapsto \int_\Omega fgdx \in \mathbb{R}.$$

We will prove that the map L defined above is a well-defined linear isometric map but omit the proof of the difficult part of the theorem, i.e. the fact that L is surjective, as this would require techniques not used elsewhere in the course and as this proof is carried out in Part C Functional Analysis.

Proof of Theorem 7.3 except for surjectivity of L. We note that since p, q are so that $\frac{1}{p} + \frac{1}{q} = 1$ we know from Hölder's inequality that the product fg of two functions $f \in L^q(\Omega)$ and $g \in L^p(\Omega)$ is integrable and

$$|\int_\Omega fg dx| \leq \|f\|_{L^q}\|g\|_{L^q}. \tag{7.1}$$

We now remark that since the integral is linear, we have that for all $f, \tilde{f} \in L^q(\Omega)$, $\lambda \in \mathbb{R}$ and any $g \in L^q(\Omega)$ that $L(f + \lambda\tilde{f})(g) = Lf(g) + \lambda L\tilde{f}(g)$, i.e. that the map $L : f \to Lf$ is linear. Similarly, given any $f \in L^q(\Omega)$ we have that for any $g, \tilde{g} \in L^p(\Omega)$ and any $\lambda \in \mathbb{R}$ that $Lf(g + \lambda\tilde{g}) = Lf(g) + \lambda Lf(\tilde{g})$ so Lf is a linear map from L^p to $\mathbb{R}$ and indeed an element of $(L^p(\Omega))^*$ as it is bounded with

$$\|Lf\|_{(L^p(\Omega))^*} = \sup_{g\in L^q, g\neq 0} \frac{|Lf(g)|}{\|g\|_{L^p}} = \sup_{g\in L^q, g\neq 0} \frac{|\int fg|}{\|g\|_{L^p}} \leq \sup_{g\in L^q, g\neq 0} \frac{\|f\|_{L^q}\|g\|_{L^p}}{\|g\|_{L^p}} = \|f\|_{L^q}$$

where we used (7.1) in the penultimate step.

We finally show that also

$$\|Lf\|_{L^p(\Omega)^*} \geq \|f\|_{L^q} \tag{7.2}$$

for every $f \in L^q$ and thus that L is indeed isometric.

This proof is a bit technical as we need to be careful with the exponents, so it can be useful to first consider special cases such as $p = q = 2$ or $p = 1, q = \infty$, where the exponents are much nicer to see the structure of the argument, before digesting the general case. We first treat the case that $p > 1$ and hence $q < \infty$.

The estimate (7.2) is trivial if $f = 0$ so suppose that $f \neq 0$. We choose $g := |f|^{q-2}f$ so that $fg = |f|^q$ and hence $L(f)(g) = \int_\Omega |f|^q dx = \|f\|_{L^q}^q$. We now note that since $\frac{1}{p} + \frac{1}{q} = 1$ we have $(q-1)p = q$ and so

$$\|g\|_{L^p} = \Big(\int |g|^p \, dx\Big)^{1/p} = \Big(\int (|f|^{(q-1)})^p \, dx\Big)^{1-\frac{1}{q}} = \Big(\int |f|^q \, dx\Big)^{\frac{1}{q}(q-1)} = \|f\|_{L^q}^{q-1}$$

which means that

$$\frac{Lf(g)}{\|g\|_{L^p}} = \frac{\|f\|_{L^q}^q}{\|f\|_{L^q}^{q-1}} = \|f\|_{L^q}$$

and hence that $\|Lf\|_{(L^p)^*} \geq \|f\|_{L^q}$ as claimed in (7.2).

If $p = 1$ and hence $q = \infty$ then we prove that for any $\varepsilon > 0$ there exists a function $g_\varepsilon \in L^1$ so that $\frac{Lf(g)}{\|g\|_{L^1}} \geq \|f\|_{L^\infty} - \varepsilon$. To this end we consider the set $A_\varepsilon := \{x : |f(x)| \geq \|f\|_{L^\infty} - \varepsilon\}$, which is measurable (and well defined upto a null set). If this set has finite measure then we define $g_\varepsilon(x) := \operatorname{sign}(f(x)) \cdot \chi_{A_\varepsilon}$ which is in $L^1(\Omega)$ with L^1-norm equal to the measure of A_ε, which by the definition of the L^∞ norm is positive. As $fg \geq (\|f\|_{L^\infty} - \varepsilon)\chi_{A_\varepsilon}$ we can thus immediately check that $Lf(g) \geq (\|f\|_{L^\infty} - \varepsilon)\|g_\varepsilon\|_{L^1}$ which gives the claimed bound. Finally, if A_ε has infinite measure, then we can replace A_ε by any subset $\tilde{A}_\varepsilon \subset A_\varepsilon$ whose measure is finite and positive and apply the above argument for the corresponding function $g_\varepsilon \in L^1$. $\square$

The analogue result for the sequence spaces is

Theorem 7.4 (Dual space of $\ell^p(\mathbb{R})$)**.** *Let $1 \le p < \infty$ and let $1 < q \le \infty$ be so that $\frac{1}{p} + \frac{1}{q} = 1$. Then $(\ell^p)^* \cong \ell^q$, and an isometric isomorphism is given by the map $L : \ell^q(\mathbb{R}) \to (\ell^p(\mathbb{R}))^*$ where for $x \in \ell^q$ the linear map $L(x) \in (\ell^p)^*$ is defined by*

$$L(x) : \ell^p(\mathbb{R}) \ni y \mapsto \sum_{j=1}^{\infty} x_j y_j \in \mathbb{R}.$$

The proof that L is a well defined isometric linear map from ℓ^q to $(\ell^p)^*$ is exactly the same as for the function spaces L^p (replacing functions by sequences and integral by sums), so we do not repeat it.

Instead we explain how one can prove surjectivity of the map L in case of $p = 1$ to show that indeed $(\ell^1)^* = \ell^\infty$:

Proof of surjectivity of $L : \ell^\infty \to (\ell^1)^*$. Given $f \in (\ell^1)^*$ we define a sequence $x = (x_j)_{j \in \mathbb{N}}$ by setting $x_j = f(e^{(j)})$ where as usual $e^{(j)}$ is the sequence with $e^{(j)}_k = \delta_{jk}$. Then

$$|x_j| \le \|f\|_{(\ell^1)^*} \|e^{(j)}\|_1 = \|f\|_{(\ell^1)^*}$$

so $x \in \ell^\infty$. We now claim that $Lx = f$. To see this we note that since both Lx and f are linear and as by construction $(Lx)(e^{(j)}) = x_j = f(e^{(j)})$, we know that $f|_Y = Lx|_Y$ for $Y := \text{span}(\{e^{(j)}, j \in \mathbb{N}\})$. As $Y \subset \ell^1$ is dense, compare with the proof of Proposition 5.6, we thus obtain from Corollary 4.2 that indeed $f = Lx$. □

The proof of surjectivity of L for general sequence spaces ℓ^p, $1 \le p < \infty$ is very similar (though one needs to be more careful with the exponents).

WARNING.

$$(\ell^\infty)^* \ncong \ell^1 \text{ and } (L^\infty(\Omega))^* \ncong L^1(\Omega).$$

While the analogue of the maps L defined in Theorems 7.3 and 7.4 also give isometric linear maps from L^1 to $(L^\infty)^*$ (respectively ℓ^1 to $(\ell^\infty)^*$) these maps are not surjective. For the sequence spaces one can show that ℓ^1 is isomorphic to the dual of a subspace of ℓ^∞, namely $(c_0)^* \cong \ell^1$, where c_0 denotes the subspace of all sequences that converge to zero, compare problem sheet 4.

To see that $L : \ell^1 \to (\ell^\infty)^*$ cannot be surjective, we consider the subspace $c \subset \ell^\infty$ of all sequences that converge and let $f : c \to \mathbb{R}$ be the map that assigns to each $x \in c$ its limit $f(x) = \lim_{n\to\infty} x_n$. Then f is clearly linear and bounded on $(c, \|\cdot\|_\infty)$, so by Hahn-Banach, has an extension $\tilde{f} \in (\ell^\infty)^*$ (Note that this is an instance where we apply Hahn-Banach to an inseparable space). But we cannot write $\tilde{f}$ in the form $\tilde{f}(x) = \sum_{j=1}^{\infty} x_j y_j$ for some $y \in L^1$ so $L(\ell^1)$ is a proper subspace of $(\ell^\infty)^*$.

7.3 Second Dual and completion

As the dual space $(X^*, \|\cdot\|_{X^*})$ of a normed space $(X, \|\cdot\|)$ is again a normed space, we can consider its dual space X^{**} which is called the second dual space or bidual space of X. The most important property of this space is that it will always contain an isometric image of the space X itself, obtained via the canonical map

$$i : X \to X^{**}, \quad i(x)(f) := f(x) \tag{7.3}$$

that maps each element x to the functional $i(x) : X^* \to \mathbb{R}$ that evaluates elements $f \in X^*$ at the point x.

Proposition 7.5. *Let $(X, \|\cdot\|)$ be a normed space and let $i : X \to X^{**}$ be the canonical map defined by (7.3). Then i is linear and isometric, i.e.*

$$\|i(x)\|_{X^{**}} = \|i\|_X.$$

WARNING. We remark that i is in general NOT surjective, e.g.

$$(L^1)^{**} \cong (L^\infty)^* \ncong L^1.$$

However, it turns out that for many important spaces the space X is isometrically isomorphic to its bidual X^{**}. Spaces for which $i(X) = X^{**}$ are called *reflexive*, and their properties will be further analysed in part C Functional analysis. Reflexivity (and also separability) is in particular relevant in applications, as it allows one to extract a subsequence of any given bounded sequence that 'converges in the weak sense' to some limit, a property that is hugely relevant as one often tries to prove the existence of a solution of a problem (be it an abstract equation on some Banach space, the existence of a minimiser in calculus of variations or a solution of a PDE) by considering a sequence of approximate solutions (or solutions of approximations of the problem) and hoping to find a subsequence of these approximate solutions that converges in some sense to a solution of the original problem.

From the characterisation of the dual spaces of ℓ^p and L^p obtained above we know in particular

- ℓ^p is reflexive for $1 < p < \infty$ as for $q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have $(\ell^p)^{**} \cong (\ell^q)^* \cong \ell^p$
- L^p is reflexive for $1 < p < \infty$ as for $q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$ we have $(L^p)^{**} \cong (L^q)^* \cong L^p$
- L^1, L^∞, ℓ^1 and ℓ^∞ are not reflexive.

Proof of Proposition 7.5. i is clearly a linear map from $X^* \to \mathbb{R}$ and as for any $f \in X^*$ with $\|f\|_{X^*} = 1$

$$|i(x)(f)| = |f(x)| \leq \|f\| \, \|x\| = \|x\|$$

we have that

$$\|i(x)\|_{X^{**}} = \sup_{f \in X^*, \|f\|_{X^*}=1} |f(x)| \leq \|x\|.$$

To see that also $\|i(x)\| \geq \|x\|$ we now choose $f \in X^*$ with $\|f\|_{X^*} = 1$ as in Proposition 6.5 so that $f(x) = \|x\|$. □

We note that this argument is essentially just a repetition of the proof of Corollary 6.6 (i) which directly gives that $\|i(x)\| = \|x\|$.

As the dual space of any normed space is complete, we know in particular that X^{**} is complete and hence that every closed subspace of X^{**} is itself a Banach space. This allows us to view any non-complete normed space as a dense subspace of a Banach space

Corollary 7.6. *Let $(X, \|\cdot\|)$ be any normed space. Then $(X, \|\cdot\|)$ is isometrically isomorphic to $i(X)$ which can be seen as dense subspace of the Banach space $(Y, \|\cdot\|_{X^{**}})$ where $Y = \overline{i(X)} \subset X^{**}$.*

A Banach space $(Y, \|\cdot\|_Y)$ into which X embedds isometrically as a dense subset is called completion of X. Such a completion is determined up to isometric isomorphisms, i.e. given any two spaces Y, $\tilde{Y}$ so that there exist isometric maps $J : X \to Y$ respectively $\tilde{J} : X \to \tilde{Y}$ for which $J(X)$ (respectively $\tilde{J}(X)$) is dense in Y (respectively $\tilde{Y}$), we have that there is a (unique) isometric isomorphism from $I : Y \to \tilde{Y}$ so that

$$\tilde{J} = I \circ J.$$

Indeed, this map I is determined as the unique extension of $\tilde{J} \circ H$, $H : J(X) \to X$ the inverse of the bijective map $J : X \to J(X)$, from the dense subspace $J(X) \subset Y$ to the whole space Y, compare Theorem 4.1.

7.4 Dual operators

We recall the following construction from Part A Linear Algebra:

Let X and Y be any vector spaces over the same field $\mathbb{F}$ and let $X' := \{L : X \to \mathbb{R} \text{ linear}\}$ and $Y' := \{L : Y \to \mathbb{F} \text{ linear}\}$ be the corresponding sets of linear functionals (so far we do not introduce any norm on X and Y, so it would also make no sense to talk about continuity).

Then we can associate to any linear map

$$T : X \to Y$$

the map

$$T' : Y' \to X'$$

where for any $f \in Y'$ we define $T'(f) \in X'$ by

$$(T'(f))(x) = f(T(x)),$$

and one easily checks that $T(f)$ is indeed linear, and thus an element of X', and that the map $f \mapsto T(f)$ is also linear.

We may now ask whether this construction works also in the setting of Functional Analysis, where we work with normed spaces instead of just vector spaces and bounded linear operators instead of just linear operators. The following proposition answers this question positively:

Proposition 7.7 (dual operator)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces over the same field $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ and let $T \in L(X, Y)$. Then the dual operator*

$$\begin{aligned} T' &: Y^* \to X^* \\ f &\mapsto T'(f) : x \mapsto T'(f)(x) := f(Tx) \end{aligned} \tag{7.4}$$

is well defined and a bounded linear operator $T' \in L(Y^*, X^*)$ with

$$\|T'\|_{L(Y^*,X^*)} = \|T\|_{L(X,Y)}.$$

Proof. As already mentioned, the fact that for each $f \in X^*$ the map $T'(f)$ is linear and that T itself is a linear operator is easily checked (and the proof is exactly the same as in the finite dimensional case that was covered in part A Linear Algebra). We first show that $T'(f) \in X^*$ with

$$\|T'(f)\|_{X^*} \leq \|T\|_{L(X,Y)} \|f\|_{Y^*} \text{ for every } f \in Y^* \tag{7.5}$$

which ensures that T' is a well defined operator in $L(Y^*, X^*)$ with $\|T'\|_{L(Y^*,X^*)} \leq \|T\|_{L(X,Y)}$. To see this we note that for every $x \in X$ with $\|x\|_X = 1$

$$|T'(f)(x)| = |f(T(x))| \leq \|f\|_{Y^*} \|Tx\|_Y \leq \|f\|_{Y^*} \|T\|_{L(X,Y)} \|x\|_X = \|f\|_{Y^*} \|T\|_{L(X,Y)}$$

so that (7.5) follows from the definition of the operator norm. To see that also $\|T'\| \geq \|T\|$ we will prove that

$$\|Tx\|_Y \leq \|T'\|_{L(Y^*,X^*)} \|x\|_X \text{ for all } x \in X \tag{7.6}$$

which implies that $\|T\| = \inf\{M : \|Tx\| \leq M\|x\|\} \leq \|T'\|$.

This estimate (7.6) trivially holds true for all $x \in \ker(T)$, so suppose that $Tx \neq 0$. Then Proposition 6.5 (which was a consequence of the Theorem of Hahn-Banach) gives us an element $f \in Y^*$ with $\|f\|_{Y^*} = 1$ so that

$$f(Tx) = \|Tx\|.$$

Hence

$$\|Tx\| = f(Tx) = (T'(f))(x) \leq \|T'\|\|fx\| \leq \|T'\|\|f\|\|x\| = \|T'\|\|x\|$$

as claimed.

□

You have seen in Part A Linear Algebra that for finite dimensional spaces there are several relations involving kernels of maps/dual maps and annihilators of the images of dual maps/maps. Many of these relations have an analogue for general normed spaces, but one needs to be careful in particular with statements that involve spaces, such as the images TX or $T'Y^*$, that are in general not closed, and such statements often require us to take the closure of the corresponding sets. Some of these relations will be proven on problem sheet 4.

Chapter 8

Spectral Theory

8.1 Operators with continuous inverses

Before we discuss the spectrum of linear operators, we make some more remarks about the invertibility of linear operators where we recall by definition $T \in L(X)$ is invertible if it is bijective and $T^{-1} \in L(X)$.

Such an operator T must hence be so that

(i) T is surjective, i.e. $TX = X$

(ii) T is injective, i.e. $Tx \neq 0$ for all $x \in X \setminus \{0\}$

(iii) T^{-1} is bounded, i.e. there exists $M \in \mathbb{R}$ so that $\|T^{-1}y\| \leq M\|y\|$.

We also remark that if T is algebraically invertible then (iii) holds if and only if

$$\exists \delta > 0 \text{ so that for all } x \in X \text{ we have } \|Tx\| \geq \delta \|x\|, \tag{8.1}$$

and note that (8.1) of course implies that T is injective.

While (8.1) of course does not imply that the map is surjective, it gives the following useful information on the image of T.

Proposition 8.1. *(closed range) Let X be a Banach space and let $T \in L(X)$ be so that (8.1) holds true. Then T is injective and $TX \subset X$ is closed. In particular, if TX is additionally dense in X then T is invertible.*

WARNING. This result is wrong if X is not assumed to be complete and we also remark that the image of general bounded linear operators from Banach spaces is not closed. As an example consider the inclusion map $i : (C[0,1], \|\cdot\|_{sup}) \to (L^1[0,1], \|\cdot\|_{L^1})$ which is a bounded linear operator whose image is the subspace of L^1 given by all continuous functions which cannot be closed in L^1 as it is a dense proper subspace of L^1.

Proof. The only statement whose proof is not trivial is that the image TX is closed which we can prove as follows: Given any sequence y_n in TX which converges $y_n \to y$ to some $y \in X$, we let

$x_n \in X$ be so that $Tx_n = y_n$. We then note that as (y_n) is a Cauchy-sequence, the assumption (8.1) implies that

$$\|x_n - x_m\| \leq \delta^{-1}\|T(x_n - x_m)\| = \delta^{-1}\|y_n - y_m\| \underset{n,m\to\infty}{\longrightarrow} 0,$$

i.e. that also (x_n) is Cauchy and thus, as X is complete, that $x_n \to x$ for some $x \in X$. As T is continuous we thus get that $y = \lim y_n = \lim Tx_n = Tx \in TX$. □

We furthermore record the following useful lemma

Lemma 8.2. *Let $(X, \|\cdot\|)$ be a normed space, $S, T \in L(X)$. Suppose that $ST = TS$ and that ST is invertible. Then also S and T are invertible.*

Proof. By symmetry it suffices to prove the claim for T and we shall prove this by an argument by contradiction. So suppose that the claim is false. Then we either have that T is not surjective, which is impossible as in this case we would have that $ST(X) = TS(X) = T(SX) \subset TX \subsetneqq X$ so ST would not be surjective, or there exists no $\delta > 0$ so that (8.1) holds. In this case we can choose $x_n \in X \setminus \{0\}$ so that $\frac{\|Tx_n\|}{\|x_n\|} \to 0$ and thus conclude that also

$$\frac{\|STx_n\|}{\|x_n\|} \leq \|S\|_{L(X)}\frac{\|Tx_n\|}{\|x_n\|} \to 0,$$

which means that (8.1) does not hold true for ST, and hence that ST does not have a bounded inverse. □

8.2 Spectrum and resolvent set

For the rest of the chapter we consider $(X, \|\cdot\|)$ to be a normed space over $\mathbb{F} = \mathbb{C}$.

Definition 15. Let $T \in L(X)$.

- Then the *resolvent set* $\rho(T)$ is defined by
$$\rho(T) := \{\lambda \in \mathbb{C} : T - \lambda \mathrm{Id} \text{ is invertible }\}. \tag{8.2}$$
$R_\lambda(T) := (T - \lambda\mathrm{Id})^{-1} \in L(X)$ is then called the resolvent operator.
- The spectrum of T is defined by
$$\sigma(T) := \mathbb{C} \setminus \rho(T).$$

We note that the resolvent set $\rho(T)$ consists of all those $\lambda \in \mathbb{C}$ for which the equation

$$Tx - \lambda x = y$$

has a unique solution for each y which furthermore depends continuously on the right hand side y.

A number $\lambda \in \mathbb{C}$ is in the spectrum if at least one of the following holds

(i) $T - \lambda\mathrm{Id}$ fails to be injective

(ii) There exists no $\delta > 0$ so that for all $x \in X$

$$\|Tx - \lambda x\| \geq \delta \|x\|$$

(iii) $T - \lambda \mathrm{Id}$ is not surjective

We can refine the definition of the spectrum accordingly:

Definition 16. Let $T \in L(X)$.

- $\lambda \in \mathbb{C}$ is called an eigenvalue if there exists $x \in X$ with $x \neq 0$ so that $Tx = \lambda x$. We call the set
$$\sigma_P(T) := \{\lambda : \lambda \text{ eigenvalue of } T\}$$
the point spectrum of T.

- $\lambda \in \mathbb{C}$ is called an approximate eigenvalue if there exists a sequence $x_n \in X$ with $\|x_n\| = 1$ so that $\|Tx_n - \lambda x_n\| \to 0$. We call the set
$$\sigma_{AP}(T) := \{\lambda : \lambda \text{ approximate eigenvalue of } T\}$$
the approximate point spectrum.

We note that every eigenvalue λ is also an approximate eigenvalue (as we may simply choose $x_n = x$ for an element of $\ker(T - \lambda \mathrm{Id})$ that is normalised to $\|x\| = 1$), so we have

$$\sigma_P(T) \subset \sigma_{AP}(T) \subset \sigma(T).$$

We also remark that λ is an approximate eigenvalue if and only if (ii) from above holds, as (ii) is equivalent to
$$\inf_{x \in X, \|x\|=1} \frac{\|Tx - \lambda x\|}{\|x\|} = 0$$
as the linearity of T means that this is infimum the same whether it is computed over all of $X \setminus \{0\}$ or just over $\{x \in X : \|x\| = 1\}$.

Remark (Other subsets of the spectrum (Off syllabus))**.**
One often also divides up the spectrum into the following disjoint subsets

- the point spectrum $\sigma_P(T)$
- the continuous spectrum
$$\sigma_C(T) := \{\lambda \in \sigma(T) : T - \lambda \mathrm{Id} \text{ is injective and } (T - \lambda \mathrm{Id})(X) \text{ is dense in } X\}$$
- the residual spectrum
$$\sigma_R(T) := \sigma(T) \setminus (\sigma_P(T) \cup \sigma_C(T))$$

and more on this subject will be discussed in Functional Analysis 2.

8.2.1 Examples

Example 1. (Operators on finite dimensional space) If X is finite dimensional then you know from Linear Algebra (and the rank-nullity theorem) that a linear map $L : X \to X$ is injective if and only if it is surjective. As all linear operators on finite dimensional spaces are continuous we thus have that in this case $\sigma(T) = \sigma_P(T)$.

Example 2. (An operator on ℓ^∞ for which $\sigma_P \neq \sigma_{AP}$) Consider the operator $T \in L(\ell^\infty)$ defined by $T(x) = (\frac{x_j}{j})_{j\in\mathbb{N}}$. Then each $\lambda = \frac{1}{j}$ is an eigenvalue as we have $T(e^{(j)}) = \frac{1}{j}e^{(j)}$ for $e^{(j)} = (\delta_{jk})_{k\in\mathbb{N}}$. While $\lambda = 0$ is not an eigenvalue it is clearly an approximate eigenvalue as e.g. $e^{(k)}$ gives a sequence in ℓ^∞ with $\|e^{(k)}\|_\infty = 1$ and $\|T(e^{(k)})\| \to 0$.

Example 3. (An Integral operator) Let $X = C([0,1])$, as always equipped with the sup norm, and consider $T \in L(X)$ defined by $Tx(t) := \int_0^t x(s)ds$.

Claim: $\sigma(T) = \{0\}$ while $\sigma_P(T) = \emptyset$

Proof: We first show that $\lambda = 0 \in \sigma(T) \setminus \sigma_P(T)$. Indeed differentiating the equation $Tx = 0$ (which is allowed as $Tx \in C^1$ for $x \in X$) we immediately get that $x(t) = 0$ for every t and hence that $\lambda = 0$ is not an eigenvalue. On the other hand for any $x \in X$ we have that $Tx(t = 0) = 0$ so $TX \subset \{x \in X : x(0) = 0\} \neq X$. So T is not surjective and thus $0 \in \sigma(T)$,

Let now $\lambda \neq 0$. Then we can use that the proof of Picard's Theorem from DE1 shows that for any $y \in C([0,1])$ the integral equation $Tx - \lambda x = y$ has a unique solution $x = (T - \lambda \mathrm{Id})^{-1}(y)$; here we note that for $y \in C^1$ the equation is equivalent to the initial value problem $x'(t) - \lambda^{-1}x(t) = -\lambda^{-1}y'(t)$ on $[0,1]$ with $x(0) = 0$, but that the proof of Picard from DE1 actually applies to give the existence of a unique solution of the integral equation also just for y continuous.

Furthermore the fact that this solution depends continuously on y can e.g. be obtained from Gronvall's lemma. Hence λ is not in the spectrum.

An alternative proof that $\sigma(T) = \{0\}$, based on the general properties of the spectrum that we prove in the following section, is carried out on problem sheet 4.

Example 4.(Shift operator on ℓ^1) Consider $T : \ell^1 \to \ell^1$ defined by $T(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots)$. We first determine the eigenvalues, i.e. the point spectrum. So suppose that $\lambda \in \mathbb{C}$ is so that for some $x \in \ell^1 \setminus \{0\}$ we have $Tx = \lambda x$. Then $T^j x = \lambda^j x$ so $x_j = (T^j(x))_1 = \lambda^j x_1$. Hence $x_1 \neq 0$ (as $x \neq 0$) and $x = x_1(1, \lambda, \lambda^2, \ldots)$ which satisfies $Tx = \lambda x$ for all values of λ, but is only an element of ℓ^1 if $|\lambda| < 1$. We hence conclude that $\sigma_P(T)$ is the open unit disc $B_1(0) \subset \mathbb{C}$.

We may now check that every point in the closed unit disc is an approximate eigenvalue and indeed that

$$\sigma(T) = \sigma_{AP}(T) = \overline{B_1(0)}$$

as Theorem 8.3, that we prove below, ensures that always $\sigma(T) \subset \overline{B_{\|T\|}(0)}$, so in the present situation where $\|T\| = 1$

$$\overline{B_1(0)} \subset \sigma_{AP}(T) \subset \sigma(T) \subset \overline{B_1(0)}$$

so all these sets need to agree.

8.3 Properties of the spectrum of general bounded linear operators on Banach spaces

Our first main result about the spectrum of bounded linear operators is

Theorem 8.3 (Properties of the Spectrum of bounded linear operators on Banach spaces)**.** *Let $(X, \|\cdot\|)$ be a complex Banach space. Then for any $T \in L(X)$ we have*

(i) The resolvent set $\rho(T)$ is open and the map

$$\rho(T) \ni \lambda \mapsto R_\lambda(T)$$

is analytic, i.e. for any $\lambda_0 \in \rho(T)$ there exists a neighbourhood U of λ_0 and 'coefficients' $A_j(\lambda_0, T) \in L(X)$ so that for every $\lambda \in U$ the resolvent operator is given by the convergent power series

$$R_\lambda(T) = \sum_{j=0}^{\infty} (\lambda - \lambda_0)^j A_j(\lambda_0, T).$$

(ii) The spectrum $\sigma(T)$ is non-empty, compact and for every $\lambda \in \sigma(T)$ we have $|\lambda| \leq \|T\|_{L(X)}$.

One of the most important aspects of the above theorem is the last part, i.e. that every bounded operator has non-empty spectrum. Here we crucially use that the vector space is over $\mathbb{C}$. The claim is not true if we were to only consider the real spectrum as you already know from Linear Algebra.

Proof. For the proof of (i) we use Lemma 2.5 and its Corollary 2.6 on the convergence of the Neumann series $\sum_{j=0}^{\infty} S^j = (\mathrm{Id} - S)^{-1}$ for operators $S \in L(X)$ with $\|S\| < 1$:

Let λ_0 be any element of the resolvent set, i.e. so that $(T - \lambda_0 \mathrm{Id})$ is invertible, and denote by $R_{\lambda_0}(T)$ its continuous inverse. Corollary 2.6 then implies that for any $S \in L(X)$ with $\|S\| < \delta := \|R_{\lambda_0}(T)\|^{-1}$ also $T - \lambda_0 \mathrm{Id} - S$ is invertible and its inverse can be written as

$$(T - \lambda_0 \mathrm{Id} + S)^{-1} = \big((T - \lambda_0 \mathrm{Id}) \cdot (\mathrm{Id} - R_{\lambda_0}(T) S)\big)^{-1} = (\mathrm{Id} - R_{\lambda_0}(T)S)^{-1} R_{\lambda_0}(T) = \sum_{j=0}^{\infty} (R_{\lambda_0}(T) S)^j R_{\lambda_0}(T),$$

where the Neumann-series converges since $\|R_{\lambda_0}(T) S\| \leq \|R_{\lambda_0}(T)\| \|S\| = \delta^{-1} \|S\| < 1$.

Given any $\lambda \in \mathbb{C}$ with $|\lambda - \lambda_0| < \delta$, we may apply this argument to $S = (\lambda - \lambda_0)\mathrm{Id}$, which has $\|S\| = |\lambda - \lambda_0|$ to obtain that $T - \lambda \mathrm{Id} = T - \lambda_0 \mathrm{Id} - S$ is invertible with inverse

$$R_\lambda(T) = (T - \lambda \mathrm{Id})^{-1} = \sum_{j=0}^{\infty} (\lambda - \lambda_0)^j R_{\lambda_0}(T)^{j+1}. \tag{8.3}$$

Hence any such $\lambda \in B_\delta(\lambda_0) \subset \mathbb{C}$ is in the resolvent set, so the resolvent set is open and the resolvent operator is analytic in λ.

To prove (ii) we first note that (i) implies that the spectrum $\sigma(T) = \mathbb{C} \setminus \rho(T)$ is closed. Furthermore, given any λ with $|\lambda| > \|T\|$ we have that $\|\frac{1}{\lambda} T\| = \frac{1}{|\lambda|} \|T\| < 1$ and we hence obtain from Lemma 2.5 that $\mathrm{Id} - \frac{1}{\lambda} T$ is invertible and hence so is $T - \lambda \mathrm{Id} = -\lambda(\mathrm{Id} - \frac{1}{\lambda} T)$, As we will later use, we furthermore have that in this case where $\lambda > \|T\|$

$$\|R_\lambda(T)\| \leq |\lambda|^{-1} \|\mathrm{Id} - \tfrac{1}{\lambda} T\| \leq |\lambda|^{-1} \sum_{j=0}^{\infty} \|\tfrac{T}{\lambda}\|^j \leq |\lambda|^{-1} \sum_{j=0}^{\infty} (\tfrac{\|T\|}{\lambda})^j = \frac{1}{|\lambda| - \|T\|}. \tag{8.4}$$

This establishes the claim that $\sigma(T) \subset \{\lambda \in \mathbb{C} : |\lambda| \leq \|T\|\}$ and hence that the spectrum is both bounded and closed, so compact.

It remains to prove that the spectrum of any operator is non-empty, which we will prove by contradiction, using both the Theorem of Hahn-Banach (applied to functionals on $L(X)$, i.e elements of $(L(X))^*$, instead of X^*) and Liouville's Theorem from Complex Analysis that the only holomorphic maps $g : \mathbb{C} \to \mathbb{C}$ which are bounded are the constant maps.

So suppose that $\sigma(T)$ is empty. Then the resolvent operator R_λ is defined on all of $\mathbb{C}$ so given any $f \in (L(X))^*$ we can define a function $g_f : \mathbb{C} \to \mathbb{C}$ by

$$g_f(\lambda) := f(R_\lambda(T)).$$

We note that this function is not only well defined, but furthermore that for any $\lambda_0 \in \mathbb{C}$ the function g_f is analytic in a neighbourhood of λ_0, namely

$$g_f(\lambda) = \sum_{j=0}^{\infty} (\lambda - \lambda_0)^j f(R_{\lambda_0}(T)^{j+1}) \tag{8.5}$$

in the neighbourhood of λ_0 where the expansion (8.3) converges. In particular, g_f is holomorphic.

We now claim that g_f is also bounded. To see this we first note that as g_f is continuous, it is bounded on any compact set, in particular on the closed disc $\overline{B_{2\|T\|}(0)}$. On the other hand, for any $\lambda \in \mathbb{C}$ with $|\lambda| \geq 2\|T\|$ we know from (8.4) that $\|R_\lambda(T)\| \leq \frac{1}{|\lambda| - \|T\|} \leq \frac{1}{\|T\|}$ and hence

$$|g_f(\lambda)| \leq \|f\|_{(L(X))^*} \|R_\lambda(T)\| \leq \|f\|_{(L(X))^*} \|T\|^{-1}$$

so that g_f is also bounded on $(B_{2\|T\|}(0))^c$.

From the Theorem of Liouville we thus obtain that g_f must be constant, $g_f(\lambda) = C_f$ for a constant that depends only on the element $f \in (L(X))^*$ used in the definition of g_f. Returning to the expansion (8.5) we thus conclude that all terms with $j \geq 1$ must be zero, i.e. that for any number $\lambda_0 \in \mathbb{C}$ and any $k \geq 2$ we have that

$$f(R_{\lambda_0}(T)^k) = 0 \text{ for every } f \in (L(X))^*.$$

But by the Theorem of Hahn-Banach, or rather its consequence that we stated in Proposition 6.5, this implies that all the operators $R_{\lambda_0}(T)^k$, $k \geq 2$, must be zero, which is of course wrong since all of these operators are powers of invertible operators and thus invertible. □

We note that the estimate that $|\lambda| \leq \|T\|$ can be refined by using

Lemma 8.4. *Let $(X, \|\cdot\|)$ be a Banach space, $T \in L(X)$. Then for any $\lambda \in \sigma(T)$ and any $j \in \mathbb{N}$ we have that $\lambda^j \in \sigma(T^j)$ and thus in particular*

$$|\lambda|^j \leq \|T^j\|.$$

Proof. We note that for every $j \in \mathbb{N}$ we can write

$$T^j - \lambda^j \mathrm{Id} = (T - \lambda \mathrm{Id})S = S(T - \lambda \mathrm{Id}) \text{ for } S := (T^{j-1} + \lambda T^{j-2} + \lambda^2 T^{j-3} + \ldots + \lambda^{j-1}),$$

where we note that these two operators commute. If λ^j is not in the spectrum of T^j, then the operator on the left is invertible and hence by Lemma 8.2 also $T - \lambda \mathrm{Id}$ must be invertible and thus λ cannot be in the spectrum of T. □

We thus know that $|\lambda| \leq \inf_{j\in\mathbb{N}} \|T^j\|^{1/j}$. Indeed, one can show that $\|T^j\|^{1/j}$ converges as $j \to \infty$ with $\lim_{j\to\infty} \|T^j\|^{1/j} = \inf_{j\in\mathbb{N}} \|T^j\|^{1/j}$ and that this number agrees with the so called *spectral radius* which is defined to be

$$r(T) := \sup\{|\lambda| : \lambda \in \sigma(T)\}.$$

Theorem 8.5. *Let $(X, \|\cdot\|)$ be a complex Banach space. Then for any $T \in L(X)$ we have*

$$r(T) = \lim_{j\to\infty} \|T^j\|^{1/j} = \inf_{j\in\mathbb{N}} \|T^j\|^{1/j}.$$

We remark that Lemma 8.4 is only a very special case of the following result, which is our second main result about the spectrum of bounded linear operators on Banach spaces:

Theorem 8.6. *Let X be a complex Banach space, $T \in L(X)$ and let p be a complex polynomial. Then*

$$\sigma(p(T)) = p(\sigma(T)) := \{p(\lambda) : \lambda \in \sigma(T)\}.$$

Here we set $p(T) := \sum_{j=0}^n a_j T^j$ if the polynomial p is given by $p(z) = \sum_{j=0}^n a_j z^j$, with the usual convention that $T^0 = \mathrm{Id}$.

Proof. We first remark that if p is constant, say $p = c \in \mathbb{R}$, then the spectrum of $p(T) = c\mathrm{Id}$ is simply $\{c\}$ while the fact that $\sigma(T)$ is non-empty implies that also $p(\sigma(T)) = \{c\}$. So suppose that p has degree $n \geq 1$ let $\mu \in \mathbb{C}$ be any given number. As we are working in $\mathbb{C}$ we can factorise $p(\cdot) - \mu$ and write it as $p(z) - \mu = \alpha(z - \beta_1(\mu)) \dots (z - \beta_n(\mu))$ for some $\alpha \neq 0$ and equally factorise

$$p(T) - \mu\mathrm{Id} = \alpha(T - \beta_1(\mu)\mathrm{Id}) \dots (T - \beta_n(\mu)\mathrm{Id}) \tag{8.6}$$

where we note that all operators on the right hand side commute which will allow us to apply Lemma 8.2.

We now note that since the zeros $\beta_j(\mu)$ of $p(\cdot) - \mu$ can be equivalently characterised as the solutions $t = \beta_j(\mu)$ of the equation $p(t) = \mu$ we have that

$$\mu \in p(\sigma(T)) \Leftrightarrow \exists j \text{ so that } \beta_j(\mu) \in \sigma(T).$$

We then note that, applying Lemma 8.2 to (8.6) yields that if $\beta_j(\mu) \in \sigma(T)$ then $p(T) - \mu\mathrm{Id}$ cannot be invertible, i.e. μ must be an element of $\sigma(p(T))$. Hence $p(\sigma(T)) \subset \sigma(p(T))$. We now prove that also $p(\sigma(T))^c \subset \sigma(p(T))^c$ and hence that $p(\sigma(T)) \subset \sigma(p(T))$. To see this we note that if $\mu \notin p(\sigma(T))$ then $\beta_j(\mu) \notin \sigma(T)$ so $T - \beta_j(\mu)\mathrm{Id}$ is invertible for all $j = 1, \dots, n$. But then (8.6) shows that $p(T) - \mu\mathrm{Id}$ is the composition of invertible operators so invertible and thus $\mu \in \rho(T) = \sigma(p(T))^c$. ∎

This theorem can in particular be applied if a given operator can be written as a polynomial of a simpler operator.

As a final result of this course, we prove that there is the following close connection between the spectrum of an operator and the spectrum of its dual operator $T' \in L(X^*)$:

Theorem 8.7. *Let $(X, \|\cdot\|)$ be a Banach space, let $T \in L(X)$ and let $T' \in L(X^*)$ be the corresponding dual operator defined by $(T'f)(x) = f(Tx)$. Then*

$$\sigma(T) = \sigma_{AP}(T) \cup \sigma_P(T').$$

Proof. By definition $\sigma_{AP}(T) \subset \sigma(T)$, so it is enough to prove

Claim 1: $\sigma_P(T') \subset \sigma(T)$

and

Claim 2: $\sigma(T) \setminus \sigma_{AP}(T) \subset \sigma_P(T')$

Proof of Claim 1: Let $\lambda \in \sigma_P(T')$. Then there exists $f \in X^*$ with $f \neq 0$ so that $T'f = \lambda f$, i.e. so that for every $x \in X$

$$0 = (T'f - \lambda f)(x) = f(Tx) - \lambda f(x) = f(Tx - \lambda x).$$

Hence the restriction $f|_Y$ of f to the image $Y = (T - \lambda \mathrm{Id})X$ of $T - \lambda \mathrm{Id}$ is zero, so as f is not the zero element, we must have that $Y \neq X$, i.e. that $T - \lambda \mathrm{Id}$ is not surjective. Thus $\lambda \in \sigma(T)$.

Proof of Claim 2: Let $\lambda \in \sigma(T) \setminus \sigma_{AP}(T)$. Then as λ is not an approximate eigenvalue of T we know that there exists some $\delta > 0$ so that $\|Tx - \lambda x\| \geq \delta \|x\|$ for all $x \in X$ which, thanks to Proposition 8.1, implies that the image $Y = (T - \lambda \mathrm{Id})(X)$ is closed. At the same time Y cannot be all of X as otherwise $T - \lambda \mathrm{Id}$ would have a bounded inverse, so Y is a proper closed subspace of X. We can thus apply Proposition 6.9, that we obtained from Hahn-Banach, to conclude that there exists some $f \in X^*$ with $\|f\| = 1$, and thus in particular $f \neq 0$, so that $f|_Y = 0$. This implies that $T'(f) = \lambda f$ and thus that $\lambda \in \sigma_P(T')$ since for every $x \in X$ we have $Tx - \lambda x \in Y$ and thus $(T'(f) - \lambda f)(x) = f(Tx - \lambda x) = 0$.

□