MATH.MA.810

Introduction to Functional Analysis

Lassi Paunonen

2024

Contents

Foreword

These lecture notes cover the material on the master's level course MATH.MA.810 "Introduction to Functional Analysis" at Tampere University. The lecture notes were written in January–February 2021, and the structure of the course (and part of the material in Chapter 1) are based very loosely on an earlier set of lecture notes "Introduction to Functional Analysis" by Seppo Pohjolainen, and with additional contributions by myself and Petteri Laakkonen during 2016–2020. The course also has a set of video lectures which are available as a playlist on Youtube (direct link here).

The main topics of the course include the theory of general *vector spaces* equipped with either a *norm* (leading to *Banach spaces*) or an *inner product* (leading to *Hilbert spaces*), and the analysis of *bounded linear operators* on such vector spaces. In particular, in many ways the theory of "bounded linear operators generalises" the theory of matrices — which are mappings between two Euclidean spaces $\mathbb{C}^m$ and $\mathbb{C}^n$ — to the situation where the spaces $\mathbb{C}^m$ and $\mathbb{C}^n$ are replaced with the more general "vector spaces". Two particularly important types of such general vector spaces are *function spaces* (that is, spaces of functions) and *spaces of infinite sequences* (or "infinite vectors"), and we will study these types of spaces in detail on this course.

The theory presented on this course also forms a foundation for further studies in functional analysis on the courses MATH.MA.830 "Advanced Functional Analysis" and MATH.APP.810 "Mathematical Control Theory".

If you discover mistakes or typographical errors in the lecture notes, or if you have any other suggestions for improvements, I would be happy to hear about them! You can contact me by email, my address is `firstname.lastname@tuni.fi`.

Lassi Paunonen,
at Tampere, Finland
6.4.2024

Contact details:

Associate Professor in Mathematics
Tampere University
Email: `firstname.lastname@tuni.fi`
Website (personal): https://lassipaunonen.wordpress.com/
Website (research group): https://sysgrouptampere.wordpress.com/
Youtube channel: https://www.youtube.com/channel/UCzAtWk1YX0yUgnU805N9jUA

Notation

The set of natural numbers	$\mathbb{N} = \{1, 2, \ldots, n, \ldots\}$
The set of integers	$\mathbb{Z} = \{\ldots, -n, \ldots, -1, 0, 1, \ldots, n, \ldots\}$
The set of rational numbers	$\mathbb{Q} = \left\{ \frac{m}{n} \;\middle|\; m, n \in \mathbb{Z},\ n \neq 0 \right\}$
The set of real numbers	$\mathbb{R}$
The set of complex numbers	$\mathbb{C}$
The real or complex absolute value	$\lvert \cdot \rvert$
A sequence	$(x_1, x_2, \ldots, x_n, \ldots) = (x_n) = (x_n)_{n=1}^{\infty}$

In addition we denote by $\sup(E)$ the **supremum** or the **least upper bound** of the set $E \subset \mathbb{R}$, i.e., a value $G \in \mathbb{R}$ such that if $M \in \mathbb{R}$ is an arbitrary upper bound[1] of the set E, then $G \leq M$.

Similarly, we denote by $\inf(E)$ the **infimum** or the **greatest lower bound** of the set $E \subset \mathbb{R}$, i.e., a value $g \in \mathbb{R}$ such that if $m \in \mathbb{R}$ is an arbitrary lower bound of the set E, then $g \geq M$.

Supremum and infimum do not always exist, but the axiom on **completeness of real numbers** implies that every non-empty subset of $\mathbb{R}$ that is bounded from above has a least upper bound. Analogously, every non-empty subset of $\mathbb{R}$ that is bounded from below has a greatest lower bound.

On this course we also use the following short-hand notation:

$$\sup_{k \in I} x_k := \sup(\{\, x_k \mid k \in I \,\}) \qquad \text{and} \qquad \sup_{a \leq t \leq b} f(t) := \sup(\{\, f(t) \mid a \leq t \leq b \,\}).$$

[1] Recall that a $M \in \mathbb{R}$ is "upper bound" of the set if $x \leq M$ for all $x \in E$.

1. Banach Spaces

In this chapter we introduce the normed spaces and Banach spaces. We begin by generalizing the Euclidean vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ to a mathematical structure called the "vector space" with addition and scalar multiplication satisfying certain axioms. In particular, we are interested in function spaces, i.e., vector spaces with functions as elements. We will introduce several vector spaces that will be encountered throughout the text. We define the dimension of a vector space and observe that most of the vector spaces, e.g., the function spaces and the sequence spaces, considered in this text are *infinite-dimensional* as opposed to the "finite-dimensional" spaces $\mathbb{R}^n$ and $\mathbb{C}^n$.

In the second main part of the chapter we equip the vector space with a **norm**. Norms are especially used in defining "the size" of an element of the vector space, and in discussing the "distance" between two elements. We especially use the norm to study the convergence of sequences on normed spaces, and define concept of **Banach spaces**, which are normed spaces which are "complete". Completeness of a normed space will be defined as the property that *every Cauchy sequence of the space is a convergent sequence*. Throughout the chapter we will encounter several examples of normed spaces, some of which have the very nice property of completeness, and some of which are not complete (but which are nevertheless important spaces).

1.1 Linear Vector spaces

Definition 1.1. The set X is a **linear space** or a **vector space** if the following operations are defined for its elements: vector addition $+$: $X \times X \to X$ and scalar multiplication $\cdot : \mathbb{C} \times X \to X$ and these operations satisfy the following axioms:

1. $x + y = y + x \quad \forall x, y \in X$
2. $x + (y + z) = (x + y) + z \quad \forall x, y, z \in X$
3. There exists a zero element $0 \in X$ so that $x + 0 = x$ for all $x \in X$
4. Every $x \in X$ has an (additive) inverse element $-x$ so that $x + (-x) = 0$
5. $\alpha(\beta x) = (\alpha\beta)x \quad \forall \alpha, \beta \in \mathbb{C},\ \forall x \in X$
6. $1x = x \quad \forall x \in X$
7. $0x = 0 \quad \forall x \in X$
8. $\alpha(x + y) = \alpha x + \alpha y \quad \forall \alpha \in \mathbb{C},\ \forall x, y \in X$
9. $(\alpha + \beta)x = \alpha x + \beta x \quad \forall \alpha, \beta \in \mathbb{C},\ \forall x \in X$

It should be noted that vector spaces can also be defined using real scalars $\alpha, \beta \in \mathbb{R}$. Then we call the vector space a **real vector space**. If not stated otherwise, we always assume that the scalars are complex numbers. We can call a vector space a **complex vector space** if we want to underline that the scalars are complex numbers.

Example 1.2. It is straightforward to verify that $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces if vector addition and scalar multiplication are defined in the usual way. ◇

The vector space can consist of elements that are very different from those vectors that we know from the elementary linear algebra. Particularly interesting vector spaces are those consisting of functions or sequences which will be considered throughout this text as standard examples of infinite-dimensional spaces (dimension is to be defined later).

Example 1.3 (The space $C(I)$ of continuous functions)**.** Let $I \subseteq \mathbb{R}$ be an arbitrary fixed interval (open or (half-)closed / bounded or unbounded). We denote by $C(I)$ the set of all functions $f : I \to \mathbb{C}$ which are continuous on I. The set $C(I)$ becomes a vector space if the addition of two elements and the scalar multiplication are defined so that

$$\begin{aligned} (f+g)(t) &= f(t) + g(t) \quad && \forall t \in I \\ (\alpha f)(t) &= \alpha f(t) && \forall t \in I \end{aligned}$$

for all $f, g \in C(I)$ and $\alpha \in \mathbb{C}$.

The interval I is fixed and changing it will result in a different space. In fact, even including or excluding one of the endpoints of the interval changes the space $C(I)$ drastically: For example, the function $f : (0,1] \to \mathbb{C}$ defined by $f(t) = 1/t$ is in $C((0,1])$, but not in $C([0,1])$ (more precisely, it is not possible to extend the function f to $[0,1]$ in such a way that the result would be continuous on $[0,1]$). ◇

Example 1.4 (The space $C(\Omega)$ of continuous functions)**.** More generally, if $\Omega \subset \mathbb{R}^n$ for some $n \in \mathbb{N}$, the set $C(\Omega)$ of continuous functions $f : \Omega \to \mathbb{C}$ becomes a vector space if we define the addition and scalar multiplication so that

$$\begin{aligned} (f+g)(z) &= f(z) + g(z) \quad && \forall z \in \Omega \\ (\alpha f)(z) &= \alpha f(z) && \forall z \in \Omega \end{aligned}$$

for all $f, g \in C(\Omega)$ and $\alpha \in \mathbb{C}$. ◇

Example 1.5 (The sequence space $\ell(\mathbb{C})$)**.** The set

$$\ell(\mathbb{C}) = \left\{ x \;\middle|\; x = (x_1, x_2, \ldots, x_n, \ldots),\, x_i \in \mathbb{C}, i \in \mathbb{N} \right\}$$

of complex sequences is a vector space when addition and scalar multiplication are defined as

$$\begin{aligned} x + y &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n, \ldots) \\ \alpha x &= (\alpha x_1, \alpha x_2, \ldots, \alpha x_n, \ldots) \end{aligned}$$

where $x = (x_1, x_2, \ldots, x_n, \ldots)$ and $y = (y_1, y_2, \ldots, y_n, \ldots)$ and $\alpha \in \mathbb{C}$. We can note that the addition and scalar multiplication in $\ell(\mathbb{C})$ are defined in exactly the same way as in $\mathbb{C}^n$. In fact, the sequence $(x_1, x_2, \ldots,) \in \ell(\mathbb{C})$ can be considered to be a *vector with an infinite number of elements* x_k.

Often on this course we will also use more compact notations for the sequence $x = (x_1, x_2, \ldots)$, namely, $x = (x_k)_{k=1}^{\infty}$, $x = (x_k)_{k \in \mathbb{N}}$, or $(x_k)_k$. ◇

Remark 1.6. To be precise, the linear vector space is the triplet $(X, +, \cdot)$. However, it is common practice not to explicitly state the addition "+" and the scalar multiplication "·" if they are clear from the context. In particular, when talking about the vector spaces of Examples 1.2–1.5 on this course we always assume that the addition and scalar multiplication are defined as above.

Definition 1.7. A non-empty subset S of a vector space X is a **subspace of** X if

$$\alpha x + \beta y \in S \qquad \forall x, y \in S \;\; \forall \alpha, \beta \in \mathbb{C}. \tag{1.1}$$

Note that since S is a subset of X, the vector addition and scalar multiplication operators of X are also defined for the elements in S. Both of these operations also appear in the condition (1.1).

Example 1.8 (The trivial subspace $\{0\}$)**.** If X is a vector space, then $S = \{0\} \subset X$ is a subspace of X. Indeed, S is not empty, since $0 \in S$. To prove (1.1), let $x, y \in S$ and $\alpha, \beta \in \mathbb{C}$ be arbitrary. Since $S = \{0\}$, we necessarily have $x = y = 0$. The vector space axiom 7 implies that $0 \cdot 0 = 0$ (meaning, "$0 \in \mathbb{C}$ times $0 \in X$ is $0 \in X$"), and axioms 5, 7, and 3 further imply

$$\alpha x + \beta y = \alpha 0 + \beta 0 = \alpha(0 \cdot 0) + \beta(0 \cdot 0) = (\alpha 0)0 + (\beta 0)0 = 0 \cdot 0 + 0 \cdot 0 = 0 + 0 = 0.$$

Since $\alpha x + \beta y = 0 \in S$ and $\alpha, \beta \in \mathbb{C}$ and $x, y \in X$ were arbitrary, (1.1) holds and $S = \{0\}$ is indeed a subspace of X.

The subspace $S = \{0\}$ is referred to as the "trivial subspace" of a vector space X, and by assuming that a subspace S of X is "non-trivial" we mean that S is assumed to contain at least one element which is not the zero element. The same terminology of "trivial" and "non-trivial" is used when discussing vector spaces (since $X = \{0\}$ is indeed a vector space). Any real or complex non-trivial vector space always contains an infinite number of elements, whereas the trivial space $\{0\}$ only contains a single element! ◇

It is straightforward to prove that all of the vector space axioms are also satisfied on any subspace S of a vector space X (when the addition and scalar multiplication are defined in the same way as in X), and therefore every subspace S is in fact a vector space as well. This opens up the possibility that we can (sometimes very easily) prove that a given set X with given operations is a vector space by identifying it as a subspace of a larger vector space $\tilde{X}$ with the same operations. This is demonstrated in the examples below. It is also trivial to show that if S_1 is a subspace of X and S_2 is a subspace of S_1, then S_2 is a subspace of X.

The next examples introduce selected important subspaces of the spaces $C(I)$ and $\ell(\mathbb{C})$. Recall that as in linear algebra, proving that S is a subspace involves two things: Verifying that S is not empty (e.g., it contains at least the zero element $0 \in X$) and that the property (1.1) holds.

Example 1.9 (Subspaces of $C(I)$)**.** In this example, $I \subset \mathbb{R}$ is a non-empty interval and the addition and scalar multiplication are defined as in Example 1.3.

(i) Let $P(I)$ be the set of polynomials defined on the interval $I \subset \mathbb{R}$, i.e.,

$$P(I) = \left\{ p : I \to \mathbb{C} \;\middle|\; p(t) = \sum_{k=0}^{n} a_k t^k \text{ for some } n \in \mathbb{N} \cup \{0\} \text{ and } \{a_k\}_{k=0}^{n} \subset \mathbb{C} \right\}.$$

Since all polynomials are continuous functions, we have $P(I) \subset C(I)$. The function $p : I \to \mathbb{C}$ satisfying $p(t) \equiv 0$ is a polynomial, and thus $P(I)$ is not empty. It is also easy to verify that a linear combination of two polynomials is again a polynomial, and thus (1.1) holds. Thus $P(I)$ is a subspace of $C(I)$, and in particular $P(I)$ is vector space when the vector addition and scalar multiplication are defined on the same way as on the space $C(I)$.

(ii) The set

$$C^n(I) = \left\{ f : I \to \mathbb{C} \;\middle|\; f^{(k)} \text{ exists and is continuous in } I \text{ for all } k = 0, 1, \dots, n \right\}$$

of all n-times continuously differentiable functions on I is a subspace of $C(I)$. In addition, we can define the space $C^\infty(I)$ of functions whose derivatives of *all* orders are continuous functions,

$$C^\infty(I) = \left\{ f : I \to \mathbb{C} \;\middle|\; f^{(k)} \text{ exists and is continuous in } I \text{ for all } k \in \mathbb{N} \right\}.$$

The functions $f \in C^\infty(I)$ are said to be "*smooth*" due to their excellent differentiability properties. Both of these spaces are non-empty, since the zero function on I is is smooth. Moreover, the condition (1.1) can be verified easily based on fundamental differential calculus of functions of a single variable.

Sometimes it is convenient to use the notation $C^0(I) = C(I)$. It follows from the definitions that $C^n(I)$ is a subspace of $C^m(I)$ whenever $n > m$. In addition, since a polynomial has derivatives of all orders we have that $P(I)$ is a subspace of $C^n(I)$ for any $n \in \mathbb{N}$ and of $C^\infty(I)$.

(iii) Consider the set

$$C_c^\infty(\mathbb{R}) = \left\{ f \in C^\infty(\mathbb{R}) \;\middle|\; \exists M > 0 : f(t) = 0 \text{ whenever } |t| > M \right\}$$

of all smooth functions on $\mathbb{R}$ with *compact support*. This term refers to the property that the "support of f", defined as

$$\operatorname{supp} f = \overline{\{ t \in \mathbb{R} \mid f(t) \neq 0 \}} \subset \mathbb{R}$$

is a compact subset of $\mathbb{R}$. Here the horizontal line denotes the closure of the set in $\mathbb{R}$, and thus $\operatorname{supp} f$ is always a closed subset of $\mathbb{R}$. By definition, each function $f \in C_c^\infty$ can have nonzero values $f(t) \neq 0$ only inside a bounded interval $I = [a, b] \subset \mathbb{R}$. We can now show that C_c^∞ is a subspace of $C^\infty(\mathbb{R})$ (and consequently also a subspace of $C^n(I)$ for all $n \in \mathbb{N} \cup \{0\}$). Since the zero function on $\mathbb{R}$ definitely has a compact support (in fact $\operatorname{supp} 0$ is the empty set), we have $0 \in C_c^\infty(\mathbb{R})$, and thus $C_c^\infty(\mathbb{R})$ is not empty. Let $f, g \in C_c^\infty(\mathbb{R})$ and $\alpha, \beta \in \mathbb{C}$ be arbitrary. By definition there exist $M_f, M_g > 0$ such that $f(t) = 0$ whenever $|t| > M_f$ and $g(t) = 0$ whenever $|t| > M_g$. Since $\alpha f + \beta g$ has continuous derivatives of all orders, and $\alpha f(t) + \beta g(t) = 0$ whenever $|t| > \max\{M_f, M_g\}$, we have $\alpha f + \beta g \in C_c^\infty(\mathbb{R})$. Thus $C_c^\infty(\mathbb{R})$ is a subspace of $C^\infty(\mathbb{R})$, and of $C(\mathbb{R})$.

(iv) Since polynomials have continuous derivatives of all orders, we also have that $P(I)$ is a subspace of $C^\infty(I)$ for any interval $I \subset \mathbb{R}$. Similarly, $P(\mathbb{R})$ is a subspace of $C^\infty(\mathbb{R})$, but not a subspace of $C_c^\infty(\mathbb{R})$, since $P(\mathbb{R}) \not\subset C_c^\infty(\mathbb{R})$.

◇

Example 1.10 (Subspaces of $\ell(\mathbb{C})$)**.** In this example, the addition and scalar multiplication are defined as in Example 1.5.

(i) Let $\ell^{fin}(\mathbb{C})$ be the set of all complex sequences that have only a finite number of non-zero terms, i.e.,

$$\ell^{fin}(\mathbb{C}) = \left\{ (x_k)_{k=1}^{\infty} \in \ell(\mathbb{C}) \;\middle|\; \exists N \in \mathbb{N} : k \geq N \implies x_k = 0 \right\}.$$

The set $\ell^{fin}(\mathbb{C})$ is non-empty, because it contains the zero sequence $(0, 0, \ldots)$. To verify that (1.1) holds, let $x = (x_k)_k \in \ell^{fin}(\mathbb{C})$, $y = (y_k)_k \in \ell^{fin}(\mathbb{C})$, and $\alpha, \beta \in \mathbb{C}$ be arbitrary. By definition of $\ell^{fin}(\mathbb{C})$, there exists $N_x, N_y \in \mathbb{N}$ such that $x_k = 0$ if $k \geq N_x$ and $y_k = 0$ if $k \geq N_y$. We have

$$\alpha x + \beta y = (\alpha x_k + \beta y_k)_{k=1}^{\infty},$$

and $\alpha x_k + \beta y_k = 0$ if $k \geq \max\{N_x, N_y\}$. This means that $\alpha x + \beta y \in \ell^{fin}(\mathbb{C})$ by definition and (1.1) holds. Thus $\ell^{fin}(\mathbb{C})$ is a subspace of $\ell(\mathbb{C})$.

(ii) A complex sequence $(x_k)_{k=1}^{\infty} = (x_1, x_2, \ldots)$ is said to be *bounded* if there exists $M > 0$ such that $|x_k| \leq M$ for all $k \in \mathbb{N}$. This property is equivalent to the condition that $\sup_{k\in\mathbb{N}} |x_k| < \infty$. The space of all "bounded sequences"

$$\ell^{\infty}(\mathbb{C}) = \left\{ (x_k)_{k=1}^{\infty} \in \ell(\mathbb{C}) \;\middle|\; (x_k)_{k=1}^{\infty} \text{ is bounded} \right\}$$

is a subspace of $\ell(\mathbb{C})$ (proved as an exercise).

(iii) For $p \geq 1$, the set $\ell^p(\mathbb{C})$ of "p-summable sequences" defined as

$$\ell^p(\mathbb{C}) = \left\{ (x_k)_{k=1}^{\infty} \in \ell(\mathbb{C}) \;\middle|\; \sum_{k=1}^{\infty} |x_k|^p < \infty \right\}$$

is a subspace of $\ell(\mathbb{C})$ (proved as an exercise).

◇

As mentioned above, the fact that $\ell^{fin}(\mathbb{C})$, $\ell^{\infty}(\mathbb{C})$ and $\ell^p(\mathbb{C})$ for $p \geq 1$ are subspaces of $\ell(\mathbb{C})$ implies that all of these spaces are vector spaces (with the same definitions of vector addition and scalar multiplication as in $\ell(\mathbb{C})$).

Proposition 1.11. *For all $1 \leq p \leq q < \infty$ we have*

$$\ell^{fin}(\mathbb{C}) \subset \ell^p(\mathbb{C}) \subset \ell^q(\mathbb{C}) \subset \ell^{\infty}(\mathbb{C}).$$

In particular, $\ell^p(\mathbb{C})$ is a subspace of $\ell^q(\mathbb{C})$ and of $\ell^{\infty}(\mathbb{C})$.

Proof. The claim regarding the subspace property follows immediately once we show that the inclusions $\ell^p(\mathbb{C}) \subset \ell^q(\mathbb{C}) \subset \ell^{\infty}(\mathbb{C})$ hold, because we saw in Example 1.10 that $\ell^p(\mathbb{C})$ is non-empty and closed under the vector addition and scalar multiplication (i.e., (1.1) holds).

To show that $\ell^q(\mathbb{C}) \subset \ell^{\infty}(\mathbb{C})$, let $x = (x_k)_{k=1}^{\infty} \in \ell^q(\mathbb{C})$ be arbitrary. Since $|x_k|^q \geq 0$ for all $k \in \mathbb{N}$, the series $\sum_{k=1}^{\infty} |x_k|^q$ can converge only if $|x_k|^q \to 0$ as $k \to \infty$. We know from elementary calculus that any convergent sequence of scalars is bounded, meaning that there exists a constant $M > 0$ such that $|x_k|^q \le M$ (and thus also $|x_k| \le \sqrt[q]{M}$) for all $k \in \mathbb{N}$. This implies $x \in \ell^\infty(\mathbb{C})$, and since $x \in \ell^q(\mathbb{C})$ was arbitrary, we have that $\ell^q(\mathbb{C}) \subset \ell^\infty(\mathbb{C})$.

To show $\ell^p(\mathbb{C}) \subset \ell^q(\mathbb{C})$, let $p \le q$ and let $(x_k)_{k=1}^\infty \in \ell^p(\mathbb{C})$ be arbitrary. As shown above, we have $\sup_{k\in\mathbb{N}} |x_k| < \infty$. Using this property we can estimate

$$\sum_{k=1}^\infty |x_k|^q = \sum_{k=1}^\infty |x_k|^{q-p}|x_k|^p \le \sum_{k=1}^\infty \Big(\sup_{n\in\mathbb{N}} |x_n|\Big)^{q-p}|x_k|^p = \underbrace{\Big(\sup_{n\in\mathbb{N}} |x_n|\Big)^{q-p}}_{<\infty} \cdot \sum_{k=1}^\infty |x_k|^p < \infty.$$

since $q - p \ge 0$ and $x \in \ell^p(\mathbb{C})$ by assumption. Thus $x \in \ell^q(\mathbb{C})$. Since $x \in \ell^p(\mathbb{C})$ was arbitrary, we have proved that $\ell^p(\mathbb{C}) \subset \ell^q(\mathbb{C})$.

Finally, if $x = (x_k)_{k=1}^\infty \in \ell^{fin}(\mathbb{C})$, there exists $K \in \mathbb{N}$ such that $x_k = 0$ for all $k \ge K$. This also implies $\sum_{k=1}^\infty |x_k|^p = \sum_{k=1}^K |x_k|^p < \infty$, and thus $x \in \ell^p(\mathbb{C})$. □

The inclusions between the spaces $\ell^p(\mathbb{C})$ for different values of $1 \le p < \infty$ are related to the rates at which the elements x_k converge to zero as $k \to \infty$. In particular, if $q > p \ge 1$ the space $\ell^q(\mathbb{C})$ is larger than $\ell^p(\mathbb{C})$, and in particular $\ell^q(\mathbb{C})$ contains sequences whose elements x_k are allowed to converge to zero at a slower rate as $k \to \infty$. For example, the elements of $x = \left(\frac{1}{\sqrt{k}}\right)_{k=1}^\infty$ converge to zero at a slower rate than the elements of $y = \left(\frac{1}{k}\right)_{k=1}^\infty$. Indeed, we have $x \notin \ell^2(\mathbb{C})$, but $y \in \ell^2(\mathbb{C})$ since

$$\sum_{k=1}^\infty \left|\frac{1}{\sqrt{k}}\right|^2 = \sum_{k=1}^\infty \frac{1}{k} = \infty \quad \text{and} \quad \sum_{k=1}^\infty \left|\frac{1}{k}\right|^2 = \sum_{k=1}^\infty \frac{1}{k^2} < \infty.$$

On the other hand, $x \in \ell^3(\mathbb{C})$ since

$$\sum_{k=1}^\infty \left|\frac{1}{\sqrt{k}}\right|^3 = \sum_{k=1}^\infty \frac{1}{k^{\frac{3}{2}}} < \infty.$$

Exercise 1.12. Show that the complex sequence $x = \left(\frac{1}{k}\right)_{k=1}^\infty$ is in $\ell^p(\mathbb{C})$ whenever $1 < p \le \infty$, but $x \notin \ell^1(\mathbb{C})$. In addition, give an example of a sequence that is in $\ell^\infty(\mathbb{C})$, but not in $\ell^p(\mathbb{C})$ for any $1 \le p < \infty$. ◇

1.1.1 Dimension of a Vector Space

In this section, we define the concept of **dimension** of a vector space, and define what an **infinite-dimensional space** means. We begin by generalizing the concept of linear independence of Euclidean space $\mathbb{R}^n$ to a general vector space. The dimension of a vector space is then defined as the number of elements in the largest possible linearly independent set in the space, and this number of elements can be infinite. The function and sequence spaces introduced in the previous section are typical examples of infinite-dimensional spaces. Later on this course we will see that some of the familiar properties of the Euclidean spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ depend crucially on the fact that these vector spaces are finite-dimensional. For example, every closed and bounded subset of $\mathbb{R}^n$ is compact, but this is *not* true for subsets of a general vector space.

Definition 1.13. Let X be a vector space. A finite set of elements $\{x_1, x_2, \dots, x_n\} \subset X$ is said to be **linearly independent** if for arbitrary scalars $\{\alpha_k\}_{k=1}^n \subset \mathbb{C}$ the equation

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$$

holds only if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. If $\{x_1, \dots, x_n\}$ is not linearly independent, it is **linearly dependent**.

An infinite set of elements $\{x_k\}_{k=1}^\infty \subset X$ is said to be **linearly independent** if every finite subset of $\{x_k\}_{k=1}^\infty$ is linearly independent.

As in linear algebra, a vector $x \in X$ is a said to be a (finite) linear combination of $\{x_1, x_2, \dots, x_n\}$ if there exist scalars $\{\alpha_k\}_{k=1}^n \subset \mathbb{C}$ such that

$$x = \sum_{k=1}^n \alpha_k x_k = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n.$$

We will also encounter *infinite* linear combinations on this course, but such objects cannot be be studied without the concept of *convergence*, which in turn requires additional structure from the vector space.

Definition 1.14. The **dimension** of the vector space X is $\dim(X) = n \in \mathbb{N}$ if X contains a linearly independent set with n elements, and if every set with $n+1$ elements of X is linearly dependent. If such a number $n \in \mathbb{N}$ does not exist, then X is **infinite-dimensional**, and we denote $\dim(X) = \infty$.

Example 1.15. For $n \in \mathbb{N}$, we have $\dim(\mathbb{R}^n) = n$ and $\dim(\mathbb{C}^n) = n$. $\diamond$

The following lemma shows that if we want to show that a space is infinite-dimensional, it suffices to show that the space contains an infinite-dimensional subspace.

Lemma 1.16. *Let S be a subspace of the vector space X. Then* $\dim(S) \le \dim(X)$. *If* $\dim(S) = \infty$ *this inequality is interpreted as "S is infinite-dimensional, then X is infinite-dimensional as well".*

Proof. Assume first that X is finite-dimensional, i.e. $\dim(X) = n \in \mathbb{N}$. Then by definition every set of $n+1$ elements of X is linearly dependent. But if we take an arbitrary set of elements $\{x_k\}_{k=1}^{n+1} \subset S$, then the fact that $\{x_k\}_{k=1}^{n+1} \subset X$ implies that $\{x_k\}_{k=1}^{n+1}$ is linearly dependent. Thus every set of $n+1$ elements of S is linearly dependent, and by definition we must have $\dim(S) \le n$. Thus $\dim(S) \le \dim(X)$ if $\dim(X) < \infty$.

On the other hand, if $\dim(S) = \infty$, the subspace S contains linearly independent sets with arbitrary numbers of elements. Since $S \subset X$, the same is true for the space X. Thus by definition we have that $\dim(X) = \infty$ as well. $\square$

Example 1.17. Let $I \subset \mathbb{R}$ be an interval with positive length.

(a) The space $P(I)$ of polynomials is infinite-dimensional. We can prove this by showing that $P(I)$ contains arbitrarily large linearly independent sets. To this end, let $n \in \mathbb{N}$ be arbitrary and consider the set $\{f_k\}_{k=1}^n \subset P(I)$ of *monomials*, so that $f_k(t) = t^{k-1}$ for $k \in \{1, \dots, n\}$. If $\{\alpha_k\}_{k=1}^n \subset \mathbb{C}$ are arbitrary scalars, then

$$\alpha_1 f_1(t) + \alpha_2 f_2(t) + \ldots + \alpha_n f_n(t) = \alpha_n t^{n-1} + \cdots + \alpha_2 t + \alpha_1$$

is a polynomial of order $n-1$. Thus the condition that

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n = 0$$

requires that the polynomial $\alpha_n t^{n-1} + \cdots + \alpha_2 t + \alpha_1$ is identically zero everywhere on I. However, due to the fundamental theorem of algebra, this polynomial can have at most $n-1$ roots on I, unless it is the zero polynomial. Thus necessarily $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$, and the set $\{f_k\}_{k=1}^n \subset \mathbb{N}$ is linearly independent. Since $n \in \mathbb{N}$ was arbitrary, the we have by definition that $\dim(P(I)) = \infty$.

(b) Part (a) and Lemma 1.16 imply that we also have

$$\begin{aligned} \dim(C(I)) &= \infty \\ \dim(C^m(I)) &= \infty \quad \text{for all } m \in \mathbb{N} \\ \dim(C^\infty(I)) &= \infty, \end{aligned}$$

since $P(I)$ is a subspace of all of these spaces.

◇

Example 1.18. Since $P(\mathbb{R}) \not\subset C_c^\infty(\mathbb{R})$, we need a separate argument to show that $\dim(C_c^\infty(\mathbb{R})) = \infty$. It is again sufficient to show that $C_c^\infty(\mathbb{R})$ contains arbitrarily large linearly independent sets. To this end, define $e_k \in C_c^\infty(\mathbb{R})$ for $k \in \mathbb{N}$ by

$$e_k(t) = \begin{cases} e^{\frac{-1}{(t-k)(k+1-t)}} & t \in (k, k+1) \\ 0 & \text{otherwise} \end{cases}$$

Note that $e_k(t)$ are indeed smooth functions, and each $e_k(t)$ is nonzero only on the interval $(k, k+1)$. Let $n \in \mathbb{N}$ be arbitrary and let $\{\alpha_k\}_{k=1}^n \subset \mathbb{C}$ be such that

$$\alpha_1 e_1(t) + \alpha_2 e_2(t) + \cdots + \alpha_n e_n(t) = 0 \qquad \text{for all } t \in \mathbb{R}.$$

If $k \in \{1, \dots, n\}$ and $t \in (k, k+1)$, the above assumption and the fact that $e_j(t) = 0$ for all $j \in \{1, \dots, n\}$ such that $j \neq k$ implies that $\alpha_k e_k(t) = 0$. But since $e_k(t) \neq 0$ by definition, we must have $\alpha_k = 0$. Since $k \in \{1, \dots, n\}$ was arbitrary, we deduce that $\alpha_1 = \ldots = \alpha_n = 0$, and thus the set $\{e_k\}_{k=1}^n \subset C_c^\infty(\mathbb{R})$ is linearly independent. Finally, since $n \in \mathbb{N}$ was arbitrary, $C_c^\infty(\mathbb{R})$ is an infinite-dimensional vector space by definition.

Lemma 1.16 again implies that any vector space which contains $C_c^\infty(\mathbb{R})$ as a subspace is infinite-dimensional. ◇

1.2 Banach Spaces

In this section we equip a vector space X with additional structure, namely, a *norm*. A norm especially allows us to discuss “sizes” of elements $x \in X$ as well as “distances” between two elements $x \in X$ and $y \in X$. This also lets us define the concept of **convergence** of sequences and series. Finally, we will define the important and desirable property of **completeness** of a normed vector space. Complete normed spaces are called **Banach spaces**. We will also study several examples of normed spaces, some of which are complete (i.e., which are Banach spaces), and some of which we find not to be complete.

Definition 1.19. Let X be a vector space. A **norm** on X is a function $\|\cdot\| : X \to [0,\infty)$ so that

(1) $\|x\| \geq 0$ for all $x \in X$. Moreover, $\|x\| = 0$ if and only if $x = 0$.

(2) $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{C}$ and $x \in X$.

(3) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$.

The pair $(X, \|\cdot\|)$ is called a **normed (linear) space**.

The axioms (1)–(3) which the norm $\|\cdot\|$ is required to satisfy are precisely the most fundamental properties of the norm on Euclidean spaces $\mathbb{R}^n$ and $\mathbb{C}^n$. Indeed, these spaces are important examples of normed spaces, as shown in the next example.

Example 1.20. The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ are normed spaces if the norm is defined to be the Euclidean norm, i.e.,

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}.$$

However, the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ can alternatively be equipped with other norms, for example $\|x\|_\infty = \max_{1\leq k\leq n} |x_k|$ and $\|x\|_1 = |x_1| + \cdots + |x_n|$ are both norms for $\mathbb{R}^n$ and $\mathbb{C}^n$. (You can verify that $\|\cdot\|_\infty$ and $\|\cdot\|_1$ satisfy the axioms above!). ◇

As discussed in the previous example, it is possible to define several different norms on the same vector space X. In fact, a norm on a normed space is *never* unique, because on any normed space $(X, \|\cdot\|)$ it is possible to define another norm $\|\cdot\|_\beta$ simply by "scaling" the original norm $\|\cdot\|$ by some constant $\beta > 0$, so that $\|x\|_\beta := \beta\|x\|$ for all $x \in X$. In this case, $(X, \|\cdot\|)$ and $(X, \|\cdot\|_\beta)$ will be two different normed spaces. Apart from such "simple" modifications of the norm, in our examples we will later see that the choice of the norm on a given vector space X can have a great effect on the properties of the resulting normed space $(X, \|\cdot\|)$.

Example 1.21. The space $\ell(\mathbb{C})$ is too large to be a normed space. However, we can define norms on subspaces of $\ell(\mathbb{C})$ studied in Example 1.10. Indeed, the space $\ell^\infty(\mathbb{C}) = \{ (x_k)_k \in \ell(\mathbb{C}) \mid \sup_{k\in\mathbb{N}}|x_k| < \infty \}$ can be equipped with the norm $\|\cdot\|_\infty$ defined by

$$\|x\|_\infty = \sup_{k\in\mathbb{N}} |x_k|, \qquad \forall x = (x_k)_k \in \ell^\infty(\mathbb{C}).$$

Moreover, for $1 \leq p < \infty$ the space $\ell^p(\mathbb{C}) = \{ (x_k)_k \in \ell(\mathbb{C}) \mid \sum_{k=1}^\infty |x_k|^p < \infty \}$ can be equipped with the norm $\|\cdot\|_p$ defined by

$$\|x\|_p = \left(\sum_{k=1}^\infty |x_k|^p \right)^{\frac{1}{p}}, \qquad \forall x = (x_k)_k \in \ell^p(\mathbb{C}).$$

Note that neither $\|x\|_\infty$ nor $\|x\|_p$ are well-defined for all sequences x in $\ell(\mathbb{C})$. Since $\ell^{fin}(\mathbb{C})$ is a subspace of $\ell^\infty(\mathbb{C})$ and $\ell^p(\mathbb{C})$ for all $1 \leq p < \infty$, these norms are can also be used to define a normed space $(\ell^{fin}(\mathbb{C}), \|\cdot\|_\infty)$ or $(\ell^{fin}(\mathbb{C}), \|\cdot\|_p)$.

In the following, we will verify that $\|\cdot\|_\infty$ is indeed a norm by proving that it satisfies the axioms (1)–(3) in Definition 1.19. The property that $(\ell^p(\mathbb{C}), \|\cdot\|_p)$ is a normed space is proved separately in Theorem 1.22.

(1) Since every sequence $x = (x_k)_{k=1}^\infty \in \ell^\infty(\mathbb{C})$ is by definition bounded, the supremum $\sup_{k\in\mathbb{N}}|x_k|$ exists and is finite for every $x = (x_k)_{k=1}^\infty \in \ell^2(\mathbb{C})$. Thus $\|x\|_\infty$ is well-defined for all $x \in \ell^\infty(\mathbb{C})$. Since $|x_k| \geq 0$ we also have $\|x\|_\infty = \sup_{k\in\mathbb{N}}|x_k| \geq 0$ for all $x \in \ell^\infty(\mathbb{C})$. Finally, for every $x \in \ell^\infty(\mathbb{C})$ we have

$$\begin{aligned} \|x\|_\infty = 0 \quad &\Leftrightarrow \quad \sup_{k\in\mathbb{N}} |x_k| = 0 \quad \Leftrightarrow \quad |x_k| = 0 \quad \forall k \in \mathbb{N} \\ &\Leftrightarrow \quad x_k = 0 \quad \forall k \in \mathbb{N} \quad \Leftrightarrow \quad x = 0. \end{aligned}$$

(2) Let $x = (x_k)_{k=1}^\infty \in \ell^\infty(\mathbb{C})$ and $\alpha \in \mathbb{C}$ be arbitrary. Since $\alpha x = (\alpha x_1, \alpha x_2, \ldots)$ we have

$$\|x\|_\infty = \sup_{k\in\mathbb{N}} |\alpha x_k| = \sup_{k\in\mathbb{N}} |\alpha||x_k| = |\alpha| \sup_{k\in\mathbb{N}} |x_k| = |\alpha|\|x\|.$$

(3) Let $x = (x_k)_{k=1}^\infty \in \ell^\infty(\mathbb{C})$ and $y = (y_k)_{k=1}^\infty \in \ell^\infty(\mathbb{C})$ be arbitrary. Using the triangle inequality for complex numbers and the property $\sup_{k\in\mathbb{N}}(a_k + b_k) \leq (\sup_{k\in\mathbb{N}} a_k) + (\sup_{k\in\mathbb{N}} b_k)$ for $a_k, b_k \geq 0$ we can estimate

$$\|x+y\|_\infty = \sup_{k\in\mathbb{N}} |x_k+y_k| \leq \sup_{k\in\mathbb{N}}(|x_k|+|y_k|) \leq \Big(\sup_{k\in\mathbb{N}} |x_k|\Big)+\Big(\sup_{k\in\mathbb{N}} |y_k|\Big) = \|x\|_\infty+\|y\|_\infty.$$

◇

Theorem 1.22. *If $p \geq 1$, then $(\ell^p(\mathbb{C}), \|\cdot\|_p)$ is a normed space. In particular, the norms $\|\cdot\|_p$ for $p \geq 1$ have the following properties:*

(a) ***(Minkowski's Inequality)*** *If $p \geq 1$, then*

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p, \qquad \forall x, y \in \ell^p(\mathbb{C}).$$

(b) ***(Hölder's Inequality)*** *If $p, q > 1$ are such that $\frac{1}{p} + \frac{1}{q} = 1$, then for all $x = (x_k)_{k=1}^\infty \in \ell^p(\mathbb{C})$ and $y = (y_k)_{k=1}^\infty \in \ell^q(\mathbb{C})$ we have*

$$\sum_{k=1}^\infty |x_k y_k| \leq \|x\|_p \|y\|_q.$$

Proof. Let $p \geq 1$. If $x \in \ell^p(\mathbb{C})$, we clearly have that $\|x\|_p$ is well-defined and $\|x\|_p \geq 0$. Moreover, $\|0\|_p = \sum_{k=1}^\infty 0^p = 0$. Conversely, if $\|x\|_p = \sum_{k=1}^\infty |x_k|^p = 0$ for some $x \in \ell^p(\mathbb{C})$, then clearly $|x_k| = 0$ for all $k \in \mathbb{N}$, which implies $x = 0$. Thus $\|\cdot\|_p$ satisfies axiom (1). To prove axiom (2), let $x \in \ell^p(\mathbb{C})$ and $\alpha \in \mathbb{C}$ be arbitrary. We have

$$\|\alpha x\|_p = \left(\sum_{k=1}^\infty |\alpha x_k|^p\right)^{\frac{1}{p}} = \left(\sum_{k=1}^\infty |\alpha|^p |x_k|^p\right)^{\frac{1}{p}} = |\alpha| \left(\sum_{k=1}^\infty |x_k|^p\right)^{\frac{1}{p}} = |\alpha|\|x\|_p.$$

Thus axiom (2) holds. Finally, axiom (3) follows directly from the Minkowski's Inequality which is verified in the remaining part of the proof.

We begin by proving the Hölder's inequality. Let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and let $x \in \ell^p(\mathbb{C})$ and $y \in \ell^q(\mathbb{C})$. We can assume $\|x\|_p > 0$ and $\|y\|_q > 0$, since the proof is trivial if $x = 0$ or $y = 0$. The proof utilises the **Young's inequality**, which states that for any scalars $a, b \geq 0$ we have $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$. We can use this inequality to estimate

$$\begin{aligned}\frac{1}{\|x\|_p\|y\|_q}\sum_{k=1}^{\infty}|x_k y_k| &= \sum_{k=1}^{\infty}\frac{|x_k|}{\|x\|_p}\cdot\frac{|y_k|}{\|y\|_q} \overset{\text{Young}}{\leq} \sum_{k=1}^{\infty}\left(\frac{|x_k|^p}{p\|x\|_p^p} + \frac{|y_k|^q}{q\|y\|_q^q}\right)\\ &= \frac{1}{p\|x\|_p^p}\underbrace{\sum_{k=1}^{\infty}|x_k|^p}_{=\|x\|_p^p} + \frac{1}{q\|y\|_q^q}\underbrace{\sum_{k=1}^{\infty}|y_k|^q}_{=\|y\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1.\end{aligned}$$

Multiplying both sides of this inequality with $\|x\|_p\|y\|_q$ shows that the Hölder's Inequality holds.

The Hölder's Inequality can be used to prove the Minkowski's Inequality in the case where $p > 1$. To this end, let $x, y \in \ell^p(\mathbb{C})$. We can assume that $x + y \neq 0$, since otherwise the estimate is trivial. A simple estimate shows that

$$\begin{aligned}\|x+y\|_p^p &= \sum_{k=1}^{\infty}|x_k + y_k|^p = \sum_{k=1}^{\infty}|x_k + y_k||x_k + y_k|^{p-1} \leq \sum_{k=1}^{\infty}(|x_k| + |y_k|)|x_k + y_k|^{p-1}\\ &= \sum_{k=1}^{\infty}|x_k||x_k + y_k|^{p-1} + \sum_{k=1}^{\infty}|y_k||x_k + y_k|^{p-1}.\end{aligned}$$

Our aim is to apply the Hölder's Inequality to the two sums on the right-hand side. To this end, we note that if we choose $q = \frac{p}{p-1} > 1$, then $\frac{1}{p} + \frac{1}{q} = \frac{1+p-1}{p} = 1$. With this choice of q we have

$$\sum_{k=1}^{\infty}||x_k + y_k|^{p-1}|^q = \sum_{k=1}^{\infty}|x_k + y_k|^{q(p-1)} = \sum_{k=1}^{\infty}|x_k + y_k|^p < \infty,$$

and thus $(|x_k + y_k|^{p-1}) \in \ell^q(\mathbb{C})$. We can thus use the Hölder's Inequality to estimate

$$\begin{aligned}\|x+y\|_p^p &= \sum_{k=1}^{\infty}|x_k||x_k + y_k|^{p-1} + \sum_{k=1}^{\infty}|y_k||x_k + y_k|^{p-1}\\ &= \|x\|_p\left(\sum_{k=1}^{\infty}|x_k + y_k|^{q(p-1)}\right)^{1/q} + \|y\|_p\left(\sum_{k=1}^{\infty}|x_k + y_k|^{q(p-1)}\right)^{1/q}\\ &= (\|x\|_p + \|y\|_p)\left(\sum_{k=1}^{\infty}|x_k + y_k|^p\right)^{\frac{p-1}{p}} = (\|x\|_p + \|y\|_p)\|x+y\|_p^{p-1}.\end{aligned}$$

Since we assumed $x+y \neq 0$ (in which case the claim is trivial), we have $\|x+y\|_p^{p-1} \neq 0$. We can thus divide both sides of the previous estimate with $\|x+y\|_p^{p-1}$ to arrive at the Minkowski's Inequality. □

Example 1.23. The space $C([a,b])$ becomes a normed space if we choose the norm to be, for example,

$$\|f\|_\infty = \sup_{t\in[a,b]}|f(t)| \qquad \text{or} \qquad \|f\|_p = \left(\int_a^b |f(t)|^p \,\mathrm{d}t\right)^{\frac{1}{p}}, \qquad p \geq 1. \tag{1.2}$$

Since the functions f in $C([a,b])$ are uniformly continuous, the expression for the norm $\|f\|_p$ is well-defined as a Riemann integral. Checking that $\|\cdot\|_\infty$ and $\|\cdot\|_1$ satisfy the axioms of the norm is straightforward and is left as an exercise. Similarly, in the case where $p > 1$, verifying the two first axioms is straightforward. The fact that $\|\cdot\|_p$ also satisfies the triangle inequality is an important result, known as the **Minkowski's Inequality** (for integrals), and it is presented in Theorem 1.24 below. $\diamond$

Theorem 1.24 (Minkowski's Inequality). *Let $\Omega \subset \mathbb{R}^n$ be a compact set (closed and bounded). For $p \geq 1$ we have*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p, \qquad \forall f, g \in C(\Omega)$$

where $\|\cdot\|_p$ is as in (1.2). Moreover, $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$ for all $f, g \in C(\Omega)$.

Proof. The proofs for the cases $p = 1$ and "$p = \infty$" are left as exercises. The proof for the case $p > 1$ is presented in detail on the course "MATH.MA.840 Measure and Integration". We will not study this proof in detail, because (this version of) Minkowski's Inequality is most of all a result in real analysis. However, if you are interested, the full proof is presented in Appendix A.1. $\square$

Example 1.25. The Minkowski's Inequality in Theorem 1.24 applies to functions of several variables. If $\Omega \subset \mathbb{R}^n$ is a closed and bounded subset of $\mathbb{R}^n$, then $C(\Omega)$ can be equipped with one of the norms

$$\|f\|_\infty = \sup_{z\in\Omega} |f(z)| \qquad \text{or} \qquad \|f\|_p = \left(\int_\Omega |f(z)|^p \,\mathrm{d}z\right)^{\frac{1}{p}}, \qquad p \geq 1. \tag{1.3}$$

The norms are well-defined for every $f \in C(\Omega)$, since the functions in $C(\Omega)$ are uniformly continuous by the Heine–Cantor Theorem (since Ω is a compact subset of $\mathbb{R}^n$) and consequently $z \mapsto f(z)$ is bounded and $z \mapsto |f(z)|^p$ is Riemann integrable. The fact that the norm $\|\cdot\|_p$ satisfies the triangle inequality follows from Theorem 1.24.

If Ω is an *unbounded* subset of $\mathbb{R}^n$ (e.g, an unbounded interval in $\mathbb{R}$), then $\|f\|_\infty$ and $\|f\|_p$ cannot be defined for all functions $f \in C(\Omega)$. However,

$$\|f\|_\infty = \sup_{t\in\mathbb{R}} |f(t)| \qquad \text{and} \qquad \|f\|_p = \left(\int_{-\infty}^{\infty} |f(t)|^p \,\mathrm{d}t\right)^{\frac{1}{p}}, \qquad p \geq 1$$

are well-defined for all functions f in the space $C_c^\infty(\mathbb{R})$ (in Example 1.9). Indeed, $(C_c^\infty(\mathbb{R}), \|\cdot\|_\infty)$ and $(C_c^\infty(\mathbb{R}), \|\cdot\|_p)$ are normed spaces. $\diamond$

The norm gives us a way to define the **size** of an element $x \in X$ on a normed space $(X, \|\cdot\|)$, we can also define the **distance** between two elements $x \in X$ and $y \in X$ to be the norm of their difference, i.e., $\|x - y\|$.

Definition 1.26. Let $(X, \|\cdot\|)$ be a normed space.

- The **open ball** $B(x_0, r)$ centered at $x_0 \in X$ and with radius $r > 0$ is defined as $B(x_0, r) = \{\, x \in X \mid \|x_0 - x\| < r \,\}$.
- A subset $A \subset X$ is said to be **bounded** if there exists $M > 0$ such that $\|x\| \leq M$ for all $x \in A$.

The fundamental topological concepts of **open** and **closed sets** are defined on a normed space $(X, \|\cdot\|)$ in the standard way using the definition of an open ball. In particular, a set $A \subset X$ is **open** if "for every $x \in A$ there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset A$" (i.e., if every $x \in A$ is a center of an open ball which is completely contained in A). Moreover, a set $A \subset X$ is **closed** if its **complement** $X \setminus A := \{\, x \in X \mid x \notin A \,\}$ is open. Alternatively, the closedness of $A \subset X$ can be defined as the property that "A contains all its so-called *accumulation points*". A more complete overview of the topological concepts on a normed space is given in Appendix A.2.

Exercise 1.27. Show that A is a bounded set if and only if there exists $M > 0$ such that $\|x - y\| \le M$ for all $x, y \in A$. ◇

One particularly important usage of a norm on a vector space X is that we can use the distance $\|x - y\|$ between two elements $x \in X$ and $y \in X$ to study the **convergence of a sequence**. A sequence on a vector space X is again an infinite ordered set of elements $(x_1, x_2, x_3, \ldots)$ of X, and we also use the alternative notations $(x_k)_{k=1}^\infty \subset X$, $(x_k)_{k\in\mathbb{N}} \subset X$, and $(x_k) \subset X$ (if the index set $k \in \mathbb{N}$ is clear from the context).

Definition 1.28. Let $(X, \|\cdot\|)$ be a normed space. The sequence $(x_k)_{k=1}^\infty \subset X$ **converges to the element** $x \in X$ if

$$\forall \varepsilon > 0 \; \exists n_\varepsilon \in \mathbb{N} : \quad \|x - x_k\| < \varepsilon \quad \text{whenever} \quad k > n_\varepsilon.$$

In this case x is the **limit** of the sequence $(x_k)_{k=1}^\infty$.

Note that the definition of the convergence of a sequence $(x_k)_{k=1}^\infty \subset X$ to $x \in X$ is equivalent to the property that the distances $\|x - x_k\|$ converge to zero as $k \to \infty$, i.e., $\lim_{k\to\infty} \|x - x_k\| = 0$. If the norm in the definition of convergence is clear from the context, it is possible to use the following alternative notations for the convergence:

$$x_n \to x, \quad \text{as} \quad k \to \infty, \qquad \text{or} \qquad \lim_{k\to\infty} x_k = x.$$

However, it is very important to be careful when discussing convergence of sequences on normed spaces, because the convergence of $(x_k)_{k=1}^\infty \subset X$ may depend crucially on the choice of the norm $\|\cdot\|$ on X. Indeed, the same sequence $(x_k)_{k=1}^\infty$ on a vector space X may converge with respect to one norm, and at the same time fail to converge with respect to another norm.

Exercise 1.29. Every norm $\|\cdot\|$ on a vector space X also satisfies the "**reverse triangle inequality**" which states that

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

The proof is exactly the same as in the case of Euclidean spaces $\mathbb{R}^n$ or $\mathbb{C}^n$ (the triangle inequality is applied to $\|x\| = \|(x - y) + y\|$ and $\|y\| = \|(y - x) + x\|$). You can easily prove that this inequality further implies that the norm $\|\cdot\| : X \to [0, \infty)$ is a continuous function. Indeed, you can show that for an arbitrary $\varepsilon > 0$ there exists $\delta > 0$ such that $\big|\|x\| - \|y\|\big| < \varepsilon$ whenever $\|x - y\| < \delta$. ◇

Lemma 1.30. *The limit of a convergent sequence is unique.*

Proof. Let $(x_k)_{k=1}^{\infty}$ be a sequence which converges in $(X, \|\cdot\|)$. Assume $y \in X$ and $z \in X$ are two of its limits. The triangle inequality implies

$$\|y - z\| = \|y - x_k + x_k - z\| \leq \|y - x_k\| + \|x_k - z\| \to 0$$

as $k \to \infty$. Since $y - z$ is independent of the index $k \in \mathbb{N}$, we must have $\|y - z\| = 0$, and the axiom (1) of the norm implies $y - z = 0$, i.e., $y = z$. □

In the following we will define the concept of a **Cauchy sequence** $(x_k)_{k=1}^{\infty}$ with the characteristic property that its elements x_k and x_m come arbitrarily close to each other for sufficiently large indices k and m.

Definition 1.31. The sequence $(x_k)_{k=1}^{\infty}$ of a normed space $(X, \|\cdot\|)$ is a **Cauchy sequence** if

$$\forall \varepsilon > 0 \; \exists n_\varepsilon \in \mathbb{N} : \quad \|x_k - x_m\| < \varepsilon \quad \text{whenever} \quad k, m > n_\varepsilon .$$

We sometimes use the terminology that "*the sequence* $(x_k)_{k=1}^{\infty}$ *is Cauchy*" if the condition in Definition 1.31 is satisfied. Moreover, this convergence property is sometimes denoted as $\|x_k - x_m\| \to 0$ as $k, m \to \infty$.

Exercise 1.32. Show that (x_n) is a Cauchy sequence, then for every $p \in \mathbb{N}$ we have $\|x_{n+p} - x_n\| \to 0$ as $n \to \infty$. Does the reverse claim hold? ◇

The condition that $(x_k) \subset X$ is a Cauchy sequence is a strictly weaker property than the convergence of (x_k) (when these two properties are defined using the same norm). Indeed, if $(x_k)_{k=1}^{\infty} \subset X$ converges to $x \in X$, then for all $k, m \in \mathbb{N}$ we have

$$\|x_k - x_m\| = \|x_k - x + x - x_m\| \leq \|x_k - x\| + \|x - x_m\| \to 0$$

as $k, m \to \infty$. Thus every convergent sequence is a Cauchy sequence. However, **the converse is not always true!** In fact, the question of whether or not all Cauchy sequences converge (to some limits inside the space) defines a fundamental property of a normed space $(X, \|\cdot\|)$ called **completeness**.

Definition 1.33 (Banach Space)**.** The normed space $(X, \|\cdot\|)$ is said to be **complete** if every Cauchy sequence $(x_k)_{k=1}^{\infty} \subset X$ converges to an element $x \in X$. A complete normed space $(X, \|\cdot\|)$ is called a **Banach space**.

One way of viewing the question of "completeness" of a normed space $(X, \|\cdot\|)$ is that Cauchy sequences are ideal candidates for sequences which *should converge* due to the property that the distances between their elements become arbitrarily small for large indices. This way, the completeness of a vector space guarantees that also the limits of such sequences indeed belong to the vector space X. The property of completeness of a vector space is often very useful in proofs due to the fact that it is typically much easier to prove that $(x_k)_{k=1}^{\infty}$ is a Cauchy sequence instead of proving directly that (x_k) converges. In particular, showing that (x_k) is Cauchy does not require knowledge of the *limit* $x \in X$ of the sequence, which would be necessary for directly showing the convergence $\|x_k - x\| \to 0$ as $k \to \infty$. However, if the underlying space is a Banach space, it is sufficient to prove that (x_k) is a Cauchy sequence, and completeness of the normed space guarantees that (x_k) converges to some limit $x \in X$.

The property of completeness of a normed space depends closely on both the vector space X and the norm $\|\cdot\|$. In particular, normed space $(X, \|\cdot\|)$ which is not a Banach space can sometimes be changed into a complete space by either (1) keeping the vector space X the same changing the norm $\|\cdot\|$, or (2) keeping the norm $\|\cdot\|$ same and making the vector space X larger. The latter case is referred to as **completing** the normed space, and the resulting larger space is the **completion** of the space $(X, \|\cdot\|)$. Of course, in both of these cases the resulting modified normed space is different from the orginal space $(X, \|\cdot\|)$.

Definition 1.34. Let $\|\cdot\|_A$ and $\|\cdot\|_B$ be two norms on a vector space X. The norms $\|\cdot\|_A$ and $\|\cdot\|_B$ are **equivalent** if there exist constants $C, c > 0$ such that

$$c\|x\|_B \leq \|x\|_A \leq C\|x\|_B, \qquad \forall x \in X.$$

Note that if two norms $\|\cdot\|_A$ and $\|\cdot\|_B$ on X are equivalent, then $(x_k)_k \subset X$ is a Cauchy sequence with respect to the norm $\|\cdot\|_A$ if and only if it is a Cauchy sequence with respect to the norm $\|\cdot\|_B$ (prove this!). Similarly $(x_k)_k \subset X$ converges to an element $x \in X$ with respect to the norm $\|\cdot\|_A$ if and only if it converges to x with respect to the norm $\|\cdot\|_B$. Because of this, if $\|\cdot\|_A$ and $\|\cdot\|_B$ are two equivalent norms on X, then $(X, \|\cdot\|_A)$ is a Banach space if and only if $(X, \|\cdot\|_B)$ is a Banach space.

1.2.1 Examples of Normed Spaces and Banach Spaces

Our first example illustrates that the convergence of a sequence is highly dependent on the choice of the norm: a sequence may converge with respect to one norm, but fail to converge with respect to another norm.

Example 1.35. Consider the sequence of continuous functions $(f_k) \subset C([-1, 1])$, where

$$f_k(t) = \begin{cases} 0 & -1 \leq t < -\frac{1}{k} \\ 1 + kt & -\frac{1}{k} \leq t < 0 \\ 1 - kt & 0 \leq t < \frac{1}{k} \\ 0 & \frac{1}{k} \leq t \leq 1 \end{cases}$$

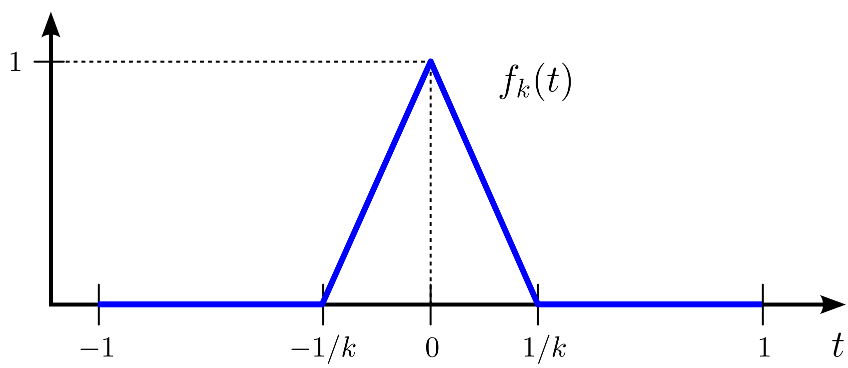

Figure 1.1: The function $f_k(\cdot)$.

A straight-forward calculation shows that

$$\|f_k - 0\|_1 = \int_{-1}^{1} |f_k(t) - 0| \,\mathrm{d}t = \int_{-1}^{1} f_k(t) \,\mathrm{d}t = \frac{1}{k} \to 0 \qquad \text{as} \quad k \to \infty$$

and thus $f_k \to 0$ in the normed space $(C([-1,1]), \|\cdot\|_1)$.

On the other hand, for all $k \in \mathbb{N}$

$$\|f_k - 0\|_\infty = \sup_{-1 \le t \le 1} |f_k(t) - 0| = f_k(0) = 1,$$

and thus the sequence $(f_k)_k$ does not converge to 0 in the space $(C([-1,1]), \|\cdot\|_\infty)$. In fact, the sequence $(f_k)_k$ does not converge at all in the space $(C([-1,1]), \|\cdot\|_\infty)$. To prove this, assume there exists $f \in C([-1,1])$ such that $\|f - f_k\|_\infty \to 0$ as $k \to \infty$. For every $t_0 \in [-1,1]$ we then also have

$$|f_k(t_0) - f(t_0)| \le \sup_{t \in [-1,1]} |f_k(t) - f(t)| = \|f_k - f\|_\infty \to 0 \qquad \text{as} \quad k \to \infty.$$

Thus the convergence of $(f_k)_k$ in the norm $\|\cdot\|_\infty$ also implies that the functions converge f_k to f "pointwise", i.e., $f_k(t) \to f(t)$ as $k \to \infty$ (in $\mathbb{C}$) for all $t \in [-1,1]$. For our sequence $(f_k)_k$ we have that $f_k(0) \to 1$ as $k \to \infty$ and $f_k(t) \to 0$ as $k \to \infty$ for all $t \in [-1,0) \cup (0,1]$. However, this implies that $\lim_{t\to 0} f(t) = 0 \neq 1 = f(0)$, which contradicts the assumption that f is continuous on $[-1,1]$. Thus $(f_k)_k$ does not converge in $(C([-1,1]), \|\cdot\|_\infty)$. ◇

Our first example of an infinite-dimensional Banach space is the space of continuous functions on a closed interval $[a,b] \subset \mathbb{R}$ equipped with the norm $\|\cdot\|_\infty$.

Theorem 1.36. *The space $(C([a,b]), \|\cdot\|_\infty)$ is a Banach space.*

Proof. Let $(f_k)_{k=1}^\infty \subset C([a,b])$ be an arbitrary Cauchy sequence, i.e. $\|f_k - f_m\|_\infty \to 0$ as $k, m \to \infty$. Our aim is to show that there exists a continuous function $f \in C([a,b])$ such that $\|f - f_k\|_\infty \to 0$ as $k \to \infty$. If this is true, we can deduce that $(C([a,b]), \|\cdot\|_\infty)$ is a Banach space. Let $\varepsilon > 0$. Since $(f_k)_{k=1}^\infty$ is a Cauchy sequence, we can choose $n_\varepsilon \in \mathbb{N}$ such that $\|f_k - f_m\|_\infty \le \varepsilon$ whenever $k, m \ge n_\varepsilon$. This also implies that for every fixed $t_0 \in [a,b]$ we have

$$|f_k(t_0) - f_m(t_0)| \le \sup_{t \in [a,b]} |f_k(t) - f_m(t)| = \|f_k - f_m\|_\infty \le \varepsilon \tag{1.4}$$

whenever $k, m \ge n_\varepsilon$. Since $\varepsilon > 0$ was arbitrary, we deduce that $(f_k(t_0))_{k=1}^\infty$ is a Cauchy sequence in $(\mathbb{C}, |\cdot|)$. But since $(\mathbb{C}, |\cdot|)$ is a complete space, this scalar Cauchy sequence converges to a limit $c_{t_0} \in \mathbb{C}$, that is, $|f_k(t_0) - c_{t_0}| \to 0$ as $k \to \infty$. We can now define a function $f : [a,b] \to \mathbb{C}$ such that $f(t_0) = c_{t_0}$ for every $t_0 \in [a,b]$. At this point we do not yet know that the function f is continuous. Thus in order to complete the proof, we still need to show that $f \in C([a,b])$ and that $\|f_k - f\|_\infty \to 0$ as $k \to \infty$. To this end, we can let $m \to \infty$ on the left-hand side of (1.4) to deduce that $|f_k(t_0) - f(t_0)| \le \varepsilon$ whenever $k \ge n_\varepsilon$. Since this is true for an arbitrary $\varepsilon > 0$, and since n_ε does not depend on $t_0 \in [a,b]$, the functions f_k converge to f *uniformly.* This property together with the Uniform Convergence Theorem implies that the limit function f is continuous.

Finally, let $\varepsilon > 0$ is arbitrary. As shown above, there exists $n_\varepsilon \in \mathbb{N}$ such that $|f_k(t_0) - f(t_0)| \leq \varepsilon$ whenever $k \geq n_\varepsilon$ and $t_0 \in [a, b]$. Because of this, we also have

$$\|f_k - f\|_\infty = \sup_{t_0 \in [a,b]} |f_k(t_0) - f(t_0)| \leq \varepsilon$$

whenever $k \geq n_\varepsilon$. Thus $\|f_k - f\|_\infty \to 0$ as $k \to \infty$. Since $(f_k)_k \subset C([a, b])$ was an arbitrary Cauchy sequence, and since we proved that $(f_k)_k$ converges to a limit $f \in C([a, b])$, we conclude that $C([a, b])$ is a Banach space. □

In the next example we consider the normed space $(C([-1, 1]), \|\cdot\|_1)$ and construct a Cauchy sequence which does not converge in this space. This implies $(C([-1, 1]), \|\cdot\|_1)$ is not complete, and therefore it is *not* a Banach space. The same construction can be generalised to show that $(C([a, b]), \|\cdot\|_p)$ is not a Banach space for any $1 \leq p < \infty$. Similarly, the spaces $(C(\Omega), \|\cdot\|_p)$ where $\Omega \subset \mathbb{R}^n$ are not complete either.

Example 1.37 ($(C([-1, 1]), \|\cdot\|_1)$ is not a Banach space)**.** Consider the sequence (f_n) in the space $(C([-1, 1]), \|\cdot\|_1)$, where (see Figure 1.2)

$$f_n(t) = \begin{cases} 0 & -1 \leq t < 0 \\ nt & 0 \leq t \leq \frac{1}{n} \\ 1 & \frac{1}{n} \leq t \leq 1. \end{cases}$$

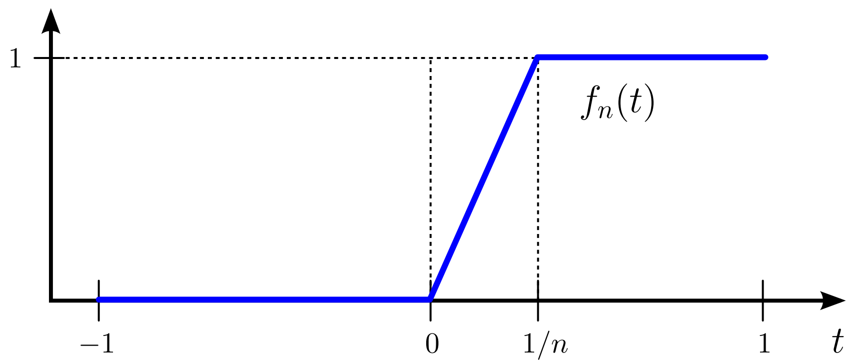

Figure 1.2: The function $f_n(\cdot)$.

Clearly,

$$\|f_n - f_m\|_1 = \int_{-1}^{1} |f_n(t) - f_m(t)| \,\mathrm{d}t \leq \frac{1}{2\min\{n, m\}} \to 0$$

as $n, m \to \infty$, and thus (f_n) is a Cauchy sequence. We will now show that (f_n) does not converge in $(C([-1, 1]), \|\cdot\|_1)$. Assume on the contrary that there exists $f \in C([-1, 1])$ such that $\lim_{n\to\infty} f_n = f$. Since $f_n(t) = 0$ for $t \in [-1, 0]$, we have

$$0 \leftarrow \|f_n - f\|_1 = \int_{-1}^{1} |f_n(t) - f(t)| \,\mathrm{d}t \geq \int_{-1}^{0} |f(t)| \,\mathrm{d}t \geq 0$$

as $n \to \infty$, which implies $\int_{-1}^{0} |f(t)| \,\mathrm{d}t = 0$. Since f is continuous, we must have $f(t) = 0$ for all $t \in [-1, 0]$. On the other hand, for any $n_0 \in \mathbb{N}$ and for all $n \geq n_0$ we similarly have

$$0 \leftarrow \|f_n - f\|_1 = \int_{-1}^{1} |f_n(t) - f(t)| \,\mathrm{d}t \geq \int_{1/n_0}^{1} |f(t) - 1| \,\mathrm{d}t \geq 0$$

as $n \to \infty$, and since f is continuous we have $f(t) = 1$ for all $t \in [1/n_0, 1]$. However, since this is true for any $n_0 \in \mathbb{N}$, we have $f(t) = 1$ for all $t \in (0, 1]$. However, this shows that f can not be continuous at $t = 0$, which is a contradiction with the assumption $f \in C([-1, 1])$. This contradiction concludes that (f_n) does not converge in $(C([-1, 1]), \|\cdot\|_1)$. ◇

The sequence spaces $(\ell^p(\mathbb{C}), \|\cdot\|_p)$ for $1 \le p < \infty$ and $(\ell^\infty(\mathbb{C}), \|\cdot\|_\infty)$ are important examples of Banach spaces. The proof of this result is also a great exercise in considering a *sequence of sequences*: we consider a Cauchy sequence $(x_k)_{k\in\mathbb{N}}$ on a sequence space, which means that every element x_k is an infinite sequence of complex numbers, i.e., $x_k = (x_k^n)_{n\in\mathbb{N}}$, where $x_k^n \in \mathbb{C}$ for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$.

Theorem 1.38. *The spaces $(\ell^\infty(\mathbb{C}), \|\cdot\|_\infty)$ and $(\ell^p(\mathbb{C}), \|\cdot\|_p)$ for $1 \le p < \infty$ are Banach spaces.*

Proof. We will prove that $(\ell^p(\mathbb{C}), \|\cdot\|_p)$ is a Banach space for $p \ge 1$, and the proof for completeness of $(\ell^\infty(\mathbb{C}), \|\cdot\|_\infty)$ is left as an exercise. Let $p \ge 1$ be arbitrary, and let $(x_k)_{k=1}^\infty \subset \ell^p(\mathbb{C})$ be a Cauchy sequence. In this sequence every element $x_k \in \ell^p(\mathbb{C})$ is a sequence of complex numbers. For these sequences, we use the notation $x_k = (x_k^n)_{n=1}^\infty = (x_k^1, x_k^2, \ldots)$. By assumption we have $\|x_k - x_m\|_p \to 0$ as $k, m \to \infty$. For every fixed $n \in \mathbb{N}$ we thus have that the nth elements x_k^n of x_k satisfy

$$|x_k^n - x_m^n| \le \left(\sum_{n=1}^\infty |x_k^n - x_m^n|^p\right)^{\frac{1}{p}} = \|x_k - x_m\|_p \to 0$$

as $k, m \to \infty$. Thus for every fixed $n \in \mathbb{N}$ the sequence $(x_k^n)_{k=1}^\infty \subset \mathbb{C}$ is a Cauchy sequence in $(\mathbb{C}, |\cdot|)$, and since $(\mathbb{C}, |\cdot|)$ is complete, there exists a limit $y^n \in \mathbb{C}$ such that $x_k^n \to y^n$ as $k \to \infty$.

We can now define a sequence $y = (y^n)_{n=1}^\infty = (y^1, y^2, \ldots)$. Our aim is to show that $y \in \ell^p(\mathbb{C})$ and that $\|x_k - y\|_p \to 0$ as $k \to \infty$. As shown in the exercises, every Cauchy sequence is bounded. Thus there exists $M > 0$ such that $\|x_k\|_p \le M$ for all $k \in \mathbb{N}$. If we let $N \in \mathbb{N}$ be arbitrary, we also have that for every $k \in \mathbb{N}$

$$\left(\sum_{n=1}^N |x_k^n|^p\right)^{\frac{1}{p}} \le \left(\sum_{n=1}^\infty |x_k^n|^p\right)^{\frac{1}{p}} = \|x_k\|_p \le M.$$

Since $\sum_{n=1}^N |x_k^n|^p$ is a finite sum, we can take a limit of this expression by letting $k \to \infty$. In particular, the above estimate implies

$$\left(\sum_{n=1}^N |y^n|^p\right)^{\frac{1}{p}} = \left(\sum_{n=1}^N |\lim_{k\to\infty} x_k^n|^p\right)^{\frac{1}{p}} = \lim_{k\to\infty} \left(\sum_{n=1}^N |x_k^n|^p\right)^{\frac{1}{p}} \le M.$$

Since $N \in \mathbb{N}$ was arbitrary and since M is independent of N, the previous estimate also implies

$$\left(\sum_{n=1}^\infty |y^n|^p\right)^{\frac{1}{p}} = \lim_{N\to\infty} \left(\sum_{n=1}^N |y^n|^p\right)^{\frac{1}{p}} \le M < \infty,$$

and thus $y \in \ell^p(\mathbb{C})$. To show that $\|x_k - y\|_p \to 0$ as $k \to \infty$, let $\varepsilon > 0$ be arbitrary. Since $(x_k)_k$ is a Cauchy sequence, we can choose $n_\varepsilon \in \mathbb{N}$ such that $\|x_k - x_m\|_p \le \varepsilon$

whenever $k, m \geq n_\varepsilon$. This also implies that for an arbitrary $N \in \mathbb{N}$ and for all $k, m \geq n_\varepsilon$ we have

$$\left(\sum_{n=1}^{N} |x_k^n - x_m^n|^p\right)^{\frac{1}{p}} \leq \left(\sum_{n=1}^{\infty} |x_k^n - x_m^n|^p\right)^{\frac{1}{p}} = \|x_k - x_m\|_p \leq \varepsilon.$$

Since $\sum_{n=1}^{N} |x_k^n - x_m^n|^p$ is a finite sum, we can take a limit $m \to \infty$ in the above estimate, which implies that for all $k \geq n_\varepsilon$ we have

$$\left(\sum_{n=1}^{N} |x_k^n - y^n|^p\right)^{\frac{1}{p}} = \left(\sum_{n=1}^{N} |x_k^n - (\lim_{m\to\infty} x_m^n)|^p\right)^{\frac{1}{p}} = \lim_{m\to\infty} \left(\sum_{n=1}^{N} |x_k^n - x_m^n|^p\right)^{\frac{1}{p}} \leq \varepsilon.$$

Since $N \in \mathbb{N}$ was arbitrary, this estimate finally implies

$$\|x_k - y\|_p = \left(\sum_{n=1}^{\infty} |x_k^n - y^n|^p\right)^{\frac{1}{p}} = \lim_{N\to\infty} \left(\sum_{n=1}^{N} |x_k^n - y^n|^p\right)^{\frac{1}{p}} \leq \varepsilon$$

for all $k \geq n_\varepsilon$. Since $\varepsilon > 0$ was arbitrary, we have that $\|x_k - y\|_p \to 0$ as $k \to \infty$. □

1.3 Completion of a Normed Space

In this section we define the concept of a "completion" of a normed space $(X, \|\cdot\|_X)$, which is (roughly speaking) the smallest complete normed space which contains X (and has the same norm $\|\cdot\|_X$).

Definition 1.39. Let $(X, \|\cdot\|)$ be a normed space. The set $A \subset X$ is **dense** in the set $B \subset X$ if

$$\forall x \in B,\ \forall \varepsilon > 0\ \exists y_\varepsilon \in A \quad \text{so that} \quad \|x - y_\varepsilon\| < \varepsilon.$$

The definition that the set A is dense in B means precisely that it is possible to "approximate" any element $x \in B$ with arbitrary accuracy with an element from the set A. As a classical example, the property that any real number $t \in \mathbb{R}$ can be approximated with arbitrary accuracy with a rational number $q \in \mathbb{Q}$ means that the set $\mathbb{Q}$ is dense in $\mathbb{R}$ (in the vector space $(\mathbb{R}, |\cdot|)$). Note that the dense set A does not necessarily need to be a subset of A, and therefore the rational numbers $\mathbb{Q}$ are also dense in set of irrational numbers $\mathbb{R} \setminus \mathbb{Q}$.

Exercise 1.40. Let $(X, \|\cdot\|)$ be a normed space. Show that the set $A \subset X$ is dense in the set $B \subset X$ if and only if for every $x \in B$ there exists a sequence $(x_k)_{k=1}^{\infty} \subset A$ such that

$$\|x - x_k\| \to 0 \qquad \text{as} \quad k \to \infty.$$

◇

Exercise 1.41. Let $(X, \|\cdot\|)$ be a normed space. Let $A \subset X$ be dense in $B \subset X$ and let B be dense in $C \subset X$. Show that A is dense in C. ◇

Definition 1.42 (Completion — Simplified Version)**.** Let $(X, \|\cdot\|_X)$ be a normed space and let $(Y, \|\cdot\|_X)$ be a Banach space. If $X \subset Y$ is a subspace of Y and if X is dense in Y, then Y is a **completion** of X.

Note that in this definition the norms on the two spaces need to be the same, and X is required to be a dense *subset* of Y. The role of the assumption that X is dense in Y is to ensure that the completion is "minimal" in the sense that Y does not contain parts which are not related to the original space X. The above definition of a "completion" of a normed space is actually a highly simplifed and restricted version. In the full definition of a completion (see, e.g., [NS82, Sec. 3.14]), the space X does not need to be contained in the space Y, but instead it needs to be "*isometrically isomorphic*" to a normed space $(\tilde{Y}, \|\cdot\|_Y)$ where $\tilde{Y}$ is a subspace of Y. This more general version of the definition is illustrated in Figure 1.3. In particular, the "isometrically isomorphic" relationship between X and $\tilde{Y}$ means that these two spaces can be considered to be the same space (in terms of properties of normed spaces), and this part of the definition of a completion is sometimes essential in identifying a completion of a space. This way, completions in the sense of Definition 1.42 are indeed "completions" also in the sense of the full definition, but it is good to remember that in the sense of its true definition, a completion Y does not always need to contain the original space X, but only its copy. In this section we will use Definition 1.42 to study completions which do contain the original space.

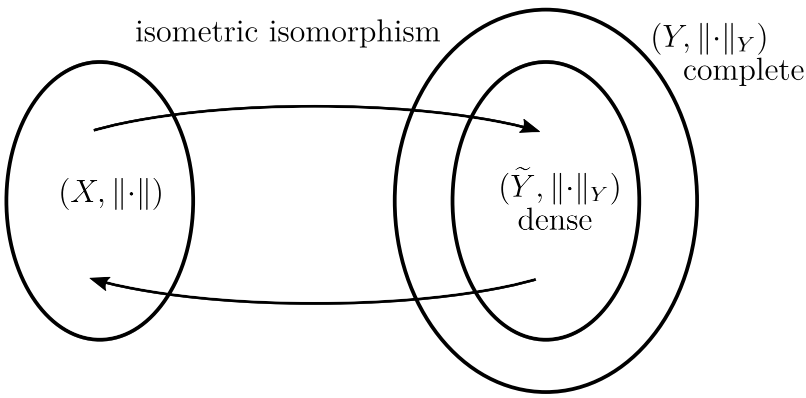

Figure 1.3: Completion $(Y, \|\cdot\|_Y)$ of a normed space $(X, \|\cdot\|)$.

Theorem 1.43 (Weierstrass Approximation Theorem)**.** *The set $P([a, b])$ is dense in $C([a, b])$ with respect to the norm $\|\cdot\|_\infty$.*

Proof. We can without loss of generality assume $a = 0$ and $b = 1$, since otherwise we can define a change of variables $s = \frac{t-a}{b-a}$. Let $f \in C([0, 1])$ be arbitrary, and define the sequence $(p_n) \in P([0, 1])$ of polynomials so that

$$p_n(t) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} t^k (1-t)^{n-k}.$$

Here $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient. The functions $p_n(t)$ are called **Bernstein polynomials**. From real analysis we know that a continuous function f on a closed interval $[a,b]$ is uniformly continuous, i.e.

$$\forall\varepsilon > 0\ \exists\delta > 0 \quad \text{such that} \quad |s-t| < \delta \;\Rightarrow\; |f(s) - f(t)| < \varepsilon.$$

Let $\varepsilon > 0$ be arbitrary, and let $\delta > 0$ be such that the above property holds. The binomial formula tells us that

$$1 = 1^n = [t + (1-t)]^n = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k},$$

and using this identity we can write for any $t \in [0,1]$ and $n \in \mathbb{N}$

$$\begin{aligned}
|f(t) - p_n(t)| &= \left| f(t) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} t^k (1-t)^{n-k} \right| \\
&= \left| f(t)\,[t + (1-t)]^n - \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} t^k (1-t)^{n-k} \right| \\
&= \left| \sum_{k=0}^{n} \left[f(t) - f\left(\frac{k}{n}\right) \right] \binom{n}{k} t^k (1-t)^{n-k} \right| \le \sum_{k=0}^{n} \left| f(t) - f\left(\frac{k}{n}\right) \right| \left| \binom{n}{k} t^k (1-t)^{n-k} \right| \\
&= \sum_{|t-\frac{k}{n}|<\delta} \left| f(t) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} t^k (1-t)^{n-k} + \sum_{|t-\frac{k}{n}|\ge\delta} \left| f(t) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} t^k (1-t)^{n-k} \\
&\le \varepsilon \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} + \sum_{\frac{(k-nt)^2}{n^2\delta^2}\ge 1} 1 \cdot 2M \binom{n}{k} t^k (1-t)^{n-k}
\end{aligned}$$

where $M = \sup_{t\in[0,1]} |f(t)|$. Using the property $1 \le \frac{(t-\frac{k}{n})^2}{\delta^2}$ in the second sum we get

$$|f(t) - p_n(t)| \le \varepsilon + \frac{2M}{n^2\delta^2} \sum_{k=0}^{n} (k-nt)^2 \binom{n}{k} t^k (1-t)^{n-k} = \varepsilon + \frac{2M}{\delta^2}\frac{1}{n} t(1-t).$$

Here the last identity follows from probability theory, and more precisely from the fact that if $Z \sim B(n,t)$ is a binomially distributed random variable, then the expected value and the variance of Z are given by $\mathbb{E}(Z) = nt$ and $\mathrm{Var}(Z) = \mathbb{E}\left[(Z - \mathbb{E}(Z))^2\right] = nt(1-t)$, and thus

$$\sum_{k=0}^{n} (k-nt)^2 \binom{n}{k} t^k (1-t)^{n-k} = \mathrm{Var}(Z) = nt(1-t).$$

If $n > \frac{M}{\delta^2\varepsilon}$, we obtain

$$|f(t) - p_n(t)| \le \varepsilon + \frac{2M}{\delta^2} \cdot \frac{1}{n} \cdot \frac{1}{4} < \varepsilon + \frac{1}{2}\varepsilon = \frac{3\varepsilon}{2},$$

i.e $\forall\varepsilon > 0\ \exists n_\varepsilon \in \mathbb{N}$ which does not depend on t, so that

$$n > n_\varepsilon \quad \Rightarrow \quad |f(t) - p_n(t)| < \frac{3\varepsilon}{2}.$$

It follows that for all $n > n_\varepsilon$ we have

$$\|f - p_n\|_\infty = \sup_{t\in[0,1]} |f(t) - p_n(t)| \le \frac{3\varepsilon}{2} < 2\varepsilon.$$

Thus $\|f - p_n\|_\infty \to 0$ as $n \to \infty$. Since $f \in C([a,b])$ was arbitrary and since we found a sequence $(p_n)_{n=1}^\infty \subset P([a,b])$ such that $p_n \to f$ as $n \to \infty$ in $(C([a,b]), \|\cdot\|_\infty)$, the set $P([a,b])$ is dense in $C([a,b])$ by Exercise 1.40. □

Corollary 1.44. *The space* $(C([a,b]), \|\cdot\|_\infty)$ *is the completion of* $(P([a,b]), \|\cdot\|_\infty)$.

Proof. Due the Weierstrass Approximation Theorem the set of polynomials $P([a,b])$ is dense in $C([a,b])$. The claim now follows directly from the fact that $C([a,b])$ is complete with respect to the norm $\|\cdot\|_\infty$. □

1.4 Lebesgue Spaces

We already learned that if $I \subset \mathbb{R}$ is a finite interval, then the space $C(I)$ of continuous functions is not complete with respect to the norm $\|\cdot\|_p$ defined by

$$\|f\|_p = \left(\int_I |f(t)|^p \,\mathrm{d}t \right)^{\frac{1}{p}},$$

for any $1 \le p < \infty$. In this section we will study the *completions* of the spaces $(C(I), \|\cdot\|_p)$. These spaces turn out to be the so-called "**Lebesgue spaces**" which are denoted by $L^p(I)$. Lebesgue spaces are studied in greater detail on the course "MATH.MA.840 Measure and Integration". This section only provides a quick overview of some of the basic properties of these spaces (and provides a very quick introduction to the main concepts to those who may have not completed the Measure and Integration course). The main purpose of the current section is to confirm that the spaces $(C(I), \|\cdot\|_p)$ *do have* well-defined completions, and to study some fundamental properties of the norms $\|\cdot\|_p$ for $1 \le p < \infty$.

The definition of the Lebesgue spaces $L^p(\Omega)$ for $1 \le p < \infty$ are given below. We also define these spaces for functions of several variables (i.e. $f : \Omega \subset \mathbb{R}^n \to \mathbb{C}$), but in our examples we will mostly consider the functions of a single variable.

Definition 1.45. For $\Omega \subset \mathbb{R}^n$ the **Lebesgue space** $L^p(\Omega)$ is defined as

$$L^p(\Omega) = \left\{ f : \Omega \to \mathbb{C} \;\middle|\; \int_\Omega |f(z)|^p \,\mathrm{d}z < \infty \right\},$$

and the norm on $L^p(\Omega)$ is defined by

$$\|f\|_p = \left(\int_\Omega |f(z)|^p \,\mathrm{d}z \right)^{\frac{1}{p}}, \qquad f \in L^p(\Omega).$$

The integrals $\int_\Omega |f(z)|^p \,\mathrm{d}z$ are defined as "Lebesgue integrals".

In the case where $\Omega \subset \mathbb{R}$ is an interval, we also use the following notation

$$\begin{aligned} L^p(I) &= L^p(a,b), \quad I = (a,b) \quad \text{or} \quad I = [a,b] \text{ ("the same space!")} \\ L^p(I) &= L^p(0,\infty), \quad I = (0,\infty) \quad \text{or} \quad I = [0,\infty) \\ L^p(I) &= L^p(-\infty,0), \quad I = (-\infty,0) \quad \text{or} \quad I = (-\infty,0] \end{aligned}$$

If $\Omega \subset \mathbb{R}^n$ is compact (closed and bounded), the space $L^p(\Omega)$ contains all functions in $C(\Omega)$ (since the Lebesgue integral is well-defined and coincides with the Riemann integral for all uniformly continuous functions), i.e., $C(\Omega) \subset L^p(\Omega)$. This also explains the notation "$\|\cdot\|_p$" for the norm on $L^p(\Omega)$, since this norm is equal to the one defined in (1.3) on the subspace $C(\Omega)$ of $L^p(\Omega)$. In addition to continuous functions, $L^p(a,b)$ contains several types of discontinuous functions, such as the limit of the Cauchy sequence in Example 1.37, which satisfies $f(t) = 0$ for $-1 < t < 0$ and $f(t) = 1$ for $0 < t < 1$.

The precise characterization of the space $L^p(\Omega)$ can be given using **measure and integration theory**. For us, there are two main complications to notice, and these are discussed in the following two examples.

Example 1.46. $L^p(a,b)$ (and similarly $L^p(\Omega)$) is not a space of **functions**: If we take any function $f : [a,b] \to \mathbb{C}$ such that $\|f\|_p < \infty$, and another function $g : (a,b) \to \mathbb{C}$ which differs from f only at a finite number of points $t_1, \ldots, t_n \in (a,b)$, then

$$\|f-g\|_p^p = \int_a^b |f(t)-g(t)| \,\mathrm{d}t = \int_{\{t_1\}} |f(t)-g(t)| \,\mathrm{d}t + \cdots + \int_{\{t_n\}} |f(t)-g(t)| \,\mathrm{d}t = 0,$$

since the integral over a single point is always 0. The axiom (1) of the norm would then require that $f - g = 0$, that is, $f = g$, which is not true since $f(t_k) \neq g(t_k)$ for $1 \leq k \leq n$. This difficulty will be avoided by **defining** the two functions f and g *to be the same function*. More generally, we will say that $f = g$ in the space $L^p(\Omega)$ if (and only if) $\|f-g\|_p = 0$. This way, we can see that instead of single functions, $L^p(\Omega)$ in fact consists of **equivalence classes of functions**, and especially the values $f(z_0)$ at individual points $z_0 \in \Omega$ do not matter when considering f as an element of $L^p(\Omega)$. ◇

Example 1.47. The Riemann integral **is not suitable** for defining the norm $\|\cdot\|_p$ on $L^p(\Omega)$: When considering the limits of Cauchy sequences of functions in $C(\Omega)$, we arrive at situations where the limit functions f may not have a well-defined Riemann integral. This is the situation, for example, for the two functions $f : (0,1) \to \mathbb{C}$

$$f(t) = \sin\left(\frac{1}{t}\right) \qquad \text{or} \qquad g(t) = \begin{cases} 1 & t \in (0,1) \text{ is rational} \\ 0 & t \in (0,1) \text{ is irrational} \end{cases}$$

(if you like, you can plot the function f and investigate its behaviour). Because of this, we instead define the norm $\|\cdot\|_{L^p}$ using the **Lebesgue integral**, which can handle integration of more complicated functions than the Riemann integral. For example, the above second function satisfies $g \in L^p(a,b)$ for any $p \geq 1$, since it has a well-defined Lebesgue integral $\int_0^1 |g(t)| \,\mathrm{d}t = 0$, due to the fact that $[a,b] \cap \mathbb{Q}$ is a "**set of measure zero**". If we compare this property to the comments in Example 1.46, then we can see that in fact $g = 0$ as a function of $L^p(a,b)$. ◇

Example 1.48. The functions in L^p-spaces for $1 \le p < \infty$ do not need to be **bounded** (i.e., they can have arbitrarily large values). For example, the function $f : (0,1) \to \mathbb{R}$ defined by $f(t) = t^{-1/2}$ belongs to $L^1(0,1)$ even though $f(t) \to \infty$ as $t \to 0$. ◇

The scale of $L^p(\Omega)$-spaces is made complete by including the case "$p = \infty$". Here $\operatorname{ess\,sup}_{z\in\Omega}|f(z)|$ is defined as the smallest upper bound of $|f(z)|$ "outside sets of measure zero", i.e., we can ignore values in any countable number of points $z \in \Omega$ just like explained in Example 1.46.

Definition 1.49. For $\Omega \subset \mathbb{R}^n$ the **Lebesgue space** $L^\infty(\Omega)$ is defined as

$$L^\infty(\Omega) = \left\{ f : \Omega \to \mathbb{C} \;\middle|\; \operatorname*{ess\,sup}_{z\in\Omega}|f(z)| < \infty \right\},$$

with norm $\|f\|_\infty = \operatorname{ess\,sup}_{z\in\Omega}|f(z)|$.

The essential supremum $\operatorname{ess\,sup}_{z\in\Omega}|f(z)|$ is equal to the supremum $\sup_{z\in\Omega}|f(z)|$ whenever the function $f : \Omega \to \mathbb{C}$ is continuous, and therefore the norm $\|\cdot\|_\infty$ coincides with the norm $\|\cdot\|_\infty$ defined in (1.3) for continuous functions in $L^\infty(\Omega)$.

The norms on Lebesgue spaces satisfy the Minkowski's Inequality (which is exactly the triangle inequality for the norm $\|\cdot\|_p$), as well as the very useful **Hölder's Inequality**. The special case where $p = q = 2$ the Hölder's Inequality is called the **Cauchy–Schwarz Inequality**.

Theorem 1.50. *Let $\Omega \subset \mathbb{R}^n$.*

- ***Hölder's Inequality:*** *Let $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, or $p = 1$ and $q = \infty$. If $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$, then $fg \in L^1(\Omega)$ and*

$$\|fg\|_1 \le \|f\|_p \|g\|_q.$$

- ***Cauchy–Schwarz Inequality:*** *If $f, g \in L^2(\Omega)$, then $fg \in L^1(\Omega)$ and*

$$\|fg\|_1 \le \|f\|_2 \|g\|_2.$$

- ***Minkowski's Inequality:*** *If $p \ge 1$ or $p = \infty$, then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p, \qquad f, g \in L^p(\Omega).$$

The completeness of $L^p(\Omega)$ in the following theorem is sometimes called the **Riesz–Fischer Theorem** (but it is worthwile to note that this name is sometimes used for another result as well).

Theorem 1.51. *The spaces $L^p(I)$ and $L^\infty(I)$ are Banach spaces.*

Proof. The proof is presented in on the course "MATH.MA.840 Measure and Integration", and can be found, for example, in the references [Rud87, Axl20]. □

1.4.1 Dense Subspaces of Lebesgue Spaces

The Lebesgue spaces are complete normed spaces, i.e., Banach spaces. In this section we will demonstrate that for $1 \leq p < \infty$ the space $L^p(\Omega)$ is also the completion of the space of continuous functions equipped with the norm $\|\cdot\|_p$. In reality, here we already require a more general version of the concept of "completion" due the fact that if $\Omega \subset \mathbb{R}^n$ is a compact set, then $C(\Omega)$ is not truly a subset of the space $L^p(\Omega)$, because the latter consists of *equivalence classes of functions* rather than functions $f : \Omega \to \mathbb{C}$. However, we can interpret $C(\Omega)$ as a subset of $L^p(\Omega)$ if we identify every function $f \in C(\Omega)$ with the equivalence class containing this particular function. In fact, this correspondence is precisely the "isometric isomorphism" in the more general definition of completions of normed spaces. As long as we keep this in mind, we can show that for a compact $\Omega \subset \mathbb{R}^n$, the space $(L^p(\Omega), \|\cdot\|_p)$ is a completion of $(C(\Omega), \|\cdot\|_p)$ by showing that for every element $f \in L^p(\Omega)$ and for every $\varepsilon > 0$ there exists a function $f_\varepsilon \in C(\Omega)$ such that $\|f - f_\varepsilon\|_p < \varepsilon$[1]. This will prove that the equivalence classes of continuous functions form a dense subspace of $L^p(\Omega)$.

Another important thing to note is that $(C(\Omega), \|\cdot\|_p)$ is in general not a normed space if $\Omega \subset \mathbb{R}^n$ is an unbounded set (for example if $\Omega = \mathbb{R}^n$, $\Omega = \{\, z = (z_1, \dots, z_n)^T \in \mathbb{R} \mid z_1 > 0 \,\}$, or $\Omega = \mathbb{R}^n \setminus B(0,1)$), since in this case the integral $\int_\Omega |f(z)|^p \,\mathrm{d}z$ can diverge for certain continuous functions $f \in C(\Omega)$. Motivated by this, in the following we define the space $C_c(\Omega)$ of functions with "compact support".

Definition 1.52. The **support** of the function $f : \Omega \subset \mathbb{R}^n \to \mathbb{C}$ is defined as

$$\operatorname{supp} f = \overline{\left\{ z \in \Omega \;\middle|\; f(z) \neq 0 \right\}}.$$

Here the vertical line denotes the closure of the set in $\mathbb{R}^n$. The space $C_c(\Omega)$ of "**functions with compact support in** Ω" is defined as

$$C_c(\Omega) = \{\, f \in C(\Omega) \mid \operatorname{supp} f \subset K \text{ for some compact set } K \subset \Omega \,\}.$$

Exercise 1.53. Let $\Omega \subset \mathbb{R}^n$. Prove that $C_c(\Omega)$ has an equivalent characterisation

$$C_c(\Omega) = \{\, f \in C(\Omega) \mid \operatorname{supp} f \text{ is compact and } \operatorname{supp} f \subset \Omega \,\}.$$

Hint: Note that $\operatorname{supp} f$ is by definition always a closed subset of $\mathbb{R}^n$. ◇

For us, the most important property of the space $C_c(\Omega)$ is that the norms $\|f\|_p$ and $\|f\|_\infty$ are clearly well defined (i.e., finite) for every $f \in C_c(\Omega)$ even when $\Omega \subset \mathbb{R}^n$ is an unbounded (measurable) subset of $\mathbb{R}^n$. However, functions with compact support are often very useful even in the case where Ω is already a bounded set. This is especially the case in the study of differential equations and "differential operators", which are topics of the course "MATH.MA.830 Advanced Functional Analysis". One of the implications of the definition is that if $\Omega \subset \mathbb{R}^n$ is an open set, then the functions $f \in C_c(\Omega)$ need to be identically zero near the boundary of the set Ω. This is illustrated in the case of open intervals in the following exercise.

Exercise 1.54. Assume $a < b$ and $\Omega = (a, b) \subset \mathbb{R}$. Show that if $f \in C_c(\Omega)$, then there exists $\varepsilon > 0$ such that $f(t) = 0$ whenever $t \in (a, a + \varepsilon)$ or $t \in (b - \varepsilon, b)$. ◇

[1] The notation "$\|f - f_\varepsilon\|_p$" here may appear sloppy, but in the case of Lebesgue spaces it is typical to identify the actual equivalence class with a function $f : \Omega \to \mathbb{C}$ which belongs to this class.

In the following we will show that the L^p-spaces are indeed completions of spaces of continuous functions. We will only present the results in the case of functions of a single variable when $\Omega \subset \mathbb{R}$ is an interval (either bounded or unbounded, and not necessarily open or closed), but corresponding results also exist for the more general case $\Omega \subset \mathbb{R}^n$.

Theorem 1.55. *Let $I \subset \mathbb{R}$ be an interval and let $1 \le p < \infty$.*

(a) $(L^p(I), \|\cdot\|_p)$ *is the completion of* $(C_c(I), \|\cdot\|_p)$.

(b) $(L^p(\mathbb{R}), \|\cdot\|_p)$ *is the completion of* $(C_c^\infty(\mathbb{R}), \|\cdot\|_p)$, *where*

$$C_c^\infty(\mathbb{R}) = \left\{ f \in C^\infty(\mathbb{R}) \;\middle|\; \operatorname{supp} f \ \text{ is compact} \right\}$$

(c) *If $I = [a, b]$, then* $(L^p(I), \|\cdot\|_p)$ *is the completion of* $(P(I), \|\cdot\|_p)$.

Proof. **Part (a):** Since $(L^p(I), \|\cdot\|_p)$ is Banach and $C_c(I) \subset L^p(I)$, it suffices to show that $C_c(I)$ is dense in $L^p(I)$ with respect to the norm $\|\cdot\|_p$. Here we will only outline the proof, the details can be found, for example, in the reference [Rud87, Thm. 3.14]. Let $f \in L^p(I)$ and $\varepsilon > 0$ be arbitrary. Our aim is to find $f_\varepsilon \in C_c(I)$ such that $\|f - f_\varepsilon\|_p < \varepsilon$. We can do this by finding three different functions $f_0, s_\varepsilon, f_\varepsilon$ which can be used to estimate

$$\|f - f_\varepsilon\|_p \le \|f - f_0\|_p + \|f_0 - s_\varepsilon\|_p + \|s_\varepsilon - f_\varepsilon\|_p < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

The first step is find a function $f_0 \in L^p(I)$ with a compact support satisfying $\|f - f_0\|_p < \varepsilon/3$. Since $\int_I |f(t)|^p \,\mathrm{d}t < \infty$, we can choose a compact set $K \subset I$ such that

$$\int_{I \setminus K} |f(t)|^p \,\mathrm{d}t < \left(\frac{\varepsilon}{3}\right)^p,$$

and if we choose $f_0 = f\chi_K$, where χ_K is the **characteristic function of** K, i.e.

$$\chi_K(t) = \begin{cases} 1 & t \in K \\ 0 & t \in I \setminus K \end{cases}$$

then

$$\|f - f_0\|_p = \left(\int_I |f(t) - f_0(t)|^p\right)^{\frac{1}{p}} = \left(\int_{I \setminus K} |f(t)|^p \,\mathrm{d}t\right)^{\frac{1}{p}} < \frac{\varepsilon}{3}.$$

The next step in the proof is to approximate the function f_0 with a "simple function" $s_\varepsilon \in L^p(I)$ in such a way that $\|f_0 - s_\varepsilon\|_p < \varepsilon/3$. By definition, a **simple function** $g : \mathbb{R} \to \mathbb{C}$ is a function that only has a fixed set of values $\{\alpha_1, \ldots, \alpha_N\}$ for some $N \in \mathbb{N}$, i.e., there exist sets $X_1, \ldots, X_N \subset I$ such that

$$g(t) = \sum_{k=1}^{N} \alpha_k \cdot \chi_{X_k}(t),$$

where $\chi_{X_k}(\cdot)$ is a characteristic function of X_k. It is important that the sets X_k are "**measurable**". However, we do not need to know the precise definition of measurability of a set (which is one of the topics on the course "MATH.MA.840 Measure and Integration"), for us it mainly means that the characteristic functions $\chi_{X_k}(\cdot)$ have well-defined (Lebesgue) integrals $\int_I \chi_{X_k}(t)\,\mathrm{d}t$.

If a function $g : I \to [0,\infty)$ has nonnegative real values, we can do a "**sampling**" to the values of g by choosing a function $\varphi_n : [0,\infty) \to [0,\infty)$ in such a way that φ_n has a distinct set of values $\{\alpha_1,\dots,\alpha_N\}$ with $N \to \infty$ as $n \to \infty$, and the parameter t is "rounded down" to the value $\varphi_n(t) \in \{\alpha_1,\dots,\alpha_N\}$ (see Figure 1.4). Then we have that $s_n(t) = (\varphi_n \circ g)(t) = \varphi_n(g(t))$ is a simple function. Moreover, it is not at all obvious (especially since f_1 is in general discontinuous!), but it follows from the theory of Lebesgue integrals that

$$\|s_n - f_0\|_p \to 0 \qquad \text{as} \quad n \to \infty.$$

This implies that we can choose a simple function s_ε such that $\|f_0 - s_\varepsilon\|_p < \varepsilon/3$. Since f_0 has compact support, we can also choose s_ε with a compact support.

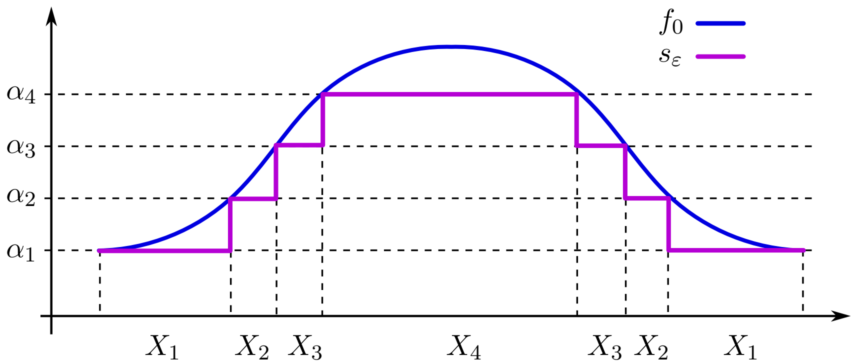

Figure 1.4: Approximation with a simple function s_ε.

Finally, it is again not at all obvious, but the discontinuous sampled function s_ε can be approximated with a continuous function with any given finite accuracy (this is due to "Lusin's Theorem", but intuitively you can also think of the jumps in the simple function being replaced with steep rises and drops). Because of this, we can choose $f_\varepsilon \in C(I)$ in such a way that $\|s_\varepsilon - f_\varepsilon\|_p < \varepsilon/3$, and since s_ε has compact support, we can choose f_ε to have the same property. Since the estimate in the beginning of the proof implies $\|f - f_\varepsilon\|_p < \varepsilon$, the proof is complete.

Part (b): We have from Part (a) that $C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ with respect to the norm $\|\cdot\|_p$. Therefore by Exercise 1.41 it suffices to show that $C_c^\infty(\mathbb{R})$ is dense in $C_c(\mathbb{R})$ with respect to the norm $\|\cdot\|_p$. To this end, let $f \in C_c(\mathbb{R})$ be arbitrary. For $\alpha > 0$, define The function $m_\alpha(t)$ has derivatives of all orders, $m_\alpha \in C_c^\infty(\mathbb{R})$, and

$$m_\alpha(t) = \begin{cases} c_\alpha \cdot e^{\frac{1}{t^2-\alpha^2}} & \text{if } |t| < \alpha \\ 0 & \text{otherwise} \end{cases}$$

where $c_\alpha = \left(\int_{-\alpha}^{\alpha} e^{\frac{1}{s^2-\alpha^2}}\,\mathrm{d}s\right)^{-1}$.

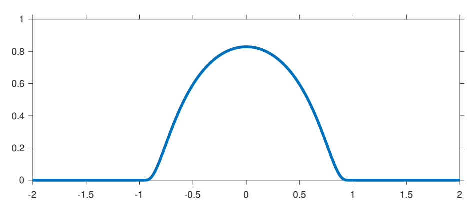

Figure 1.5: Plot of m_α for $\alpha = 1$.

$\int_{-\alpha}^{\alpha} m_\alpha(t)\,\mathrm{d}t = 1$. Define

$$f_\alpha(t) = \int_{\mathbb{R}} f(t-s) m_\alpha(s)\,\mathrm{d}s.$$

A change of variables $u = t - s$ shows that

$$f_\alpha(t) = \int_{\mathbb{R}} f(u) m_\alpha(t-u)\,\mathrm{d}u. \tag{1.5}$$

Because $m_\alpha \in C_c^\infty(\mathbb{R})$ and $f \in C_c(\mathbb{R})$, the above formula implies that we also have $f_\alpha \in C_c^\infty(\mathbb{R})$. Since f has a compact support by assumption, f is uniformly continuous on $\mathbb{R}$, and thus for any $\varepsilon > 0$ there exists an $\alpha > 0$ such that

$$|f(t-s) - f(t)| < \varepsilon \text{ whenever } |s| < \alpha.$$

Using this, for any $t \in \mathbb{R}$ we can estimate the difference $|f_\alpha(t) - f(t)|$ by (recall that $\int_{\mathbb{R}} m_\alpha(u)\,\mathrm{d}u = 1$, and that $m_\alpha(u) = 0$ if $|u| \geq \alpha$)

$$\begin{aligned} |f_\alpha(t) - f(t)| &= \left| \int_{\mathbb{R}} [f(t-u) - f(t)] m_\alpha(u)\,\mathrm{d}u \right| \\ &\leq \int_{\mathbb{R}} |f(t-u) - f(t)| m_\alpha(u)\,\mathrm{d}u \\ &= \int_{-\alpha}^{\alpha} |f(t-u) - f(t)| m_\alpha(u)\,\mathrm{d}u < \varepsilon. \end{aligned}$$

We examine the problem now from the point of view of the p-norm. Since f and f_α have compact supports, we can choose $a, b \in \mathbb{R}$ such that $a < b$. The formula (1.5) implies that $\operatorname{supp} f_\alpha \subset [a-\alpha, b+\alpha]$, because if $t > b+\alpha$, then $m_\alpha(t-u)$ in the integral is nonzero only if $|t-u| < |\alpha|$, in which case necessarily $u > b$ and $f(u) = 0$ (the case $t < a - \alpha$ can be analysed similarly). Thus we have

$$\|f_\alpha - f\|_p^p = \int_{\mathbb{R}} |f_\alpha(t) - f(t)|^p\,\mathrm{d}t = \int_{a-\alpha}^{b+\alpha} |f_\alpha(t) - f(t)|^p\,\mathrm{d}t \leq \int_{a-\alpha}^{b+\alpha} \varepsilon^p\,\mathrm{d}t = \varepsilon^p(b-a+2\alpha)$$

so that $\|f_\alpha - f\|_p \leq \varepsilon(b-a+2\alpha)^{1/p}$. Thus the norm $\|f_\alpha - f\|_p$ can be made arbitrarily small by choosing a sufficiently small $\alpha > 0$. Therefore $(C_c^\infty(\mathbb{R}), \|\cdot\|_p)$ is dense in $(C_c(\mathbb{R}), \|\cdot\|_p)$ and the claim holds.

Part (c): Since $[a, b]$ is a finite and closed interval, we have $C_c([a, b]) = C([a, b])$ (we can choose $K = [a, b]$ in the definition of $C_c([a, b])$). Since by Part (a), $(L^p(a, b), \|\cdot\|_p)$ is the completion of $(C([a, b]), \|\cdot\|_p)$, it is sufficient to show that $P([a, b])$ is dense in $C([a, b])$ with respect to the norm $\|\cdot\|_p$. We already have from the Weirstrass Approximation Theorem (in Theorem 1.43) that $P([a, b])$ is dense in $C([a, b])$ with respect to the norm $\|\cdot\|_\infty$. Since the interval $[a, b]$ has finite length, for any $f \in C([a, b])$ and $p \in P([a, b])$ we can estimate

$$\begin{aligned} \|f - p\|_p &= \left(\int_a^b |f(t) - p(t)|^p\,\mathrm{d}t \right)^{1/p} \leq \left(\int_a^b \sup_{s \in [a,b]} |f(s) - p(s)|^p\,\mathrm{d}t \right)^{1/p} \\ &= \left(\int_a^b 1\,\mathrm{d}t \right)^{1/p} \sup_{s \in [a,b]} |f(s) - p(s)| = (b-a)^{1/p} \|f - p\|_\infty. \end{aligned}$$

This estimate and the denseness of $(P([a, b]), \|\cdot\|_\infty)$ in $(C([a, b]), \|\cdot\|_\infty)$ directly imply that any function $f \in C([a, b])$ can be approximated in the norm $\|\cdot\|_p$ with arbitrary accuracy with elements $p \in P([a, b])$. Thus $(P([a, b]), \|\cdot\|_p)$ is dense in $(C([a, b]), \|\cdot\|_p)$. This completes the proof that $(L^p(a, b), \|\cdot\|_p)$ is the completion of $(P([a, b]), \|\cdot\|_p)$. $\square$

2. Linear Operators

In this chapter we study functions (or “mappings”) $T : X \to Y$ between two normed spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, meaning that an element $x \in X$ is mapped to a element $y = T(x) \in Y$. We in particular focus on functions which are *linear* in the sense that

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \qquad \forall \alpha, \beta \in \mathbb{C},\ \forall x, y \in X.$$

Such linear mappings, which are traditionally called *linear operators*, have a very well-formed theory and in this chapter we will begin to study their most fundamental properties. We begin by defining **linear operators** more precisely.

Definition 2.1. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. A mapping $T : X \to Y$ is a **linear operator** if

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \qquad \forall \alpha, \beta \in \mathbb{C},\ \forall x, y \in X. \tag{2.1}$$

More generally, the linear operator T may be defined only for elements $x \in D(T)$ where $D(T)$ is a subspace of X, and we denote $T : D(T) \subset X \to Y$. In this case $D(T)$ is called the **domain** of T and the identity in (2.1) is required to hold for all $\alpha, \beta \in \mathbb{C}$ and $x, y \in D(T)$.

For a linear operator $T : D(T) \subset X \to Y$ it is customary to write Tx instead of $T(x)$ for $x \in D(T)$.

We begin by introducing examples of linear operators between different types of normed spaces. We will also encounter these same operators several times throughout our course.

Example 2.2. As we recall from linear algebra, a matrix $A \in \mathbb{C}^{m\times n}$ can be interpreted as a mapping between two Euclidean spaces $\mathbb{C}^n$ and $\mathbb{C}^m$. More precisely, if $A \in \mathbb{C}^{m\times n}$ is a fixed matrix, we can define $T : \mathbb{C}^n \to \mathbb{C}^m$ as the operation of multiplication of $x \in \mathbb{C}^n$ (from the left) with the matrix A, i.e., $Tx = Ax \in \mathbb{C}^m$ for all $x \in \mathbb{C}^n$. Then T is indeed a linear operator because the linearity of the matrix multiplication implies that

$$T(\alpha x + \beta y) = A(\alpha x + \beta y) = \alpha Ax + \beta Ay = \alpha Tx + \beta Ty$$

for all $\alpha, \beta \in \mathbb{C}$ and $x, y \in \mathbb{C}^n$.

Because every matrix $A \in \mathbb{C}^{m\times n}$ defines a linear operator, the theory of linear operators can (in some sense) be considered as an extension of the theory of matrices. Indeed, on this course we will see that linear operators have many features which closely resemble the familiear properties of matrices. On the other hand, we will also see that parts of the theory of linear operators are very “rich” compared to the corresponding theory of matrices. This is for example the case with *spectral theory* of linear operators (Chapter 4) which generalises the study of eigenvalues and eigenvectors of matrices. ◇

Example 2.3 (Point evaluation)**.** Let $\Omega \subset \mathbb{R}^n$ be a closed and bounded set, and let $X = C(\Omega)$ with the norm $\|\cdot\|_\infty$. For a fixed $z_0 \in \Omega$ we can define the operator $T_{z_0} : X \to \mathbb{C}$ which "evaluates" a function $f \in C(\Omega)$ at the point z_0, i.e., $T_{z_0} f = f(z_0) \in \mathbb{C}$. Since the functions in $C(\Omega)$ are continuous, every $f \in C(\Omega)$ has a well-defined value at the point z_0, and thus $T_{z_0} f$ is well-defined for every $f \in X$. The linearity of the operator T_{z_0} follows from the definitions of the addition and scalar multiplication in $C(\Omega)$. Indeed, for all $\alpha, \beta \in \mathbb{C}$ and $f, g \in X$ we have

$$\begin{aligned} T_{z_0}(\alpha f + \beta g) &= (\alpha f + \beta g)(z_0) = (\alpha f)(z_0) + (\beta g)(z_0) \\ &= \alpha f(z_0) + \beta g(z_0) = \alpha T_{z_0} f + \beta T_{z_0} g. \end{aligned}$$

We can also define the point evaluation operator T_{z_0} as an operator from $X = L^p(\Omega)$ for $p \geq 1$ (with the norm $\|\cdot\|_p$) to $\mathbb{C}$. However, since the functions $f \in L^p(\Omega)$ do not have uniquely defined values at any particular points, the expression "$T_{z_0} f$" cannot be defined for all functions $f \in X$! Instead, we need to choose the domain $D(T_{z_0})$ of the operator T_{z_0} in such a way that $T_{z_0} f$ is well-defined for all $f \in D(T_{z_0})$. Since $\Omega \subset \mathbb{R}^n$ is compact (closed and bounded), we have $C(\Omega) \subset L^p(\Omega)$, and we can for example choose $D(T_{z_0}) = C(\Omega)$. In this case $f(z_0)$ is again well-defined for all $f \in D(T_{z_0})$ and the linearity of the operator $T_{z_0} : D(T_{z_0}) \subset X \to \mathbb{C}$ follows directly from our earlier computation. ◇

Example 2.4 (Shift operators on $\ell^p(\mathbb{C})$ and $\ell^\infty(\mathbb{C})$)**.** Let $X = \ell^p(\mathbb{C})$ for $p \geq 1$ or $X = \ell^\infty(\mathbb{C})$. For every infinite sequence $(x_k)_{k=1}^\infty \in X$ we can define the operators of *shifting* all the elements of a sequence either "to the left" or "to the right". More precisely, we can define the *right shift operator* $S_r : X \to X$ such that

$$S_r x = (0, x_1, x_2, x_3, \ldots), \qquad x = (x_k)_{k=1}^\infty$$

and the *left shift operator* $S_l : X \to X$ such that

$$S_l x = (x_2, x_3, x_4 \ldots), \qquad x = (x_k)_{k=1}^\infty.$$

Clearly both $S_r x$ and $S_l x$ are well-defined infinite sequences of complex numbers. In order to show that S_r and S_l are well-defined mappings from X to X, we want to show that $S_r x \in X$ and $S_l x \in X$ whenever $x \in X$. Since $X = \ell^p(\mathbb{C})$ for $p \geq 1$ or $X = \ell^\infty(\mathbb{C})$, our aim is to show that $\|S_r x\|_X < \infty$ and $\|S_l x\|_X < \infty$ for all $x \in X$.

Let us first consider the case where $X = \ell^p(\mathbb{C})$ for $p \geq 1$. Let $x = (x_k)_{k=1}^\infty \in X$ be arbitrary and denote $S_r x = y = (y_k)_{k=1}^\infty$. By definition, we have $y_1 = 0$ and $y_k = x_{k-1}$ for $k \geq 2$. We then have

$$\|S_r x\|_p = \|y\|_p = \left(\sum_{k=1}^\infty |y_k|^p\right)^{\frac{1}{p}} = \left(\sum_{k=2}^\infty |x_{k-1}|^p\right)^{\frac{1}{p}} = \left(\sum_{k=1}^\infty |x_k|^p\right)^{\frac{1}{p}} = \|x\|_p < \infty,$$

and thus $S_r x \in X$. On the other hand, if $x = (x_k)_{k=1}^\infty \in X$ is arbitrary and we denote $S_l x = y = (y_k)_{k=1}^\infty$, then $y_k = x_{k+1}$ for $k \in \mathbb{N}$. Because of this, we have

$$\|S_l x\|_p = \left(\sum_{k=1}^\infty |y_k|^p\right)^{\frac{1}{p}} = \left(\sum_{k=1}^\infty |x_{k+1}|^p\right)^{\frac{1}{p}} = \left(\sum_{k=2}^\infty |x_k|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^\infty |x_k|^p\right)^{\frac{1}{p}} = \|x\|_p < \infty,$$

which implies $S_l x \in X$. Thus $S_r : X \to X$ and $S_l : X \to X$ are well-defined when $X = \ell^p(\mathbb{C})$ with $p \geq 1$. Repeating the arguments in the case where $X = \ell^\infty(\mathbb{C})$ is left as an exercise.

Finally, the linearity of the operators S_r and S_l follow from the definitions of addition and scalar multiplication for sequences. Indeed, if $\alpha, \beta \in \mathbb{C}$ and $x, y \in X$ are arbitrary, we have

$$\begin{aligned} S_r(\alpha x + \beta y) &= (0, \alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \dots) \\ &= \alpha(0, x_1, x_2, \dots) + \beta(0, y_1, y_2, \dots) \\ &= \alpha S_r x + \beta S_r y \end{aligned}$$

and

$$\begin{aligned} S_l(\alpha x + \beta y) &= (\alpha x_2 + \beta y_2, \alpha x_3 + \beta y_3, \dots) \\ &= \alpha(x_2, x_3, \dots) + \beta(y_2, y_3, \dots) \\ &= \alpha S_l x + \beta S_l y. \end{aligned}$$

Thus both $S_r : X \to X$ and $S_l : X \to X$ are linear operators. ◇

Exercise 2.5. In Example 2.4, prove that $S_r x \in \ell^\infty(\mathbb{C})$ and $S_l x \in \ell^\infty(\mathbb{C})$ for all $x \in \ell^\infty(\mathbb{C})$. ◇

Exercise 2.6. If $T : X \to X$, the n**th power** $T^n : X \to X$ of T is defined as the composition $T^n x = T(T^{n-1}x) = \cdots = T(T(\cdots(Tx)))$ for $x \in X$. Alternatively, it is possible to denote $T^n = TT \cdots T$ (where T appears n times). Write down the definitions of the powers S_r^2, S_l^2, S_r^3, and S_l^3 for the shift operators in Example (2.4). ◇

Example 2.7 (Differential operators)**.** Differentiation of a function is a linear operation, and we can indeed define an operator T which differentiates a given function $f : [a, b] \to \mathbb{C}$, i.e. $Tf = f'$ (the values of this function for $t \in [a, b]$ can be denoted by $(Tf)(t) = f'(t)$).

In order to make the operator well-defined, we need to choose the spaces X and Y in a suitable way. For example, if our aim is to consider continuously differentiable functions, we can choose $X = C^1([a, b])$ (the space of functions having continuous derivative, and at $t = a$ and $t = b$ the derivatives are defined as the "one-sided" derivatives) with (for example) the norm $\|\cdot\|_\infty$ and we can let $Y = C([a, b])$ with the norm $\|\cdot\|_\infty$. Due to our choises, we have that $Tf = f' \in C([a, b]) = Y$ for all $f \in C^1([a, b]) = Y$, and thus T is well-defined as an operator $T : X \to Y$. The operator T is also linear, since for all $\alpha, \beta \in \mathbb{C}$ and for all $f, g \in X = C^1([a, b])$ the linearity of differentiation implies

$$T(\alpha f + \beta g) = (\alpha f + \beta g)' = \alpha f' + \beta g' = \alpha Tf + \beta Tg.$$

In the case of operators which map functions $f : [a, b] \to \mathbb{C}$ to other functions $g : [a, b] \to \mathbb{C}$ (and in other similar instances) it is often convenient to be able to choose the spaces X and Y to be the same space, i.e., $X = Y$. Especially, the property $X = Y$ will allow us to later consider the *spectrum* of the operator (Chapter 4), or define *powers* of the operator (similarly as in Exercise 2.6). In the case of the differential operator, we can achieve this by defining $X = Y = C([a, b])$ with the norm $\|\cdot\|_\infty$, and defining the operator $T : D(T) \subset X \to X$ on the domain $D(T) = C^1([a, b])$ which is a subspace of X. ◇

2.1 Bounded Linear Operators

We will now focus our attention to an important class of linear operators, namely, **bounded linear operators**.

Definition 2.8. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. The linear operator $T : D(T) \subset X \to Y$ is **bounded** if there exists $M \geq 0$ such that

$$\|Tx\|_Y \leq M\|x\|_X, \qquad \forall x \in D(T).$$

The linearity of an operator $T : D(T) \subset X \to Y$ has very strong consequences especially for the continuity of the operator (in the sense of continuity of a function from X to Y). More precisely, the mapping $T : D(T) \subset X \to Y$ is defined to be **continuous at the point** $x \in D(T)$ if *for every* $\varepsilon > 0$ *there exists* $\delta_x > 0$ *such that*

$$\|Tx - Ty\|_Y < \varepsilon \qquad \text{whenever} \quad y \in D(T) \text{ and } \|x - y\|_X < \delta_x.$$

Here the value $\delta_x > 0$ is allowed to depend on $x \in D(T)$. In particular, as shown in the next theorem, a linear operator $T : D(T) \subset X \to Y$ is continuous at the single point $x = 0$ if and only if it is a bounded operator, and in this case it is also "uniformly continuous" everywhere on $D(T)$. In uniform continuity, the value "$\delta_x > 0$" can be chosen independently of $x \in D(T)$. The theorem below also explains why both terms **bounded linear operator** and **continuous operator** appear in the literature and that both of them refer to the same class of operators.

Theorem 2.9. *Let* $(X, \|\cdot\|_X)$ *and* $(Y, \|\cdot\|_Y)$ *be normed spaces and let* $T : D(T) \subset X \to Y$ *be a linear operator. The following are equivalent.*

(i) T *is continuous at* $x = 0$.

(ii) T *is uniformly continuous on* $D(T)$, *i.e.,* for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|Tx - Ty\|_Y < \varepsilon \qquad \textit{whenever} \quad x, y \in D(T) \textit{ and } \|x - y\|_X < \delta.$$

(iii) T *is a bounded operator.*

Proof. **(i) ⇒ (ii):** Assume that T is continuous at $x = 0$ and let $\varepsilon > 0$ be arbitrary. Since $T0 = 0$, by definition of continuity there exists $\delta_0 > 0$ such that

$$\|Ty\|_Y < \varepsilon \qquad \text{whenever} \quad y \in D(T) \text{ and } \|y\|_X < \delta_0. \tag{2.2}$$

Our aim is to show that we can directly choose $\delta = \delta_0 > 0$ in the definition of uniform continuity. Assume that $x \in D(T)$ and $y \in D(T)$ satisfy $\|x - y\|_X < \delta_0$. The linearity of T implies that $T(x - y) = Tx - Ty$, and thus (2.2) implies

$$\|Tx - Ty\|_Y = \|T(x - y)\|_Y < \varepsilon$$

since $\|x - y\|_X < \delta_0$. Since $\varepsilon > 0$ was arbitrary, this shows that $T : D(T) \subset X \to Y$ is uniformly continuous.

(ii) ⇒ (iii): Assume $T : D(T) \subset X \to Y$ is uniformly continuous. Thus for $\varepsilon = 1$ we can find $\delta > 0$ such that (choosing $x = 0$) $\|Ty\|_Y < 1$ whenever $y \in D(T)$ and

$\|y\|_X < \delta$. Now let $x \in D(T)$ be arbitrary. We can assume $x \neq 0$, since otherwise $Tx = 0$ and the claim is trivial. In order to use the continuity of T, our aim is to define a vector $y \in D(T)$ satisfying $\|y\|_X < \delta$ by "scaling" the norm of the vector of x. In particular, if we define $y = \alpha x \in D(T)$ with $\alpha = \delta/(2\|x\|_X)$, then we have

$$\|y\|_X = \|\alpha x\|_X = |\alpha|\|x\|_X = \frac{\delta}{2\|x\|_X}\|x\|_X = \delta/2 < \delta.$$

Because of this, we have

$$\|Ty\|_Y < 1 \quad\Leftrightarrow\quad \|T(\alpha x)\|_Y < 1 \quad\Leftrightarrow\quad |\alpha|\|Tx\|_Y < 1 \quad\Leftrightarrow\quad \|Tx\|_Y < \frac{1}{|\alpha|} = \frac{2}{\delta}\|x\|_X.$$

Since $x \in D(T)$ was arbitrary, we have shown that $\|Tx\|_Y \leq M\|x\|_X$ with $M = 2/\delta > 0$ for all $x \in D(T)$, and thus the operator T is bounded.

(iii) $\Rightarrow$ (i): This part is left as an exercise. □

Exercise 2.10. Prove the implication from (iii) to (i) in Theorem 2.9. ◇

Remark 2.11. In the case of mappings between different normed spaces, for example, $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, it is *very important* to distinguish between the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$! Indeed, if the vector spaces X and Y are different, there is no reason why the norm $\|x\|_Y$ would be defined for elements $x \in X$ (see, e.g., Example 2.3). Moreover, even if X and Y are the same vector space, the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ may be different, and for example different choices of norms on these spaces define which operators $T : X \to Y$ are bounded and which operators are not.

This being said, throughout the course we very often use the notation $\|x\|$ for $\|x\|_X$ in the case where the choice of the norm is uniquely defined and clear from the context. For example, if $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are two normed spaces, then we may write the definition of the boundedness of $T : D(T) \subset X \to Y$ as

$$\exists M \geq 0 : \qquad \|Tx\| \leq M\|x\|, \qquad \forall x \in D(T).$$

This is because "$\|Tx\|$" only makes sense if it is exactly $\|Tx\|_Y$, and similarly "$\|x\|$" can only mean the norm $\|x\|_X$. However, we will only use this shorter notation if we are not simultaneously considering multiple norms on the vector spaces X and Y, and sometimes we write $\|\cdot\|_X$ and $\|\cdot\|_Y$ simply for additional emphasis or clarity.

Example 2.12. The **identity operator** $I : X \to X$ mapping every element of X to itself, i.e. $Ix = x$ for all $x \in X$, is a bounded linear operator. Indeed,

$$\|Ix\|_X = \|x\|_X, \qquad \forall x \in X$$

implies that the condition in Definition 2.8 holds with $M = 1$. Similarly, we can study the boundedness of the **zero operator** $O : X \to Y$ mapping every element of X to the zero element of Y, i.e., $Ox = 0$. We have

$$\|Ox\|_Y = \|0\|_Y = 0 = 0 \cdot \|x\|_X, \qquad \forall x \in X.$$

Thus the condition in Definition 2.8 holds with $M = 0$ and $O : X \to Y$ is bounded. ◇

Before considering more examples of bounded (and unbounded) operators, we will also define the **space** of bounded linear operators from X to Y.

Definition 2.13. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. We define **the space of bounded linear operators** $B(X,Y)$ as

$$B(X,Y) = \{\, T : X \to Y \mid T \text{ is a bounded linear operator} \,\}.$$

The space $B(X,Y)$ becomes a vector space with addition $S+T : X \to Y$ and scalar multiplication $\alpha T : X \to Y$ defined by

$$(S+T)x = Sx + Tx,$$
$$(\alpha T)x = \alpha T x,$$

for all $S, T \in B(X,Y)$, $\alpha \in \mathbb{C}$, and $x \in X$.

Linear operators with $D(T) = X$ and $Y = \mathbb{C}$ have a special name: A linear operator $T : X \to \mathbb{C}$ is called a **linear functional**, and the space of **bounded linear functionals** $B(X, \mathbb{C})$ is called the **dual space of** X. The point evaluation operator $T_{z_0} : C(\Omega) \to \mathbb{C}$ in Example 2.3 is an example of a linear functional (in Example 2.18 we will see that it is also a **bounded** linear functional).

Exercise 2.14. Prove that $S + T : X \to Y$ and $\alpha T : X \to Y$ in Definition 2.13 are bounded linear operators, i.e., $S + T \in B(X,Y)$ and $\alpha T \in B(X,Y)$. ◇

In the case where the two normed spaces are the same, we denote $B(X,X)$ by $B(X)$. The zero element $O \in B(X,Y)$ of the vector space is the zero operator in Example 2.12 (mapping every element $x \in X$ to the zero element $0 \in Y$, i.e., $Ox = 0 \in Y$ for all $x \in X$). The space $B(X,Y)$ in fact becomes a normed space with the **operator norm** defined below.

Definition 2.15. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. Then the **(operator) norm** of $T \in B(X,Y)$ is defined as

$$\|T\|_{B(X,Y)} = \sup_{\|x\|_X \le 1} \|Tx\|_Y.$$

Exercise 2.16. Prove that $(B(X,Y), \|\cdot\|_{B(X,Y)})$ is a normed space (i.e., prove that $\|\cdot\|_{B(X,Y)}$ is a norm on $B(X,Y)$). ◇

The definition of the operator norm indicates that $\|T\|$ can be seen as the maximal size of the norm of the image Tx of any vector $x \in X$ with norm $\|x\| \le 1$. The definition also directly implies that if $\|x\| \le 1$, then $\|Tx\| \le \|T\|_{B(X,Y)}$. This further implies that if $x \ne 0$ and if we define $y = \frac{1}{\|x\|}x$, then $\|x\|y = x$, $\|y\| = 1$, and

$$\|Tx\| = \| \|x\| Ty\| = \|x\| \|Ty\| \le \|T\| \|x\|.$$

Because of this, in the Definition 2.8 it is in particular always possible to choose $M = \|T\| \ge 0$.

Exercise 2.17. Prove that if $T \in B(X,Y)$ and if $M \ge 0$ is such that $\|Tx\| \le M\|x\|$ for all $x \in X$, then $\|T\| \le M$. ◇

The most typical way to compute the norm $\|T\|$ of a linear operator is to first find (as small as possible) upper bound $M \geq 0$ such that $\|Tx\| \leq M\|x\|$ (which implies that necessarily $\|T\| \leq M$). If it is then possible to find an element $y \in X$ such that $\|Ty\| = M\|y\|$, we can deduce $\|T\| = M$. More generally, in the second part it is sufficient to find a sequence $(x_k)_{k=1}^\infty \subset X$ such that $x_k \neq 0$ for all $k \in \mathbb{N}$ and

$$\frac{\|Tx_k\|}{\|x_k\|} \to M, \qquad \text{as} \quad k \to \infty.$$

Example 2.18. Let $\Omega \subset \mathbb{R}^n$ be a closed and bounded set, let $z_0 \in \Omega$, and consider the point evaluation operator $T_{z_0} : X \to \mathbb{C}$ in Example 2.3. We will show that T_{z_0} is a bounded linear operator when we choose $X = C(\Omega)$ with the norm $\|\cdot\|_\infty$. Indeed, with this choice of X, we have that for any $f \in C(\Omega)$

$$\|Tf\|_{\mathbb{C}} = |f(z_0)| \leq \sup_{z\in\Omega} |f(z)| = \|f\|_\infty.$$

Thus $T_{z_0} \in B(X, \mathbb{C})$ is bounded and in particular $\|T_{z_0} f\| \leq \|f\|_\infty$ for all $f \in X$. To compute the norm $\|T_{z_0}\|$ we can note that if we choose $f \in C(\Omega)$ to be the constant function such that $f(z) = 1$ for all $z \in \Omega$, then

$$\|T_{z_0} f\| = |f(z_0)| = 1 = \sup_{z\in\Omega} |f(z)| = \|f\|_\infty,$$

and thus for this function we have $\|Tf\| = 1 \cdot \|f\|$, which implies together with our earlier estimate implies that $\|T\| = 1$.

On the other hand, the operator T_{z_0} is *not bounded* if we define it instead as an operator $T_{z_0} : D(T_{z_0}) \subset L^p(\Omega) \to \mathbb{C}$ with domain $D(T_{z_0}) = C(\Omega)$. This is because it is possible to define a sequence $(f_k)_{k=1}^\infty \subset C(\Omega)$ of continuous functions which all have p-norm equal to $\|f_k\|_p = 1$, but whose values $f_k(z_0)$ at the point $z_0 \in \Omega$ increase without bound as $k \to \infty$. Indeed, this can be done by defining suitable "hat functions" centered at the point z_0 and with decreasing radii (this is especially simple if the point $z_0 \in \Omega$ is in the interior of Ω). ◇

Example 2.19. In this example we will show that the differential operator $T : C^1([a, b]) \to C([a, b])$ in Example 2.7 is not bounded (and thus is an *unbounded operator*). Since both the spaces $C^1([a, b])$ and $C([a, b])$ use the norm $\|\cdot\|_\infty$, our aim is to find a sequence $(f_k)_{k=1}^\infty$ of functions such that the maximal values $f_k'(t)$ become increasingly large when compared to the maximal values $f_k(t)$ of the functions themselves. Ideal functions for this purpose are functions whose values are bounded, but which behave in an "*oscillatory*" manner, such as trigonometric functions $\sin(\omega t)$ and $\cos(\omega t)$ for $\omega > 0$. Indeed, if we define $f_k(t) = \sin(kt)$, then $(f_k)_k \subset C^1([a, b])$, but for high values of k the oscillation of $f_k(t)$ becomes more and more rapid. This is directly reflected in the norm $\|Tf_k\|$, since

$$\|Tf_k\|_\infty = \sup_{t\in[a,b]} |f_k'(t)| = \sup_{t\in[a,b]} |\frac{d}{dt} \sin(kt)| = \sup_{t\in[a,b]} |k \sin(kt)| = k.$$

On the other hand, we have

$$\|f_k\|_\infty = \sup_{t\in[a,b]} |f_k(t)| = \sup_{t\in[a,b]} |\sin(kt)| = 1$$

for all $k \in \mathbb{N}$. Since $(f_k)_k \subset C^1([a,b])$ is a sequence such that $\|f_k\|_\infty = 1$ and $\|Tf_k\|_\infty \to \infty$ as $k \to \infty$, the operator T is not bounded (it is not possible to find a constant $M \geq 0$ such that the condition of Definition 2.8 would hold).

Whether or not an operator between two vector spaces is bounded or not is highly dependent on the choices of the norms (i.e., the boundedness is indeed a property of the operator between two *normed spaces*). The differential operator in this example becomes a bounded operator if we change the norm on the space $C^1([a,b])$ to be defined as $\|f\|_{C^1} = \|f\|_\infty + \|f'\|_\infty$. Indeed, it is a straighforward exercise to verify that $\|\cdot\|_{C^1}$ is a norm on $C^1([a,b])$, and for all $f \in C^1([a,b])$ we have

$$\|Tf\|_\infty = \|f'\|_\infty \leq \|f\|_\infty + \|f'\|_\infty = \|f\|_{C^1}.$$

Thus $T \in B(C^1([a,b]), C([a,b]))$ with the choice $\|\cdot\|_{C^1}$ of the norm on $C^1([a,b])$. ◇

Exercise 2.20. Let $X = \ell^p(\mathbb{C})$ with $p \geq 1$, or $X = \ell^\infty(\mathbb{C})$. Prove that the shift operators S_r and S_l in Example 2.4 satisfy $S_r \in B(X)$ and $S_r \in B(X)$. Moreover, compute $\|S_l\|$ and $\|S_r\|$. ◇

Example 2.21. Assume $(t,s) \mapsto k(t,s) \in C(\Omega)$ where $\Omega = [a,b] \times [a,b] \subset \mathbb{R}^2$. We can define an **integral operator** T on the space $X = L^p(a,b)$ with $1 \leq p < \infty$ so that for every $f \in X$ the function $Tf : [a,b] \to \mathbb{C}$ is defined by

$$(Tf)(t) = \int_a^b k(t,s) f(s)\,\mathrm{d}s, \qquad \forall t \in [a,b].$$

The function $k(\cdot,\cdot)$ is called the **kernel** of the integral operator. The expression Tf is well-defined for all $f \in X$, since the function $k(\cdot,\cdot)$ is continuous on $[a,b] \times [a,b]$, and thus uniformly continuous, and thus for all $f \in L^p(a,b)$ we have that $k(t,\cdot)f(\cdot) \in L^p(a,b) \subset L^1(a,b)$ for every fixed $t \in [a,b]$.

Our aim is to show that $T \in B(X,X)$. Our first task is to verify that the mapping $T : X \to X$ is well-defined in the sense that $Tf \in X$ for all $f \in X$. We will not consider the measurability of Tf, but we will show that Tf is p-integrable. Since $k(\cdot,\cdot) \in C([a,b]^2)$, there exists a constant $C > 0$ such that $|k(t,s)| \leq C$ for all $t,s \in [a,b]$. Let $f \in X = L^p(a,b)$ be arbitrary. We can estimate

$$\begin{aligned}\int_a^b |(Tf)(t)|^p\,\mathrm{d}t &= \int_a^b \left|\int_a^b k(t,s)f(s)\,\mathrm{d}s\right|^p \mathrm{d}t \leq \int_a^b \left(\int_a^b |k(t,s)||f(s)|\,\mathrm{d}s\right)^p \mathrm{d}t \\ &\leq \int_a^b \left(\int_a^b C|f(s)|\,\mathrm{d}s\right)^p \mathrm{d}t = C^p \left(\int_a^b 1\,\mathrm{d}t\right)\left(\int_a^b |f(s)|\,\mathrm{d}s\right)^p = C^p(b-a)\|f\|_1^p.\end{aligned}$$

Since $f \in L^p(a,b) \subset L^1(a,b)$, this estimate shows that Tf is p-integrable. Since $f \in X$ was arbitrary, we deduce that T is well-defined as a mapping $T : X \to X$. In order to show that T is bounded, we need to show that there exists $M > 0$ such that

$$\|Tf\|_p \leq M\|f\|_p, \qquad \forall f \in X = L^p(a,b).$$

Let $f \in X$ be arbitrary. Our previous estimate shows that $\|Tf\|_p \leq C(b-a)^{1/p}\|f\|_1$. If $p = 1$, this estimate is exactly of the correct form, and we can choose $M = C(b-a)^{1/p} = C(b-a)$. On the other hand, if $1 < p < \infty$, we can use the Hölder's inequality (Theorem 1.50) to estimate the norm $\|f\|_1$ with the norm $\|f\|_p$. Indeed, if we define

$q = p/(p-1)$ and define $g : [a,b] \to \mathbb{C}$ such that $g(s) = 1$ for all $s \in [a,b]$, then $g \in L^q(a,b)$ and Hölder's inequality implies

$$\|f\|_1 = \|gf\|_1 \le \|g\|_q \|f\|_p = \left(\int_a^b 1^q \,\mathrm{d}s\right)^{\frac{1}{q}} \|f\|_p = (b-a)^{\frac{1}{q}} \|f\|_p.$$

Combining this with our ealier estimate for $\|Tf\|_p$ shows that

$$\|Tf\|_p \le C(b-a)^{1/p}\|f\|_1 \le C(b-a)^{1/p+1/q}\|f\|_p = C(b-a)\|f\|_p.$$

We can therefore choose $M = C(b-a)$ also in the case where $1 < p < \infty$. Proving the linearity of the operator T is left as an exercise, and after this step we can conclude that $T \in B(X,X)$. Finally, we note that our estimate $\|Tf\|_p \le C(b-a)\|f\|_p$ for all $f \in X$ also implies that the operator norm of T satisfies $\|T\| \le C(b-a)$. ◇

Exercise 2.22. Show that the integral operator $T : X \to X$ in Example 2.21 is linear. Moreover, find the kernel $k(\cdot,\cdot)$ which corresponds to the integral operator $T \in B(X)$ which is defined so that

$$(Tf)(t) = \int_a^t f(s)ds, \qquad \forall f \in L^p(a,b)$$

for all $t \in [a,b]$. ◇

We already confirmed that $(B(X,Y), \|\cdot\|_{B(X,Y)})$ is a normed space. This space is complete whenever the space Y is a Banach space.

Theorem 2.23. *Let $(X, \|\cdot\|_X)$ be a normed space and assume $(Y, \|\cdot\|_Y)$ is a Banach space. The space $(B(X,Y), \|\cdot\|_{B(X,Y)})$ is a Banach space.*

Proof. Let $(T_k)_{k=1}^\infty \subset B(X,Y)$ be an arbitrary Cauchy sequence. Our aim is to show that $(T_k)_k$ converges in $B(X,Y)$, i.e., there exists $T \in B(X,Y)$ such that $\|T_k - T\| \to 0$ as $k \to \infty$. For every fixed $x \in X$ we have that

$$\|T_k x - T_m x\| = \|(T_k - T_m)x\| \le \|T_k - T_m\|\|x\| \to 0 \tag{2.3}$$

as $k,m \to \infty$, and thus $(T_k x)_{k=1}^\infty \subset Y$ is a Cauchy sequence in the space $(Y, \|\cdot\|_Y)$ for every $x \in X$. Since $(Y, \|\cdot\|_Y)$ was assumed to be Banach space, every one of these sequences converges in Y, i.e., for every $x \in X$ there exists $y_x \in Y$ such that $\|T_k x - y_x\| \to 0$ as $k \to \infty$. We will now define an operator $T : X \to Y$ so that $Tx = y_x$ for all $x \in X$. Our aim is to show that $T \in B(X,Y)$ and $\|T_k - T\| \to 0$ as $k \to \infty$. The operator T is linear because for all $\alpha, \beta \in \mathbb{C}$ and $x, y \in X$ we have

$$T(\alpha x + \beta y) = \lim_{k\to\infty} T_k(\alpha x + \beta y) = \alpha \lim_{k\to\infty} T_k x + \beta \lim_{k\to\infty} T_k y = \alpha Tx + \beta Ty.$$

In order to show that T is bounded and that $T_k \to T$ as $k \to \infty$, we take an arbitrary $\varepsilon > 0$. Since $(T_k)_k$ is a Cauchy sequence, we can choose $n_\varepsilon \in \mathbb{N}$ such that $\|T_k - T_m\| \le \varepsilon$ for all $k, m \ge n_\varepsilon$. The estimate in (2.3) implies that for all $k, m \ge n_\varepsilon$ and $x \in X$ we have

$$\|T_k x - T_m x\| \le \|T_k - T_m\|\|x\| \le \varepsilon\|x\|.$$

Since the norm on a normed space is always continuous as a mapping $\|\cdot\|_Y : Y \to \mathbb{C}$, we further have that for every $x \in X$ and $m \geq n_\varepsilon$

$$\|(T - T_m)x\| = \|Tx - T_m x\| = \left\|(\lim_{k\to\infty} T_k x) - T_m x\right\| = \lim_{k\to\infty} \|T_k x - T_m x\| \leq \varepsilon \|x\|.$$

Thus $T - T_m : X \to Y$ is in particular a bounded operator if $m \geq n_\varepsilon$, and since $T = (T - T_m) + T_m$ where $T - T_m \in B(X,Y)$ and $T_m \in B(X,Y)$, we also have $T \in B(X,Y)$. Indeed, this follows directly from Exercise 2.14 where we showed that $B(X,Y)$ is a vector space, and therefore closed under addition of operators. Moreover, the above estimate also implies that the operator norm of $T - T_m$ satisfies (see Exercise 2.17)

$$\|T - T_m\| \leq \varepsilon, \qquad \forall m \geq n_\varepsilon. \tag{2.4}$$

Since $\varepsilon > 0$ was arbitrary, and there exists $n_\varepsilon \in \mathbb{N}$ such that (2.4) holds, we have that $\|T_m - T\| \to 0$ as $m \to \infty$. Since $(T_k)_{k=1}^\infty$ was an arbitrary Cauchy sequence in $B(X,Y)$, we conclude that $(B(X,Y), \|\cdot\|_{B(X,Y)})$ is complete. □

2.2 Invertible Operators

For a linear operator $T : D(T) \subset X \to Y$, the **range** $\mathrm{Ran}(T)$ (the "space of values $y = Tx$ of T") and the **kernel** or **null space** $\mathrm{Ker}(T)$ (the "space where the values of T are zero") are defined as follows.

Definition 2.24. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces and let $T : D(T) \subset X \to Y$ be a linear operator. The **range** $\mathrm{Ran}(T)$ and **kernel** (or **null space**) $\mathrm{Ker}(T)$ of T are defined as

$$\mathrm{Ran}(T) = \{\, y \in Y \mid y = Tx \text{ for some } x \in D(T) \,\}$$
$$\mathrm{Ker}(T) = \{\, x \in D(T) \mid Tx = 0 \,\}.$$

The operator T is said to be

- **injective** (or **one-to-one**) if $\mathrm{Ker}(T) = \{0\}$.
- **surjective** (or **onto**) if $\mathrm{Ran}(T) = Y$.
- **bijective** (or **one-to-one and onto**) if it is both injective and surjective.

The range $\mathrm{Ran}(T)$ is a subspace of Y and the kernel T is a subspace of $D(T)$. Indeed, we have $T0 = 0$, and thus $0 \in \mathrm{Ran}(T)$ and $0 \in \mathrm{Ker}(T)$. Moreover, if $y_1, y_2 \in \mathrm{Ran}(T)$ and $\alpha, \beta \in \mathbb{C}$ are arbitrary, then by definition there exist $x_1, x_2 \in D(T)$ such that $y_1 = Tx_1$ and $y_2 = Tx_2$. Because of this, the linearity of T implies

$$\alpha y_1 + \beta y_2 = \alpha T x_1 + \beta T x_2 = T(\alpha x_1 + \beta x_2) \in \mathrm{Ran}(T).$$

Thus $\mathrm{Ran}(T)$ is a subspace of Y. Moreover, if $x_1, x_2 \in \mathrm{Ker}(T)$ and $\alpha, \beta \in \mathbb{C}$ are arbitrary, we have

$$T(\alpha x_1 + \beta x_2) = \alpha T x_1 + \beta T x_2 = \alpha 0 + \beta 0 = 0,$$

which implies that $\alpha x_1 + \beta x_2 \in \mathrm{Ker}(T)$. This implies that $\mathrm{Ker}(T)$ is a subspace of $D(T)$.

For linear operators, the definition injectivity in Definition 2.24 coincides perfectly with its more general definition as the property that *distinct elements* $x_1 \neq x_2$ *are mapped to distinct elements* $Tx_1 \neq Tx_2$. Indeed, if the linear operator $T : D(T) \subset X \to Y$ has the property $\mathrm{Ker}(T) = \{0\}$ and if $x_1, x_2 \in D(T)$ are such that $Tx_1 = Tx_2$, then

$$T(x_1 - x_2) = Tx_1 - Tx_2 = 0 \quad \Rightarrow \quad x_1 - x_2 \in \mathrm{Ker}(T) = \{0\} \quad \Rightarrow \quad x_1 - x_2 = 0$$

which implies that $x_1 = x_2$. On the other hand, the injectivity of T in the more general sense implies that *only the zero element* $0 \in D(T)$ *can be mapped to the zero element* $0 \in Y$, which means that necessarily $\mathrm{Ker}(T) = \{0\}$.

Exercise 2.25. Prove that if $D(T) = X$, then $\mathrm{Ker}(T)$ is a closed subspace of X. One convenient way of doing this is to take an arbitrary sequence $(x_k)_{k=1}^{\infty} \subset \mathrm{Ker}(T)$ which converges in X, and to show that the limit of the sequence is in $\mathrm{Ker}(T)$. ◇

Definition 2.26. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces and let $T : D(T) \subset X \to Y$ be a linear operator. A mapping $S : \mathrm{Ran}(T) \subset Y \to X$ is an **inverse** of T if

$$\begin{aligned} STx &= x && \forall x \in D(T) \\ TSy &= y && \forall y \in \mathrm{Ran}(T). \end{aligned}$$

If such an S exists, then T is said to be **invertible**, and inverse S is denoted by $T^{-1} = S$.

The operator $T : D(T) \subset X \to Y$ is said to be **boundedly invertible** if T has an inverse $T^{-1} \in B(Y, X)$.

Our definition of bounded invertibility in particular requires that T^{-1} is defined on Y, and thus $\mathrm{Ran}(T) = Y$. There are differences in this terminology in the literature, and sometimes T is defined to be boundedly invertibly if its inverse T^{-1} is a bounded as an operator $T : \mathrm{Ran}(T) \subset Y \to X$. The following theorem shows that the inverse of a linear operator is always linear, and that every injective operator has a well-defined and unique inverse.

Theorem 2.27. *Let* $(X, \|\cdot\|_X)$ *and* $(Y, \|\cdot\|_Y)$ *be normed spaces and let* $T : D(T) \subset X \to Y$ *be a linear operator. Then the following hold.*

(a) *The inverse of an operator* T *is unique (if it exists).*

(b) *The inverse* T^{-1} *(if it exists) is a linear operator.*

(c) *The operator* T *has an inverse if and only if it is injective.*

Proof. **Part (a):** Let $S_1 : D(T) \subset Y \to X$ and $S_2 : D(T) \subset Y \to X$ be two inverses of T. For every $y \in \mathrm{Ran}(T)$ there exists $x \in D(T)$ such that $y = Tx$, and we have

$$S_1 y = S_1 T x = x = S_2 T x = S_2 y.$$

Thus $S_1 = S_2$.

Part (b): Left as an exercise.

Part (c): If the operator T is injective, for every $y \in \mathrm{Ran}(T)$ there exists a unique $x \in D(T)$ such that $y = Tx$. We can define a mapping $S : \mathrm{Ran}(T) \subset Y \to X$ such that for every $y \in \mathrm{Ran}(T)$ the value Sy is defined to be this unique $x \in X$ (this is often called the *algebraic inverse* of T). By definition we then have that if $y \in \mathrm{Ran}(T)$ and $x \in D(T)$ is such that $y = Tx$, then

$$TSy = Tx = y.$$

Moreover, if $x \in D(T)$, we also have $STx = x$, since $x \in D(T)$ is the unique element which maps to $Tx \in \mathrm{Ran}(T)$. Thus S is an inverse of T in the sense of Definition 2.26.

Assume now that T has an inverse $T^{-1} : \mathrm{Ran}(T) \subset Y \to X$. If $x \in \mathrm{Ker}(T)$ is arbitrary, then the properties of T^{-1} in Definition 2.26 and the linearity of T^{-1} (part (b)) imply that

$$x = S\underbrace{Tx}_{=0} = S0 = 0.$$

Since $x \in \mathrm{Ker}(T)$ was arbitrary, we have that $\mathrm{Ker}(T) = \{0\}$ and the operator T is injective. □

Exercise 2.28. Consider the shift operators S_r and S_l in Example 2.4. Show that S_r is injective but not surjective, and that S_l is surjective but not injective. Moreover, show that $S_r^{-1} = S_l$. ◇

Exercise 2.29. Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, and $(Z, \|\cdot\|_Z)$ be normed spaces and assume $S : D(S) \subset X \to Y$ and $T : Y \to Z$ are bijective. Show that the composition $ST : D(T) \subset X \to Z$ is well-defined and invertible, and that its inverse satisfies $(ST)^{-1} = T^{-1}S^{-1}$. ◇

Example 2.30. In this example we demonstrate that a bounded operator does not necessarily have a bounded inverse. Let $X = \ell^p(\mathbb{C})$ for some $p \geq 1$ and consider a **multiplication operator** $T \in B(X)$ defined so that

$$Tx = \left(\frac{x_k}{k}\right)_{k=1}^{\infty}, \qquad \forall x = (x_k)_{k=1}^{\infty} \in X.$$

Thus the application of the operator T on x multiplies each element x_k of the sequence $x = (x_k)_k$ with $1/k$. The multiplication operator is well-defined as an operator $T : X \to X$, since for every $x \in X = \ell^p(\mathbb{C})$ we have

$$\|Tx\|_p = \left(\sum_{k=1}^{\infty}\left|\frac{x_k}{k}\right|^p\right)^{\frac{1}{p}} \leq \left(\max_{k\in\mathbb{N}}\left\{\frac{1}{k^p}\right\}\sum_{k=1}^{\infty}|x_k|^p\right)^{\frac{1}{p}} = \left(\sum_{k=1}^{\infty}|x_k|^p\right)^{\frac{1}{p}} = \|x\|_p,$$

and thus $Tx \in X$ for all $x \in X$. It is a straightforward exercise to verify that T is linear, and the above estimate also shows that $T \in B(X)$ with $\|T\| \leq 1$. Moreover, the operator T is injective, since for any $x \in X$ the property $Tx = 0$ implies that $x_k/k = 0$ for all $k \in \mathbb{N}$, and thus $x_k = 0$ for all $k \in \mathbb{N}$ as well. However, the operator T is not surjective. In fact, our aim is to show that $\mathrm{Ran}(T)$ is a "proper" subspace of X (meaning that $\mathrm{Ran}(T) \neq X$), namely,

$$\mathrm{Ran}(T) = \{\, (y_k)_{k=1}^{\infty} \in X \mid (ky_k)_{k=1}^{\infty} \in X \,\}. \tag{2.5}$$

Indeed, every element $y \in \operatorname{Ran}(T)$ belongs to the space on the right-hand side of (2.5), since $y = Tx$ for some $x = (x_k)_{k=1}^\infty \in X$, and thus

$$(ky_k)_{k=1}^\infty = \left(k \cdot \frac{x_k}{k}\right)_{k=1}^\infty = (x_k)_{k=1}^\infty \in X.$$

On the other hand, if we take an arbitrary $y \in X$ such that $(ky_k)_{k=1}^\infty \in X$ and define sequence $x = (x_k)_{k=1}^\infty$ such that $x_k = ky_k$ for all $k \in \mathbb{N}$, then our assumptions imply that $x = (x_k)_k = (ky_k)_k \in X$, and

$$y = (y_k)_k = \left(\frac{ky_k}{k}\right)_k = \left(\frac{x_k}{k}\right)_k = Tx \in \operatorname{Ran}(T).$$

Thus $\operatorname{Ran}(T)$ is given by (2.5), and $\operatorname{Ran}(T) \neq X$.

Since $T \in B(X)$ is injective, it has a well-defined inverse $T^{-1} : \operatorname{Ran}(T) \subset X \to X$. It's quite easy to guess that the inverse operator should have the effect of multiplying every element y_k of a sequence with k, and indeed if we define $D(S) = \operatorname{Ran}(T)$ and

$$Sy = (ky_k)_{k=1}^\infty, \qquad \forall y = (y_k)_{k=1}^\infty \in D(S),$$

then for all $x \in X$ and $y \in \operatorname{Ran}(T)$ we have

$$\begin{aligned} STx &= S\left(\frac{x_k}{k}\right)_k = \left(k \cdot \frac{x_k}{k}\right)_k = (x_k)_k = x \\ TSy &= T\,(ky_k)_k = \left(\frac{ky_k}{k}\right)_k = (y_k)_k = y. \end{aligned}$$

Thus $T^{-1} = S$. ◇

Remark 2.31. The multiplication operators on $\ell^p(\mathbb{C})$ and $\ell^\infty(\mathbb{C})$, such as the one in Example 2.30, are often called **infinite diagonal matrices**. Indeed, if we interpret the sequences $x = (x_k)_{k=1}^\infty \in \ell^p(\mathbb{C})$ to be "infinite column vectors", then the operator T in Example 2.30 has the effect of multiplication of the vector x (from the left) with an "infinite" diagonal matrix with diagonal elements $1/k$ for $k \in \mathbb{N}$.

The following theorem shows that a bijection between two Banach spaces is always boundedly invertible. The assumption of the completeness of the spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ is important in this result (in fact, we could make any injective operator $T : D(T) \subset X \to Y$ bijective by simply redefining it as an operator $T : D(T) \subset X \to \operatorname{Ran}(T)$, since $(\operatorname{Ran}(T), \|\cdot\|_Y)$ is a normed space).

Theorem 2.32 (Bounded Inverse Theorem)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. Every bijective operator $T \in B(X, Y)$ has a bounded inverse $T^{-1} \in B(Y, X)$.*

Proof. The theorem is very important, but its proof is unfortunately outside the scope of our course. You can find it, for example, from [Rud87, Thm. 5.10], and it appears often together with the so-called *Closed Graph Theorem* and *Open Mapping Theorem* (these three results are in fact equivalent). □

The Bounded Inverse Theorem also has a more general version where the operator T is not necessarily a bounded operator, but instead a "*closed operator*". However, the above version is very suitable for our purposes. The Bounded Inverse Theorem is especially imporant in **spectral theory** (Chapter 4), which studies "*the ways in which an operator can fail to have a bounded inverse*".

Definition 2.33 (Isometric Isomorphism)**.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. An operator $T \in B(X,Y)$ is an **isometric isomorphism** if T is boundedly invertible and if

$$\|Tx\|_Y = \|x\|_X, \qquad \forall x \in X. \tag{2.6}$$

If an isometric isomorphism $T \in B(X,Y)$ exists, then the spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are said to be **isometrically isomorphic**.

The isometric isomorphism of $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ means that we can consider X and Y to be the "same normed space" by identifying the elements $x \in X$ and $y = Tx \in Y$. In particular, the isometric isomorphism between the spaces preserves the sizes of the elements ($\|y\|_Y = \|Tx\|_Y = \|x\|_X$) and the distances between two elements ($\|y_1 - y_2\|_Y = \|T(x_1 - x_2)\|_Y = \|x_1 - x_2\|_X$). However, it is important to note that the isometric isomorphism is allowed to lose any other structure of the space: For example, the spaces $(\mathbb{C}, |\cdot|)$ and $(\mathbb{R}^2, \|\cdot\|_2)$ are isometrically isomorphic (via the mapping $T(a+ib) = (a,b)^T \in \mathbb{R}^2$), but for example defining a multiplication between two elements is possible only on $\mathbb{C}$, and not on $\mathbb{R}^2$ [NS82, p. 258].

Exercise 2.34. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. Prove that if $T \in B(X,Y)$ is **isometric** in the sense that (2.6) holds, then T is an isometric isomorphism between $(X, \|\cdot\|_X)$ and $(\mathrm{Ran}(T), \|\cdot\|_Y)$. ◇

2.3 The Hahn–Banach Theorem

In this section we study the Hahn–Banach Theorem, which is a fundamentally important result in functional analysis. This result appears in the literature in many forms and has several useful corollaries, some of which we will also utilise in the next section when considering "dual spaces" in greater detail. The result deals with *extensions* of linear operators defined below.

Definition 2.35. Let X and Y be vector spaces and let $T : D(T) \subset X \to Y$ and $S : D(S) \subset X \to Y$ be linear operators. The operator S is an **extension** of T if $D(T) \subset D(S)$ and $Sx = Tx$ for all $x \in D(T)$. Conversely, in this situation T is a **restriction** of S.

Recall that a *linear functional* is a linear operator from a vector space X to the space of complex numbers $\mathbb{C}$. The main result of this section shows that any bounded linear functional ψ defined on a subspace Y of a normed space X can be extended to the full space X without increasing the norm of the functional. This theorem is the version of the Hahn–Banach Theorem "on a normed space". A more general (and classical) version of this result is also applicable in the situation where X is a vector space and the values of the functional ψ are bounded from above by a *seminorm* (see, e.g., [TL80, Sec. III.2], [Kre89, Sec. 4.2–4.3]). Note that in the case of linear functionals, it is customary to use slightly different notation than for other linear operators: linear functionals are often denoted, for example, with lower case letters $f, g : X \to \mathbb{C}$, or with Greek letters $\phi, \psi : X \to \mathbb{C}$, and in this case the parentheses are not removed from the notation, i.e., we denote "$f(x)$" or "$\psi(x)$" for $x \in X$.

Theorem 2.36 (The Hahn–Banach Theorem (on a Normed Space))**.** *Let* $(X, \|\cdot\|)$ *be a normed space and let* Y *be a subspace of* X *(considered with the same norm). Every bounded linear functional* $\psi \in B(Y, \mathbb{C})$ *has an extension* $\phi \in B(X, \mathbb{C})$ *satisfying*

$$\|\phi\|_{B(X,\mathbb{C})} = \|\psi\|_{B(Y,\mathbb{C})}.$$

The proof of the Hahn–Banach Theorem is presented separately in Section 2.5. It is definitely advanced material, and you are encouraged to study it, but *you don't need to memorise it*! The proof has a few quite interesting aspects: First of all, this is one instance (and perhaps the only one) where we encounter the *real* vector spaces on our course, since the standard proof of the result first verifies that the result is true for a real vector space X and a *real linear* functional $\psi : X \to \mathbb{R}$, and subsequently uses this to prove the generalised version of the theorem for complex vector spaces. Moreover, the proof makes use of *Zorn's Lemma* which is a fundamental mathematical building block in set theory (and in particular equivalent to the *Axiom Of Choice*). In our proof, Zorn's Lemma is used to show that among all the possible extensions of ψ, there is a "maximal" extension ϕ which is defined on the whole space X.

In the rest of this section we focus on two important consequences of the Hahn–Banach Theorem. The first such corollary is that for any nonzero element $x \in X$ of a normed space X, we can find a bounded linear functional ϕ with norm $\|\phi\| = 1$ such that the functional ϕ evaluated at x produces the norm of x, i.e., $\phi(x) = \|x\|$. In this case ϕ is called a **norming functional** of x. This fairly simple looking result is quite important in our study of the "dual space" $B(X, \mathbb{C})$ in the next section. We'll especially use Lemma 2.37 to show that the space $B(X, \mathbb{C})$ is always non-trivial (i.e., it always contains nonzero elements) whenever X itself is not trivial.

Lemma 2.37. *Let* $(X, \|\cdot\|)$ *be a normed space. For every* $x \in X$ *satisfying* $x \neq 0$ *there exists* $\phi_x \in B(X, \mathbb{C})$ *such that*

$$\phi_x(x) = \|x\|,$$

and $\|\phi_x\| = 1$*. In particular,*

$$\|x\| = \sup_{\|\phi\| \leq 1} |\phi(x)|.$$

Proof. Let $x \in X$, $x \neq 0$, be fixed and consider the subspace $Y = \{ \alpha x \mid \alpha \in \mathbb{C} \}$ of X. Define a linear functional $\psi_x : Y \to \mathbb{C}$ by $\psi_x(\alpha x) = \alpha\|x\|$ for all $\alpha \in \mathbb{C}$. Then for all $y = \alpha x \in Y$ we have

$$|\psi_x(y)| = |\alpha\|x\|| = |\alpha|\|x\| = \|\alpha x\| = \|y\|,$$

and thus $\psi_x \in B(Y, \mathbb{C})$ with norm $\|\psi_x\| = 1$. The Hahn–Banach Theorem implies that ψ_x has an extension $\phi_x \in B(X, \mathbb{C})$ satisfying $\|\phi_x\| = \|\psi_x\| = 1$. Moreover, since $x \in Y$ has the form $x = \alpha x$ with $\alpha = 1$, we have

$$\phi_x(x) = \psi_x(x) = 1 \cdot \|x\| = \|x\|,$$

as required.

To prove the last identity in the statement, we can first note that $|\phi(x)| \leq \|\phi\|\|x\|$ implies that $\sup_{\|\phi\|\leq 1}|\phi(x)| \leq \|x\|$, and since the norming functional $\phi_x \in B(X,\mathbb{C})$ satisfies $\|\phi_x\| = 1$, we also have

$$\sup_{\|\phi\|\leq 1} |\phi(x)| \geq |\phi_x(x)| = \|x\|. \qquad \square$$

Another very useful consequence of the Hahn–Banach Theorem is that we can use bounded linear functionals to "test the denseness" of a subspace Z of a normed space X. Indeed, the next lemma implies that we can prove that a subspace Z of X is dense in X by letting $\phi \in B(X,\mathbb{C})$ be an arbitrary functional such that $\phi(x) = 0$ for all $x \in Z$, and then proving that ϕ is then necessarily the zero functional.

Lemma 2.38. *The subspace Z of a normed space X. Then Z is dense in X if and only if for every $\phi \in B(X,\mathbb{C})$*

$$\phi(x) = 0, \quad \forall x \in Z \qquad \textit{only if} \quad \phi = 0. \tag{2.7}$$

Proof. To prove that "only if" part, assume that Z is dense in X. If $\phi \in B(X,\mathbb{C})$ is such that $\phi(x) = 0$ for all $x \in Z$, then the continuity of ϕ implies that $\phi(x) = 0$ also for all $x \in X$, and thus $\phi = 0$.

To prove the "if" part, assume that (2.7) holds. Assume on the contrary that Z is not dense in X. Then there exists $x_0 \in X$ and $\varepsilon > 0$ such that $\|x_0 - z\| \geq \varepsilon$ for all $z \in Z$. Our aim is to show that there exists $\phi \in B(X,\mathbb{C})$ such that $\phi(x) = 0$ for all $x \in Z$ but $\phi(x_0) \neq 0$, which is a contradiction with our assumption that (2.7) holds. To this end, we consider the subspace $Y = \{\, x + \alpha x_0 \mid x \in Z, \alpha \in \mathbb{C} \,\}$ of X and define a linear functional $\psi : Y \to \mathbb{C}$ so that $\psi(x + \alpha x_0) = \alpha$ for all $x \in Z$ and $\alpha \in \mathbb{C}$. It is easy to verify that ψ is linear. To show that $\psi \in B(Y,\mathbb{C})$ we first note that for all $x \in Z$ we have $\psi(x) = 0$ (corresponding to $\alpha = 0$). Moreover, if $x \in Z$ and $\alpha \neq 0$, then $-\alpha^{-1}x \in Z$ and thus

$$|\psi(x + \alpha x_0)| = |\alpha| = \frac{|\alpha|}{\varepsilon} \cdot \varepsilon \leq \frac{|\alpha|}{\varepsilon}\|\alpha^{-1}x + x_0\| = \frac{1}{\varepsilon}\|x + \alpha x_0\|.$$

Thus $\psi \in B(Y,\mathbb{C})$ with norm $\|\psi\| \leq 1/\varepsilon$. The Hahn-Banach theorem implies that ψ has an extension $\phi \in B(X,\mathbb{C})$. Since ϕ is an extension of ψ, we in particular have

$$\phi(x_0) = \psi(x_0) = \psi(0 + 1 \cdot x_0) = 1 \neq 0,$$

and for all $x \in X \subset Y$

$$\phi(x) = \psi(x) = \psi(x + 0 \cdot x_0) = 0.$$

Thus $\phi(x) = 0$ for all $x \in Z$, but $\phi \neq 0$ since $\phi(x_0) \neq 0$. Because of this, condition (2.7) does not hold, which is a contradiction with our original assumption. $\square$

2.4 The Dual Space

In this section we study the "dual space" of a normed space, which consists of the bounded linear functionals on the space. Since $(\mathbb{C}, |\cdot|)$ is a complete space, Theorem 2.23 especially implies that the dual space of a normed space $(X, \|\cdot\|)$ is always a Banach space.

Definition 2.39. Let $(X, \|\cdot\|_X)$ be a normed space. The **dual space** X' of X is the space of bounded linear functionals on X, i.e., $X' := B(X, \mathbb{C})$. The dual space is a Banach space $(X', \|\cdot\|_{X'})$ with the norm $\|\cdot\|_{X'} = \|\cdot\|_{B(X,\mathbb{C})}$.

As we already saw in the last section, linear functionals are often denoted slightly differently than other operators between normed spaces, for example, with lower case letters $f, g \in X'$, or with Greek letters $\phi, \psi \in X'$. Moreover, it is customary to keep the parentheses when writing these functions, such as in "$f(x)$" or "$\phi(x)$" for $x \in X$. Another very common notation in the case of dual spaces is the "**dual pairing**" $\langle x, \phi \rangle_{X,X'}$, which consists of a vector $x \in X$ and a functional $\phi \in X'$. This notation simply means that "the functional $\phi \in X'$ is evaluated at $x \in X$", i.e.

$$\langle x, \phi \rangle_{X,X'} := \phi(x), \qquad \forall x \in X.$$

Our investigation of the Hahn–Banach Theorem in the previous section gives us important information regarding dual spaces of normed spaces. In particular, the dual space X' of a non-trivial normed space X (i.e., $X \neq \{0\}$) *always* contain nonzero functionals (and are thus non-trivial). This is a consequence of Lemma 2.37, since if there exists $x \in X$ such that $x_0 \neq 0$, this result shows us there also exists $\phi \in B(X, \mathbb{C}) = X'$ such that $\phi(x_0) = \|x_0\| > 0$, and thus $\phi \neq 0$. Another way to interpret this same consequence is to note that the space X' **separates the points of** X in the sense that "if $x_1, x_2 \in X$ are such that $x_1 \neq x_2$, then there exists $\phi \in X'$ such that $\phi(x_1) \neq \phi(x_2)$" (this follows from choosing $x_0 = x_1 - x_2$ above).

We can now take look at particular examples of bounded linear functionals $\phi \in X'$ in the case of different spaces X.

Example 2.40. Let $\Omega \subset \mathbb{R}^n$ be a closed and bounded set, and let $X = C(\Omega)$ with the norm $\|\cdot\|_\infty$. In Examples 2.3 and 2.18 we studied the *point evaluation operator* T_{z_0}, which is defined by $T_{z_0} f = f(z_0) \in \mathbb{C}$ (where $z_0 \in \Omega$ is fixed). We proved that $T_{z_0} \in B(X, \mathbb{C})$, and thus T_{z_0} is a bounded linear functional on X, i.e., $T_{z_0} \in X'$. ◇

Exercise 2.41. Let $X = \ell^p(\mathbb{C})$ for some $p \geq 1$ or $p = \infty$. Show that if $n \in \mathbb{N}$ is fixed, then the operator $\phi_n : X \to \mathbb{C}$ defined by

$$\phi_n(x) = x_n, \qquad \forall x = (x_k)_{k\in\mathbb{N}} \in X$$

is a bounded linear functional on X, i.e., $\phi_n \in X'$. (You can note that this in a way a "sequence version" of the point evaluation in Example 2.40.) ◇

Example 2.42. Let $I \subset \mathbb{R}$ be an interval and let $X = L^p(I)$ for some $1 < p < \infty$. Moreover, if we choose $q = p/(p-1)$, then $1/p + 1/q = 1$. In this example we will show that for any $g \in L^q(I)$ we can define a bounded linear functional ϕ_g on X with the formula

$$\phi_g(f) = \int_I f(t)g(t)\,\mathrm{d}t, \qquad \forall f \in L^p(I).$$

To show this, let $g \in L^q(I)$ be fixed. The integral formula for $\phi_g(f)$ is well-defined for any $f \in L^p(I)$, since the Hölder's inequality in Theorem 1.50 implies that

$$\int_I |f(t)g(t)|\,\mathrm{d}t = \|fg\|_1 \leq \|f\|_p \|g\|_q < \infty.$$

Thus $fg \in L^1(I)$ and $\phi_g(f) = \int_I f(t)g(t)\,\mathrm{d}t \in \mathbb{C}$ for any $f \in X$. In order to show that $\phi_g : X \to \mathbb{C}$ is linear, let $\alpha_1, \alpha_2 \in \mathbb{C}$ and $f_1, f_2 \in X$ be arbitrary. Using the linearity of integration

$$\begin{aligned}\phi_g(\alpha_1 f_1 + \alpha_2 f_2) &= \int_I (\alpha_1 f_1 + \alpha_2 f_2)(t) g(t)\,\mathrm{d}t = \int_I (\alpha_1 f_1(t) + \alpha_2 f_2(t)) g(t)\,\mathrm{d}t \\ &= \alpha_1 \int_I f_1(t) g(t)\,\mathrm{d}t + \alpha_2 \int_I f_2(t) g(t)\,\mathrm{d}t = \alpha_1 \phi_g(f_1) + \alpha_2 \phi_g(f_2),\end{aligned}$$

which shows that ϕ_g is linear. Finally, to show that ϕ_g is bounded we can again use the Hölder's inequality in Theorem 1.50. Indeed, for any $f \in L^p(I)$ we can estimate

$$|\phi_g(f)| = \left| \int_I f(t) g(t)\,\mathrm{d}t \right| \le \int_I |f(t) g(t)|\,\mathrm{d}t = \|fg\|_1 \le \|g\|_q \|f\|_p.$$

This shows that ϕ_g is indeed bounded and $\|\phi_g\|_{B(X,\mathbb{C})} \le \|g\|_q$. Thus ϕ_g is a bounded linear functional on X, i.e., $\phi_g \in X'$, and $\|\phi_g\|_{X'} = \|\phi_g\|_{B(X,\mathbb{C})} \le \|g\|_q$. ◇

Exercise 2.43. Let $X = \ell^p(\mathbb{C})$ for some $1 < p < \infty$. Moreover, let $1 < q < \infty$ be such that $1/p + 1/q = 1$ and let $y \in \ell^q(\mathbb{C})$ be fixed. Show that the operator $\phi_y : X \to \mathbb{C}$ defined by

$$\phi_y(x) = \sum_{k=1}^{\infty} x_k y_k, \qquad \forall x = (x_k)_{k \in \mathbb{N}} \in X$$

is a bounded linear functional on X, i.e., $\phi_y \in X'$. (You can note that this is a "sequence version" of the linear functional defined in Example 2.42.) ◇

It is possible to explicitly characterise and identify dual spaces of certain normed spaces X. We can especially do this for the Lebesgue spaces $L^p(\Omega)$ and for the sequence spaces $\ell^p(\mathbb{C})$. In this we also customarily consider two normed space X and Y to be the same space if they are isometrically isomorphic (Definition 2.33), and similarly we denote "$X \subset Y$" whenever X is isometrically isomorphic with a subspace Z of Y. The theorem below characterises the duals of Lebesgues spaces $L^p(I)$ (except the space $L^\infty(I)$ whose characterisation is way beyong the scope of our course!). Note that in particular that the space $L^2(I)$ is its own dual, i.e., $(L^2(I))' = L^2(I)$. This is a fundamental property of **Hilbert spaces** which we will study later on this course.

Theorem 2.44. *Let $I \subset \mathbb{R}$ be an interval and let $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. Then the dual space of $L^p(I)$ is (isometrically isomorphic to) $L^q(I)$. Moreover, $(L^1(I))' = L^\infty(I)$ but $(L^\infty(I))' \ne L^1(I)$.*

Proof. We will only prove the statements of the theorem for the case $1 < p < \infty$. To this end, let $p > 1$, denote $X = L^p(I)$, and define $q = p/(p-1)$ (in which case $1/p + 1/q = 1$). In Example 2.42 we saw that for any $g \in L^q(I)$ we can define a bounded linear functional $\phi_g \in X'$ with the formula

$$\phi_g(f) = \int_I f(t) g(t)\,\mathrm{d}t, \qquad \forall f \in L^p(I).$$

A much deeper result, the so-called **Radon–Nikodym Theorem** implies that actually *every functional $\phi \in (L^p(I))'$ is of the form $\phi = \phi_g$ for some $g \in L^q(I)$* (see, e.g., [Rud87, Thm. 6.16]). This already gives a strong indication that the space $(L^p(I))'$ could be identified with $L^q(I)$. In the remaining part of the proof we will confirm this by showing that the mapping $T : L^q(I) \to (L^p(I))'$ which maps $g \in L^q(I)$ to $Tg = \phi_g \in X'$ is an isometric isomorphism.

We already saw that $\|Tg\|_{X'} = \|\phi_g\|_{X'} \leq \|g\|_q$, and thus $T \in B(L^q(I), X')$ with norm $\|T\| \leq \|g\|_q$. Moreover, the fact that T is surjective follows from the property that any functional $\phi \in X'$ is of the form $\phi = \phi_g = Tg$ for some $g \in L^q(I)$ (as implied by the Radon–Nikodym Theorem). Since $L^q(I)$ and X' are both Banach spaces, Exercise 2.34 shows that T is an isometric isomorphism if it satisfies the condition (2.6), i.e., if $\|\phi_g\|_{X'} = \|g\|_q$ for all $g \in L^q(I)$. To show this, let $g \in L^q(I)$ be arbitrary. Since we already know that $\|\phi_g\|_{X'} \leq \|g\|_q$, we only need to find a function $f \in L^p(I)$ such that $|\phi_g(f)| = \|f\|_p \|g\|_q$. To this end, we define $f(\cdot)$ such that $f(t) = \overline{g(t)}|g(t)|^{q/p-1}$ for (almost) all $t \in I$ such that $g(t) \neq 0$, and $f(t) = 0$ if $g(t) = 0$. We then have

$$\|f\|_p = \left(\int_I |f(t)|^p dt \right)^{\frac{1}{p}} = \left(\int_I |g(t)|^{p \cdot \frac{q}{p}} dt \right)^{\frac{1}{p}} = \|g\|_q^{q/p} < \infty,$$

and thus $f \in L^p(I)$. Moreover, $q/p + 1 = q(1/p + 1/q) = q$ shows that

$$\begin{aligned} |\phi_g(f)| &= \left| \int_I g(t)\overline{g(t)}|g(t)|^{q/p-1} \,\mathrm{d}t \right| = \int_I |g(t)|^{q/p+1} \,\mathrm{d}t \\ &= \int_I |g(t)|^q \,\mathrm{d}t = \|g\|_q^q = \|g\|_q^{q/p}\|g\|_q = \|f\|_p\|g\|_q. \end{aligned}$$

This completes the proof of the property $\|\phi_g\|_{X'} = \|g\|_q$, and thus T is an isometric isomorphism between $(L^p(I))'$ and $L^q(I)$. □

A similar result holds for the sequence spaces $\ell^p(\mathbb{C})$ for $p \geq 1$ and $p = \infty$.

Theorem 2.45. *Let $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. The dual space of $\ell^p(\mathbb{C})$ is (isometrically isomorphic to) $\ell^q(\mathbb{C})$. Moreover, $(\ell^1(\mathbb{C}))' = \ell^\infty(\mathbb{C})$ but $(\ell^\infty(\mathbb{C}))' \neq \ell^1(\mathbb{C})$.*

Proof. The proof is left as an exercise. □

The following result shows that a normed space X is always a subspace of its "second dual" $X'' = (X')'$ (again, more precisely in the sense that X is isometrically isomorphic to a subspace of X''). This result is another important consequence of the Hahn–Banach Theorem, and its corollary in Lemma 2.37.

Theorem 2.46. *Let $(X, \|\cdot\|)$ be a normed space. The **second dual** (or **bidual**) $X'' := (X')'$ of X has the property $X \subset X''$.*

Proof. For any fixed $x \in X$ we can define a functional $\psi_x : X' \to \mathbb{C}$ which *evaluates* every functional $\phi \in X'$ at the point x, i.e.,

$$\psi_x(\phi) = \phi(x) \in \mathbb{C}, \qquad \forall \phi \in X'.$$

Showing that ψ_x is a bounded linear functional on X', and that "**evaluation map**" $J : X \to X''$ defined by $Jx = \psi_x$ for all $x \in X$ is an isometric isomorphism between X and $\mathrm{Ran}(J) \subset X''$ are left as exercises (also see Exercise 2.34). □

The isometric mapping $J : X \to X''$ in the proof of Theorem 2.46 is an important tool and it has a special name.

Definition 2.47. Let $(X, \|\cdot\|)$ be a normed space. The **evaluation map** $J : X \to X''$ is defined by $Jx = \psi_x$ for all $x \in X$, where $\psi_x \in X''$ is such that

$$\psi_x(\phi) = \phi(x), \qquad \forall \phi \in X'.$$

Exercise 2.48. Show that evaluation map $J : X \to X''$ in Definition 2.47 is well-defined (meaning that $Jx \in X''$ for all $x \in X$) and that it is an isometric isomorphism between $(X, \|\cdot\|_X)$ and $(\operatorname{Ran}(J), \|\cdot\|_{X''})$. ◇

In a general situation, the space X'' can be strictly larger than the space X (this is for example the case with $L^1(\Omega)$ and $\ell^1(\mathbb{C})$). On the other hand, if X has the property that *every* functional $\psi \in X''$ is necessarily of the form $\psi = \psi_x$ for some $x \in X$, then the space is called **reflexive**. This property can be described equivalently as the property that the evaluation map $J : X \to X''$ is surjective. By Exercise 2.48 this immediately implies that a reflexive space is isometrically isomorphic with its second dual X'', but in fact the condition of reflexivity is strictly stronger, and it is in particular possible that a space X is isometrically isomorphic to X'' without being reflexive [Jam51]! Note that since X'' (as a dual of a normed space) is always complete, a normed space X with the property $X'' = X$ is always a Banach space.

Definition 2.49. A Banach space $(X, \|\cdot\|)$ is called **reflexive** if the evaluation map $J : X \to X''$ is surjective, i.e., $\operatorname{Ran}(J) = X''$.

In many ways, reflexive spaces are a very nicely behaving subclass of normed spaces, and there are several particular results which are only true on reflexive spaces. Among the Lebesgue spaces $L^p(\Omega)$ and the sequence spaces $\ell^p(\mathbb{C})$, the reflexive spaces are exactly those with $1 < p < \infty$.

Theorem 2.50. *Let $I \subset \mathbb{R}$ be an interval. The spaces $L^p(I)$ and $\ell^p(\mathbb{C})$ are reflexive if and only if $1 < p < \infty$.*

Proof. Assume $1 < p < \infty$. To show that $X = L^p(I)$ is reflexive, let $q = p/(p-1)$. The fairly simple idea in the proof is to use the isometric isomorphism between the Lebesgue space and its dual in the proof of Theorem 2.44 twice — once for $L^p(I)$ and once for $L^q(I)$ — but the full argument is quite delicate.

Let $T \in B(L^q(I), X')$ be the isometric isomorphism in the proof of Theorem 2.44. Then by definition T^{-1} maps every $\phi \in X'$ to a function $T^{-1}\phi \in L^q(I)$, and

$$\phi(f) = \int_I f(t)(T^{-1}\phi)(t)\,\mathrm{d}t, \qquad \forall \phi \in X',\ \forall f \in L^p(I). \tag{2.8}$$

Now let $\psi \in X''$ be arbitrary. Our aim is to show that ψ is in the range of the evaluation map $J : X \to X''$, which is equivalent to showing that ψ is of the form $\psi(\phi) = \phi(h)$ for some $h \in L^p(I)$ and for all $\phi \in X'$. If we define the composition map $\varphi = \psi \circ T$ (i.e., $\varphi(g) = \psi(Tg)$ for all $g \in L^q(I)$), then $\varphi \in B(L^q(I), \mathbb{C}) = (L^q(I))'$. The isometric isomorphism in the proof of Theorem 2.44 (now applied on $L^q(I)$) shows that there exists a function $h \in L^p(I)$ such that

$$\varphi(g) = \int_I g(t)h(t)\,\mathrm{d}t, \qquad \forall g \in L^q(I).$$

For an arbitrary $\phi \in X'$ the above identity together with (2.8) imply that

$$\psi(\phi) = \psi(TT^{-1}\phi) = \varphi(T^{-1}\phi) = \int_I (T^{-1}\phi)(t)h(t)\,\mathrm{d}t = \phi(h).$$

Thus ψ is indeed in the range of the evaluation map, and since $\psi \in X''$ was arbitrary, we have that $L^p(I)$ is reflexive by definition.

Proving that $\ell^p(\mathbb{C})$ is reflexive if $1 < p < \infty$ is left as an exercise. On the other hand, $(L^1(I))'' = (L^\infty(I))' \neq L^1(I)$ and $(\ell^1(\mathbb{C}))'' = (\ell^\infty(\mathbb{C}))' = c_0(\mathbb{C}) \neq \ell^1(\mathbb{C})$ immediately imply that $L^1(I)$ and $\ell^1(\mathbb{C})$ cannot be reflexive. In addition, the second duals of $L^\infty(I)$ and $\ell^\infty(\mathbb{C})$ are also strictly larger than the spaces themselves. □

Exercise 2.51. Prove that $\ell^p(\mathbb{C})$ is reflexive if $1 < p < \infty$. ◇

2.4.1 Completion of a Normed Space Revisited [Optional]

We already discussed completions of a normed space in Section 1.3. We can come back to this topic now that we are familiar with the concept of isometric isomorphisms. The full definition of a completion is stated in the following.

Definition 2.52 (Completion — Full Version)**.** Let $(X, \|\cdot\|_X)$ be a normed space and let $(Y, \|\cdot\|_X)$ be a Banach space. If X is isometrically isomorphic with a dense subspace of Y, then Y is a **completion** of X.

Our main motivation for returning to the topic of completions in connection with dual spaces is that the evaluation map $J : X \to X''$ gives us the means to construct the completion of any normed space. In fact, as the theorem shows, the completion of a normed space is always unique up to isometric isomorphisms between different completions. There are different ways to construct the completion, but the guiding principle is that the completion should contain the "missing" limits of the Cauchy sequences of X. However, as we already discussed before, identifying these limits may either be difficult (when there is no intuitive larger superspace for X) or straightforward (the space X is already a subspace of a larger complete space) based on the properties of the space. The dual spaces provide a way to guarantee the existence of these missing elements, since X is isometrically isomorphic with $\mathrm{Ran}(J)$, which is a subspace of the Banach space X'' containing the limits of all the Cauchy sequences in $\mathrm{Ran}(J)$ (or equivalently, in X).

Theorem 2.53. *Every normed linear space $(X, \|\cdot\|)$ has a completion which is unique up to isometric isomorphisms. A particular completion of $(X, \|\cdot\|)$ is given by*

$$\left(\overline{\mathrm{Ran}(J)}, \|\cdot\|_{X''}\right)$$

(the closure of $\mathrm{Ran}(J)$ in X'').

Proof. To prove that $(\overline{\mathrm{Ran}(J)}, \|\cdot\|_{X''})$ is a completion of X we first note that this space is complete because $\overline{\mathrm{Ran}(J)}$ is by definition a closed subspace of the Banach space X''. Moreover, as shown in the proof of Theorem 2.46, the space X is isometrically isomorphic to $\mathrm{Ran}(J) \subset \overline{\mathrm{Ran}(J)}$, which is a dense subspace of $\overline{\mathrm{Ran}(J)}$.

The uniqueness of the completion is based on the property that a bounded linear operator defined on a dense subspace of a Banach space can be extended uniquely to the full space. We do not present the proof here (see, e.g., [NS82, Sec. 5.9, Problem 6]). □

2.5 The Proof of the Hahn–Banach Theorem [Optional]

In this section we present the proof of the Hahn–Banach Theorem. The proof is also a bit long (by our standards), so to clarify its structure we divide it into three parts:

Part I: Constructing a particular extension $\psi_1 : Y_1 \to \mathbb{R}$ of $\psi : Y \to \mathbb{R}$ on $Y_1 \neq Y$.

Part II: Using Zorn's Lemma to show that the set of all extensions of ψ has a "maximal extension" defined on all of X (this maximal extension will be our ϕ).

Part III: Extending the result from the case of a real vector space X and $\psi : X \to \mathbb{R}$ to a complex vector space X and $\psi : X \to \mathbb{C}$.

Proof of Theorem 2.36.
Part I:
Assume first that X is a real vector space (i.e., Definition 1.1 is satisfied with $\mathbb{R}$ in place of $\mathbb{C}$) and that for a subspace $Y \subset X$, the mapping $\psi : Y \to \mathbb{R}$ is *real-linear* in the sense that $\psi(\alpha x + \beta y) = \alpha\psi(x) + \beta\psi(y)$ for all $\alpha, \beta \in \mathbb{R}$ and $x, y \in Y$. We can assume $Y \neq X$, since otherwise we can choose $\phi = \psi$. Because of this, we can choose $x_0 \in X$ such that $x_0 \notin Y$, and define a subspace $Y_1 = \{\, x + \lambda x_0 \mid x \in Y,\ \lambda \in \mathbb{R} \,\}$ of X. Moreover, for a fixed $\alpha \in \mathbb{R}$ we can define a mapping $\psi_1 : Y_1 \to \mathbb{R}$ by

$$\psi_1(x + \lambda x_0) = \psi(x) + \lambda\alpha, \qquad \forall x \in Y,\ \lambda \in \mathbb{R}.$$

It is an easy exercise to show that $\psi_1 : Y_1 \to \mathbb{R}$ is a (real-)linear functional. Moreover, ψ_1 is an extension of ψ, since for all $x \in Y$ we have

$$\psi_1(x) = \psi_1(x + 0x_0) = \psi(x) + 0\alpha = \psi(x).$$

In the next step, our aim is to show that it is possible to choose $\alpha \in \mathbb{R}$ in such a way that $\|\psi_1\| = \|\psi\|$ (here and in the rest of the proof we denote $\|\psi\| = \|\psi\|_{B(Y,\mathbb{C})}$, $\|\psi_1\| = \|\psi_1\|_{B(Y_1,\mathbb{C})}$, and so on). Since ψ_1 is an extension of ψ, we always have $\|\psi_1\| \geq \|\psi\|$ because

$$\|\psi_1\| = \sup_{x \in Y_1} |\psi_1(x)| \geq \sup_{x \in Y} |\psi_1(x)| = \sup_{x \in Y} |\psi(x)| = \|\psi\|.$$

Thus (by definition of ψ_1) it is sufficient to show that $|\psi(x) + \lambda\alpha| \leq \|\psi\|\|x + \lambda x_0\|$ for all $x \in Y$ and $\lambda \in \mathbb{R} \setminus \{0\}$ (note that for $\lambda = 0$ the inequality holds by assumption). Moreover, choosing $x = -\lambda y \in Y$ for $y \in Y$ in the previous inequality, we arrive at another equivalent condition for $\alpha \in \mathbb{R}$, namely,

$$|\psi(-\lambda y) + \lambda\alpha| \leq \|\psi\|\|-\lambda y + \lambda x_0\|, \qquad \forall y \in Y, \lambda \in \mathbb{R} \setminus \{0\}.$$

We can modify this condition using the linearity of ψ. Indeed, for all $y \in Y$ and $\lambda \in \mathbb{R} \setminus \{0\}$ we have

$$\begin{aligned} & |\psi(-\lambda y) + \lambda\alpha| \leq \|\psi\| \|-\lambda y + \lambda x_0\| \\ \Leftrightarrow \quad & |\lambda||-\psi(y) + \alpha| \leq |-\lambda| \|\psi\| \|y - x_0\| \\ \Leftrightarrow \quad & |-\psi(y) + \alpha| \leq \|\psi\| \|y - x_0\| \\ \Leftrightarrow \quad & \psi(y) - \|\psi\| \|y - x_0\| \leq \alpha \leq \psi(y) + \|\psi\| \|y - x_0\|. \end{aligned}$$

Because the last condition is required to hold for all $y \in Y$, we can see that our choice of $\alpha \in \mathbb{R}$ is possible if (and only if) all the real intervals $[\psi(y) - \|\psi\| \|y - x_0\|, \psi(y) + \|\psi\| \|y - x_0\|] \subset \mathbb{R}$ for different $y \in Y$ contain at least one common point. This is satisfied, if all of the upper limits of these different intervals are greater than equal to all of the lower limits of these intervals, i.e., if we have

$$\psi(y) - \|\psi\| \|y - x_0\| \leq \psi(x) + \|\psi\| \|x - x_0\|, \qquad \forall x, y \in Y \tag{2.9}$$

But for arbitrary $x, y \in Y$ we can estimate

$$\begin{aligned} \psi(y) - \psi(x) = \psi(y - x) \leq \|\psi\| \|y - x\| &= \|\psi\| \|y - x_0 + x_0 - x\| \\ &\leq \|\psi\| \|y - x_0\| + \|\psi\| \|x_0 - x\| \end{aligned}$$

which immediately implies that (2.9) holds. This completes our proof that it is possible to choose $\alpha \in \mathbb{R}$ in such a way that the extended functional $\psi_1 \in B(Y_1, \mathbb{R})$ satisfies $\|\psi_1\| = \|\psi\|$.

Part II:

If $Y_1 = X$, our proof is complete, and the desired extension ϕ is ψ_1. On the other hand, if $Y_1 \neq X$, we could repeat our earlier process to find a functional ψ_2 defined on a larger subspace Y_2 of X. However, in the case of general infinite-dimensional vector spaces, repeating this process (possibly) infinite number of times **would not** necessarily lead to an extension ϕ defined on all of X! Instead, we can prove the existence of ϕ using Zorn's Lemma:

Lemma 2.54 (Zorn's Lemma)**.** *Let P be a nonempty partially ordered set with the property that every completely ordered subset of P has an upper bound in P. Then P contains at least one maximal element.*

To apply this result, our aim is to set up a partial order on the set of all extensions of ψ with the same norm, and define the partial ordering so that $\varphi_1 \prec \varphi_2$ if φ_2 is an extension of φ_1. Our motivation is that the "maximal element" in Zorn's Lemma will then be our functional ϕ. More precisely, let P be the set of all $\varphi : D(\varphi) \subset X \to \mathbb{R}$ which are extensions of $\psi : Y \to \mathbb{R}$ (i.e., $Y \subset D(\varphi) \subset X$ and $\varphi(x) = \psi(x)$ for all $x \in Y$) with $D(\varphi) \neq Y$ and which satisfy $\|\varphi\| = \|\psi\|$. For $\varphi_1, \varphi_2 \in P$ we define that $\varphi_1 \prec \varphi_2$ if $D(\varphi_1) \subset D(\varphi_2)$ and φ_2 is an extension of φ_1. This relation $\prec$ defines a partial ordering on P and P is not empty, since the extension $\psi_1 : Y_1 \subset X \to \mathbb{R}$ we constructed above satisfies $\psi_1 \in P$.

To apply Zorn's Lemma, we need to show that "every completely ordered subset of P has an upper bound". To this end, let Q be a completely ordered subset of P (meaning that for any $\varphi_1, \varphi_2 \in Q$ we have either $\varphi_1 \prec \varphi_2$ or $\varphi_2 \prec \varphi_1$). Our aim is to show that the set Q has an upper bound $\varphi \in P$, meaning that $\varphi_1 \prec \varphi$ for all

$\varphi_1 \in Q$. Due to our definitions, $\varphi : D(\varphi) \subset X \to \mathbb{R}$ is required to be a functional which satisfies $\|\varphi\| = \|\psi\|$ and which is an extension of all functionals φ_1 in Q. To this end, define $D(\varphi)$ to be the union of all domains $D(\varphi_1)$ for different $\varphi_1 \in Q$, which is defined as

$$D(\varphi) = \bigcup_{\varphi_1 \in Q} D(\varphi_1) := \{\, x \in X \mid x \in D(\varphi_1) \text{ for some } \varphi_1 \in Q \,\}.$$

The fact that $D(\varphi)$ is a subspace of X is verified as an exercise. We can now define a functional $\varphi : D(\varphi) \subset X \to \mathbb{R}$ so that for every $x \in D(\varphi)$ we have $\varphi(x) = \varphi_1(x)$ if $x \in D(\varphi_1)$. Showing that the value $\varphi(x)$ is uniquely defined for all $x \in D(\varphi)$ and that φ is indeed a (real-)linear functional satisfying $\|\varphi\| = \|\psi\|$ are left as exercises. Since every $\varphi_1 \in Q$ is an extension of ψ, we in particular have that $Y \subset D(\varphi)$ and $\varphi(x) = \psi(x)$ for $x \in Y$. Thus φ is indeed an extension of ψ, and we have $\varphi \in P$. Moreover, we can similarly deduce that φ is an extension of every functional in Q, i.e., $\varphi_1 \prec \varphi$ for all $\varphi_1 \in Q$, and thus φ is an "upper bound" of Q.

Zorn's Lemma now implies that the set P must have (at least one) maximal element $\varphi \in P$, meaning that $\varphi : D(\varphi) \subset X \to \mathbb{R}$ is an extension of every $\varphi_1 \in P$, and therefore also an extension of our original functional $\psi : Y \to \mathbb{R}$. It remains to argue that necessarily $D(\varphi) = X$. However, if we had $D(\varphi) \neq X$, then we could further extend φ (exactly as we constructed ψ_1) to arrive at a functional $\tilde{\varphi}$ which is an extension of both ψ and φ, and $D(\varphi) \subsetneq D(\tilde{\varphi})$. However, this would contradict the property that φ is "maximal", since $\tilde{\varphi} \not\prec \varphi$. Thus $D(\varphi) = X$ and $\varphi : X \to \mathbb{R}$ satisfies the properties in the theorem.

<u>Part III:</u>
Now consider the case where X is a complex vector space and $\psi \in B(Y, \mathbb{C})$. We first note that both X is *also* a real vector space, since it trivially satisfies the conditions of Definition 1.1 for all scalars in $\mathbb{R}$, and similarly Y is a also subspace of the real vector space X. As an exercise (Exercise 2.55) you will show that $\varphi \in B(Z, \mathbb{C})$ is a (complex-)linear functional on the complex vector space Z if and only if it has the form

$$\varphi(x) = \varphi_R(x) - i\varphi_R(ix), \qquad \forall x \in Z,$$

where $\varphi_R \in B(Z, \mathbb{R})$ is a real-linear functional on the real vector space Z. Furthermore, you will show that $\|\varphi\| = \|\varphi_R\|$. These results imply that our original functional $\psi \in B(Y, \mathbb{C})$ satisfies $\psi(x) = \psi_R(x) - i\psi_R(ix)$ for all $x \in Y$, where $\psi_R \in B(Y, \mathbb{R})$ is a real-linear functional satisfying $\|\psi_R\| = \|\psi\|$. By Parts I and II, ψ_R has an extension $\phi_R \in B(X, \mathbb{R})$ with norm $\|\phi_R\| = \|\psi_R\|$. Thus if we define a linear functional $\phi : X \to \mathbb{C}$ such that $\phi(x) = \phi_R(x) - i\phi_R(ix)$ for all $x \in X$, then for all $x \in Y$ we have

$$\phi(x) = \phi_R(x) - i\phi_R(ix) = \psi_R(x) - i\psi_R(ix) = \psi(x),$$

and thus $\phi : X \to \mathbb{C}$ is an extension of ψ. Finally, since $\phi_R \in B(X, \mathbb{R})$ implies $\phi \in B(X, \mathbb{C})$ and

$$\|\phi\| = \|\phi_R\| = \|\psi_R\| = \|\psi\|$$

our proof is complete. □

Exercise 2.55. Let X be a complex vector space. Prove that $\varphi \in B(X, \mathbb{C})$ is a linear functional if and only if there exists a (real-)linear functional $\varphi_R \in B(X, \mathbb{C})$ such that

$$\varphi(x) = \varphi_R(x) - i\varphi_R(ix), \qquad \forall x \in X.$$

Also prove that $\|\varphi_R\| = \|\varphi\|$. ◇

Exercise 2.56. Consider the completely ordered subset Q of P and the mapping $\varphi \in P$ in the proof of the Hahn–Banach Theorem. Prove that $D(\varphi)$ is a subspace of the (real) vector space X. Moreover, prove that the values of $\varphi : D(\varphi) \subset X \to \mathbb{R}$ are uniquely defined, and that $\varphi \in B(D(\varphi), \mathbb{R})$ is a (real-)linear functional satisfying $\|\varphi\| = \|\psi\|$. ◇

3. Hilbert Spaces

In this chapter we take a look at a special class of Banach spaces having additional structure in the form of an **inner product**, which is closely connected to the norm of the space. In addition to the *sizes* and *distances* (determined by the norm), the inner product allows us to also consider *angles* between two elements of vector space, and in particular define **orthogonality** of vectors. More generally as well, on inner product spaces we tend to meet various *geometric* concepts such as the Pythagorean theorem and similar identities, orthogonal complements, and orthogonal projections onto subspaces.

Definition 3.1. Let X be a vector space. An **inner product** on X is a function $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{C}$ so that

(a) $\langle x,x\rangle \geq 0$ and $\langle x,x\rangle = 0$ if and only if $x = 0$

(b) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x,z\rangle + \beta\langle y,z\rangle$ for all $x,y,z \in X$ and $\alpha,\beta \in \mathbb{C}$

(c) $\langle x,y\rangle = \overline{\langle y,x\rangle}$ for all $x,y \in X$

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

Sometimes we use the notation $\langle\cdot,\cdot\rangle_X$ to distinguish the inner product on the space X from different inner products. The definition of an inner product has a similar nature as the definition of a norm in Definition 1.19. In fact, every inner product $\langle\cdot,\cdot\rangle$ defines a norm, and therefore every inner product space $(X,\langle\cdot,\cdot\rangle)$ is also a normed space.

Definition 3.2. Let $(X,\langle\cdot,\cdot\rangle)$ be an inner product space. The **norm $\|\cdot\|$ induced by the inner product** is defined by

$$\|x\| = \sqrt{\langle x,x\rangle} \qquad \forall x \in X.$$

If we do not state otherwise, we always by default consider the inner product space $(X,\langle\cdot,\cdot\rangle)$ also to be a normed space *with the norm induced by the inner product* (as opposed to X being equipped with some other norm).

Every inner product defines a norm, but the converse is not true and only a relatively small portion of normed spaces are inner product spaces. This will be demonstrated in the examples in this chapter, but before looking more closely at concrete inner product spaces, we will first justify that the norm induced by the inner product is indeed a norm. This follows fairly directly (using exactly the same proof as in the case of Euclidean spaces) from the following generalised Cauchy–Schwarz Inequality, which holds for all inner products.

Lemma 3.3 (Cauchy–Schwarz Inequality)**.** *Let* $(X, \langle\cdot,\cdot\rangle)$ *be an inner product space. Then*

$$|\langle x, y\rangle| \le \|x\|\|y\| \qquad \forall x, y \in X,$$

where $\|\cdot\|$ *is the norm induced by the inner product, i.e,.* $\|z\| = \sqrt{\langle z, z\rangle}$ *for* $z \in X$*.*

Proof. If $\langle y, y\rangle = 0$, then $y = 0$ and the claim holds. Assume now that $y \neq 0$. For every $\alpha \in \mathbb{C}$ the properties of the inner product imply

$$\begin{aligned} 0 &\le \langle x - \alpha y, x - \alpha y\rangle = \langle x, x\rangle - \alpha\langle y, x\rangle - \overline{\alpha}\langle x, y\rangle + \alpha\overline{\alpha}\langle y, y\rangle \\ &= \|x\|^2 - \alpha\langle y, x\rangle - \overline{\alpha\langle y, x\rangle} + |\alpha|^2\|y\|^2 \\ &= \|x\|^2 - 2\,\mathrm{Re}\Big(\alpha\langle y, x\rangle\Big) + |\alpha|^2\|y\|^2. \end{aligned}$$

Since $\|y\|^2 = \langle y, y\rangle \neq 0$, we can choose $\alpha = \langle x, y\rangle/\|y\|^2 \in \mathbb{C}$. Then $\alpha\langle y, x\rangle = |\langle x, y\rangle|^2/\|y\|^2 \in \mathbb{R}$, and therefore the above inequality implies

$$\begin{aligned} 0 &\le \|x\|^2 - 2\,\mathrm{Re}\Big(\alpha\langle y, x\rangle\Big) + |\alpha|^2\|y\|^2 \\ &= \|x\|^2 - 2\frac{|\langle x, y\rangle|^2}{\|y\|^2} + \frac{|\langle x, y\rangle|^2}{\|y\|^4}\|y\|^2 \\ &= \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}. \end{aligned}$$

This immediately implies $|\langle x, y\rangle|^2 \le \|x\|^2\|y\|^2$. □

Exercise 3.4. Prove that the norm induced by the inner product on $(X, \langle\cdot,\cdot\rangle)$ is a norm on X. ◇

Exercise 3.5. Prove that two vectors $x, y \in X$ with $x \neq 0$ satisfy the Cauchy–Schwarz Inequality in Lemma 3.3 as an equality, i.e., $|\langle x, y\rangle| = \|x\|\|y\|$, if and only if $y = \alpha x$ for some $\alpha \in \mathbb{C}$. (Hint: In the "only if"-part, you can prove that $\|y - \alpha x\|^2 = 0$ when $\alpha = e^{i\theta}\|y\|/\|x\|$ where $\theta \in [0, 2\pi]$ is chosen in a suitable way). ◇

Exercise 3.6. Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Prove that for every fixed $y \in X$, the mappings $\langle\cdot, y\rangle : X \to \mathbb{C}$ and $\langle y, \cdot\rangle : X \to \mathbb{C}$ are linear functionals. Moreover, prove that the mapping $\langle\cdot,\cdot\rangle : X \to X \to \mathbb{C}$ is continuous when the space $X \times X$ is equipped with the norm $\|\cdot\|_{X\times X}$ satisfying $\|(x, y)\|^2_{X\times X} = \|x\|^2_X + \|y\|^2_X$ for all $x, y \in X$. Note that the norm on $X \times X$ is induced by the inner product $\langle(x_1, y_1), (x_2, y_2)\rangle_{X\times X} = \langle x_1, y_1\rangle_X + \langle x_2, y_2\rangle_X$. ◇

When an inner product space is complete (i.e., a Banach space) with respect to the norm induced by the inner product, it is called a **Hilbert space**. Alternatively, a Hilbert space is a Banach space whose norm is induced by an inner product.

Definition 3.7. An inner product space $(X, \langle\cdot,\cdot\rangle)$ is a **Hilbert space** if it is complete with respect to the norm induced by the inner product.

Our primary examples of inner product and Hilbert spaces come from the Lebesgue spaces $L^p(\Omega)$ and the sequence spaces $\ell^p(\mathbb{C})$. In both cases, the spaces are Hilbert spaces precisely for the exponent $p = 2$. As a further point of reference we can also immediately note that on the Euclidean spaces the Euclidean norm $\|\cdot\|_2$ is induced by the inner product $\langle x, y\rangle = \sum_{k=1}^n x_k \overline{y_k}$ for $x, y \in \mathbb{C}^n$. Since $\mathbb{C}^n$ and $\mathbb{R}^n$ are complete with respect to the Euclidean norm, both of these spaces are also Hilbert spaces.

Example 3.8 (Function Spaces)**.** Let $\Omega \subset \mathbb{R}^n$ be a closed and bounded set. We can define an inner product on the space $X = C(\Omega)$ by

$$\langle f, g\rangle = \int_\Omega f(z)\overline{g(z)}\,\mathrm{d}z, \qquad \forall f, g \in C(\Omega). \tag{3.1}$$

Since $z \mapsto f(z)\overline{g(z)}$ is uniformly continuous, the inner product $\langle f, g\rangle$ is well-defined for all $f, g \in X$. The three axioms of the inner product can be verified in the following way. For every $f \in C(\Omega)$ we have

$$\langle f, f\rangle = \int_\Omega f(z)\overline{f(z)}\,\mathrm{d}z = \int_\Omega |f(z)|^2\,\mathrm{d}z = \|f\|_2^2 \geq 0,$$

and the properties of the $\|\cdot\|_2$-norm on $C(\Omega)$ imply that $\langle f, f\rangle = \|f\|_2^2 = 0$ if and only if $f = 0$. Thus the first axiom is satisfied. On the other hand, if $f, g, h \in C(\Omega)$ and $\alpha, \beta \in \mathbb{C}$ are arbitrary, we have

$$\begin{aligned}\langle \alpha f + \beta g, h\rangle &= \int_\Omega (\alpha f(z) + \beta g(z))\overline{h(z)}\,\mathrm{d}z \\ &= \alpha \int_\Omega f(z)\overline{h(z)}\,\mathrm{d}z + \beta \int_\Omega g(z)\overline{h(z)}\,\mathrm{d}z = \alpha\langle f, h\rangle + \beta\langle g, h\rangle.\end{aligned}$$

Finally, for all $f, g \in C(\Omega)$ we have

$$\langle f, g\rangle = \int_\Omega f(z)\overline{g(z)}\,\mathrm{d}z = \int_\Omega \overline{g(z)\overline{f(z)}}\,\mathrm{d}z = \overline{\int_\Omega g(z)\overline{f(z)}\,\mathrm{d}z} = \overline{\langle g, f\rangle}.$$

Thus (3.1) is an inner product, and $(C(\Omega), \langle\cdot,\cdot\rangle)$ is an inner product space. However, this space is not a Hilbert space, since as we saw in Section 1.2.1, the $C(\Omega)$ is not complete with respect to the norm $\|\cdot\|_2$. But it is perhaps not surprising that the *completion* of this normed space, namely $L^2(\Omega)$, is in fact a Hilbert space!

Indeed, if $\Omega \subset \mathbb{R}^n$ is an open or closed set, we can define an inner product on $X = L^2(\Omega)$ similarly as above, by

$$\langle f, g\rangle = \int_\Omega f(z)\overline{g(z)}\,\mathrm{d}z, \qquad \forall f, g \in L^2(\Omega). \tag{3.2}$$

The fact that the inner product $\langle f, g\rangle$ is well-defined for all $f, g \in L^2(\Omega)$ follows directly from the Hölder inequality (Theorem 1.50) with $p = q = 2$, which implies that

$$\int_\Omega |f(z)\overline{g}(z)|\,\mathrm{d}z \leq \|f\|_2\|\overline{g}\|_2 = \|f\|_2\|g\|_2 < \infty,$$

which implies that $z \mapsto f(z)\overline{g(z)} \in L^1(\Omega)$ and the Lebesgue integral in (3.2) exists. Similarly as above, we again have $\langle f, f\rangle = \|f\|_2^2$ for all $f \in L^2(\Omega)$, which in particular implies that $\langle f, f\rangle \geq 0$ and $\langle f, f\rangle = 0$ if and only if $f = 0$ (in $L^2(\Omega)$). The other two axioms of Definition 3.1 can be verified exactly as in the case of $C(\Omega)$. Thus $\langle\cdot,\cdot\rangle$ in (3.2) is indeed an inner product and $(L^2(\Omega), \langle\cdot,\cdot\rangle)$ is an inner product space. Since the inner product induces the natural norm $\|\cdot\|_2$ on $L^2(\Omega)$, and the space $L^2(\Omega)$ is complete with respect to this norm, $L^2(\Omega)$ is a Hilbert space.

Of all the spaces $L^p(\Omega)$ with different $p \geq 1$ or $p = \infty$, only the space $L^2(\Omega)$ is a Hilbert space. It is possible to define inner products on different spaces as well (in particular, $L^p(\Omega) \subset L^2(\Omega)$ if $\Omega \subset \mathbb{R}^n$ is compact and $p > 2$ or $p = \infty$), but these spaces are not complete normed spaces. Moreover, it is possible to define other inner products on the space $L^2(\Omega)$. The most trivial way is through scaling, i.e., by defining $\langle f, g\rangle_{new} = \alpha\langle f, g\rangle$ for some $\alpha > 0$. More generally, if $w : \Omega \to \mathbb{R}$ is a continuous function such that there exist $w_0, w_1 > 0$ such that $0 < w_0 \leq w(z) \leq w_1$, then

$$\langle f, g\rangle_w = \int_\Omega f(z)\overline{g(z)}w(z)\,\mathrm{d}z, \qquad \forall f, g \in L^2(\Omega)$$

defines an inner product on $L^2(\Omega)$. This is often called a **weighted inner product**, and the function $w(\cdot)$ is its "weight (function)". It is straighforward to show that $(L^2(\Omega), \langle\cdot,\cdot\rangle_w)$ is also a Hilbert space due to the fact that the norm induced by $\langle\cdot,\cdot\rangle_w$ is *equivalent* with the natural norm on $L^2(\Omega)$ in the sense of Definition 1.34. ◇

Example 3.9 (Sequence Spaces)**.** We can define an inner product on $\ell^2(\mathbb{C})$ using the formula

$$\langle x, y\rangle = \sum_{k=1}^{\infty} x_k\overline{y_k},$$

for all $x = (x_k)_{k=1}^\infty \in \ell^2(\mathbb{C})$ and $y = (y_k)_{k=1}^\infty$. Indeed, the inner product is well-defined, since the Hölder inequality in (with $p = q = 2$) in Theorem 1.22 implies that

$$\sum_{k=1}^{\infty}|x_k\overline{y_k}| = \sum_{k=1}^{\infty}|x_k y_k| \leq \|x\|_2\|y\|_2 < \infty$$

and thus the series in the definition of $\langle x, y\rangle$ is absolutely convergent. To prove that $\langle\cdot,\cdot\rangle$ is really an inner product on $\ell^2(\mathbb{C})$, we first note that if $x \in \ell^2(\mathbb{C})$, then

$$\langle x, x\rangle = \sum_{k=1}^{\infty} x_k\overline{x_k} = \sum_{k=1}^{\infty}|x_k|^2 = \|x\|_2^2 \geq 0.$$

Thus $\langle\cdot,\cdot\rangle$ (provided that it is an inner product) induces the norm $\|\cdot\|_2$ of $\ell^2(\mathbb{C})$. In particular, the properties of the norm also imply that $\langle x, x\rangle = 0$ if and only if $x = 0$.

If $\alpha, \beta \in \mathbb{C}$ and $x, y, z \in \ell^2(\mathbb{C})$ are arbitrary, then

$$\langle \alpha x + \beta y, z\rangle = \sum_{k=1}^{\infty}(\alpha x_k + \beta y_k)\overline{z_k} = \alpha\sum_{k=1}^{\infty} x_k\overline{z_k} + \beta\sum_{k=1}^{\infty} y_k\overline{z_k} = \alpha\langle x, z\rangle + \beta\langle y, z\rangle,$$

and

$$\langle x, y\rangle = \sum_{k=1}^{\infty} x_k\overline{y_k} = \sum_{k=1}^{\infty}\overline{\overline{x_k}y_k} = \overline{\sum_{k=1}^{\infty}\overline{x_k}y_k} = \langle y, x\rangle.$$

Thus $\langle\cdot,\cdot\rangle$ satisfies the axioms of an inner product. Moreover, since this inner product induces the norm $\|\cdot\|_2$, and since $(\ell^2(\mathbb{C}), \|\cdot\|_2)$ is a complete normed space, we have that $(\ell^2(\mathbb{C}), \langle\cdot,\cdot\rangle)$ is a Hilbert space.

Similarly as in the case of the Lebesgue spaces, the space $\ell^2(\mathbb{C})$ is the only Hilbert space among the spaces of p-summable sequences $\ell^p(\mathbb{C})$ for $1 \leq p < \infty$ and $p = \infty$. ◇

3.1 Orthogonality and The Riesz Representation Theorem

In this section we continue to study the special properties of Hilbert spaces. In particular, the inner product allows us to consider *angles* and *orthogonality* between vectors and subspaces. As the main results of this section we will see that a Hilbert space can be decomposed into a sum of any closed subspace and its "orthogonal complement". This strong property also leads to the remarkable Riesz Representation Theorem which presents a simple characterisation of all bounded linear functionals on a Hilbert space!

On the Euclidean space $\mathbb{R}^n$, the inner product (which is exactly the "dot product") can be used to compute the *angle* between two vectors $x, y \in \mathbb{R}^n$ using the *Law of Cosines*, which states that the smallest angle θ between $x \neq 0$ and $y \neq 0$ satisfies

$$\cos\theta = \frac{\langle x, y\rangle}{\|x\|\|y\|}. \tag{3.3}$$

Motivated by this, we can also use the formula (3.3) to define the **angle** θ between two vectors $x, y \in X$ on a Hilbert space X whenever the inner product between these vectors satisfies $\langle x, y\rangle \in \mathbb{R}$. Moreover, in the more general case where $\langle x, y\rangle \in \mathbb{C}$, the angle between x and y can sometimes be defined by replacing $\langle x, y\rangle$ with $\mathrm{Re}\langle x, y\rangle \in \mathbb{R}$ on the right-hand side of (3.3). There are two especially important cases of angles between vectors, namely, when the vectors are **orthogonal** (when the angle between them is $\theta = \pi/2$) or when they are **parallel** (or **collinear**), i.e., when $y = \alpha x$ for some $\alpha \in \mathbb{C}$. Since $\cos(\pi/2) = 0$, in view of (3.3) the orthogonality of two vectors $x, y \in X$ can be defined using the inner product as the condition $\langle x, y\rangle = 0$.

Definition 3.10. Let X be an inner product space.

- Elements $x, y \in X$ are **orthogonal** if $\langle x, y\rangle = 0$. In this case we denote $x \perp y$.
- Subsets $M \subset X$ and $N \subset X$ are **orthogonal** if $\langle x, y\rangle = 0$ for all $x \in M$ and $y \in N$. In this case we denote $M \perp N$.
- For a subset $M \subset X$, the **orthogonal complement** $M^\perp$ of M is defined as

$$M^\perp = \left\{ y \in X \;\middle|\; \langle x, y\rangle = 0 \;\; \forall x \in M \right\}.$$

Exercise 3.11. Let X be an inner product space and let $M \subset X$. Show that $M^\perp$ is a closed subspace of X. Moreover, show that $\overline{M}^\perp = M^\perp$ ($\overline{M}$ denotes the closure of M). (Hint: For every $x \in \overline{M}$ there exists $(x_k)_{k=1}^\infty \subset M$ such that $x_k \to x$ as $k \to \infty$). ◇

Exercise 3.12. Let M be a subspace of a Hilbert space X. Prove $M \cap M^\perp = \{0\}$. ◇

Exercise 3.13. Let X be an inner product space. Prove that for every $y \in X$ we have

$$\|y\| = \sup_{\|x\| \le 1} |\langle y, z\rangle|.$$

◇

If two vectors x and y in an inner product space are orthogonal, then they satisfy the familiar Pythagorean Theorem.

Lemma 3.14. *If X is an inner product space and $x \in X$ and $y \in X$ are orthogonal, then*

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

Proof. If $x \perp y$, then $\langle x, y\rangle = \langle y, x\rangle = 0$, and thus a direct computation shows

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2. \qquad \square$$

Another similar example of the geometric properties of a Hilbert space is the following Parallelogram Law, which has the same form as the corresponding result in plane geometry (concerning proportions of parts in a parallelogram).

Lemma 3.15 (Parallelogram Law)**.** *The norm induced by the inner product on X satisfies*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \qquad \forall x, y \in X.$$

Proof. Left as an exercise. $\square$

Sums, direct sums, and orthogonal sums of two subspaces of an inner product space can be defined exactly as in the case of Euclidean spaces.

Definition 3.16. Let M and N be subspaces of a vector space X.

(a) The **sum** $M+N$ of M and N is defined as

$$M+N = \left\{ z \in X \;\middle|\; z = x+y,\ x \in M,\ y \in N \right\}. \tag{3.4}$$

(b) If $M \cap N = \{0\}$, then $M+N$ is a **direct sum**, and it is denoted by $M \oplus N$.

(c) If X is an inner product space and $M \perp N$, then $M+N$ is an **orthogonal sum** and it is denoted by $M \oplus_\perp N$.

Note that in each case (a)–(c) in Definition 3.16 the actual set, $M+N$, $M \oplus N$, and $M \oplus_\perp N$ is *exactly the same set* in (3.4). The "direct sum" and the notation $M \oplus N$ simply means that subspaces M and N have the additional property $M \cap N = \{0\}$, and similarly in the "orthogonal sum" and the notation $M \oplus_\perp N$ means that M and N have the property $M \perp N$. Similarly as on Euclidean spaces, it is easy to show that in direct and orthogonal sums every element has a *unique* representation $z = x+y$ with $x \in M$ and $y \in N$. Moreover, if M and N are closed subspaces of X, then the orthogonal sum $M \oplus_\perp N$ is also a closed subspace of X [NS82, Thm. 5.15.5] (in contrast, a direct sum $M \oplus N$ is not in general closed [NS82, Exer. 11 of Sec. 5.17]).

Remark 3.17. Note that the notations for direct and orthogonal sums of subspaces vary quite a bit in the literature! Take care to always make sure that you are using the right concept.

Hilbert spaces have the following very strong (again geometric) property which shows that the full space can be decomposed in to the sum of *any* closed subspace M and its orthogonal complement. This property is radically different from the case of Banach spaces, where it is possible that if M is a closed subspace, then there may not exist a closed subspace N such that $M \cap N = \{0\}$ and $X = M+N$.

Theorem 3.18. *If M is a closed subspace of a Hilbert space X, then*

$$X = M \oplus_\perp M^\perp.$$

The proof of Theorem 3.18 can be found, for example, in [NS82, Thm. 5.15.6], and it is also presented in detail in Appendix A.3.

Exercise 3.19. Let X be a Hilbert space.

(a) Show that if M and N are closed subspaces of X such that $X = M \oplus_\perp N$, then $N = M^\perp$.

(b) Show that if M is a subspace of X, then $(M^\perp)^\perp = \overline{M}$ (the closure of M). (Hint: Use Exercise 3.11 and part (a)).

(c) Show that a subspace $M \subset X$ is dense in X if and only if $M^\perp = \{0\}$. (Hint: Note that $N = N \oplus_\perp \{0\}$ for any subspace N of X).

◇

The possibility of decomposing a Hilbert space into the sum of a closed subspace and its orthogonal complement also leads to one of the most fundamentally important result on Hilbert spaces, namely, the Riesz Representation Theorem. This result shows that *all* bounded linear functionals on a Hilbert space X have a one-to-one correspondence with elements of X. In particular, a Hilbert space X is isometrically isomorphic with itself (in a slightly generalised sense, see the comments before Theorem 3.22). As a corollary of this theorem, every Hilbert space is reflexive (see Exercise 3.23).

Theorem 3.20 (Riesz Representation Theorem)**.** *Let ϕ be a bounded linear functional on a Hilbert space X. There exists a unique $z \in X$ such that*

$$\phi(x) = \langle x, z\rangle, \qquad \forall x \in X. \tag{3.5}$$

Moreover, $\|\phi\|_{B(X,\mathbb{C})} = \|z\|_X$.

Proof. Let $\phi \in B(X,\mathbb{C})$ be arbitrary. Define $M = \operatorname{Ker}(\phi) = \{\, x \in X \mid \phi(x) = 0\,\}$. Since ϕ is bounded, M is a closed subspace of X. If $M = X$, we have $\phi = 0$ and we can choose $z = 0$. On the other hand, if $M \neq X$, then we have $X = M \oplus_\perp M^\perp$ by Theorem 3.18.

To narrow down our choice of $z \in X$, we will first argue that we must have $z \in M^\perp$. To this end, we note that the element $z \in X$ has a unique representation $z = x_z + y_z \in M \oplus_\perp M^\perp$ such that $x_z \in M = \operatorname{Ker}(\phi)$ and $y_z \in M^\perp$. If $z \in X$ satisfies the conditions of the theorem, then the fact that $\phi(x_z) = 0$ and $x_z \perp y_z$ imply that

$$0 = \phi(x_z) = \langle x_z, z\rangle = \langle x_z, x_z + y_z\rangle = \langle x_z, x_z\rangle + \langle x_z, y_z\rangle = \|x_z\|^2.$$

Thus $x_z = 0$, and we indeed have $z = y_z \in M^\perp$. We can now turn to finding the correct element $z \in M^\perp$. Even though we will not prove it here, it turns out that the space $M^\perp$ is in fact a one-dimensional subspace of X, meaning that it *only contains a single linearly independent vector.* Because of this, we always have $z = \alpha y_0$, where $y_0 \in M^\perp$ is an arbitrary vector with $\|y_0\| = 1$, and we only need to find the suitable constant $\alpha \in \mathbb{C} \setminus \{0\}$. To find the correct value of $\alpha \in \mathbb{C}$, we can require that the identity (3.5) holds for $x = y_0$, which implies (using $\langle y_0, y_0 \rangle = \|y_0\|^2 = 1$)

$$\phi(y_0) = \langle y_0, z \rangle = \langle y_0, \alpha y_0 \rangle = \overline{\alpha} \langle y_0, y_0 \rangle = \overline{\alpha} \qquad \Leftrightarrow \qquad \alpha = \overline{\phi(y_0)}.$$

Thus we *should* have $z = \overline{\phi(y_0)} y_0$ in the theorem.

We still need to prove that (3.5) holds for our choice $z = \overline{\phi(y_0)} y_0$ and for all $x \in X$. To this end, we let $x \in X$ be arbitrary and we study the difference $\phi(x) - \langle x, z \rangle$, where $z = \overline{\phi(y_0)} y_0$. Using again the property that $\langle y_0, y_0 \rangle = \|y_0\|^2 = 1$ and the axioms of the inner product, we get

$$\begin{aligned} \phi(x) - \langle x, z \rangle &= \phi(x) \cdot 1 - \langle x, \overline{\phi(y_0)} y_0 \rangle = \phi(x) \langle y_0, y_0 \rangle - \phi(y_0) \langle x, y_0 \rangle && \text{(3.6a)} \\ &= \langle \phi(x) y_0 - \phi(y_0) x, y_0 \rangle. && \text{(3.6b)} \end{aligned}$$

In order to show that $\phi(x) - \langle x, z \rangle = 0$, we need to show that the inner product in the last expression is equal to zero, which is equivalent to the property that $\phi(x) y_0 - \phi(y_0) x$ and y_0 are orthogonal. Since we know that $y_0 \in M^\perp$, it is sufficient to show that $\phi(x) y_0 - \phi(y_0) x \in M = \operatorname{Ker}(\phi)$. Using the linearity of ϕ, a direct computation shows that

$$\phi(\phi(x) y_0 - \phi(y_0) x) = \phi(x) \phi(y_0) - \phi(y_0) \phi(x) = 0,$$

which indeed means that $\phi(x) y_0 - \phi(y_0) x \in M = \operatorname{Ker}(\phi)$. Because of this, (3.6) implies that $\phi(x) - \langle x, z \rangle = 0$. Since $x \in X$ was arbitrary, we have proved that (3.5) holds.

To prove the claim regarding the norms of ϕ and z, we can note that (3.5) and Exercise 3.13 imply $\|\phi\|_{B(X,\mathbb{C})} = \sup_{\|x\| \leq 1} |\phi(x)| = \sup_{\|x\| \leq 1} |\langle x, z \rangle| = \|z\|$.

Finally, to prove that $z \in X$ is unique, let $z_1, z_2 \in X$ such that $\phi(x) = \langle x, z_1 \rangle = \langle x, z_2 \rangle$ for all $x \in X$. Then for $x = z_1 - z_2$ we have

$$0 = \langle x, z_1 \rangle - \langle x, z_2 \rangle = \langle x, z_1 - z_2 \rangle = \|z_1 - z_2\|^2,$$

which implies $z_1 = z_2$. □

You can also compare the Riesz Representation Theorem to the "Radon–Nikodym Theorem" which we employed in describing the dual spaces of Lebesgue space in the proof of Theorem 2.44. As we recall, this result stated that if $1 < p < \infty$, then every bounded linear functional $\phi \in (L^p(I))'$ has the form

$$\phi(f) = \int_I f(t) g(t) \, \mathrm{d}t, \qquad \forall f \in L^p(I)$$

for some fixed $g \in L^q(I)$ where $q = p/(p-1)$. We can see that in the case where $p = 2$ (when $L^2(I)$ is a Hilbert space), this result is precisely equivalent to the Riesz Representation Theorem due to the fact that $q = 2$ (the element "z" being $\overline{g} \in L^2(I)$). On the other hand, the Radon–Nikodym Theorem shows that also those Lebesgue spaces $L^p(I)$ for $1 < p < \infty$ which are *not* Hilbert spaces have a similar characterisation of bounded linear functionals as in the Riesz Representation Theorem, but in this case the function g does not necessarily need to a member of the original space $L^p(I)$, but instead belongs to a different space $L^q(I)$.

Corollary 3.21. *If X is a Hilbert space, then the extension ϕ in the Hahn-Banach theorem (Theorem 2.36) is unique.*

Proof. By the Riesz-Representation theorem, the extension ϕ has the form $\phi(x) = \langle x, z\rangle$ for some unique element $z \in X$, and thus it is unique. □

The Riesz Representation Theorem also implies that every Hilbert space is reflexive. The proof of this result (in Theorem 3.22 below) is based on the fact that we can define a mapping $T : X \to X'$ such that for every $z \in X$ the value $\phi = Tz$ is the functional satisfying

$$\phi(x) = \langle x, z\rangle, \qquad \forall x \in X. \tag{3.7}$$

Indeed, the Cauchy–Schwarz Inequality in Lemma 3.3 immediately implies that ϕ is a bounded linear functional on X, and therefore the mapping T is well-defined. Moreover, the Riesz Representation Theorem implies that every bounded linear functional is of this form and that $\|\phi\|_{X'} = \|z\|_Z$, and thus this mapping is surjective, injective, and isometric (since $\|Tz\|_{X'} = \|\phi\|_{X'} = \|z\|_X$). However, the mapping is in fact not linear, but instead it is "*antilinear*" (sometimes called "*conjugate linear*"), since for all $\alpha, \beta \in \mathbb{C}$ and $z_1, z_2 \in X$ the properties of the inner product imply

$$T(\alpha z_1 + \beta z_2) = \langle\cdot, \alpha z_1 + \beta z_2\rangle = \langle\cdot, \alpha z_1 + \beta z_2\rangle = \overline{\alpha}\langle\cdot, z_1\rangle + \overline{\beta}\langle\cdot, z_2\rangle = \overline{\alpha}Tz_1 + \overline{\beta}Tz_2.$$

However, the fact that the mapping $T : X \to X'$ is antilinear instead of linear does not give rise to any difficulties as long as we take the additional complex conjugations into account, and in all other ways the mapping T behaves like a **linear** isometric isomorphism. In particular, the mapping T has a bounded inverse which is also antilinear and isometric (since the isometry of T implies $\|T^{-1}\phi\|_X = \|TT^{-1}\phi\|_{X'} = \|\phi\|_{X'}$).

Theorem 3.22. *Every Hilbert space is reflexive.*

Proof. Left as an exercise (Exercise 3.23). □

Exercise 3.23. In this exercise we prove that every Hilbert space is reflexive. For this, you can complete the following two steps below. Here $T : X \to X'$ is the mapping described above such that $Tz = \phi$ satisfies (3.7).

(a) Show that X' is a Hilbert space with the inner product defined by

$$\langle\phi, \psi\rangle_{X'} = \langle T^{-1}\psi, T^{-1}\phi\rangle_X.$$

(b) Use part (a) to prove that X is reflexive (Hint: Apply the Riesz Representation Theorem on X').

◇

3.2 The Adjoint of a Bounded Operator

We can now turn to investigating some special properties of operators on Hilbert spaces. We begin by defining the concept of an **adjoint** of a bounded linear operator

$T \in B(X,Y)$ between two Hilbert spaces $(X, \langle\cdot,\cdot\rangle_X)$ and $(Y, \langle\cdot,\cdot\rangle_Y)$. To this end, we first note that for every such operator and for every fixed $y \in Y$ we can define a bounded linear functional ϕ_y on X using the formula

$$\phi_y(x) = \langle Tx, y\rangle_Y, \qquad \forall x \in X.$$

Indeed, ϕ_y is clearly well-defined and linear since $T \in B(X,Y)$. Moreover, for all $x \in X$ the Cauchy–Schwarz Inequality implies that

$$|\phi_y(x)| = |\langle Tx, y\rangle_Y| \le \|Tx\|_Y \|y\|_Y \le \|T\|_{B(X,Y)} \|x\|_X \|y\|_Y,$$

which implies that $\phi_y \in B(X,\mathbb{C})$ and $\|\phi_y\| \le \|T\|\|y\|$. Since X is a Hilbert space and ϕ_y is a bounded linear functional, the Riesz Representation Theorem implies that there exists a $z_y \in X$ (depending on $y \in Y$) such that

$$\langle Tx, y\rangle_Y = \phi_y(x) = \langle x, z_y\rangle_X, \qquad \forall x \in X. \tag{3.8}$$

Since the element $z_y \in X$ in the Riesz Representation Theorem is unique, we can define a *mapping* $T^* : Y \to X$ *from the element* $y \in Y$ *to* $z_y \in X$, simply by defining $T^* y = z_y$ for all $y \in Y$. As shown below, it turns out that this mapping is linear and bounded, and its operator norm is the same as the norm of T. Using the relationship $T^* y = z_y$, the identity (3.8) can be written in the form

$$\langle Tx, y\rangle_Y = \langle x, T^* y\rangle_X, \qquad \forall x \in X,\ \forall y \in Y. \tag{3.9}$$

This formula is of course very familiar for us from linear algebra as one fundamental property of the conjugate transpose of a matrix $A \in \mathbb{C}^{m\times n}$. Indeed, if the matrix $A \in \mathbb{C}^{m\times n}$ is interpreted as a linear operator $A \in B(\mathbb{C}^n, \mathbb{C}^m)$, then its adjoint operator corresponds to the matrix $A^* \in B(\mathbb{C}^m, \mathbb{C}^n)$. Finally, it turns out that T^* is also the *unique* operator in $B(Y,X)$ which satisfies the identity (3.9).

Definition 3.24. Let X and Y be Hilbert spaces and $T \in B(X,Y)$. The operator $T^* \in B(Y,X)$ satisfying (3.9) is called the **adjoint (operator)** of T.

Exercise 3.25. Prove that the adjoint T^* of $T \in B(X,Y)$ is uniquely defined, i.e., that if there are two linear operators $S_1 : Y \to X$ and $S_2 : Y \to X$ satisfying

$$\langle Tx, y\rangle_Y = \langle x, S_1 y\rangle_X = \langle x, S_2 y\rangle_X, \qquad \forall x \in X,\ \forall y \in Y,$$

then $S_1 = S_2$ (Hint: Show that $S_1 y = S_2 y$ for all $y \in Y$). ◇

The uniqueness of the adjoint also gives us a way of *finding* the adjoint of $T \in B(X,Y)$: If we can find $S : Y \to X$ such that $\langle Tx, y\rangle = \langle x, Sy\rangle$ for all $x \in X$ and $y \in Y$, then $T^* = S$.

Exercise 3.26. Prove that $T^* \in B(Y,X)$ (i.e., T^* is linear and bounded). ◇

The fundamental properties of the adjoint operator are collected in the following theorem.

Theorem 3.27. *Let X and Y be Hilbert spaces. Every $T \in B(X,Y)$ has a uniquely defined adjoint operator $T^* \in B(Y,X)$. The operators T and T^* satisfy*

$$\langle Tx, y\rangle_Y = \langle x, T^*y\rangle_X, \qquad \forall x \in X,\ \forall y \in Y$$

and $\|T^*\| = \|T\|$.

Proof. The construction at the beginning of the section based on the Riesz Representation Theorem showed that every $T \in B(X,Y)$ has (at least one) adjoint, and this adjoint operator is unique by Exercise 3.25. Moreveor, $T^* \in B(Y,X)$ by Exercise 3.26. Finally, for $y \in Y$ we can use (3.9) to estimate

$$\|T^*y\|^2 = \langle T^*y, T^*y\rangle = \langle TT^*y, y\rangle \le \|TT^*y\|\|y\| \le \|T\|\|T^*y\|\|y\|,$$

which implies $\|T^*\| \le \|T\|$. On the other hand, if $x \in X$, the identity (3.9) similarly implies

$$\|Tx\|^2 = \langle Tx, Tx\rangle = \langle T^*Tx, x\rangle \le \|T^*Tx\|\|x\| \le \|T^*\|\|Tx\|\|x\|,$$

which implies $\|T\| \le \|T^*\|$. Combining the two inequalities we get $\|T^*\| = \|T\|$. □

Example 3.28 (Adjoints of the Shift Operators on $\ell^2(\mathbb{C})$)**.** In this example we consider the right and left shift operators in 2.4 on $X = \ell^2(\mathbb{C})$. Recall that the *right shift operator* $S_r : X \to X$ was defined by

$$S_r x = (0, x_1, x_2, x_3, \ldots), \qquad x = (x_k)_{k=1}^\infty \in X$$

and the *left shift operator* $S_l : X \to X$ by

$$S_l x = (x_2, x_3, x_4 \ldots), \qquad x = (x_k)_{k=1}^\infty \in X.$$

These operators were shown to be bounded in Exercise 2.20, i.e., $S_r, S_l \in B(X)$. To identify the adjoint S_r^* of the right shift operator, our aim is to find a linear operator $T : X \to X$ such that $\langle S_r x, y\rangle = \langle x, Ty\rangle$ for all $x, y \in X$. We then have from Exercise 3.25 that $S_r^* = T$. By definition of the shift operators and the inner product on $X = \ell^2(\mathbb{C})$, for all $x, y \in X$ we have (using $(S_r x)_1 = 0$ and $(S_r x)_k = x_{k-1}$ for $k \ge 2$)

$$\langle S_r x, y\rangle = \sum_{k=1}^\infty (S_r x)_k \overline{y_k} = \sum_{k=2}^\infty x_{k-1}\overline{y_k} = \sum_{j=1}^\infty x_j \overline{y_{j+1}} = \sum_{j=1}^\infty x_j \overline{(S_l y)_j} = \langle x, S_l y\rangle.$$

Since $x, y \in X$ were arbitrary, the above identity implies that in fact $S_r^* = S_l$! Moreover, since the same identity also implies

$$\langle S_l y, x\rangle = \overline{\langle x, S_l y\rangle} = \overline{\langle S_r x, y\rangle} = \langle y, S_r x\rangle, \qquad \forall x, y \in X,$$

we also have $S_l^* = S_r$. Because of this, the right and left shift operators on $\ell^2(\mathbb{C})$ are each others' adjoint operators. ◇

Example 3.29 (Multiplication Operator)**.** Let $\Omega \subset \mathbb{R}^n$ be an open or closed set and consider the Hilbert space $X = L^2(\Omega)$. For a fixed complex-valued and continuous function $q \in C(\Omega)$ satisfying $\sup_{z\in\Omega}|q(z)| < \infty$ we can define a **multiplication operator** $M_q : X \to X$ so that for every $f \in L^2(\Omega)$ the value $M_q f \in L^2(\Omega)$ is a function such that

$$(M_q f)(z) = q(z)f(z), \qquad \text{for (almost) all } z \in \Omega.$$

The map $M_q : X \to X$ is well-defined, since for all $f \in L^2(\Omega)$ the estimate

$$\int_\Omega |q(z)f(z)|^2 \,\mathrm{d}z = \Bigl(\sup_{z\in\Omega} |q(z)|\Bigr)^2 \int_\Omega |f(z)|^2 \,\mathrm{d}z < \infty$$

implies that $M_q f \in L^2(\Omega)$. Moreover, it is easy to show that $M_q : X \to X$ is linear, and the previous estimate also implies that

$$\|M_q f\|_2 = \Bigl(\int_\Omega |q(z)f(z)|^2 \,\mathrm{d}z\Bigr)^{\frac{1}{2}} \le \Bigl(\sup_{z\in\Omega} |q(z)|\Bigr)\|f\|_2,$$

and thus $M_q \in B(X)$ with norm $\|M_q\| \le \sup_{z\in\Omega}|q(z)|$.

In exactly the same way, we could have defined M_q more generally on $X = L^p(\Omega)$. In this example, however, we are especially interested in computing the adjoint operator of M_q. To this end, let $f, g \in L^2(\Omega)$ be arbitrary. By definition, we have

$$\langle M_q f, g\rangle = \int_\Omega q(z)f(z)\overline{g(z)} \,\mathrm{d}z = \int_\Omega f(z)\overline{\overline{q(z)}g(z)} \,\mathrm{d}z = \langle f, M_{\overline{q}} g\rangle,$$

where $M_{\overline{q}} \in B(X)$ is the multiplication operator corresponding to the function $\overline{q(\cdot)} \in C(\Omega)$ and satisfying $\sup_{z\in\Omega}|\overline{q(z)}| = \sup_{z\in\Omega}|q(z)| < \infty$. By Exercise 3.25 we have that $M_q^* = M_{\overline{q}}$. ◇

Example 3.30 (Adjoint of a Functional $\phi \in X'$)**.** If X is a Hilbert space, we can compute the adjoint of a bounded linear functional $\phi \in B(X, \mathbb{C})$ on X. For this, we can use the Riesz Representation Theorem which implies that there exists a unique $z_\phi \in X$ such that $\phi(x) = \langle x, z_\phi\rangle_X$ for all $x \in X$. We now have $Y = \mathbb{C}$ in the definition of the adjoint. This Euclidean space is equipped with the inner product $\langle y_1, y_2\rangle = y_1\overline{y_2}$. Thus if we let $x \in X$ and $y \in \mathbb{C}$ be arbitrary, we have that

$$\langle \phi(x), y\rangle_{\mathbb{C}} = \phi(x)\overline{y} = \langle x, z_\phi\rangle_X \overline{y} = \langle x, y z_\phi\rangle_X.$$

The last expression is indeed of the form $\langle x, Sy\rangle_X$, where the operator $S : \mathbb{C} \to X$ is defined by $Sy = y z_\phi$ for all $y \in \mathbb{C}$. Thus by Exercise 3.25 we have that $\phi^* \in B(\mathbb{C}, X)$ is such that

$$\phi^* y = y z_\phi, \qquad \forall y \in \mathbb{C}.$$

Sometimes this relationship is denoted more compactly as $\phi^* = z_\phi$, and in this notation it is customary to think of $B(\mathbb{C}, X)$ and X as the same space. ◇

The adjoints of sums and compositions of two operators also satisfy identities which are familiar from linear algebra.

Lemma 3.31. *Let X, Y, and Z be Hilbert spaces.*

(a) *If $T \in B(X,Y)$, then $(T^*)^* = T$.*

(b) *If $T, S \in B(X,Y)$, then $(T+S)^* = T^* + S^*$.*

(c) *If $T \in B(X,Y)$ and $\alpha \in \mathbb{C}$, then $(\alpha T)^* = \overline{\alpha} T^*$.*

(d) *If $T \in B(X,Y)$ and $S \in B(Y,Z)$, then $(ST)^* = T^* S^*$.*

(e) *If $T \in B(X,Y)$ is boundedly invertible, then also T^* is boundedly invertible and $(T^*)^{-1} = (T^{-1})^*$.*

Proof. Left as an exercise. □

Example 3.32. In Example 3.28, instead of the separate argument for showing that $S_l^* = S_r$, we could have alternatively used part (a) of Lemma 3.31 together with the property $S_r^* = S_l$. Indeed, these would have allowed us to more directly deduce $S_l^* = (S_r^*)^* = S_r$. ◇

Exercise 3.33. Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Prove that $\|T^*T\| = \|T\|^2$. ◇

The following result describes the nice relationships between the ranges and kernels of a bounded linear operator and its adjoint.

Theorem 3.34. *Let X and Y be Hilbert spaces and let $T \in B(X,Y)$. Then*

$$\overline{\mathrm{Ran}(T)} = \mathrm{Ker}(T^*)^\perp \qquad \textit{and} \qquad \mathrm{Ker}(T) = \mathrm{Ran}(T^*)^\perp$$

(where $\overline{\mathrm{Ran}(T)}$ is the closure of $\mathrm{Ran}(T) \subset Y$). In particular, the spaces X and Y can be decomposed as

$$X = \mathrm{Ker}(T) \oplus_\perp \overline{\mathrm{Ran}(T^*)} \qquad \textit{and} \qquad Y = \mathrm{Ker}(T^*) \oplus_\perp \overline{\mathrm{Ran}(T)}.$$

Proof. We begin by proving that $\mathrm{Ker}(T) = \mathrm{Ran}(T^*)^\perp$. For every $x \in X$ we have

$$\begin{aligned} x \in \mathrm{Ker}(T) \quad &\Leftrightarrow \quad Tx = 0 \\ &\Leftrightarrow \quad \langle Tx, y\rangle = 0 \quad \forall y \in Y \\ &\Leftrightarrow \quad \langle x, T^*y\rangle = 0 \quad \forall y \in Y \\ &\Leftrightarrow \quad \langle x, z\rangle = 0 \quad \forall z \in \mathrm{Ran}(T^*) \\ &\Leftrightarrow \quad x \in \mathrm{Ran}(T^*)^\perp, \end{aligned}$$

and thus $\mathrm{Ker}(T) = \mathrm{Ran}(T^*)^\perp$. Together with Exercise 3.19(b) and Lemma 3.31 this identity applied to $T^* \in B(Y,X)$ also implies that

$$\mathrm{Ker}(T^*)^\perp = (\mathrm{Ran}((T^*)^*)^\perp)^\perp = (\mathrm{Ran}(T)^\perp)^\perp = \overline{\mathrm{Ran}(T)}.$$

The decompositions $X = \mathrm{Ker}(T) \oplus_\perp \overline{\mathrm{Ran}(T^*)}$ and $Y = \mathrm{Ker}(T^*) \oplus_\perp \overline{\mathrm{Ran}(T)}$ now follow directly from Theorem 3.18. □

A relatively small but quite important class of operators the coincide with their own adjoints, i.e., $T^* = T$. Such operators are called **self-adjoint**. Note that since for $T \in B(X,Y)$, the adjoint T^* is by definition an operator from Y to X, the condition "$T^* = T$" in particular requires that $Y = X$.

Definition 3.35. Let X be a Hilbert space. The operator $T \in B(X)$ is **self-adjoint** if $T^* = T$.

Exercise 3.36. Show that if $T \in B(X)$ is self-adjoint, then $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in X$. In fact, also the converse statement holds (if you want to prove this, you can study the condition $\langle Tx, x\rangle \in \mathbb{R}$ in the case where $x = y + \alpha z$ for suitable $\alpha \in \mathbb{C}$ for example $\alpha = 1$ and $\alpha = i$, to prove that $\mathrm{Re}\langle Ty, z\rangle = \mathrm{Re}\langle y, Tz\rangle$ and $\mathrm{Im}\langle Ty, z\rangle = \mathrm{Im}\langle y, Tz\rangle$). ◇

Example 3.37. The multiplication operator $M_q \in B(X)$ in Example 3.29 is self-adjoint if (and only if) the function $q(\cdot)$ is real-valued, i.e., $q(z) \in \mathbb{R}$ for all $z \in \Omega$. Indeed, in this case we have $M_q^* = M_{\overline{q}} = M_q$. ◇

Theorem 3.38. *Let X be a Hilbert space and let $T \in B(X)$ be self-adjoint. Then*

$$\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|.$$

Proof. **[Optional]** Denote $M = \sup_{\|x\|=1}|\langle Tx, x\rangle|$. Since $|\langle Tx, x\rangle| \le \|Tx\|\|x\| \le \|T\|\|x\|^2$ for all $x \in X$, we clearly have $M \le \|T\|$ due to the Cauchy–Schwarz Inequality. To show that $\|T\| \le M$, we first note that $|\langle Tx, x\rangle| \le M\|x\|^2$ for all $x \in X$. If $x, y \in X$, then a direct computation shows that

$$\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle = 2\langle Tx, y\rangle + 2\langle Ty, x\rangle.$$

If we choose $y = \alpha Tx$ for some $\alpha > 0$, then the above identity, the self-adjointness of T, and the definition of M further imply

$$\begin{aligned}4\alpha\|Tx\|^2 &= 4\alpha\langle Tx, Tx\rangle = 2\alpha\langle Tx, Tx\rangle + 2\alpha\langle T^*Tx, x\rangle = 2\langle Tx, \alpha Tx\rangle + 2\langle T(\alpha Tx), x\rangle \\ &= \langle T(x+\alpha Tx), x+\alpha Tx\rangle - \langle T(x-\alpha Tx), x-\alpha Tx\rangle \\ &\le |\langle T(x+\alpha Tx), x+\alpha Tx\rangle| + |\langle T(x-\alpha Tx), x-\alpha Tx\rangle| \\ &\le M\|x+\alpha Tx\|^2 + M\|x-\alpha Tx\|^2.\end{aligned}$$

Applying the Parallelogram Law in Lemma 3.15 to the last expression shows that

$$2\alpha\|Tx\|^2 \le M\left(\|x\|^2 + \|\alpha Tx\|^2\right) = M\left(\|x\|^2 + \alpha^2\|Tx\|^2\right).$$

If $x \in X$ is such that $Tx \neq 0$, we can choose $\alpha = \|x\|/\|Tx\| > 0$, and the above inequality implies

$$2\frac{\|x\|}{\|Tx\|}\|Tx\|^2 \le M\left(\|x\|^2 + \frac{\|x\|^2}{\|Tx\|^2}\|Tx\|^2\right) \qquad \Leftrightarrow \qquad 2\|x\|\|Tx\| \le 2M\|x\|^2,$$

or $\|Tx\| \le M\|x\|$. On the other hand, if $Tx = 0$, then also $\|Tx\| = 0$, and therefore we have $\|Tx\| \le M\|x\|$ for all $x \in X$. This immediately implies that $\|T\| \le M$, and the proof is complete. □

Example 3.39. In this example we study the integral operator in Example 2.21 on the space $X = L^2(a,b)$. We assume the kernel satisfies $(t,s) \mapsto k(t,s) \in C([a,b]\times[a,b])$ and define the integral operator $T \in B(X)$ by

$$(Tf)(t) = \int_a^b k(t,s)f(s)\,\mathrm{d}s, \qquad \forall t \in [a,b].$$

In Example 2.21 we saw that (now for $p = q = 2$)

$$\|T\| \le \left(\int_a^b \int_a^b |k(t,s)|^2 dtds\right)^{\frac{1}{2}} = \|k(\cdot,\cdot)\|_{L^2(\Omega)} < \infty,$$

where $\|k(\cdot,\cdot)\|_{L^2(\Omega)}$ is the norm of $k(\cdot,\cdot)$ on the space $L^2(\Omega)$ with $\Omega = [a,b]\times[a,b]$. We can first identify the adjoint of T. If $f, g \in L^2(a,b)$, then

$$\begin{aligned}\langle Tf, g\rangle &= \int_a^b (Tf)(t)\overline{g(t)}\,\mathrm{d}t = \int_a^b \left(\int_a^b k(t,s)f(s)\,\mathrm{d}s\right)\overline{g(t)}\,\mathrm{d}t \\ &= \int_a^b\int_a^b k(t,s)f(s)\overline{g(t)}\,\mathrm{d}s\,\mathrm{d}t = \int_a^b\int_a^b f(s)\overline{\overline{k(t,s)}g(t)}\,\mathrm{d}t\,\mathrm{d}s \\ &= \int_a^b f(s)\overline{\left(\int_a^b \overline{k(t,s)}g(t)\,\mathrm{d}t\right)}\,\mathrm{d}s = \langle f, Sg\rangle\end{aligned}$$

(the order of integration can be changed using Fubini's Theorem) where S is an operator defined as

$$(Sg)(t) = \int_a^b \overline{k(s,t)}g(s)ds, \qquad \forall g \in L^2(a,b).$$

Thus $S = T^*$ by Exercise 3.25, and T^* is therefore an integral operator with kernel $(t,s) \mapsto \overline{k(s,t)}$. This also leads to the condition for the integral operator T to be self-adjoint: $T^* = T$ if (and in fact, only if) the kernel satisfies $k(t,s) = \overline{k(s,t)}$ for all $t, s \in [a,b]$. ◇

3.3 Orthonormal Bases in Separable Hilbert Spaces [Optional]

In this section we explore the concept of a "*basis*" of a Banach or a Hilbert space X. As we recall from linear algebra, a "basis" of the Euclidean space $\mathbb{C}^n$ (or $\mathbb{R}^n$) is a linearly independent set of vectors $\{e_k\}_{k=1}^n \subset \mathbb{C}^n$ such that every vector $x \in \mathbb{C}^n$ can be represented in the form

$$x = \sum_{k=1}^n \alpha_k e_k$$

with a unique set of "*coordinates*" $\{\alpha_k\}_{k=1}^n \subset \mathbb{C}$. In addition, the basis $\{e_k\}_{k=1}^n$ is an *orthonormal basis*, if $\|e_k\| = 1$ and $e_k \perp e_j$ for all $k, j \in \{1,\dots,n\}$ such that $k \ne j$.

First of all, Banach and Hilbert spaces (or more generally, infinite-dimensional vector spaces) do not in general have "bases" in this same sense. Moreover, by definition of the dimension of a vector space in Definition 1.14, it is necessary that any "basis" of an infinite-dimensional vector space X contains an infinite number of elements. Because of this, we need to be particularly careful in how we define the concept of a basis in order to ensure that the concept is sufficiently strong to be useful but at the same time sufficiently general so that we are not focusing on too narrow class of cases. On this course we will focus only on **Schauder bases**, which contain a countable number of elements and which have the (very strong) property that *every vector $x \in X$ can be expressed as an infinite linear combination of the basis vectors.* The existence of a Schauder basis on a Banach or a Hilbert space is always a special property of the particular space, but we will see that some of our familiar examples do possess Schauder bases[1].

The Schauder basis can be used to represent elements of the space X as *infinite linear combinations* of the basis elements. For this purpose, we need the definition of the convergence of an *infinite series* on a normed space.

Definition 3.40. Let $(X, \|\cdot\|)$ be a normed space and let $(x_k)_{k=1}^\infty \subset X$. The **series**

$$\sum_{k=1}^\infty x_k$$

converges on X if the sequence $(s_n)_{n=1}^\infty \subset X$ of its partial sums $s_n := \sum_{k=1}^n x_k$ converges on X. If $s_n \to s \in X$ as $n \to \infty$, then we define

$$\sum_{k=1}^\infty x_k = s.$$

Exercise 3.41. Assume $(X, \|\cdot\|)$ is a Banach space. Show that a series $\sum_{k=1}^\infty x_k$ on X converges if it **converges absolutely**, i.e., if

$$\sum_{k=1}^\infty \|x_k\| < \infty.$$

(This result is a *purely sufficient* condition for convergence of a series, and there are many important series which converge but fail to converge absolutely! Moreover, the result holds if and only if X is a Banach space.) ◇

Exercise 3.42. Assume $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed spaces and $T \in B(X, Y)$. Show that if $(x_k)_{k=1}^\infty \subset X$ is such that the series $\sum_{k=1}^\infty x_k$ converges on X, then

$$\sum_{k=1}^\infty Tx_k = T \sum_{k=1}^\infty x_k$$

(i.e., the series on the left-hand side converges in Y and its value is given by Tx, where $x = \sum_{k=1}^\infty x_k$). This result means that we can (formally) "*take a bounded operator out of a convergent series*". ◇

[1] More generally, every infinite-dimensional Banach space does possess a more general form of a "basis", a *Hamel basis*, but this basis always contains an uncountable number of elements

We can now define the concepts of a Schauder basis on a Banach space, and orthonormal basis on a Hilbert space. In the definition, δ_{kj} denotes the Kronecker delta, i.e.,

$$\delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j. \end{cases}$$

Definition 3.43. Let X be a normed space. A sequence $(e_k)_{k=1}^{\infty} \subset X$ is a **Schauder basis** of X if for every $x \in X$ there exists a unique sequence $(\alpha_k)_{k=1}^{\infty} \subset \mathbb{C}$ such that

$$x = \sum_{k=1}^{\infty} \alpha_k e_k. \tag{3.10}$$

If X is an inner product space, then $(e_k)_{k=1}^{\infty} \subset X$ is an **orthonormal basis** if it is a Schauder basis of X and if $\langle e_k, e_j \rangle = \delta_{kj}$ for all $k, j \in \mathbb{N}$.

The scalars $\{\alpha_k\}_k$ in the representation (3.10) are called the **coordinates** of the vector $x \in X$ in the basis $(e_k)_{k=1}^{\infty}$. By definition, a Schauder basis is an orthonormal basis if $\|e_k\| = 1$ for all $k \in \mathbb{N}$, and in addition $e_k \perp e_j$ for all $k, j \in \mathbb{N}$ such that $k \neq j$.

Example 3.44. Consider the space $X = \ell^p(\mathbb{C})$ and define $(e_k)_{k=1}^{\infty} \subset X$ such that $e_k = (e_k^j)_{j=1}^{\infty}$ with $e_k^k = 1$ and $e_k^j = 0$ if $j \neq k$. This set of vectors obviously corresponds to the canonical Euclidean basis of $\mathbb{C}^n$ and $\mathbb{R}^n$. Our aim is to show that $(e_k)_{k=1}^{\infty}$ is also a Schauder basis in $\ell^p(\mathbb{C})$, and an orthonormal basis in $\ell^2(\mathbb{C})$.

If $x = (x_k)_{k=1}^{\infty} \in \ell^p(\mathbb{C})$ is arbitrary, then our aim is to show that (3.10) holds for some scalars α_k. Due to the structure of the vectors $e_k \in \ell^p(\mathbb{C})$ (the kth element is equal to 1 and the others are equal to 0), we can already see that if (3.10) holds, then necessarily $\alpha_k = x_k$ for all $k \in \mathbb{N}$ (only the kth term on the right-hand side of (3.10) affects the kth element of the series). Because we have such a good guess for the coordinates α_k, we can aim to directly show that the partial sums $s_n = \sum_{k=1}^{n} \alpha_k e_k$ converge to x in X. Indeed, for all $n \in \mathbb{N}$ we have that

$$\begin{aligned} \|x - s_n\|_p^p = \left\| x - \sum_{k=1}^{n} x_k e_k \right\|_p^p &= \|x - (x_1, \ldots, x_n, 0, 0, \ldots)\|_p^p \\ &= \|(0, \ldots, 0, x_{n+1}, x_{n+2}, \ldots)\|_p^p = \sum_{k=n+1}^{\infty} |x_k|^p \to 0 \end{aligned}$$

as $n \to \infty$ since $x \in \ell^p(\mathbb{C})$. Since $x \in X$ was arbitrary, we have shown that every element of X has a representation (3.10) with unique coordinates $(\alpha_k)_{k=1}^{\infty} \in \mathbb{C}$, and thus $(e_k)_{k=1}^{\infty}$ is by definition a Schauder basis of $X = \ell^p(\mathbb{C})$.

In the case where $X = \ell^2(\mathbb{C})$, we can note that $\|e_k\|_2 = 1$ for all $k \in \mathbb{N}$, and (using the definition of the inner product in Example 3.9), we have that if $k \neq j$, then

$$\langle e_k, e_j \rangle = \sum_{n=1}^{\infty} e_k^n e_j^n = 0$$

since in each term either $e_k^n = 0$ or $e_j^n = 0$. Thus $(e_k)_{k=1}^{\infty}$ is by definition an orthonormal basis of $\ell^2(\mathbb{C})$. ◇

Exercise 3.45. Let X be a Hilbert space and assume the sequence $(e_k)_{k=1}^{\infty} \subset X$ is *orthonormal*, i.e., $\langle e_k, e_j \rangle = \delta_{kj}$ for all $k, j \in \mathbb{N}$. Show that if $(\alpha_k)_{k=1}^{\infty} \subset \mathbb{C}$, then the series $\sum_{k=1}^{\infty} \alpha_k e_k$ converges in X if and only if $(\alpha_k)_{k=1}^{\infty} \in \ell^2(\mathbb{C})$. Moreover, prove that

$$\left\| \sum_{k=1}^{\infty} \alpha_k e_k \right\|^2 = \sum_{k=1}^{\infty} |\alpha_k|^2.$$

◇

Exercise 3.46 (Bessel's Inequality)**.** Let X be an inner product space and assume $(e_k)_{k=1}^{\infty} \subset X$ is *orthonormal*, i.e., $\langle e_k, e_j \rangle = \delta_{kj}$ for all $k, j \in \mathbb{N}$. Prove that

$$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 \leq \|x\|^2, \qquad \forall x \in X.$$

(Hint: Expand $\left\| \sum_{k=1}^{N} \langle x, e_k \rangle e_k - x \right\|^2$ to show $\sum_{k=1}^{N} |\langle x, e_k \rangle|^2 \leq \|x\|^2$, then let $N \to \infty$). ◇

As mentioned above, not all Banach or even Hilbert spaces have Schauder bases. In the case of Hilbert spaces, the precise condition required for the existence of an orthonormal basis is the **separability** of the space.

Definition 3.47. A normed space X is **separable** if it contains a countable and dense subset.

If a normed space X has a Schauder basis $(e_k)_{k=1}^{\infty}$, it is necessarily a separable (this can be shown by considering, for example, that the subset of all finite linear combinations of $(e_k)_k$ with coefficients α_k whose real and imaginary parts are rational numbers). The question whether or not the converse statement is true for complete normed spaces, i.e., if every separable Banach space has a Schauder basis, was a long-standing and famous open problem (posed by Stefan Banach in 1932). This problem was solved in 1972 by Per Enflo, who constructed a separable (and even reflexive!) Banach space which does not have a Schauder basis [Enf73]. On the other hand, as the following theorem states, every separable Hilbert space does have a Schauder basis, and even an orthonormal basis.

Theorem 3.48. *Every nontrivial separable Hilbert space has an orthonormal basis.*

The proof of Theorem 3.48 is not very difficult, but we will not consider it on this course. Instead, we will limit ourselves to presenting selected fundamental properties and characterisations of an orthonormal basis. In particular, the representation (3.11) in the following theorem is a **generalised Fourier series expansion** of $x \in X$. This terminology is fully justified, since in fact the classical Fourier series studied are in fact infinite linear combinations of (for example) the sequences $(e^{i2\pi k \cdot/\tau})_{k=1}^{\infty}$, or $(\cos(2\pi k \cdot /\tau))_{k=1}^{\infty}$, or $(\sin(2\pi k \cdot /\tau))_{k=1}^{\infty}$, which are precisely orthonormal bases of the Hilbert space $L^2(0, \tau)$!

Theorem 3.49. *Let X be a Hilbert space and assume $(e_k)_{k=1}^{\infty} \subset X$ is an orthonormal sequence, i.e., $\langle e_k, e_j \rangle = \delta_{kj}$ for all $k, j \in \mathbb{N}$. Then the following are equivalent.*

(a) *The sequence $(e_k)_{k=1}^{\infty}$ is an orthonormal basis of X.*

(b) (Fourier Series Expansion) *Every $x \in X$ has the representation*

$$x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k. \tag{3.11}$$

(c) *For every $x \in X$ we have*

$$\|x\|^2 = \sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2.$$

(d) *If $x \in X$ is such that $\langle x, e_k \rangle = 0$ for all $k \in \mathbb{N}$, then $x = 0$.*

Proof. We will only present the "easy" parts of the proof. The complete proof for the last implication "(d)⇒(a)" can be found, for example, in [NS82, Thm. 5.17.8].

(a)⇒(b): Assume $(e_k)_{k=1}^{\infty}$ is an orthonormal basis of X and let $x \in X$. By definition, there exist $(\alpha_k)_{k=1}^{\infty} \subset \mathbb{C}$ such that

$$x = \sum_{k=1}^{\infty} \alpha_k e_k.$$

The claim in (b) holds if $\alpha_k = \langle x, e_k \rangle$ for all $k \in \mathbb{N}$. To show this, let $n \in \mathbb{N}$ be arbitrary. The mapping $y \mapsto \langle y, e_n \rangle$ is a bounded linear functional on X, and thus by Exercise 3.42 we have that

$$\langle x, e_n \rangle = \left\langle \sum_{k=1}^{\infty} \alpha_k e_k, e_n \right\rangle = \sum_{k=1}^{\infty} \alpha_k \langle e_k, e_n \rangle = \sum_{k=1}^{\infty} \alpha_k \delta_{kn} = \alpha_n$$

since the sequence $(e_k)_{k=1}^{\infty}$ is orthonormal.

(b)⇒(c): Let $x \in X$ and assume (3.11) holds. By Exercise 3.45 we have that

$$\|x\|^2 = \left\| \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k \right\|^2 = \sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2.$$

(c)⇒(d): If $x \in X$ is such that $\langle x, e_k \rangle = 0$ for all $k \in \mathbb{N}$, then by part (c) we have

$$\|x\|^2 = \sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 = \sum_{k=1}^{\infty} 0 = 0,$$

and thus $x = 0$. □

Example 3.50 (The Complex Fourier Basis)**.** Consider $X = L^2(0, 1)$ and define a sequence $(e_k)_{k=-\infty}^{\infty} \subset X$ (i.e., a doubly infinite sequence $(\ldots, e_{-2}, e_{-1}, e_0, e_1, e_2, \ldots)$)

such that $e_k(t) = e^{i2\pi kt}$ for all $k \in \mathbb{Z}$ and $t \in [0,1]$. The sequence is orthonormal, since for all $k, n \in \mathbb{Z}$ we have

$$\langle e_k, e_n \rangle = \int_0^1 e_k(t)\overline{e_n(t)}\,\mathrm{d}t = \int_0^1 e^{i2\pi kt}e^{-i2\pi nt}\,\mathrm{d}t = \int_0^1 e^{i2\pi(k-n)t}\,\mathrm{d}t = \begin{cases} 0 & k \neq n \\ 1 & k = n. \end{cases}$$

The theory of Fourier series implies that if $f \in L^2(0,1)$, then

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kt}$$

(in the sense that the series on the right-hand side converges in $L^2(0,1)$), where the coefficients $(c_k)_{k=-\infty}^{\infty} \subset \mathbb{C}$ are defined as

$$c_k = \int_0^1 f(t)e^{-i2\pi kt}\,\mathrm{d}t = \int_0^1 f(t)\overline{e_k(t)}\,\mathrm{d}t = \langle f, e_k \rangle.$$

Thus the complex Fourier basis functions $(e_k)_{k=-\infty}^{\infty}$ are in fact an orthonormal basis of the Hilbert space $L^2(0,1)$. ◇

The existence of an orthonormal basis has the consequence that in fact every infinite-dimensional and separable Hilbert space is in fact isometrically isomorphic to the space $\ell^2(\mathbb{C})$. The proof of this very nice result is a consequence of Theorem 3.49 and the fact that by definition of a Schauder basis, there is a one-to-one correspondence between every element $x \in X$ and the coordinates $(\alpha_k)_{k=1}^{\infty} \subset X$ in the basis $(e_k)_{k=1}^{\infty}$.

Theorem 3.51 (Riesz–Fischer Theorem). *Every separable and infinite-dimensional Hilbert space is isometrically isomorphic to $\ell^2(\mathbb{C})$.*

Proof. Let $(e_k)_{k=1}^{\infty} \subset X$ be an orthonormal basis of X. Define $T : X \to \ell^2(\mathbb{C})$ so that

$$Tx = \big(\langle x, e_k \rangle\big)_{k=1}^{\infty}, \qquad \forall x \in X.$$

The mapping is well-defined, since Theorem 3.49(c) implies that for all $x \in X$

$$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 = \|x\|_X^2 < \infty,$$

and thus $Tx \in \ell^2(\mathbb{C})$. Moreover, this same result also implies that $\|Tx\|_{\ell^2}^2 = \|x\|_X^2$, and thus $T \in B(X, \ell^2(\mathbb{C}))$ is isometric and in particular injective. It is also easy to show that T is linear. Finally, to show that T is surjective, let $(\alpha_k) \subset \ell^2(\mathbb{C})$. By Exercise (3.45) the series $\sum_{k=1}^{\infty} \alpha_k e_k$ converges in X. For all $n \in \mathbb{N}$ we have

$$\left\langle \sum_{k=1}^{\infty} \alpha_k e_k, e_n \right\rangle = \sum_{k=1}^{\infty} \alpha_k \langle e_k, e_n \rangle = \sum_{k=1}^{\infty} \alpha_k \delta_{kn} = \alpha_n$$

and thus

$$T\left(\sum_{k=1}^{\infty} \alpha_k e_k\right) = (\alpha_n)_{n=1}^{\infty}$$

by definition. Since $(\alpha_k)_{k=1}^{\infty} \in \ell^2(\mathbb{C})$ was arbitrary, we have that T is surjective. Thus T has a bounded inverse by Theorem 2.32, and it is by definition an isometric isomorphism between X and $\ell^2(\mathbb{C})$. □

4. Spectral Theory

This final chapter is devoted to the study of the *spectrum* of a bounded linear operator $T \in B(X)$. For a bounded operator $T \in B(X)$ on a Banach space, the **spectrum** $\sigma(T)$ of T is defined as the set of complex numbers $\lambda \in \mathbb{C}$ such that the operator

$$\lambda I - T \in B(X)$$

***does not** have a bounded inverse* $(\lambda I - T)^{-1} \in B(X)$. Here $I \in B(X)$ denotes the identity operator on X. On the other hand, the **resolvent set** $\rho(T)$ of T is defined as those values $\lambda \in \mathbb{C}$ such that this operator has a bounded inverse $(\lambda I - T)^{-1} \in B(X)$ (and thus $\rho(T) = \mathbb{C} \setminus \sigma(T)$).

Definition 4.1. Let X be a Banach space and let $T \in B(X)$. The **resolvent set** $\rho(T)$ of T is defined as

$$\rho(T) = \{\, \lambda \in \mathbb{C} \mid \lambda I - T \in B(X) \text{ is boundedly invertible} \,\}.$$

The **spectrum** $\sigma(T)$ of T is defined as $\sigma(T) = \mathbb{C} \setminus \rho(T)$.

The spectrum of an operator describes many important features and properties of an operator, and its study can be used to gain valuable information on how the operator behaves as a mapping $T : X \to X$. The spectrum and the resolvent are most directly connected to the solvability of the *operator equations* of the form

$$(\lambda I - T)x = y, \tag{4.1}$$

where $y \in X$ is given "*data*" and $x \in X$ is the unknown solution of the linear equation. The spectral analysis of the operator T can be used to study the existence, uniqueness and continuity properties of the solution x of (4.1), and in important special cases it can also be used to construct the solution $x \in X$ explicitly. Because of this, spectral theory of linear operators is an essential tool in the analysis of differential equations (when T is a differential operator[1]) and integral equations (when T is an integral operator).

When $\lambda \in \mathbb{C}$ is in the resolvent set of T, i.e., $\lambda \in \rho(T)$, the operator $\lambda I - T$ has a bounded inverse $(\lambda I - T)^{-1} \in B(X)$ and

> *for every $y \in X$ the operator equation (4.1) has a unique solution $x \in X$ depending continuously on $y \in X$.*

[1] This topic is considered in greater detail on the course "MATH.MA.830 Advanced Functional Analysis".

Indeed, for any $y \in X$ the equation (4.1) has a solution $x = (\lambda I - T)^{-1}y$, and this solution is unique because $\lambda I - T$ is injective (if $(\lambda I - T)x_1 = y = (\lambda I - T)x_2$, then injectivity implies $x_1 = x_2$). In our setting the property "x depends continuously on y" means that if $y_1, y_2 \in X$ and if $x_1, x_2 \in X$ are the solutions of the equations $(\lambda I - T)x_1 = y_1$ and $(\lambda I - T)x_2 = y_2$, then

$$\|x_1 - x_2\| = \|(\lambda I - T)^{-1}y_1 - (\lambda I - T)^{-1}y_2\| \leq \|(\lambda I - T)^{-1}\|\|y_1 - y_2\|.$$

This estimate implies that *a small change in the data $y \in X$ will result in a (relatively) small change in the solution $x \in X$.*

Example 4.2. The spectrum of a linear operator $T \in B(X)$ generalises the concept of *eigenvalues* of a square matrix $A \in \mathbb{C}^{n\times n}$. Indeed, $\lambda \in \mathbb{C}$ is by definition an eigenvalue of A if there exists $x \in \mathbb{C}^n$, $x \neq 0$, such that $(\lambda I - A)x = 0$. Thus $\lambda \in \mathbb{C}$ is an eigenvalue of A precisely if the matrix $\lambda I - A$ is not injective, and therefore does not have an inverse. On the other hand, for any $\lambda \in \mathbb{C}$ which is not an eigenvalue of A, the square matrix $\lambda I - A$ has full rank, which means that $\lambda I - A$ is non-singular and has the inverse matrix $(\lambda I - A)^{-1} \in \mathbb{C}^{n\times n}$. ◇

In the case of linear operators on infinite-dimensional spaces, spectral theory is very *rich* compared to the analysis of eigenvalues of a matrix. Indeed, in the next section we will learn that even though $T \in B(X)$ can have *eigenvalues* and *eigenvectors*, these do not usually provide the full picture of the spectral properties of a linear operator. This is due to the fact that on an infinite-dimensional space the operator $\lambda I - T$ can "fail to have a bounded inverse" in several different ways, whereas a matrix $\lambda I - A$ has an inverse matrix if and only if it is not injective. In our examples we will see that an operator may not have any eigenvalues at all, or alternatively it may even have an uncountably infinite number of eigenvalues covering a domain of the complex plane $\mathbb{C}$.

4.1 The Parts of the Spectrum

The spectrum $\sigma(T)$ of a bounded linear operator can be decomposed into parts based on the more precise properties of the operator $\lambda I - T$.

Definition 4.3 (Parts of The Spectrum)**.** Let X be a Banach space and let $T \in B(X)$. The spectrum $\sigma(T)$ consists of three distinct parts:

(a) The **point spectrum** $\sigma_p(T)$ is defined as

$$\sigma_p(T) = \{\, \lambda \in \mathbb{C} \mid \lambda I - T \text{ is not injective} \,\}.$$

The values $\lambda \in \sigma_p(T)$ are called **eigenvalues** of T, and $x \in \operatorname{Ker}(\lambda I - T)$ such that $x \neq 0$ are **eigenvectors** of T (corresponding to λ).

(b) The **residual spectrum** $\sigma_r(T)$ is defined as

$$\sigma_r(T) = \{\, \lambda \in \mathbb{C} \mid \lambda I - T \text{ is injective and } \overline{\operatorname{Ran}(\lambda I - T)} \neq X \,\}.$$

(c) The **continuous spectrum** $\sigma_c(T)$ is defined as

$$\sigma_c(T) = \{\, \lambda \in \mathbb{C} \mid \lambda I - T \text{ is injective, } \overline{\operatorname{Ran}(\lambda I - T)} = X, \text{ but } \operatorname{Ran}(\lambda I - T) \neq X \,\}.$$

The motivation for calling $\lambda \in \sigma_p(T)$ and $x \in \operatorname{Ker}(\lambda I - T)$ with $x \neq 0$ *eigenvalues* and *eigenvectors* is quite easy to see, since these elements satisfy

$$(\lambda I - T)x = 0 \qquad \Leftrightarrow \qquad Tx = \lambda x.$$

The three parts of the spectrum are indeed pairwise disjoint and they cover the whole spectrum, meaning that every $\lambda \in \sigma(T)$ belongs to precisely one of the sets $\sigma_p(T)$, $\sigma_r(T)$, and $\sigma_c(T)$. We leave the verification of this property as an exercise.

Exercise 4.4. Show that $\sigma(T) = \sigma_p(T) \cup \sigma_r(T) \cup \sigma_c(T)$ and that every $\lambda \in \sigma(T)$ belongs to precisely one of the sets $\sigma_p(T)$, $\sigma_r(T)$, and $\sigma_c(T)$.

To show that the union of $\sigma_p(T)$, $\sigma_r(T)$, and $\sigma_c(T)$ is precisely the spectrum of T, you can recall that by the Bounded Inverse Theorem (Theorem 2.32) $\lambda I - T \in B(X)$ has a bounded inverse if it is injective and surjective. Conversely, the existence of a bounded inverse implies that $\lambda I - T$ must be both injective and surjective. $\diamond$

To determine if a given complex number $\lambda \in \mathbb{C}$ belongs to the resolvent set $\rho(T)$ or to one of the parts $\sigma_p(T)$, $\sigma_r(T)$, and $\sigma_c(T)$ of the spectrum, we can take the following steps:

Step 1: If $\lambda I - T$ is not injective, then $\lambda \in \sigma_p(T)$.

Step 2: If $\lambda I - T$ is injective, but $\operatorname{Ran}(\lambda I - T)$ is not dense in X, then $\lambda \in \sigma_r(T)$.

Step 3: If $\lambda I - T$ is injective and $\operatorname{Ran}(\lambda I - T)$ is dense in X, but $\operatorname{Ran}(\lambda I - T) \neq X$, then $\lambda \in \sigma_c(T)$.

Step 4: If $\lambda I - T$ is injective and $\operatorname{Ran}(\lambda I - T) = X$, then $\lambda I - T$ is boundedly invertible by Theorem 2.32, and thus $\lambda \in \rho(T)$.

This process is also illustrated in Figure 4.1 from [NS82, Fig. 6.5.1].

The following result shows that the continuous spectrum of $T \in B(X)$ can alternatively be defined as the set of $\lambda \in \mathbb{C}$ for which $\lambda I - T$ is injective and has dense range, but the inverse operator $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is not bounded.

Theorem 4.5. *Let X be a Banach space and let $T \in B(X)$. Then $\lambda \in \sigma_c(T)$ if and only if $\lambda I - T$ is injective, $\overline{\operatorname{Ran}(\lambda I - T)} = X$, and the operator $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is unbounded.*

Proof. Assume first that $\lambda I - T$ is injective, $\overline{\operatorname{Ran}(\lambda I - T)} = X$ and the operator $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is unbounded. Then the definitions imply that we cannot have $\lambda \in \sigma_p(T)$ ($\lambda I - T$ is injective), $\lambda \in \sigma_r(T)$ ($\overline{\operatorname{Ran}(\lambda I - T)} = X$), or $\lambda \in \rho(T)$ (the inverse $(\lambda I - T)^{-1}$ is not a bounded operator). Because of this, we necessarily have $\lambda \in \sigma_c(T)$.

On the other hand, assume $\lambda \in \sigma_c(T)$. Then by definition the operator $\lambda I - T$ is injective, $\overline{\operatorname{Ran}(\lambda I - T)} = X$, and $\operatorname{Ran}(\lambda I - T) \neq X$. The injectivity implies that $\lambda I - T$ has a well-defined algebraic inverse $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$. Assume on the contrary that this inverse is a bounded operator, i.e., there exists $M > 0$ such that $\|(\lambda I - T)^{-1} y\| \leq M\|y\|$ for all $y \in \operatorname{Ran}(\lambda I - T)$. Our aim is to show that $\operatorname{Ran}(\lambda I - T) = X$, which will contradict our assumption that $\lambda \in \sigma_c(T)$. Let $y \in X$ be arbitrary. Since $\overline{\operatorname{Ran}(\lambda I - T)} = X$, there exists a sequence $(y_k)_{k=1}^{\infty} \subset \operatorname{Ran}(\lambda I - T)$

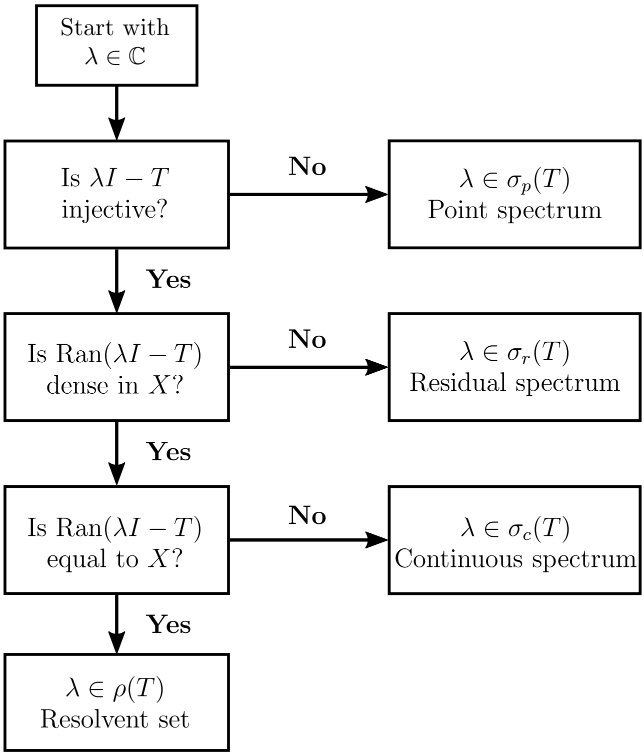

Figure 4.1: Parts of the spectrum [NS82, Fig. 6.5.1].

such that $y_k \to y$ as $k \to \infty$ (in X). By definition of $\mathrm{Ran}(\lambda I - T)$, there exist $(x_k)_{k=1}^{\infty} \subset X$ such that $y_k = (\lambda I - T)x_k$ for all $k \in \mathbb{N}$. Using the boundedness of the inverse $(\lambda I - T)^{-1} : \mathrm{Ran}(\lambda I - T) \subset X \to X$ we have

$$\begin{aligned} \|x_k - x_m\| &= \|(\lambda I - T)^{-1}(\lambda I - T)(x_k - x_m)\| \\ &= \|(\lambda I - T)^{-1}(y_k - y_m)\| \\ &\leq M\|y_k - y_m\| \to 0, \qquad \text{as} \quad k, m \to \infty \end{aligned}$$

since $(y_k)_{k=1}^{\infty}$ is a Cauchy sequence (as a convergent sequence), and thus $(x_k)_{k=1}^{\infty} \subset X$ is a Cauchy sequence as well. Since X is a Banach space, the sequence $(x_k)_{k=1}^{\infty} \subset X$ converges to a limit $x \in X$. Using $y_k = (\lambda I - T)x_k$ we then have that

$$\begin{aligned} \|y - (\lambda I - T)x\| &= \|y - y_k + y_k - (\lambda I - T)x\| \\ &= \|(y - y_k) + [(\lambda I - T)x_k - (\lambda I - T)x]\| \\ &\leq \|y - y_k\| + \|(\lambda I - T)(x_k - x)\| \\ &\leq \|y - y_k\| + \|\lambda I - T\|\|x_k - x\| \longrightarrow 0, \qquad \text{as} \quad k \to \infty \end{aligned}$$

since $y_k \to y$ as $k \to \infty$ and $x_k \to x$ as $k \to \infty$. Since $\|y - (\lambda I - T)x\|$ is independent of k, we must have $y = (\lambda I - T)x \in \mathrm{Ran}(\lambda I - T)$. Since $y \in X$ was arbitrary, we have shown that $\mathrm{Ran}(\lambda I - T) = X$, which contradicts our assumption $\lambda \in \sigma_c(T)$. Because of this our original claim holds. □

The following lemma provides a useful sufficient condition for proving the unboundedness of the inverse $(\lambda I - T)^{-1} : \mathrm{Ran}(\lambda I - T) \subset X \to X$ in Theorem 4.5.

In the literature the sequence $(x_k)_{k=1}^\infty$ appearing in Lemma 4.6 is referred to as an **approximate eigenvector** of T (corresponding to λ). Combining Theorem 4.5 and Lemma 4.6 we can see that $\lambda \in \mathbb{C}$ is in the continuous spectrum of T if and only if $\lambda I - T$ is injective, $\overline{\operatorname{Ran}(\lambda I - T)} = X$ and T has an approximate eigenvector $(x_k)_{k=1}^\infty$ corresponding to λ.

Lemma 4.6. *Let X be a Banach space and let $T \in B(X)$ and $\lambda \notin \sigma_p(T)$. The inverse $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is unbounded if and only if there exists $(x_k)_{k=1}^\infty \subset X$ such that $\|x_k\| = 1$ for all $k \in \mathbb{N}$ and*

$$\|(\lambda I - T)x_k\| \to 0, \qquad \text{as} \quad k \to \infty.$$

Proof. Since $\lambda \notin \sigma_p(T)$ by assumption, the inverse $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is well-defined and $(\lambda I - T)^{-1}(\lambda I - T)x = x$ for all $x \in X$.

Assume first that $(x_k)_{k=1}^\infty \subset X$ is as in the statement of the lemma. Then $(\lambda I - T)x_k \neq 0$ for all $k \in \mathbb{N}$ since $\lambda I - T$ is injective. If we define

$$y_k = \frac{1}{\|(\lambda I - T)x_k\|}(\lambda I - T)x_k \in \operatorname{Ran}(\lambda I - T), \qquad \forall k \in \mathbb{N},$$

then $\|y_k\| = 1$ for all $k \in \mathbb{N}$, but

$$\begin{aligned} \|(\lambda I - T)^{-1} y_k\| &= \|(\lambda I - T)^{-1}\frac{(\lambda I - T)x_k}{\|(\lambda I - T)x_k\|}\| = \frac{\|x_k\|}{\|(\lambda I - T)x_k\|} \\ &= \frac{1}{\|(\lambda I - T)x_k\|} \to \infty \end{aligned}$$

as $k \to \infty$. Thus the operator $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is not bounded.

Conversely, if $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is unbounded we can find $(y_k)_{k=1}^\infty \subset \operatorname{Ran}(\lambda I - T)$ such that $\|y_k\| = 1$ for all $k \in \mathbb{N}$ and $\|(\lambda I - T)^{-1}y_k\| \geq k$. Defining $x_k = \|(\lambda I - T)^{-1}y_k\|^{-1}(\lambda I - T)^{-1}y_k$, we have $\|x_k\| = 1$ for all $k \in \mathbb{N}$, and

$$\|(\lambda I - T)x_k\| = \|(\lambda I - T)\frac{(\lambda I - T)^{-1}y_k}{\|(\lambda I - T)^{-1}y_k\|}\| = \frac{\|y_k\|}{\|(\lambda I - T)^{-1}y_k\|} \leq \frac{1}{k} \to 0$$

as $k \to \infty$. □

Example 4.7 (The operators $I \in B(X)$ and $O \in B(X)$)**.** As our first example, we will study the spectrum of the identity operator $I \in B(X)$ and the zero operator $O \in B(X)$ on a Banach space X. For any $\lambda \in \mathbb{C}$ we have

$$\lambda I - I = (\lambda - 1)I,$$

and this operator is boundedly invertible if and only if $\lambda - 1 \neq 0$. Thus $\rho(I) = \mathbb{C} \setminus \{1\}$. On the other hand, if $\lambda = 1$, then the operator $\lambda I - I = O$ is not injective, and thus $1 \in \sigma_p(I)$. In fact, every nonzero element $x \in X$, $x \neq 0$, is an eigenvector of I corresponding to the eigenvalue $\lambda = 1$, since $\operatorname{Ker}(1I - I) = \operatorname{Ker}(O) = X$.

Similarly, that fact that for all $\lambda \in \mathbb{C}$ we have

$$\lambda I - O = \lambda I$$

implies that $\rho(O) = \mathbb{C} \setminus \{0\}$, and since $0I - O = O$ is not injective, we have $0 \in \sigma_p(O)$. ◇

Example 4.8 (Multiplication Operator on $\ell^p(\mathbb{C})$ for $1 \le p \le \infty$)**.** In this example we study the spectrum of an *multiplication operator* on $X = \ell^p(\mathbb{C})$ with $1 \le p < \infty$. We already considered such a "infinite diagonal matrix" in Example 2.30. We let $(\lambda_k)_{k=1}^\infty \in \ell^\infty(\mathbb{C})$ and define $T \in B(X)$ such that

$$Tx = (\lambda_k x_k)_{k=1}^\infty \in X, \qquad \forall x = (x_k)_{k=1}^\infty \in X.$$

We will show that the spectrum of the operator satisfies $\sigma(T) = \overline{\{ \lambda_k \mid k \in \mathbb{N} \}}$, i.e., the closure of the set of points $\{ \lambda_k \mid k \in \mathbb{N} \} \subset \mathbb{C}$. More precisely, each of the points $\lambda_k \in \mathbb{C}$ for $k \in \mathbb{N}$ turns out to be an eigenvalue of the operator T (similarly as in the case of diagonal matrices), and those accumulation points of the set $\{ \lambda_k \mid k \in \mathbb{N} \}$ which do not belong to the set are in the continuous spectrum $\sigma_c(T)$ (which does not appear in the case of matrices). In this example we will focus on these limit points, and leave the study of the eigenvalues and the resolvent set of T as an exercise (Exercise 4.9).

Let $\lambda \in \mathbb{C}$ be such that $\lambda \in \overline{\{ \lambda_k \mid k \in \mathbb{N} \}}$ but $\lambda \notin \{ \lambda_k \mid k \in \mathbb{N} \}$ (i.e., λ is a limit point of the set). Our aim is to prove that $\lambda \in \sigma_c(T)$ by showing that $\overline{\mathrm{Ran}(\lambda I - T)} = X$ but $\mathrm{Ran}(\lambda I - T) \neq X$. To prove that $\mathrm{Ran}(\lambda I - T)$ is dense in X, we will show that the space

$$\ell^{fin}(\mathbb{C}) = \left\{ (x_k)_{k=1}^\infty \in \ell(\mathbb{C}) \;\middle|\; \exists N \in \mathbb{N} : k \ge N \Rightarrow x_k = 0 \right\}$$

of sequences with at most finite number of nonzero elements is contained in $\mathrm{Ran}(\lambda I - T)$. Since $\ell^{fin}(\mathbb{C})$ is dense in $\ell^p(\mathbb{C})$ (verifying this is a straightforward exercise), this will prove that $\overline{\mathrm{Ran}(\lambda I - T)} = X$. To this end, let $y = (y_k)_{k=1}^\infty \in \ell^{fin}(\mathbb{C})$ be arbitrary. By definition there exists $N \in \mathbb{N}$ such that $y_k = 0$ for all $k \ge N$. Since we assumed that $\lambda \notin \{ \lambda_k \mid k \in \mathbb{N} \}$, we have $\lambda - \lambda_k \neq 0$ for all $k \in \mathbb{N}$. Thus we can write

$$y = (y_k)_k = \left(\frac{\lambda - \lambda_k}{\lambda - \lambda_k} y_k \right)_k = (\lambda I - T) \left(\frac{y_k}{\lambda - \lambda_k} \right)_k = (\lambda I - T)(x_k)_k$$

where we have defined $x_k = y_k/(\lambda - \lambda_k)$ for all $k \in \mathbb{N}$. We have $x_k = 0$ for all $k \ge N$, and thus $(x_k)_{k=1}^\infty \in \ell^{fin}(\mathbb{C}) \subset \ell^p(\mathbb{C})$. Thus the above computation shows that $y \in \mathrm{Ran}(\lambda I - T)$. Since $y \in \ell^{fin}(\mathbb{C})$ was arbitrary, we indeed have $\ell^{fin}(\mathbb{C}) \subset \mathrm{Ran}(\lambda I - T) \subset X$. The denseness of $\ell^{fin}(\mathbb{C})$ in X now implies that $\overline{\mathrm{Ran}(\lambda I - T)} = X$ (and in particular $\lambda \notin \sigma_r(T)$).

Finally, in showing $\lambda \in \sigma_c(T)$ we can use Theorem 4.5 and Lemma 4.6. These results show that since $\lambda \notin \sigma_p(T)$ and $\mathrm{Ran}(\lambda I - T)$ is dense in X, we have $\lambda \in \sigma_c(T)$ if (and only if) there exists $(x_n)_{n=1}^\infty \subset X$ such that $\|x_n\| = 1$ for all $n \in \mathbb{N}$ and $\|(\lambda I - T)x_n\| \to 0$ as $n \to \infty$. Since λ is by assumption a limit point of the set $\{ \lambda_k \mid k \in \mathbb{N} \}$, there exist a sequence $(\lambda_{k_n})_{n=1}^\infty \subset \{ \lambda_k \mid k \in \mathbb{N} \}$ such that $\lambda_{k_n} \to \lambda$ as $n \to \infty$. If we now define $x_n = e_{k_n} \in \ell^p(\mathbb{C})$ (where e_{k_n} is the sequence whose k_nth element is 1 and the other elements are zero), then clearly $\|x_n\| = \|e_{k_n}\| = 1$ for all $n \in \mathbb{N}$. Moreover, $(\lambda I - T)x_n$ is a sequence whose k_nth element is equal to $\lambda - \lambda_{k_n}$, and the other elements are zero. Thus we have

$$\|(\lambda I - T)x_n\|_p = |\lambda - \lambda_{k_n}| \to 0, \qquad \text{as} \quad n \to \infty.$$

By Theorem 4.5 and Lemma 4.6 we deduce that $\lambda \in \sigma_c(T)$. Since we let $\lambda \in \overline{\{ \lambda_k \mid k \in \mathbb{N} \}} \setminus \{ \lambda_k \mid k \in \mathbb{N} \}$ be arbitrary, $\overline{\{ \lambda_k \mid k \in \mathbb{N} \}} \setminus \{ \lambda_k \mid k \in \mathbb{N} \} \subset \sigma_c(T)$. However, by Exercise 4.9 below all other points of $\mathbb{C}$ belong to either $\sigma_p(T) = \{ \lambda_k \mid k \in \mathbb{N} \}$ or $\rho(T) = \mathbb{C} \setminus \overline{\{ \lambda_k \mid k \in \mathbb{N} \}}$. Therefore the complete description of the spectrum is

$$\begin{aligned}
\sigma_p(T) &= \{ \lambda_k \mid k \in \mathbb{N} \} \\
\sigma_c(T) &= \overline{\{ \lambda_k \mid k \in \mathbb{N} \}} \setminus \{ \lambda_k \mid k \in \mathbb{N} \} \\
\sigma_r(T) &= \varnothing \\
\rho(T) &= \mathbb{C} \setminus \overline{\{ \lambda_k \mid k \in \mathbb{N} \}}.
\end{aligned}$$

◇

Exercise 4.9. Prove that the spectrum of the operator $T \in B(X)$ in Example 4.8 satisfies $\sigma_p(T) = \{ \lambda_k \mid k \in \mathbb{N} \}$ and $\mathbb{C} \setminus \overline{\{ \lambda_k \mid k \in \mathbb{N} \}} \subset \rho(T)$. ◇

Exercise 4.10 (Rank One Operators)**.** Let X be a Banach space. If $T \in B(X)$ is such that $\dim(\operatorname{Ran}(T)) = 1$, then $\operatorname{Ran}(T) = \{ \alpha y \mid \alpha \in \mathbb{C} \}$ for some fixed $y \neq 0$. Thus every $T \in B(X)$ with $\dim(\operatorname{Ran}(T)) = 1$ has the has the form

$$Tx = \phi(x)y, \qquad \forall x \in X,$$

where $y \in X$ and $\phi \in X'$ are such that $y \neq 0$ and $\phi \neq 0$. Describe the parts of the spectrum and the resolvent set of such operators T. (You can begin by considering the point spectrum and treat the cases $\lambda = 0$ and $\lambda \neq 0$ separately). ◇

4.2 Properties of the Spectrum

For $\lambda \in \rho(T)$ the bounded inverse $(\lambda I - T)^{-1} \in B(X)$ has a special name. This **resolvent operator** $R(\lambda, T)$ appears often in various proofs related to spectral theory of linear operators.

Definition 4.11. Let X be a Banach space and let $T \in B(X)$. For $\lambda \in \rho(T)$ the **resolvent operator** $R(\lambda, T)$ is defined as

$$R(\lambda, T) = (\lambda I - T)^{-1} \in B(X).$$

In this section we will investigate selected general properties of the spectrum and the resolvent set of a bounded operator $T \in B(X)$. In particular, we will learn that the spectrum $\sigma(T)$ is always a compact (closed and bounded) subset of the complex plane, and correspondingly the resolvent set $\rho(T)$ is an open set. Moreover, both $\sigma(T)$ and $\rho(T)$ are nonempty sets (we will only prove the latter property, see [TL80, Thm. V.3.2] for the proof of $\sigma(T) \neq \varnothing$).[2] The proofs of these results utilise the **Neumann Series**, which tells us that the operator $I - T$ is boundedly invertible whenever $T \in B(X)$ satisfies $\|T\| < 1$. This result generalises the familiar *geometric series* $(1 - z)^{-1} = \sum_{n=0}^{\infty} z^n$ for real or complex number z with $|z| < 1$.

[2]These properties and the fact that $\sigma(T)$ is a bounded subset of $\mathbb{C}$ are true for all bounded operators $T \in B(X)$. The concept of *spectrum* can also be generalised to the case of possibly unbounded operators $T : D(T) \subset X \to X$ (this topic is studied on "MATH.MA.830 Advanced Functional Analysis"). For unbounded operators it is possible that either $\sigma(T) = \varnothing$ or $\rho(T) = \varnothing$, or that both $\sigma(T)$ and $\rho(T)$ are unbounded subsets of $\mathbb{C}$.

Theorem 4.12. *Assume that X is a Banach space and that $T \in B(X)$ satisfies $\|T\| < 1$. Then the operator $I - T$ is boundedly invertible and its inverse is given by the* ***Neumann Series***

$$(I-T)^{-1} = \sum_{n=0}^{\infty} T^n$$

(the series converges in the operator norm $\|\cdot\|_{B(X)}$). The norm of the inverse satisfies

$$\|(I-T)^{-1}\| \le \frac{1}{1-\|T\|}.$$

Proof. Since $\|T\| < 1$, we have

$$\sum_{n=0}^{\infty} \|T^n\| \le \sum_{n=0}^{\infty} \|T\|^n = \frac{1}{1-\|T\|} \tag{4.2}$$

(the sum of a geometric series of real numbers) and thus the series $\sum_{n=0}^{\infty} T^n$ is absolutely convergent. Since $B(X)$ is a Banach space, this implies that the series converges in $B(X)$ (see Exercise 3.41), and thus $\sum_{n=0}^{\infty} T^n = S$ for some $S \in B(X)$.

To prove that $S = (I-T)^{-1}$ we need to show that $S(I-T) = I$ and $(I-T)S = I$. We can do this by studying the partial sums of the series S, i.e., $S_N = \sum_{n=0}^{N} T^n$ for $N \in \mathbb{N}$. We have

$$(I-T)S_N = \sum_{n=0}^{N} (I-T)T^n = \sum_{n=0}^{N} (T^n - T^{n+1}) = I - T^{N+1}$$
$$S_N(I-T) = \sum_{n=0}^{N} T^n(I-T) = \sum_{n=0}^{N} (T^n - T^{n+1}) = I - T^{N+1}$$

(all but the first and last terms in the sums cancel out). Since $\|T\| < 1$, we have $\|T^{N+1}\| \le \|T\|^{N+1} \to 0$ as $N \to \infty$, and therefore

$$\|S_N(I-T) - I\| = \|T^{N+1}\| \to 0$$
$$\|(I-T)S_N - I\| = \|T^{N+1}\| \to 0$$

as $N \to \infty$. On the other hand, since $S_N \to S$ as $N \to \infty$, we also have

$$\|S_N(I-T) - S(I-T)\| = \|(S_N - S)(I-T)\| \le \|S_N - S\|\|I-T\| \to 0$$
$$\|(I-T)S_N - (I-T)S\| = \|(I-T)(S_N - S)\| \le \|I-T\|\|S_N - S\| \to 0$$

as $N \to \infty$. Thus (using the uniqueness of the limit of a convergent series)

$$S(I-T) = \lim_{N\to\infty} S_N(I-T) = I \qquad \text{and} \qquad (I-T)S = \lim_{N\to\infty} (I-T)S_N = I,$$

and we indeed have $(I-T)^{-1} = S = \sum_{n=0}^{\infty} T^n \in B(X)$.

It remains to show that $\|(I-T)^{-1}\| \le (1-\|T\|)^{-1}$ The continuity of the norm, the triangle inequality, and the estimate (4.2) imply that

$$\|(I-T)^{-1}\| = \left\| \lim_{N\to\infty} \sum_{n=0}^{N} T^n \right\| = \lim_{N\to\infty} \left\| \sum_{n=0}^{N} T^n \right\| \le \lim_{N\to\infty} \sum_{n=0}^{N} \|T^n\| = \sum_{n=0}^{\infty} \|T^n\| \le \frac{1}{1-\|T\|}.$$

□

Even though the Neumann series "only" concerns the inverse of a very specific type of operator, $I - T$ with $\|T\| < 1$, and the series itself is typically not even useful in computing this inverse, the Neumann series can be used as quite a versatile theoretical tool. Indeed, Theorem 4.12 becomes very useful through the following identities:

$$\lambda I - T = \lambda \left(I - \frac{1}{\lambda} T \right), \qquad \text{(Proof of Theorem 4.13)}$$

$$\lambda I - (T + S) = (\lambda I - T)\left(I - R(\lambda, T)S\right), \qquad \text{(Theorems 4.13 and 4.18)}$$

The following result shows that the spectrum $\sigma(T)$ is a compact (closed and bounded) subset of $\mathbb{C}$ and the resolvent set $\rho(T)$ is open.

Theorem 4.13. *Let X be a Banach space and let $T \in B(X)$. The spectrum $\sigma(T)$ is closed (as a subset of $\mathbb{C}$) and the resolvent set $\rho(T)$ is open. The spectrum $\sigma(T)$ is a bounded set contained in the closed disk $\{ \lambda \in \mathbb{C} \mid |\lambda| \le \|T\| \}$ (with center $0 \in \mathbb{C}$ and radius $\|T\|$), and $\rho(T) \neq \varnothing$.*

Proof. We begin by showing that the resolvent set $\rho(T)$ is open (which will also imply that $\sigma(T) = \mathbb{C} \setminus \rho(T)$ is closed). To this end, let $\lambda \in \rho(T)$ be fixed. We will show that $\lambda + \mu \in \rho(T)$ if $\mu \in \mathbb{C}$ is such that $|\mu|$ is sufficiently small. We can use $(\lambda I - T)R(\lambda, T) = I$ to write

$$(\lambda + \mu)I - T = (\lambda I - T) + \mu I = (\lambda I - T)\left(I + \mu R(\lambda, T)\right).$$

Here $\lambda I - T$ is boundedly invertible, and Theorem 4.12 implies that also the operator $I + \mu R(\lambda, T) = I - (-\mu R(\lambda, T))$ has a bounded inverse whenever

$$\|-\mu R(\lambda, T)\| < 1 \qquad \Leftrightarrow \qquad |\mu| < \frac{1}{\|R(\lambda, T)\|}.$$

Thus by Exercise 2.29 the operator $(\lambda + \mu)I - T$ is boundedly invertible, i.e., $\lambda + \mu \in \rho(T)$, whenever $|\mu| < 1/\|R(\lambda, T)\|$. Since $\lambda \in \rho(T)$ was arbitrary, we have proved that $\rho(T)$ is open and $\sigma(T)$ is closed in $\mathbb{C}$.

In the second part of the proof we will show that $|\lambda| \le \|T\|$ for all $\lambda \in \sigma(T)$. To this end, let $\lambda \in \mathbb{C}$ be such that $|\lambda| > \|T\|$. If we write

$$\lambda I - T = \lambda \left(I - \frac{1}{\lambda} T \right),$$

then $\|\frac{1}{\lambda} T\| = \|T\|/|\lambda| < 1$ together with Theorem 4.12 and Exercise 2.29 imply that $\lambda I - T$ is boundedly invertible, and thus $\lambda \in \rho(T)$. Since $\lambda \in \mathbb{C}$ with $|\lambda| > \|T\|$ was arbitrary, we have that $|\lambda| \le \|T\|$ for all $\lambda \in \sigma(T)$. Since $\|T\| < \infty$, this also implies that $\sigma(T) \neq \mathbb{C}$, and therefore $\rho(T) = \mathbb{C} \setminus \sigma(T) \neq \varnothing$. □

Exercise 4.14. Let X be a Banach space and let $T \in B(X)$. Use Theorem 4.12 to derive an expression for the resolvent $R(\lambda, T)$ when $|\lambda| > \|T\|$, and prove that

$$\|R(\lambda, T)\| \le \frac{1}{|\lambda| - \|T\|}, \qquad \text{whenever} \quad |\lambda| > \|T\|.$$

Further deduce that $\|R(\lambda, T)\| \to 0$ as $|\lambda| \to \infty$. ◇

Example 4.15 (The Left Shift Operator S_l on $\ell^p(\mathbb{C})$)**.** In this example we will study the spectrum of the *left shift operator* in Example 2.4 and on the space $X = \ell^p(\mathbb{C})$ with $1 \le p < \infty$. The operator $S_l \in B(X)$ was defined as

$$S_l x = (x_2, x_3, x_4 \dots), \qquad x = (x_k)_{k=1}^\infty.$$

The boundedness of S_l was considered in Exercise 2.20, and we in fact have $\|S_l\| = 1$. By Theorem 4.13 the spectrum $\sigma(S_l)$ is contained in the closed disk centered at $0 \in \mathbb{C}$ and with radius 1. In particular, $\{ \lambda \in \mathbb{C} \mid |\lambda| > 1 \} \subset \rho(S_l)$. Our aim is to show that $\sigma(S_l)$ is precisely this closed disk, i.e., $\sigma(S_l) = \{ \lambda \in \mathbb{C} \mid |\lambda| \le 1 \}$, and that every $\lambda \in \mathbb{C}$ with $|\lambda| < 1$ is an eigenvalue of S_l.

By definition, $\lambda \in \mathbb{C}$ is an eigenvalue of S_l if there exists a nonzero sequence $x = (x_k)_{k=1}^\infty \in X$, $x \ne 0$, such that $(\lambda I - S_l)x = 0$, or $S_l x = \lambda x$. Using the definition of S_l we have

$$(x_2, x_3, x_4, \dots) = (\lambda x_1, \lambda x_2, \lambda x_3, \dots),$$

and thus the elements x_k of the eigenvector x must satisfy

$$\begin{aligned} x_2 &= \lambda x_1 \\ x_3 &= \lambda x_2 = \lambda^2 x_1 \\ x_4 &= \lambda x_3 = \lambda^3 x_1 \\ &\vdots \end{aligned}$$

i.e., $x_k = \lambda^{k-1} x_1$ for all $k \in \mathbb{N}$. Thus every eigenvector $x \in X$ corresponding to $\lambda \in \mathbb{C}$ has the form $x = x_1(1, \lambda, \lambda^2, \lambda^3, \dots)$ with some $x_1 \in \mathbb{C} \setminus \{0\}$. Correspondingly, every vector of the form $x = x_1(1, \lambda, \lambda^2, \lambda^3, \dots)$ with $x_1 \ne 0$ is an eigenvector of S_l corresponding to $\lambda \in \mathbb{C}$ *provided that x belongs to the space X*. We can already see that the sizes of the elements $|\lambda^k| = |\lambda|^k$ of x decay very rapidly if $|\lambda| < 1$. Indeed, if $|\lambda| < 1$, we have $|\lambda|^p < 1$, and

$$\sum_{k=1}^\infty |\lambda^{k-1} x_1|^p = |x_1|^p \sum_{k=1}^\infty |\lambda|^{p(k-1)} = |x_1|^p \sum_{k=1}^\infty (|\lambda|^p)^{k-1} = |x_1|^p \frac{1}{1 - |\lambda|^p} < \infty$$

(the sum of the geometric series). Thus for every $\lambda \in \mathbb{C}$ with $|\lambda| < 1$ the vector $x = (1, \lambda, \lambda^2, \dots) \in X$, $x \ne 0$, satisfies $(\lambda I - S_l)x = 0$, and thus $\{ \lambda \in \mathbb{C} \mid |\lambda| < 1 \} \subset \sigma_p(S_l)$.

On the other hand, if $\lambda \in \mathbb{C}$ is such that $|\lambda| = 1$, then the absolute value of every element of the sequence $x = x_1(1, \lambda, \lambda^2, \lambda^3, \dots)$ with $x_1 \ne 0$ is equal to $|x_1|$, and therefore $x \notin X = \ell^p(\mathbb{C})$. Because every eigenvector of S_l is required to have this particular form, we can deduce that $\lambda \notin \sigma_p(S_l)$ if $|\lambda| = 1$. Because of this, we have $\sigma_p(S_l) = \{ \lambda \in \mathbb{C} \mid |\lambda| < 1 \}$. More detailed analysis in Exercise 4.16 will reveal that the boundary of the disk belongs to the continuous spectrum of S_l, i.e., $\sigma_p(S_l) = \{ \lambda \in \mathbb{C} \mid |\lambda| = 1 \}$. ◇

Exercise 4.16. Prove that $\sigma_c(S_l) = \{ \lambda \in \mathbb{C} \mid |\lambda| = 1 \}$ and $\sigma_r(S_l) = \varnothing$ for the left shift operator in Example 4.15. (Show that the $\lambda I - S_l$ has dense range when $|\lambda| = 1$ and use the other spectral properties of S_l). ◇

Example 4.17. Even though we only defined the spectrum $\sigma(T)$ and $\rho(T)$ for a bounded operator $T \in B(X)$, we will now take a quick look at the spectrum of an unbounded operator $T : D(T) \subset X \to X$. We will do this to illustrate how Theorem 4.13 depends on the assumption of boundedness of the operator. To this end, we consider the differential operator $T : D(T) \subset C([0,1]) \to C([0,1])$ with domain $D(T) = C^1([0,1])$ defined by $Tf = f'$ for all $f \in D(T)$. If we want to find the eigenvalues $\lambda \in \mathbb{C}$ and corresponding eigenvectors (or eigen*functions*) $f \in D(T)$, we arrive at the differential equation

$$(\lambda I - T)f = 0 \qquad \Leftrightarrow \qquad f'(t) = \lambda f(t), \quad t \in [0,1].$$

For any $\lambda \in \mathbb{C}$ this differential equation has a solution $f(t) = ce^{\lambda t}$ with $c \in \mathbb{C}$ satisfying $f \in D(T) = C^1([0,1])$ and $f \neq 0$ (when $c \neq 0$). Thus every $\lambda \in \mathbb{C}$ is an eigenvalue of T, which implies that $\sigma(T) = \sigma_p(T) = \mathbb{C}$ and $\rho(T) = \varnothing$. Differential operators can have various different kinds of spectral properties, and these properties are studied in greater detail on the course "MATH.MA.830 Advanced Functional Analysis". ◇

The following theorem is a "**perturbation result**" which concerns the change of the spectrum of the operator T when this operator is "*perturbed*" to $T + S$.

Theorem 4.18. *Let X be a Banach space and let $T \in B(X)$ and $\lambda \in \rho(T)$. If $S \in B(X)$ is such that $\|S\| < 1/\|R(\lambda,T)\|$, then $\lambda \in \rho(T+S)$. Morover, the difference between the resolvents of T and $S+T$ satisfies*

$$\|R(\lambda, T+S) - R(\lambda,T)\| \leq \frac{\|S\|\|R(\lambda,T)\|^2}{1 - \|R(\lambda,T)\|\|S\|}.$$

Proof. Let $\lambda \in \rho(T)$ be fixed and let $S \in B(X)$ be such that $\|S\| < 1/\|R(\lambda,T)\|$. Using $(\lambda I - T)R(\lambda,T) = I$ we can write

$$\lambda I - (T+S) = (\lambda I - T)\left(I - R(\lambda,T)S\right).$$

Theorem 4.12 and Exercise 2.29 imply that $\lambda I - (T+S)$ is boundedly invertible since $\lambda \in \rho(T)$ and $\|R(\lambda,T)S\| \leq \|R(\lambda,T)\|\|S\| < 1$ By Exercise 2.29 and Theorem 4.12, the resolvent operator $R(\lambda, T+S)$ can be written in the form

$$R(\lambda, T+S) = (I - R(\lambda,T)S)^{-1} R(\lambda,T) = \left(\sum_{n=0}^{\infty} (R(\lambda,T)S)^n\right) R(\lambda,T)$$

Using this expression to prove the bound for the norm $\|R(\lambda, T+S) - R(\lambda,T)\|$ is left as an exercise. □

Exercise 4.19. Complete the proof of Theorem 4.18 by showing that

$$\|R(\lambda, T+S) - R(\lambda,T)\| \leq \frac{\|S\|\|R(\lambda,T)\|^2}{1 - \|R(\lambda,T)\|\|S\|}$$

whenever $\|S\| < 1/\|R(\lambda,T)\|$. ◇

4.2.1 Further Properties of the Resolvent Operator [Optional]

Sometimes it is very useful to consider the resolvent operator $R(\lambda, T) = (\lambda I - T)^{-1} \in B(X)$ as a **function of the complex parameter** $\lambda \in \rho(T)$. According to this interpretation, the mapping $R(\cdot, T) : \rho(T) \subset \mathbb{C} \to B(X)$ is a (nonlinear) "*operator-valued function of a complex parameter* λ". The following theorem shows that the resolvent operator is a continuous function of $\lambda \in \rho(T)$[3].

Theorem 4.20. *Let X be a Banach space and let $T \in B(X)$. The function $\lambda \mapsto R(\lambda, T) \in B(X)$ is continuous on $\rho(T)$.*

Proof. Let $\mu \in \rho(T)$ be fixed. We will show that $\|R(\lambda, T) - R(\mu, T)\| \to 0$ as $\lambda \to \mu$. For this we can use the perturbation result in Theorem 4.18. Indeed, for $\lambda \in \mathbb{C}$ we can write

$$\lambda - T = \lambda - \mu + \mu - T = \mu - (T + (\mu - \lambda)I) = \mu - (T + S) \tag{4.3}$$

where we have defined $S = (\mu - \lambda)I \in B(X)$. Since $\|S\| = |\lambda - \mu|$, Theorem 4.18 implies that if $\lambda \in \mathbb{C}$ is such that $|\lambda - \mu| < 1/\|R(\mu, T)\|$, then $\lambda \in \rho(T)$ and (using (4.3))

$$\begin{aligned}\|R(\lambda, T) - R(\mu, T)\| = \|R(\mu, T + S) - R(\mu, T)\| &\leq \frac{\|S\|\|R(\mu, T)\|^2}{1 - \|R(\mu, T)\|\|S\|} \\ &= \frac{|\lambda - \mu|\|R(\mu, T)\|^2}{1 - \|R(\mu, T)\||\lambda - \mu|} \to 0\end{aligned}$$

as $\lambda \to \mu$. Thus $\lambda \mapsto R(\lambda, T)$ is continuous at $\mu \in \rho(T)$. Since $\mu \in \rho(T)$ was arbitrary, the proof is complete. □

Note that the perturbation result in Theorem 4.18 also implies continuity of the resolvent $R(\lambda, T)$ "with respect to the operator T". To see this, we can let $T \in B(X)$ and $\lambda \in \rho(T)$ be fixed and consider a sequence $(T_k)_{k=1}^\infty \subset B(X)$ such that $\|T_k - T\| \to 0$ as $k \to \infty$. Theorem 4.18 then implies that $\lambda \in \rho(T_k)$ for all sufficiently large $k \in \mathbb{N}$, and we have

$$\|R(\lambda, T_k) - R(\lambda, T)\| \leq \frac{\|T_k - T\|\|R(\lambda, T)\|^2}{1 - \|R(\lambda, T)\|\|T_k - T\|} \to 0$$

as $k \to \infty$.

Exercise 4.21 (The Resolvent Identity)**.** Let X be a Banach space and let $T \in B(X)$. Prove that for all $\lambda, \mu \in \rho(T)$ the resolvent operators of T satisfy

$$R(\lambda, T) - R(\mu, T) = (\mu - \lambda)R(\lambda, T)R(\mu, T) \qquad \textbf{(Resolvent Identity)}$$

and

$$R(\lambda, T)R(\mu, T) = R(\mu, T)R(\lambda, T).$$

◇

[3] In fact, much more is true! The mapping $\lambda \mapsto R(\lambda, T)$ is in fact *analytic* on $\rho(T)$, meaning that it has well-defined complex derivatives of all orders. But in order to establish this property, it is also necessary to define precisely what analyticity means for an operator-valued function.

4.3 Spectral Properties of Self-Adjoint Operators

In the remaining sections of the chapter we will take a closer look at some special properties of operators on a Hilbert space X. We begin by investigating self-adjoint operators (i.e., $T \in B(X)$ satisfying $T^* = T$) in this section, and move on to studying the spectra of **compact operators** in the next section. Our ultimate goal is the **spectral theorem** for compact self-adjoint operators which is presented in Section 4.5. Both special classes of self-adjoint operators and compact operators have very strong spectral theories, and on this course we only have time to study a fraction of the results which are available for these types of operators. However, we will keep our main goal in mind, and focus on those results which are required in proving the spectral theorem at the end of this chapter. Additional results on the nice spectral properties of compact and self-adjoint operators can be found, for example, in [NS82, Ch. 6], [TL80, Ch. V], and [Kre89, Ch. 8]. Throughout the rest of this chapter we will rely heavily on the Hilbert space results which we studied in Chapter 3.

Theorem 4.22. *Let X be a Hilbert space and let $T \in B(X)$ be self-adjoint. The spectrum of T is a subset of the real line, $\sigma(T) \subset [-\|T\|, \|T\|] \subset \mathbb{R}$, and $\sigma_r(T) = \varnothing$.*

Proof. We begin by showing that the residual spectrum of T is empty. We can do this by showing that for every $\lambda \in \mathbb{C}$ the injectivity of $\lambda I - T$ implies that $\mathrm{Ran}(\lambda I - T)$ is dense in X. To this end, let $\lambda \in \mathbb{C}$ and assume $\mathrm{Ker}(\lambda I - T) = \{0\}$. Exercise 3.19(c) tells us that $\overline{\mathrm{Ran}(\lambda I - T)} = X$ if and only if $\mathrm{Ran}(\lambda I - T)^\perp = \{0\}$. Because of this, we let $y \in \mathrm{Ran}(\lambda I - T)^\perp$ be arbitrary and aim to show that $y = 0$. By definition, we have (using $T^* = T$)

$$\begin{aligned}\langle (\lambda I - T)x, y\rangle = 0 \qquad \forall x \in X \qquad &\Leftrightarrow \qquad \langle x, (\overline{\lambda} I - T)y\rangle = 0 \qquad \forall x \in X\\ &\Leftrightarrow \qquad (\overline{\lambda} I - T)y = 0.\end{aligned}$$

However, the self-adjointness of T also implies

$$\begin{aligned}0 &= \|(\overline{\lambda} I - T)y\|^2 = \langle (\overline{\lambda} I - T)y, (\overline{\lambda} I - T)y\rangle = \langle (\lambda I - T)(\overline{\lambda} I - T)y, y\rangle\\ &= \langle (\overline{\lambda} I - T)(\lambda I - T)y, y\rangle = \langle (\lambda I - T)y, (\lambda I - T)y\rangle = \|(\lambda I - T)y\|^2.\end{aligned}$$

This immediately implies that $y = 0$, since $\lambda I - T$ was assumed to be injective. Since $y \in \mathrm{Ran}(\lambda I - T)^\perp$ was arbitrary, we have $\mathrm{Ran}(\lambda I - T)^\perp = \{0\}$, and thus $\mathrm{Ran}(\lambda I - T)$ is dense in X. This completes the proof that $\sigma_r(T) = \varnothing$.

The property $\sigma_r(T) = \varnothing$ implies that any spectral point of T must either be an eigenvalue or be in the continuous spectrum of T. Because of this, for proving $\sigma(T) \subset \mathbb{R}$ it is sufficient to show that if $\lambda \notin \mathbb{R}$, then $\lambda \notin \sigma_p(T)$ and $\lambda \notin \sigma_c(T)$. This part of the proof is left as an exercise (Exercise 4.23). Finally, the property $\sigma(T) \subset \mathbb{R}$ combined with Theorem 4.13 implies that $\sigma(T) \subset [-\|T\|, \|T\|]$. □

Exercise 4.23. Let X be a Hilbert space and let $T \in B(X)$ be self-adjoint. In this exercise we will complete the proof of Theorem 4.22 by showing that $\sigma(T) \subset \mathbb{R}$. We can assume to already know that $\sigma_r(T) = \varnothing$.

(a) Show that $\lambda \in \mathbb{C}$ satisfies $\lambda \in \sigma(T)$ if and only if there exists $(x_k)_{k=1}^\infty \subset X$ such that $\|x_k\| = 1$ for all $k \in \mathbb{N}$ and $\|(\lambda I - T)x_k\| \to 0$ as $k \to \infty$. (Hint: Results in Section 4.1).

(b) Show that $\lambda \in \rho(T)$ for all $\lambda \in \mathbb{C}$ with $\operatorname{Im}\lambda \neq 0$. (Hint: For $b \neq 0$, expand $\|((a+ib)I - T)x\|^2$ to show that the sequence in part (a) cannot exist).

◇

The following theorem shows that the spectrum of a self-adjoint operator T is not only contained in the interval $[-\|T\|, \|T\|]$, but at least one of the endpoints of the interval is in the spectrum $\sigma(T)$. This immediately implies that if $\sigma(T) = \{0\}$, then also $\|T\| = 0$ and therefore $T = O$. This latter property really requires the assumption of self-adjointness, since for example the matrix $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ satisfies $\sigma(A) = \sigma_p(A) = \{0\}$, but of course $A \neq 0$.

Theorem 4.24. *Let X be a Hilbert space and let $T \in B(X)$ be self-adjoint. Then $\lambda \in \sigma(T)$ where $\lambda = \|T\|$ or $\lambda = -\|T\|$ (or both). In particular, if $\sigma(T) = \{0\}$, then $T = O$.*

Proof. The second claim follows immediately from the first one. We have from Theorem 3.38 that $\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|$. By definition of the supremum, there exists a sequence $(x_k)_{k=1}^{\infty} \subset X$ such that $\|x_k\| = 1$ for all $k \in \mathbb{N}$ and $|\langle Tx_k, x_k\rangle| \to \|T\|$ as $k \to \infty$. Since T is self-adjoint, $\langle Tx_k, x_k\rangle \in \mathbb{R}$ (by Exercise 3.36), and thus there necessarily exists a subsequence $(x_{k_n})_{n=1}^{\infty}$ such that $\langle Tx_{k_n}, x_{k_n}\rangle \to \lambda$ where $\lambda = \|T\|$ or $\lambda = -\|T\|$. We then have (using $\langle x_{k_n}, Tx_{k_n}\rangle = \langle Tx_{k_n}, x_{k_n}\rangle$ and $\|x_{k_n}\| = 1$)

$$\begin{aligned}
\|(\lambda I - T)x_{k_n}\|^2 &= \langle (\lambda I - T)x_{k_n}, (\lambda I - T)x_{k_n}\rangle \\
&= \lambda^2 \|x_{k_n}\|^2 - 2\lambda\langle Tx_{x_n}, x_{k_n}\rangle + \|Tx_{k_n}\|^2 \\
&\leq \lambda^2 - 2\lambda\langle Tx_{x_n}, x_{k_n}\rangle + \|T\|^2 \\
&= 2\lambda^2 - 2\lambda\langle Tx_{x_n}, x_{k_n}\rangle \longrightarrow 2\lambda^2 - 2\lambda^2 = 0
\end{aligned}$$

as $n \to \infty$. Thus $(x_{k_n})_{n=1}^{\infty}$ is a sequence such that $\|x_{k_n}\| = 1$ for all $n \in \mathbb{N}$ and $\|(\lambda I - T)x_{k_n}\| \to 0$ as $n \to \infty$. Because of this, either $\lambda I - T$ is not injective, or the inverse $(\lambda I - T)^{-1} : \operatorname{Ran}(\lambda I - T) \subset X \to X$ is unbounded by Lemma 4.6. In both cases we have $\lambda \in \sigma(T)$. □

Lemma 4.25. *Let X be a Hilbert space and let $T \in B(X)$ be self-adjoint. If $\lambda, \mu \in \sigma_p(T)$ are such that $\lambda \neq \mu$, then $\operatorname{Ker}(\lambda I - T) \perp \operatorname{Ker}(\mu I - T)$.*

Proof. Assume $\lambda, \mu \in \sigma_p(T)$ are such that $\lambda \neq \mu$ and let $x \in \operatorname{Ker}(\lambda I - T)$ and $y \in \operatorname{Ker}(\mu I - T)$. By Theorem 4.22 we have $\lambda, \mu \in \mathbb{R}$. Thus

$$\langle Tx, y\rangle = \langle \lambda x, y\rangle = \lambda\langle x, y\rangle.$$

However, we can alternatively use $T^* = T$ to compute

$$\langle Tx, y\rangle = \langle x, Ty\rangle = \langle x, \mu y\rangle = \mu\langle x, y\rangle,$$

and thus $\lambda\langle x, y\rangle = \mu\langle x, y\rangle$. But since $\lambda \neq \mu$ by assumption, we must have $\langle x, y\rangle = 0$. Since $x \in \operatorname{Ker}(\lambda I - T)$ and $y \in \operatorname{Ker}(\mu I - T)$ were arbitrary, we indeed have $\operatorname{Ker}(\lambda I - T) \perp \operatorname{Ker}(\mu I - T)$. □

Example 4.26. In this example we continue the study of the integral operator in Examples 2.21 and 3.39 on the space $X = L^2(a,b)$. We assume the kernel satisfies $(t,s) \mapsto k(t,s) \in C(\Omega)$ where $\Omega = [a,b] \times [a,b] \subset \mathbb{R}^2$ and define $T \in B(X)$ by

$$(Tf)(t) = \int_a^b k(t,s)f(s)\,\mathrm{d}s, \qquad \forall t \in [a,b].$$

In the previous examples we already saw that T satisfies $\|T\| \le \|k(\cdot,\cdot)\|_{L^2(\Omega)}$ and that T is self-adjoint if $k(t,s) = \overline{k(s,t)}$ for all $t,s \in [a,b]$.

In this example we describe a process for computing the *nonzero* eigenvalues of T in the case where the kernel $k(\cdot,\cdot)$ is **separable** in the sense that it has the structure

$$k(t,s) = \sum_{j=1}^n k_j(t)h_j(s)$$

for some $n \in \mathbb{N}$ and for some functions $(k_j)_{j=1}^n \subset C([a,b])$ and $(h_j)_{j=1}^n \subset C([a,b])$ where $(k_j)_{j=1}^n$ and $(h_j)_{j=1}^n$ are both linearly independent sets. In this example we focus on the situation with $n = 2$ where $k(t,s) = k_1(t)h_1(s) + k_2(t)h_2(s)$ and the generalisation of the method is completed in Exercise 4.28.

If $\lambda \in \mathbb{C} \setminus \{0\}$ is an eigenvalue of T, then there exists $f \in X$ with $f \neq 0$ such that $(\lambda I - T)f = 0$, i.e., for all $t \in [a,b]$ we have

$$\begin{aligned}
\lambda f(t) = (Tf)(t) = \int_a^b k(t,s)f(s)\,\mathrm{d}s &= \int_a^b (k_1(t)h_1(s) + k_2(t)h_2(s))f(s)\,\mathrm{d}s \\
&= k_1(t)\int_a^b h_1(s)f(s)\,\mathrm{d}s + k_2(t)\int_a^b h_2(s)f(s)\,\mathrm{d}s \\
&= k_1(t)\langle h_1, \overline{f}\rangle_{L^2} + k_2(t)\langle h_2, \overline{f}\rangle_{L^2}
\end{aligned}$$

(where we denote $\overline{f}$ for the function with values $\overline{f(t)}$) Since $\lambda \neq 0$ and since the integrals in the last expression are scalars, the above identity implies that f is a linear combination of the (linearly independent) functions k_1 and k_2, i.e., $f = \alpha_1 k_1 + \alpha_2 k_2$ for some $\alpha_1, \alpha_2 \in \mathbb{C}$. If we substitute this form of f to the above equation, we see that

$$\begin{aligned}
& \lambda f = \langle h_1, \overline{f}\rangle k_1 + \langle h_2, \overline{f}\rangle k_2 \\
\Leftrightarrow \quad & \lambda(\alpha_1 k_1 + \alpha_2 k_2) = \langle h_1, \overline{\alpha_1 k_1 + \alpha_2 k_2}\rangle k_1 + \langle h_2, \overline{\alpha_1 k_1 + \alpha_2 k_2}\rangle k_2 \\
\Leftrightarrow \quad & \lambda\alpha_1 k_1 + \lambda\alpha_2 k_2 = (\alpha_1\langle h_1, \overline{k_1}\rangle + \alpha_2\langle h_1, \overline{k_2}\rangle)k_1 + (\alpha_1\langle h_2, \overline{k_1}\rangle + \alpha_2\langle h_2, \overline{k_2}\rangle)k_2.
\end{aligned}$$

Since k_1 and k_2 were assumed to be linearly independent, this last identity holds precisely if the coefficients of the functions k_1 and k_2 are identical on both sides of the equation. This condition leads to the pair of equations

$$\begin{cases} \lambda\alpha_1 = \alpha_1\langle h_1, \overline{k_1}\rangle + \alpha_2\langle h_1, \overline{k_2}\rangle \\ \lambda\alpha_2 = \alpha_1\langle h_2, \overline{k_1}\rangle + \alpha_2\langle h_2, \overline{k_2}\rangle \end{cases} \qquad \Leftrightarrow \qquad \begin{bmatrix} \langle h_1, \overline{k_1}\rangle & \langle h_1, \overline{k_2}\rangle \\ \langle h_2, \overline{k_1}\rangle & \langle h_2, \overline{k_2}\rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \lambda \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}.$$

Thus $\lambda \in \mathbb{C} \setminus \{0\}$ and $[\alpha_1, \alpha_2]^T \in \mathbb{C}^2$ are eigenvalues and corresponding eigenvectors of the matrix with elements

$$\langle h_j, \overline{k_m}\rangle = \int_a^b h_j(s)k_m(s)\,\mathrm{d}s, \qquad \text{for} \quad j,m \in \{1,2\}.$$

The above arguments now show that every solution $\lambda \neq 0$ of this eigenvalue problem is an eigenvalue of T as well, and the corresponding eigenvector is given by $f = \alpha_1 k_1 + \alpha_2 k_2$.

Note that in the general case the eigenvalue problem for λ and $[\alpha_1, \alpha_2]^T \in \mathbb{C}^2$ can have either 2, 1, or even 0 nonzero eigenvalues $\lambda \neq 0$. Indeed, it is quite possible that the coefficient matrix is the zero matrix even for nonzero functions k_j and h_m (if $\overline{k_j}$ and h_m are orthogonal for $j, m \in \{1, 2\}$).

Finally, we note that T is in particular self-adjoint if $k_1, k_2, h_1, h_2 \in C([a, b])$ are such that $h_1(t) = \beta_1 \overline{k_1(t)}$ and $h_2(t) = \beta_2 \overline{k_2(t)}$ for some $\beta_1, \beta_2 \in \mathbb{R} \setminus \{0\}$. In this case we have $k(t, s) = \beta_1 k_1(t)\overline{k_1(s)} + \beta_2 k_2(t)\overline{k_2(s)}$ and the eigenvalue problem has the form

$$A = \begin{bmatrix} \langle h_1, \overline{k_1} \rangle & \langle h_1, \overline{k_2} \rangle \\ \langle h_2, \overline{k_1} \rangle & \langle h_2, \overline{k_2} \rangle \end{bmatrix} = \begin{bmatrix} \beta_1 \langle \overline{k_1}, \overline{k_1} \rangle & \beta_1 \langle \overline{k_1}, \overline{k_2} \rangle \\ \beta_2 \langle \overline{k_2}, \overline{k_1} \rangle & \beta_2 \langle \overline{k_2}, \overline{k_2} \rangle \end{bmatrix} = \begin{bmatrix} \beta_1 \|k_1\|^2 & \beta_1 \langle k_2, k_1 \rangle \\ \beta_2 \langle k_1, k_2 \rangle & \beta_2 \|k_2\|^2 \end{bmatrix}.$$

It turns out that this matrix has the full number of linearly independent eigenvectors, and it does not have a zero eigenvalue. To see that the latter claim holds, it is sufficient to note that if A is singular, then its determinant $\det(A) = \beta_1\beta_2(\|k_1\|^2\|k_2\|^2 - |\langle k_1, k_2 \rangle|^2)$ is zero. But this means that the Cauchy–Schwarz Inequality is "satisfied as an equality", and by Exercise 3.5 this would imply that $k_1 = \beta k_2$ for some $\beta \in \mathbb{C}$, which contradicts the assumption that k_1 and k_2 are linearly independent. ◇

Exercise 4.27. Apply the method in Example 4.26 to find the eigenvalues and normalised eigenvectors of the intergral operator on $X = L^2(0, 1)$ with $k(t, s) = 1 - 3ts$ for $t, s \in [0, 1]$. Verify that the eigenfunctions are orthogonal. You can compute the necessary integrals and find the eigenvalues and eigenvectors using Matlab, Wolfram Alpha, or some other software (after all, this is not a course on linear algebra!). ◇

Exercise 4.28. Generalise the process of finding eigenvalues of an integral operator in Example 4.26 to the case $k(t, s) = \sum_{j=1}^{n} k_j(t)h_j(s)$ with $n \in \mathbb{N}$. ◇

4.4 Spectral Properties of Compact Operators

In this section we study the spectrum of another special class of operators, namely, **compact operators**.

Definition 4.29. Let X and Y be Banach spaces. An operator $T \in B(X, Y)$ is **compact** if every bounded sequence $(x_k)_{k=1}^{\infty} \subset X$ has a subsequence $(x_{k_n})_{n=1}^{\infty}$ such that $(Tx_{k_n})_{n=1}^{\infty} \subset Y$ converges in Y.

Compact operators can also be defined in a more general situation where X and Y are not necessarily complete spaces[4].

[4] Our definition also already utilises a characterisation of compactness of a set on a normed space in terms of bounded sequences containing convergent subsequences (what is known as *sequential compactness*). In the literature, the definition of a compact operator is often presented in another form which states that $T \in B(X, Y)$ (with Y Banach) is a compact operator if *for every bounded set* $A \subset X$ *the closure* $\overline{T(A)} := \overline{\{ Tx \mid x \in A \}}$ *is a compact set in* Y.

Lemma 4.30. *Let X, Y, and Z be Banach spaces.*

(a) *If $T, S \in B(X,Y)$ are compact and $\alpha \in \mathbb{C}$, then $\alpha T \in B(X,Y)$ and $T + S \in B(X,Y)$ are compact operators.*

(b) *If $T \in B(X,Y)$ and $S \in B(Y,Z)$, then $ST \in B(X,Z)$ is compact if either T or S is compact.*

(c) *If $(T_k)_{k=1}^{\infty} \subset B(X,Y)$ and $T \in B(X,Y)$ are such that T_k are compact for all $k \in \mathbb{N}$ and $\|T_k - T\| \to 0$ as $k \to \infty$, then T is compact.*

Proof. Left as an exercise. □

Exercise 4.31. Complete the proof of Lemma 4.30. ◇

Exercise 4.32. Let X be a Banach space, let $T \in B(X)$ be compact, and assume there exists a closed subspace Z of X such that $Tx \in Z$ for all $x \in Z$. Define $S : Z \to Z$ as the restriction of T to Z, i.e., $Sx = Tx$ for all $x \in Z$. Prove that $S \in B(Z)$ is compact[5]. ◇

Example 4.33 (Finite Rank Operators)**.** Let X be a Hilbert space. An operator $T \in B(X)$ is said to have **finite rank** if there exists a linearly independent set $\{y_k\}_{k=1}^{\infty} \subset X$ and $\{z_k\}_{k=1}^{\infty} \subset X$ such that T has the structure

$$Tx = \sum_{k=1}^{n} \langle x, z_k \rangle y_k, \qquad \forall x \in X. \tag{4.4}$$

Since every $y \in \mathrm{Ran}(T)$ is a linear combination of the fixed vectors $\{y_k\}_{k=1}^{n} \subset X$, and therefore $\dim(\mathrm{Ran}(T)) \le n < \infty$. In fact, any operator $T \in B(X)$ with $\mathrm{Ran}(T) < \infty$ can be written in the form (4.4) for some $n \in \mathbb{N}$, $\{y_k\}_{k=1}^{n} \subset X$ and $\{z_k\}_{k=1}^{n} \subset X$. Our aim is to show that T is compact. By Lemma 4.30(a) the sums of compact operators is compact. Therefore it is sufficient to prove that if $y, z \in X$ are fixed and if we define $T_0 \in B(X)$ so that $T_0 x = \langle x, z \rangle y$ for all $x \in X$, then $T_0 \in B(X)$ is compact (T_0 is a "rank one operator" such as the one we already considered in Exercise 4.10).

If $(x_k)_{k=1}^{\infty} \subset X$ is an arbitrary bounded sequence, then $(\langle x_k, z \rangle)_{k=1}^{\infty} \subset \mathbb{C}$ is a bounded sequence in $\mathbb{C}$. But in the space of complex numbers, every bounded sequence has a convergent subsequence, and therefore there exist a subsequence $(x_{k_n})_{n=1}^{\infty}$ and $\alpha \in \mathbb{C}$ such that $\langle x_{k_n}, z \rangle \to \alpha$ as $n \to \infty$. Because of this, we also have

$$\|T_0 x_{k_n} - \alpha y\| = \|\langle x_{k_n}, z \rangle y - \alpha y\| = |\langle x_{k_n}, z \rangle - \alpha| \|y\| \to 0, \qquad \text{as} \quad n \to \infty.$$

Thus the sequence $(T_0 x_{k_n})_{n=1}^{\infty} \subset X$ converges to $\alpha y \in X$, and T_0 is compact. ◇

Lemma 4.34. *Let X be a Banach space. The identity operator $I \in B(X)$ is compact if and only if* $\dim X < \infty$*.*

Proof. We will only present the proof in a simplified case where X is a Hilbert space (the proof for the result in a normed space X case can be found for example in [TL80,

[5]In the case where $Tx \in Z$ for all $x \in Z$, the subspace Z is said to be **invariant under** T. In this case the restriction S of T to Z is typically called the **part of** T in Z.

Thm. II.3.6]). If $\dim X < \infty$, then $I \in B(X)$ is a "finite rank operator" such as the ones studied in Example 4.33 and thus compact. On the other hand, if $\dim X = \infty$, we can let $(x_k)_{k=1}^\infty \subset X$ be an orthonormal sequence, meaning that $\|x_k\| = 1$ and $\langle x_k, x_m \rangle = 0$ for all $k, m \in \mathbb{N}$ such that $k \neq m$ (such a sequence can be constructed for example by applying the finite-dimensional Gram–Schmidt process to finite but increasingly large sets of linearly independent vectors). For all $k \neq m$ we then have

$$\|x_k - x_m\|^2 = \|x_k\|^2 - \langle x_k, x_m \rangle - \langle x_m, x_k \rangle + \|x_m\|^2 = \|x_k\|^2 + \|x_m\|^2 = 2.$$

This implies that even though $(x_k)_{k=1}^\infty$ is a bounded sequence, the sequence $(Ix_k)_{k=1}^\infty$ cannot contain any convergent subsequence. Thus $I \in B(X)$ is not compact. □

The following theorem shows that a compact operator can have at most finite number of linearly independent eigenvectors associated to any nonzero eigenvalue. This result concerns *only* nonzero eigenvalues, since for example the zero operator $T = O \in B(X)$ on an infinite-dimensional space X is clearly compact, but it has an infinite number of eigenvectors associated to its only eigenvalue $\lambda = 0$.

Theorem 4.35. *Let X be a Banach space and let T be a compact operator. Then for every $\lambda \neq 0$ the space* $\mathrm{Ker}(\lambda I - T)$ *is finite-dimensional.*

Proof. For every $x \in \mathrm{Ker}(\lambda I - T)$ we have $(\lambda I - T)x = 0$, and $\lambda \neq 0$ implies that $\lambda^{-1}Tx = x \in \mathrm{Ker}(\lambda I - T)$. Since $\lambda^{-1}T \in B(X)$ is a compact operator (by Lemma 4.30(a)) and since $\mathrm{Ker}(\lambda I - T)$ is a closed subspace of X (see Exercise 2.25), the restriction of $\lambda^{-1}T$ to $\mathrm{Ker}(\lambda I - T)$ is compact by Exercise 4.32. However, $\lambda^{-1}Tx = x$ for all $x \in \mathrm{Ker}(\lambda I - T)$ implies that this restriction is in fact the identity operator on $\mathrm{Ker}(\lambda I - T)$. Therefore Lemma 4.34 implies that $\dim(\mathrm{Ker}(\lambda I - T)) < \infty$. □

The following lemma shows that a compact operator on an infinite-dimensional space cannot have a bounded inverse. This in particular implies that if $\dim X = \infty$ and $T \in B(X)$ is a compact operator, then 0 is always in the spectrum of T!

Lemma 4.36. *Assume X is a Banach space with* $\dim X = \infty$ *and let $T \in B(X)$ be a compact operator. If T is injective, then $T^{-1} : \mathrm{Ran}(T) \subset X \to X$ is unbounded.*

Proof. Let $\dim X = \infty$ and assume on the contrary that $T^{-1} : \mathrm{Ran}(T) \subset X \to X$ is bounded, i.e., there exists $M > 0$ such that $\|T^{-1}x\| \leq M\|x\|$ for all $x \in \mathrm{Ran}(T)$. Let $(x_k)_{k=1}^\infty \subset X$ be an arbitrary bounded sequence. Since T is compact, this sequence has a subsequence $(x_{k_n})_{n=1}^\infty$ such that $(Tx_{k_n})_{n=1}^\infty \subset X$ converges in X. For all $n, m \in \mathbb{N}$ we have

$$\|x_{k_n} - x_{k_m}\| = \|T^{-1}T(x_{k_n} - x_{k_m})\| \leq M\|Tx_{k_n} - Tx_{k_m}\| \to 0, \qquad \text{as } n, m \to \infty$$

since $(Tx_{k_n})_{n=1}^\infty$ is a Cauchy sequence. Therefore $(x_{k_n})_{n=1}^\infty$ is a Cauchy sequence as well, and since X is complete, the sequence $(x_{k_n})_{n=1}^\infty$ converges in X. But since $(x_k)_{k=1}^\infty \subset X$ was an arbitrary bounded sequence on X, we have shown that *every bounded sequence has a convergent subsequence.* However, this property is equivalent to the identity operator of X being compact. By Lemma 4.34 this can only happen if $\dim X < \infty$, which is a contradiction. □

There are no restrictions on the *type* of the spectral point $0 \in \sigma(T)$ of a compact operator T, meaning that 0 can be either be an eigenvalue, in the residual spectrum or in the continuous spectrum. In the next exercise we will see an interesting example of a compact operator whose spectrum consists of the single point $0 \in \mathbb{C}$ which is not an eigenvalue.

Exercise 4.37. Let $X = \ell^2(\mathbb{C})$ and consider the operator $T \in B(X)$ defined by

$$Tx = \left(0, \frac{x_1}{1}, \frac{x_2}{2}, \frac{x_3}{3}, \dots\right), \qquad \forall x = (x_k)_{k=1}^\infty \in X.$$

Prove T is a compact operator and that $\sigma(T) = \sigma_r(T) = 0$. ◇

Example 4.38 (Integral Operator)**.** The integral operators studied in Examples 2.21, 3.39 and 4.26 are imporant examples of compact operator. Indeed, if we assume the kernel of T satisfies $(t,s) \mapsto k(t,s) \in C([a,b] \times [a,b])$ and define $T \in B(X)$ by

$$(Tf)(t) = \int_a^b k(t,s) f(s)\,\mathrm{d}s, \qquad \forall t \in [a,b],$$

then T is a compact operator. The proof of the compactness of T is fairly challenging, and because of this it is presented separately later in Example 4.44 (and this proof is optional material on the course). ◇

Example 4.39 (Multiplication Operator on $\ell^p(\mathbb{C})$)**.** In this example we consider the *multiplication operator* on $X = \ell^p(\mathbb{C})$ with $1 \le p < \infty$ which we have already studied in Examples 2.30 and 4.8. We let $(\lambda_k)_{k=1}^\infty \in \ell^\infty(\mathbb{C})$ and define $T \in B(X)$ such that

$$Tx = (\lambda_k x_k)_{k=1}^\infty \in X, \qquad \forall x = (x_k)_{k=1}^\infty \in X.$$

We will show that T is compact if $\lambda_k \to 0$ as $k \to \infty$. We can do this by showing that T is a limit of a sequence of finite rank operators $(T_N)_{N \in \mathbb{N}}$. Indeed, if we can define $T_N \in B(X)$ such that $\lim_{N\to\infty} \|T_N - T\| = 0$ and $\dim \operatorname{Ran}(T_N) < \infty$ for all $N \in \mathbb{N}$, then each T_N is compact by Example 4.33, and Lemma 4.30(c) implies that T must be compact.

For any fixed $N \in \mathbb{N}$, we can define T_N so that

$$T_N x = \left(\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_N x_N, 0, 0, \dots\right)$$

for all $x = (x_k)_k \in X$. Thus the first N components of the sequence $T_N x$ are the same as those in Tx, and the rest are zeros. We have $T_N \in B(X)$ since T_N is also a multiplication operator (corresponding to the sequence $(\lambda_1, \dots, \lambda_N, 0, 0, \dots) \in \ell^\infty(\mathbb{C})$). Moreover, we have

$$\operatorname{Ran}(T_N) = \{\, y \mid y = T_N x, \text{ for some } x \in X \,\} \subset \{\, (y_k)_k \mid y_k = 0 \text{ for } k > N \,\}.$$

Since the last set is an N-dimensional subspace of X, Lemma 1.16 implies that $\dim \operatorname{Ran}(T_N) \le N < \infty$. Thus T_N is indeed a finite rank operator, and thus compact.

To show that $\|T_N - T\| \to 0$ as $N \to \infty$, let $x \in X$. We can estimate

$$\begin{aligned}
\|(T_N - T)x\|_X^p &= \left\| \left(\lambda_1 x_1, \dots, \lambda_N x_N, 0, \dots\right) - \left(\lambda_1 x_1, \dots, \lambda_N x_N, \lambda_{N+1} x_{N+1}, \dots\right) \right\|_{\ell^p}^p \\
&= \left\| \left(0, \dots, 0, \lambda_{N+1} x_{N+1}, \lambda_{N+2} x_{N+2}, \dots\right) \right\|_{\ell^p}^p = \sum_{k=N+1}^\infty |\lambda_k|^p |x_k|^p \\
&\le \left(\sup_{k>N} |\lambda_k|^p \right) \sum_{k=N+1}^\infty |x_k|^p \le \left(\sup_{k>N} |\lambda_k| \right)^p \sum_{k=1}^\infty |x_k|^p = \left(\sup_{k>N} |\lambda_k| \right)^p \|x\|^p.
\end{aligned}$$

This inequality further implies that the operator norm $\|T_N - T\|$ satisfies

$$\|T_N - T\| = \sup_{\|x\|\leq 1} \|(T_N - T)x\|_X \leq \sup_{\|x\|\leq 1} \left(\|x\| \sup_{k>N} |\lambda_k| \right) = \sup_{k>N} |\lambda_k| \to 0$$

as $N \to \infty$, since we assumed that $\lim_{k\to\infty} |\lambda_k| = 0$. This convergence together with Lemma 4.30(c) shows that T is compact. ◇

4.5 The Spectral Theorem for Compact Self-Adjoint Operators

In this final section of the chapter we will prove the "*spectral theorem*" for compact self-adjoint operators. This result (presented in Theorem 4.41) shows that every compact self-adjoint operator can be represented using its eigenvalues and corresponding orthogonal eigenvectors. This is not at all a common property for operators on infinite-dimensional spaces, since—as we have already seen—operators do not in general need to have eigenvalues at all. The spectral theorem is a very natural culmination of our investigation of the spectral properties of self-adjoint operators and compact operators, and the proofs of the two main results of this section utilise every main property which we proved for these two classes of operators in Sections 4.3 and 4.4, as well as several others results from Chapters 3 and 4.

We begin by presenting a more detailed description of spectrum (besides the point $0 \in \mathbb{C}$) of a compact and self-adjoint operator. This important result in Theorem 4.40 is in fact also true for all compact operators on Banach spaces (and even normed spaces which are not complete), but requires a slightly longer proof without the assumption that T is self-adjoint (see, for example [TL80, Thm. V.7.10] or [NS82, Thm. 6.10.2]). Our assumption that T is self-adjoint simplifies the proof, since we especially do not need to separately prove that $\lambda \notin \sigma_r(T)$ for all $\lambda \neq 0$. We also use this additional assumption and "Hilbert space techniques" in proving that the only possible accumulation point of $\sigma(T)$ is at $0 \in \mathbb{C}$, but there are also other ways to prove this same property (see, e.g. [NS82, Cor. 6.10.5], [Kre89, Thm. 8.3-1], or [TL80, Thm. V.7.10]).

Theorem 4.40. *Let X be a Hilbert space and let $T \in B(X)$ be a compact self-adjoint operator. For every $\lambda \neq 0$ we have*

- *either $\lambda \in \rho(T)$, or*
- *$\lambda \in \sigma_p(T)$ and $\dim(\mathrm{Ker}(\lambda I - T)) < \infty$.*

The only possible accumulation point of $\sigma(T)$ is $0 \in \mathbb{C}$.

Proof. We will first show that the continuous spectrum $\sigma_c(T)$ does not contain any nonzero values $\lambda \neq 0$. To this end, assume on the contrary that $\lambda \in \sigma_c(T)$ and $\lambda \neq 0$. By Lemma 4.6 there exists a sequence $(x_k)_{k=1}^\infty \subset X$ such that $\|x_k\| = 1$ for all $k \in \mathbb{N}$ and $\|(\lambda I - T)x_k\| \to 0$ as $k \to \infty$. Our aim is to show that λ is in fact an eigenvalue (i.e., $\lambda I - T$ is not injective), which will contradict our assumption that $\lambda \in \sigma_c(T)$. Since T is compact and since the sequence $(x_k)_{k=1}^\infty$ is bounded, there exists a subsequence $(x_{k_n})_{n=1}^\infty$ and a limit $x \in X$ such that $Tx_{k_n} \to x$ as $n \to \infty$. We will show that x is an eigenvector corresponding to λ. We have

$$\|\lambda x_{k_n} - x\| = \|\lambda x_{k_n} - Tx_{k_n} + Tx_{k_n} - x\| \leq \|(\lambda - T)x_{k_n}\| + \|Tx_{k_n} - x\| \to 0$$

as $n \to \infty$. Thus $\lambda x_{k_n} \to x$ as $n \to \infty$, and since $\lambda \neq 0$, we also have $x_{k_n} \to \lambda^{-1}x$ as $n \to \infty$. A direct computation now shows that

$$Tx = \lambda T\left(\frac{1}{\lambda}x\right) = \lambda T\left(\lim_{n\to\infty} x_{k_n}\right) = \lambda \cdot \lim_{n\to\infty} Tx_{k_n} = \lambda x.$$

Therefore we indeed have $(\lambda I - T)x = 0$, and since we also have

$$\|x\| = \|\lim_{n\to\infty} \lambda x_{k_n}\| = |\lambda| \lim_{n\to\infty} \|x_{k_n}\| = |\lambda| > 0,$$

we conclude that $\lambda I - T$ is not injective. This contradicts our assumption that $\lambda \in \sigma_c(T)$, and thus completes the proof that $\lambda \notin \sigma_c(T)$ for all $\lambda \neq 0$.

Since T is self-adjoint, we have $\sigma_r(T) = \varnothing$ by Theorem 4.22. Thus for all $\lambda \neq 0$ we have either $\lambda \in \rho(T)$ or $\lambda \in \sigma_p(T)$, and in the latter case $\operatorname{Ker}(\lambda I - T)$ is finite-dimensional by Theorem 4.35.

In the final part of the proof we will show that $\sigma(T)$ cannot have any accumulation points other than $\lambda = 0$. Since $\sigma(T)$ is closed by Theorem 4.13, any accumulation point is necessarily in $\sigma(T)$, and thus it is sufficient to show that $\lambda \in \sigma(T)$ cannot be an accumulation point of $\sigma(T)$ if $\lambda \neq 0$. To this end, let $\lambda \in \sigma(T) \setminus \{0\}$ be arbitrary. We will show that λ has a neighbourhood with no other spectral points of T, i.e., that there exists $\varepsilon > 0$ such that $\mu \in \rho(T)$ whenever $\mu \in \mathbb{C}$ and $0 < |\lambda - \mu| < \varepsilon$. Since T is compact and $\lambda \neq 0$, we have $\lambda \in \sigma_p(T)$ and $\operatorname{Ker}(\lambda I - T)$ is a finite-dimensional (and thus closed) subspace of X by Theorem 4.35. Denote $M = \operatorname{Ker}(\lambda I - T)$. We will split the operator T into two parts according to the decomposition $X = M \oplus M^\perp$. More precisely, we will show that there exist $S_1 \in B(M)$ and $S_2 \in B(M^\perp)$ such that if $x = x_1 + x_2 \in X$ with $x_1 \in M$ and $x_2 \in M^\perp$, then

$$Tx = S_1x_1 + S_2x_2. \tag{4.5}$$

We first note that if $x \in M = \operatorname{Ker}(\lambda I - T)$, then $Tx = \lambda x \in M$. Thus if we define S_1 as the restriction of T to M, then $S_1 \in B(M)$ and in fact $S_1 = \lambda I$. On the other hand, if $x \in M^\perp$ and if $y \in M$ is arbitrary, the self-adjointness of T implies that $\langle Tx, y\rangle = \langle x, Ty\rangle = 0$ (since $Ty \in M$), and thus $Tx \in M^\perp$ as well. Thus if we define S_2 as the restriction of T to $M^\perp$, we have $S_2 \in B(M^\perp)$. The property (4.5) does indeed hold, since if $x = x_1 + x_2$ where $x_1 \in M$ and $x_2 \in M^\perp$, then $Tx = Tx_1 + Tx_2 = S_1x_1 + S_2x_2$.

The operator $\lambda I - S_2$ is injective, since if $x_2 \in M^\perp$ is such that $0 = (\lambda I - S_2)x_2 = (\lambda I - T)x_2$, then also $x_2 \in M = \operatorname{Ker}(\lambda I - T)$, and $M \cap M^\perp = \{0\}$ implies $x_2 = 0$. Since $S_2 \in B(M^\perp)$ is compact by Exercise 4.32, the first part of the proof shows that necessarily $\lambda \in \rho(S_2)$. Because the resolvent set is open (Theorem 4.13), there exists $\varepsilon > 0$ such that $\mu \in \rho(S_2)$ and $\mu I - S_2$ is injective whenever $\mu \in \mathbb{C}$ is such that $|\lambda - \mu| < \varepsilon$. Now let $\mu \in \mathbb{C}$ be such that $0 < |\lambda - \mu| < \varepsilon$. We will show that $\mu I - T$ is injective. Indeed, if $x \in \operatorname{Ker}(\mu I - T)$, then $x = x_1 + x_2$ for some $x_1 \in M$ and $x_2 \in M^\perp$ and we have

$$0 = (\mu I - T)x = (\mu I - S_1)x_1 + (\mu I - S_2)x_2 = \underbrace{(\mu - \lambda)x_1}_{\in M} + \underbrace{(\mu I - S_2)x_2}_{\in M^\perp}$$

$$\Rightarrow \quad \begin{cases} (\mu - \lambda)x_1 = 0 \\ (\mu I - S_2)x_2 = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases} \quad \Rightarrow \quad x = 0$$

since $\lambda \neq \mu$ and since $\mu I - S_2$ is injective. But since T is compact, the first part of the proof and the injectivity of $\mu I - T$ imply that $\mu \in \rho(T)$. Because of this, we have proved that there exists $\varepsilon > 0$ such that $\mu \in \rho(T)$ whenever $0 < |\lambda - \mu| < \varepsilon$, and therefore λ cannot be an accumulation point of $\sigma(T)$. □

The result in Theorem 4.40 leads to a powerful result known as the **Fredholm Alternative** (and Theorem 4.40 is even called the *Fredholm Alternative Theorem* in some sources). This result concerns the solvability of the linear operator equation

$$(\lambda I - T)x = y, \tag{4.6}$$

where $T \in B(X)$ is a compact operator. The Fredholm Alternative (as an immediate consequence of Theorem 4.40) states that for any $\lambda \neq 0$, precisely one of two "*alternatives*" must hold:

(a) The equation (4.6) has a unique solution $x \in X$ for any vector $y \in X$, or

(b) the homogeneous equation $(\lambda I - T)z = 0$ has a nonzero solution $z \in X$.

These cases obviously correspond to the cases where either $\lambda \in \rho(T)$ or $\lambda \in \sigma_p(T)$, respectively. In the latter case, the number of linearly independent solutions z of $(\lambda I - T)z = 0$ is finite and the linear equation (4.6) has a solution if and only if $y \perp z$ for all $z \in \operatorname{Ker}(\lambda I - T)$ [NS82, Sec. 6.11, Ex. 2]. Finally, in the case (b) it also is possible to show that every solution of (4.6) has the form $x = x_0 + z$ where $(\lambda I - T)x_0 = y$ and $z \in \operatorname{Ker}(\lambda I - T)$ [NS82, Sec. 6.11, Ex. 2].

The most classical application of the Fredholm Alternative is the study of the integral equations, especially the *Fredholm equation of the second kind*, namely,

$$g(t) = f(t) - \mu \int_a^b k(t,s) f(s)\,\mathrm{d}s, \qquad t \in [a,b],$$

which correspond to the case where T is an integral operator and $\lambda = 1/\mu$. These types of integral equations were studied extensively by Fredholm himself. In addition, the Fredholm Alternative is also a very important tool in the study of linear partial differential equations, especially *elliptic equations*.

The next theorem is the main result of this section. This "**spectral theorem**" shows that every compact self-adjoint operator can be represented using its eigenvalues and corresponding orthonormal eigenvectors (i.e., the eigenvectors $\{e_k\}_k$ are such that e_k and e_j are orthogonal for all $k \neq j$, and $\|e_k\| = 1$ for all k). This representation can (and should!) be compared to the *spectral decomposition* of a Hermitian matrix. Indeed, the form of the operator T in Theorem 4.41 has precisely the form of an "*eigenfunction expansion*", but in the case of a compact and self-adjoint operator, the expansion may contain an infinite number of terms.

Theorem 4.41 (Spectral Theorem for Compact Self-Adjoint Operators)**.** *Let X be a Hilbert space and assume $T \in B(X)$ is a compact self-adjoint operator and $T \neq O$. Then there exists $N \in \mathbb{N}$ or $N = \infty$ such that*

$$Tx = \sum_{k=1}^{N} \lambda_k \langle x, e_k \rangle e_k, \qquad \forall x \in X,$$

where $\lambda_k \in \mathbb{R} \setminus \{0\}$ are nonzero eigenvalues of T and the following hold:

- *The values $|\lambda_k|$ are non-increasing and the multiplicity of each λ_k is finite.*
- *$\{e_k\}_{k=1}^N$ is an orthonormal set of eigenvectors so that $Te_k = \lambda_k e_k$ for all k.*

If $N = \infty$, then $\lambda_k \to 0$ as $k \to \infty$.

Proof. Since T is bounded and self-adjoint, we have that $\sigma_p(T) \subset [-\|T\|, \|T\|] \subset \mathbb{R}$ by Theorem 4.22. Moreover, since T is compact, Theorem 4.40 implies that every $\lambda \in \sigma(T)$ with $\lambda \neq 0$ is an eigenvalue of T and $\operatorname{Ker}(\lambda I - T)$ is a finite-dimensional subspace of X, and that $0 \in \mathbb{C}$ is the only (possible) accumulation point of $\sigma(T)$. Because of this, we can organise the nonzero eigenvalues of T into a sequence $(\lambda_k)_{k=1}^N$ (with $N \in \mathbb{N}$ or $N = \infty$) in the order of descending $|\lambda_k|$ in such a way that each eigenvalue λ_k in the sequence is repeated $\dim(\operatorname{Ker}(\lambda_k I - T))$ times. Since $\operatorname{Ker}(\lambda_k I - T)$ is finite-dimensional for every $k \in \mathbb{N}$, we can choose an orthonormal basis on each of these spaces (using the standard Gram–Schmidt process). We can further define a sequence $(e_k)_{k=1}^N \subset X$ to consists of the orthonormal basis vectors of the spaces $\operatorname{Ker}(\lambda_k I - T)$ which are listed in the same order as the eigenvalues $(\lambda_k)_{k=1}^N$. More precisely, we define $(e_k)_{k=1}^N \subset X$ so that

$$\{e_j\}_{j=k_0}^{k_0+n} \quad \text{orthonormal basis of } \operatorname{Ker}(\lambda_{k_0} I - T) \ \text{ if } \lambda_j = \lambda_{k_0} \text{ for } j \in \{k_0, \ldots, k_0 + n\}.$$

By construction, we have that $Te_k = \lambda_k e_k$ for all k. Moreover, since the subspaces $\operatorname{Ker}(\lambda_k I - T)$ and $\operatorname{Ker}(\lambda_m I - T)$ are orthogonal if $\lambda_k \neq \lambda_m$ by Lemma 4.25, the sequence $(e_k)_{k=1}^N$ is orthonormal, i.e., $\langle e_k, e_l \rangle = \delta_{kl}$ where δ_{kl} is the Dirac delta. Since by Theorem 4.40 the spectrum $\sigma(T)$ cannot have any accumulation points other than $0 \in \mathbb{C}$, we must have $\lambda_k \to 0$ as $k \to \infty$ if $N = \infty$.

In the second part of the proof we will show that $(e_k)_{k=1}^N$ is an orthonormal basis of $\operatorname{Ran}(T)$ (in the sense of Definition 3.43). We will do this by verifying that condition (d) of Theorem 3.49 holds, i.e., "if $y \in \operatorname{Ran}(T)$ is such that $\langle y, e_k \rangle = 0$ for all $k \in \{1, \ldots, N\}$, then necessarily $y = 0$". Let $y \in \operatorname{Ran}(T)$ be an arbitrary vector satisfying $\langle y, e_k \rangle = 0$ for all k. Then there exists $x \in X$ such that $y = Tx$ and the self-adjointness of T implies (using $\lambda_k \neq 0$)

$$\begin{aligned} \langle y, e_k \rangle = 0 \quad \forall k \qquad &\Leftrightarrow \qquad \langle Tx, e_k \rangle = 0 \quad \forall k \\ &\Leftrightarrow \qquad \langle x, Te_k \rangle = 0 \quad \forall k \\ &\Leftrightarrow \qquad \lambda_k \langle x, e_k \rangle = 0 \quad \forall k \\ &\Leftrightarrow \qquad \langle x, e_k \rangle = 0 \quad \forall k. \end{aligned}$$

If we define M as the subspace of all finite linear combinations of $\{e_k\}_{k=1}^N$, i.e.,

$$M = \{\, z = \sum_{k=1}^{n} \alpha_k e_k \mid n \in \mathbb{N},\ \alpha_1, \ldots, \alpha_n \in \mathbb{C} \,\},$$

then the equivalences above imply that $x \in M^\perp$. We want to show that $Tz \in M^\perp$ for all $z \in M^\perp$. To this end, let $z \in M^\perp$ be arbitrary. If $n \in \mathbb{N}$ and $(\alpha_k)_{k=1}^n \subset \mathbb{C}$, then

$$\langle Tz, \sum_{k=1}^n \alpha_k e_k \rangle = \langle z, T \sum_{k=1}^n \alpha_k e_k \rangle = \langle z, \sum_{k=1}^n \alpha_k T e_k \rangle = \langle z, \sum_{k=1}^n \alpha_k \lambda_k e_k \rangle = 0$$

since $\sum_{k=1}^n (\alpha_k \lambda_k) e_k \in M$ and since $z \in M^\perp$ by assumption. Thus we indeed have $Tz \in M^\perp$ for all $z \in M^\perp$ and we can therefore consider the restriction S of T to the closed subspace $M^\perp$ (defined by $Sz = Tz$ for all $z \in M^\perp$) as an operator $S \in B(M^\perp)$. We have from Exercise 4.32 that S is compact, and the self-adjointness of T implies that also $S^* = S$. We will next show that $\sigma(S) = \{0\}$, which will imply $S = O$ by Theorem 4.24. To do this, let $\lambda \in \sigma(S)$ be such that $\lambda \neq 0$. Theorem 4.40 implies that λ is an eigenvalue of S and thus there exists $z \in M^\perp$, $z \neq 0$, such that $(\lambda I - S)z = 0$. But since S is a restriction of T, we also have

$$(\lambda I - T)z = (\lambda I - S)z = 0,$$

which implies $z \in \operatorname{Ker}(\lambda I - T)$. But since $\lambda \neq 0$, every eigenvector corresponding to λ can be expressed as a linear combination of $\{ e_k \mid k \in \mathbb{N} \text{ such that } \lambda_k = \lambda \}$ (the orthonormal basis of $\operatorname{Ker}(\lambda I - T)$) and thus $z \in M$. We therefore have $z \in M \cap M^\perp = \{0\}$, which contradicts our assumption that $z \neq 0$. This means that $\sigma(S) = \{0\}$, and since S is self-adjoint, we have $S = O$ by Theorem 4.24. Since our original $y \in \operatorname{Ran}(T)$ satisfies $y = Tx$ with $x \in M^\perp$, we finally have $y = Tx = Sx = Ox = 0$. Because of this, Theorem 3.49 implies that $\{e_k\}_{k=1}^N$ is an orthonormal basis of $\operatorname{Ran}(T)$.

The final part of the proof is very simple: Since $\{e_k\}_{k=1}^N$ is an orthonormal basis of $\operatorname{Ran}(T)$, Theorem 3.49(b) implies that for any $x \in X$ the vector $Tx \in \operatorname{Ran}(T)$ has the "generalised Fourier series expansion" of the form (3.11), and thus

$$Tx = \sum_{k=1}^N \langle Tx, e_k \rangle e_k = \sum_{k=1}^N \langle x, Te_k \rangle e_k = \sum_{k=1}^N \langle x, \lambda_k e_k \rangle e_k = \sum_{k=1}^N \lambda_k \langle x, e_k \rangle e_k.$$

Since $x \in X$ was arbitrary, the proof is complete. □

The spectral representation of T in Theorem 4.41 also gives us a way to express the solution of the linear equation

$$(\lambda I - T)x = y$$

for $\lambda \neq 0$ whenever $\lambda \notin \sigma_p(T)$. Indeed, in Exercise 4.42 you will verify that if $\lambda \notin \sigma_p(T)$, then the inverse of $\lambda I - T$ has an explicit formula in terms of $\{\lambda_k\}_{k=1}^N$ and $\{e_k\}_{k=1}^N$, and the unique solution x of the linear equation is given by

$$x = (\lambda I - T)^{-1} y = \frac{1}{\lambda} y + \sum_{k=1}^N \frac{\lambda_k}{\lambda(\lambda - \lambda_k)} \langle y, e_k \rangle e_k,$$

where $N \in \mathbb{N}$ or $N = \infty$ depending on the number of nonzero eigenvalues of T.

Exercise 4.42. Prove that if T is a compact self-adjoint operator with the spectral representation in Theorem 4.41 and if $\lambda \in \rho(T)$, then

$$(\lambda I - T)^{-1} y = \frac{1}{\lambda} y + \sum_{k=1}^N \frac{\lambda_k}{\lambda(\lambda - \lambda_k)} \langle y, e_k \rangle e_k, \qquad \forall y \in X.$$

You do not need to prove that the operator determined by the right-hand side is bounded, though this can be done using the Bessel's Inequality (Exercise 3.46). ◇

Exercise 4.43. Consider the intergral operator on $X = L^2(0,1)$ with $k(t,s) = 1-3ts$ for $t,s \in [0,1]$ (Exercise 4.27).

(a) Construct the spectral decomposition of T.

(b) Determine the values $\mu \neq 0$ for which the Fredholm equation of the second kind

$$g(t) = f(t) - \mu \int_0^1 (1-3ts) f(s)\,\mathrm{d}s$$

has a unique solution for every $g \in L^2(0,1)$, and write down the formula for the solution $f(t)$.

◇

4.5.1 Compactness of Integral Operators [Optional]

Example 4.44 (Integral Operator)**.** In this example we prove that the integral operator studied in Example 4.38 (and in Examples 2.21, 3.39 and 4.26) is compact. We assume the kernel of T satisfies $(t,s) \mapsto k(t,s) \in C([a,b]\times[a,b])$ and define $T \in B(X)$ by

$$(Tf)(t) = \int_a^b k(t,s) f(s)\,\mathrm{d}s, \qquad \forall t \in [a,b].$$

We will prove the compactness of the operator T by proving that it can be obtained as a limit $T = \lim_{N\to\infty} T_N$ of compact operators. In particular, we will define T_N as finite rank operators such as the ones studied in Example 4.33. To this end, for $k \in \mathbb{Z}$ we define

$$e_k(t) = \frac{1}{\sqrt{b-a}} e^{i2\pi k \frac{(t-a)}{(b-a)}}$$

Then the sequence $(e_k)_{k=-\infty}^{\infty} \subset L^2(a,b)$ (i.e., $(\ldots, e_{-2}, e_{-1}, e_0, e_1, e_2, \ldots)$) consists of the complex Fourier basis functions on the interval $[a,b]$ and in particular it is an orthonormal basis of $L^2(a,b)$ (this basis is otherwise exactly as in Example 3.50 but it is scaled from the interval $[0,1]$ to $[a,b]$). In Exercise 4.45 we will see that if we further define $e_{kj} \in L^2([a,b]\times[a,b])$ such that $e_{kj}(t,s) = e_k(t)e_j(s)$ for all $k,j \in \mathbb{Z}$, then $(e_{kj})_{(k,j)\in\mathbb{Z}^2}$ with index set $\mathbb{Z}^2 = \{\,(k,j) \mid k,j \in \mathbb{Z}\,\}$ is an orthonormal basis of $L^2([a,b]\times[a,b])$. Because $k(\cdot,\cdot) \in L^2([a,b]\times[a,b])$, we then have from Theorem 3.49 that

$$k(t,s) = \sum_{k,j\in\mathbb{Z}} \langle k, e_{kj}\rangle_{L^2(\Omega)} e_{kj}(t,s), \tag{4.7}$$

where $\Omega = [a,b]\times[a,b]$. We will define the operators $T_N \in B(X)$ based on *truncations* of the above infinite sum. More precisely, for $N \in \mathbb{N}$ we define

$$k_N(t,s) = \sum_{|k|,|j|\leq N} \langle k, e_{kj}\rangle_{L^2(\Omega)} e_{kj}(t,s) \tag{4.8}$$

and define $T_N \in B(X)$ such that

$$(T_N f)(t) = \int_a^b k_N(t,s) f(s)\,\mathrm{d}s, \qquad \forall t \in [a,b].$$

Since $k_N(\cdot,\cdot) \in C([a,b]\times[a,b])$, our earlier analysis of integral operators in Example 2.21 shows that $T_N \in B(X)$. It turns out that the definition of $k_N(\cdot,\cdot)$ has the consequence that T_N are finite rank operators, and therefore compact by Example 4.33. Indeed, using the forms of $k_N(t,s)$ and $e_{kj}(t,s) = e_k(t)e_j(s)$ we can see that

$$\begin{aligned}(T_N f)(t) &= \int_a^b k_N(t,s)f(s)\,\mathrm{d}s = \int_a^b \sum_{|k|,|j|\le N} \langle k, e_{kj}\rangle_{L^2(\Omega)} e_{kj}(t,s)f(s)\,\mathrm{d}s \\ &= \sum_{|k|,|j|\le N} \langle k, e_{kj}\rangle_{L^2(\Omega)} e_k(t)\int_a^b e_j(s)f(s)\,\mathrm{d}s \\ &= \sum_{|k|,|j|\le N} \langle f, e_j\rangle_{L^2(a,b)}\langle k, e_{kj}\rangle_{L^2(\Omega)} e_k(t).\end{aligned}$$

If we define $z_{kj} = e_j \in L^2(a,b)$ and $y_{kj} = \langle k, e_{kj}\rangle_{L^2(\Omega)} e_k \in L^2(a,b)$ for all $k,j \in \{-N,\dots,N\}$, then

$$T_N f = \sum_{|k|,|j|\le N} \langle f, z_{kj}\rangle y_{kj}, \qquad \forall f \in L^2(a,b).$$

Thus T_N is is indeed a finite rank operator and therefore compact by Example 4.33.

Finally, we need to show that $\|T - T_N\| \to 0$ as $N \to \infty$. For an arbitrary $f \in L^2(a,b)$ and for all $t \in [a,b]$ we have

$$((T - T_N)f)(t) = (Tf)(t) - (T_N f)(t) = \int_a^b (k(t,s) - k_N(t,s))f(s)\,\mathrm{d}s$$

and thus $T - T_N$ is an integral operator with kernel $k(\cdot,\cdot) - k_N(\cdot,\cdot) \in C([a,b]\times[a,b])$. Because of this, we have from Example 2.21 that

$$\|T - T_N\| \le \left(\int_a^b \int_a^b |k(t,s) - k_N(t,s)|^2 dtds\right)^{\frac{1}{2}} = \|k(\cdot,\cdot) - k_N(\cdot,\cdot)\|_{L^2(\Omega)}.$$

Comparing (4.7) and (4.8) shows that

$$k(t,s) - k_N(t,s) = \sum_{\substack{|k|>N \text{ or} \\ |j|>N}} \langle k, e_{kj}\rangle_{L^2(\Omega)} e_{kj}(t,s).$$

Since $(e_{kj})_{(k,j)\in\mathbb{Z}^2}$ is an orthonormal basis of $L^2(\Omega)$, Theorem 3.49(c) implies that $\sum_{k,j\in\mathbb{Z}} |\langle k, e_{kj}\rangle|^2 = \|k(\cdot,\cdot)\|_{L^2}^2 < \infty$ and therefore (again by Theorem 3.49(c))

$$\|T - T_N\| \le \|k(\cdot,\cdot) - k_N(\cdot,\cdot)\|_{L^2(\Omega)} = \sum_{k,j\in\mathbb{Z}} |\langle k - k_N, e_{kj}\rangle|^2 = \sum_{\substack{|k|>N \text{ or} \\ |j|>N}} |\langle k, e_{kj}\rangle|^2 \to 0$$

as $N \to \infty$. Since T_N are compact, Lemma 4.30(c) finally implies that T is compact.

The results on the spectrum of compact operators in this section now imply that any nonzero eigenvalue of T can have at most finite number of linearly independent eigenfunctions, and if the inverse T^{-1} exists, then it is an unbounded operator. The latter is not very surprising to us, since the natural inverse of an integral operator is a differential operator. ◇

Exercise 4.45. Let $e_k \in L^2(a,b)$ be defined as

$$e_k(t) = \frac{1}{\sqrt{b-a}} e^{i2\pi k \frac{(t-a)}{(b-a)}}, \qquad \forall k \in \mathbb{Z},$$

and define $e_{kj}(t,s) = e_k(t)e_j(s)$ for all $k, j \in \mathbb{Z}$. Prove that $(e_{kj})_{(k,j)\in\mathbb{Z}^2}$ with index set $\mathbb{Z}^2 = \{\, (k,j) \mid k, j \in \mathbb{Z} \,\}$ is an orthonormal basis of $L^2([a,b] \times [a,b])$. (Hint: Verify that $(e_{kj})_{(k,j)\in\mathbb{Z}^2}$ is orthonormal and that the property in Theorem 3.49(d) holds). $\diamond$

A. Additional Results

A.1 Proof of the Minkowski Inequality

In this section we show that the p-norm defined in (1.2) on $C(\Omega)$ where $\Omega \subset \mathbb{R}^n$ is a compact set (closed and bounded) satisfies the triangle inequality. This result is known as the **Minkowski Inequality**. The proof is based on Lemmas A.1 (Young's Inequality for positive real numbers) and A.2 (Hölder's Inequality for continuous functions).

Lemma A.1 (Young's Inequality)**.** *Let $p > 1$ and $q > 1$ be real numbers satisfying $\frac{1}{p} + \frac{1}{q} = 1$. If $a > 0$ and $b > 0$, then*

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Proof. Consider the illustration of the function $y = x^{p-1}$ with $a^{p-1} < b$ in Figure A.1. The area of the rectangle with sides a and b is ab. The area of I is $\frac{1}{p}a^p$ (integrate

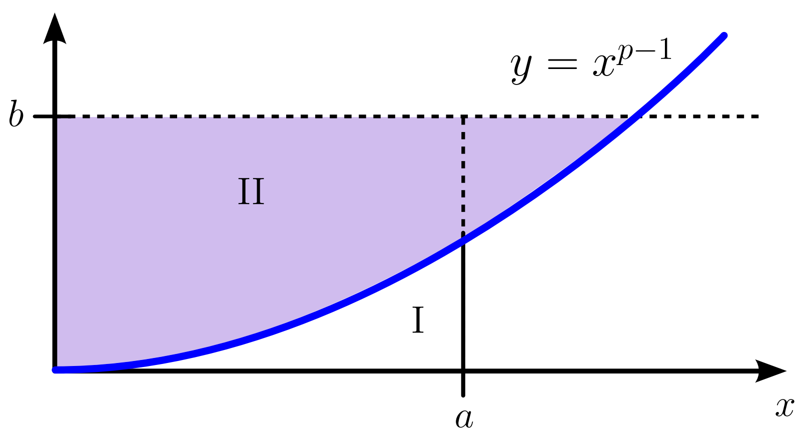

Figure A.1: The function $y = x^{p-1}$ with $a^{p-1} < b$.

the function $y = x^{p-1}$ from $x = 0$ to $x = a$) and the area of II is $\frac{1}{q}b^q$ (integrate the function $x = y^{1/(p-1)} = y^{q-1}$ (since $1/p + 1/q = 1$) from $y = 0$ to $y = b$). Thus, we obtain Young's inequality. If $a^{p-1} \geq b$ the proof works in the same way. □

Lemma A.2 (Hölder's Inequality)**.** *Let $\Omega \subset \mathbb{R}^n$ be compact and let $p > 1$, $q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then for every $f, g \in C(\Omega)$ we have*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Proof. Note first that $fg \in C(\Omega)$ whenever $f, g \in C(\Omega)$, since Ω is compact and f and g are uniformly continuous. The claim obviously holds if $f = 0$ or $g = 0$. We can thus assume $\|f\|_p > 0$ and $\|g\|_q > 0$. By Young's Inequality we have that for all $z \in \Omega$

$$\frac{|f(z)||g(z)|}{\|f\|_p\|g\|_q} \leq \frac{1}{p}\frac{|f(z)|^p}{\|f\|_p^p} + \frac{1}{q}\frac{|g(z)|^q}{\|g\|_q^q},$$

so that

$$\int_\Omega \frac{|f(z)g(z)|}{\|f\|_p\|g\|_q}\,\mathrm{d}z \leq \frac{1}{p} + \frac{1}{q} = 1,$$

which implies the claim $\|fg\|_1 = \int_\Omega |f(z)g(z)|\,\mathrm{d}z \leq \|f\|_p\|g\|_q$. □

Remark A.3. If $p = q = 2$, Hölder's inequality is called **Cauchy–Schwarz Inequality**.

Theorem A.4 (Minkowski's Inequality)**.** *Let $\Omega \subset \mathbb{R}^n$ be compact and let $p \geq 1$. Then*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p, \qquad \forall f, g \in C(\Omega).$$

Proof. The case $p = 1$ is straightforward to verify directly. Assume $p > 1$ and let $f, g \in C(\Omega)$. We can assume $\|f+g\|_p \neq 0$, since otherwise the claim is trivial. If we define $q = \frac{p}{p-1} > 1$, we have $\frac{1}{q} = 1 - \frac{1}{p}$. We can use the Hölder's Inequality to estimate

$$\begin{aligned}
\|f+g\|_p^p = \int_\Omega |f(z)+g(z)|^p\,\mathrm{d}z &= \int_\Omega |f(z)+g(z)||f(z)+g(z)|^{p-1}\,\mathrm{d}z \\
&\leq \int_\Omega |f(z)||f(z)+g(z)|^{p-1}\,\mathrm{d}z + \int_\Omega |g(z)||f(z)+g(z)|^{p-1}\,\mathrm{d}z \\
&\leq \|f\|_p\|(f+g)^{p-1}\|_q + \|g\|_p\|(f+g)^{p-1}\|_q \\
&= (\|f\|_p + \|g\|_p)\,\|(f+g)^{p-1}\|_q.
\end{aligned}$$

Because $q(p-1) = p$, we have $p-1 = \frac{p}{q}$, and the right-hand side of the above inequality is exactly

$$(\|f\|_p + \|g\|_p)\left(\int_\Omega |f(z)+g(z)|^p\,\mathrm{d}z\right)^{1-1/p} = (\|f\|_p + \|g\|_p)\,\|f+g\|_p^{p-1}.$$

Dividing both sides of the complete inequality by $\|f+g\|_p^{p-1}$ (which is nonzero by assumption) we obtain the claim

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

□

A.2 Normed Space Topology

In this section we will give a quick summary of the main topological concepts on a normed space X. In particular, we will define open and closed sets and the closure of a set, and summarise their main properties.

Definition A.5. Let $(X, \|\cdot\|)$ be a normed space.

- The **open ball** $B(x_0, r)$ centered at $x_0 \in X$ and with radius $r > 0$ is defined as $B(x_0, r) = \{ x \in X \mid \|x_0 - x\| < r \}$.
- A subset $A \subset X$ is **open** if for every $x \in A$ there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset A$.
- A point $x_0 \in X$ is an **accumulation point** of A if for every $\varepsilon > 0$ there exists $x \in A$ such that $x \in B(x_0, \varepsilon)$ and $x \neq x_0$.
- The set $A \subset X$ is **closed** if it contains all its accumulation points.
- The **closure** $\overline{A} \subset X$ of $A \subset X$ is defined as

$$\overline{A} = \{ x \in X \mid x \in A \text{ or } x \text{ is an accumulation point of } A \}.$$

- A subset $A \subset X$ is said to be **bounded** if there exists $M > 0$ such that $\|x\| \leq M$ for all $x \in A$.

Note that an accumulation point of $A \subset X$ does not in general belong to A! The definitions imply that a subset A of X is open if and only if its **complement** $X \setminus A = \{ x \in X \mid x \notin A \}$ is closed.

The definitions also imply that every set $A \subset X$ is dense in its closure $\overline{A}$. This also implies that for every $x \in \overline{A}$ there exists a sequence $(x_k)_{k=1}^{\infty} \subset A$ such that $\|x_k - x\| \to 0$ as $k \to \infty$ (see Exercise 1.40).

A.3 Decomposition of Hilbert Spaces

In this section we present the proof of the decomposition $X = M \oplus_{\perp} M^{\perp}$ of a Hilbert space X into the sum of its closed subspace M and its orthogonal complement. The proof is based on the following Minimum Norm Theorem, which also holds in a more general case where the subspace M is replaced with a "closed convex set" $K \subset X$.

Theorem A.6 (Minimum Norm Theorem)**.** *Let X be a Hilbert space and let M be a closed subspace of X. For every $x_0 \in X$ there exists a unique $y_0 \in X$ satisfying*

$$\|x_0 - y_0\| \leq \|x_0 - y\| \qquad \forall y \in M.$$

Proof. Denote $\delta = \inf\{ \|x_0 - y\| \mid y \in M \}$. By definition of the infimum, there exists a sequence $(y_k)_{k=1}^{\infty} \subset M$ such that $\|x_0 - y_k\| \to 0$ as $k \to \infty$. To show that $(y_k)_k$ is a Cauchy sequence, we can use the Parallelogram Law (Lemma 3.15) to estimate

$$\begin{aligned}
\|y_k - y_m\|^2 &= \|(y_k - x_0) + (x_0 - y_m)\|^2 \\
&= 2\|y_k - x_0\|^2 + 2\|x_0 - y_m\|^2 - \|y_k + y_m - 2x_0\|^2 \\
&= 2\|y_k - x_0\|^2 + 2\|x_0 - y_m\|^2 - 4\Big\|\underbrace{\frac{1}{2}(y_k + y_m)}_{\in M} - x_0\Big\|^2 \\
&\leq 2\|y_k - x_0\|^2 + 2\|x_0 - y_m\|^2 - 4\delta^2.
\end{aligned}$$

The last expression converges to zero as $k, m \to \infty$, and thus $(y_k)_k$ is a Cauchy sequence on X. Since X is a Hilbert space, this sequence has a limit $y_0 \in X$, and since M is a closed subspace, we have $y_0 \in M$. The continuity of the norm further implies that

$$\|x_0 - y_0\| = \|x_0 - \lim_{k\to\infty} y_k\| = \lim_{k\to\infty} \|x_0 - y_k\| = \delta.$$

To prove that y_0 is unique, assume $z_0 \in M$ is such that that $\|x_0 - z_0\| = \delta$. The Parallelogram Law then implies that

$$\begin{aligned}\|y_0 - z_0\|^2 &= \|(y_0 - x_0) + (x_0 - z_0)\|^2 \\ &= 2\|y_0 - x_0\|^2 + 2\|x_0 - z_0\|^2 - 4\left\|\frac{1}{2}(y_0 + z_0) - x_0\right\|^2 \\ &\le 2\delta^2 + 2\delta^2 - 4\delta^2 = 0\end{aligned}$$

and thus $z_0 = y_0$. □

Theorem A.7. *If M is a closed subspace of a Hilbert space X, then*

$$X = M \oplus_\perp M^\perp.$$

Proof. Let $z \in X$ be arbitrary. Our aim is to show that $z = x + y$ for some $x \in M$ and $y \in M^\perp$. The Minimum Norm Theorem implies that there exists a unique $x \in M$ so that $\|z - x\| \le \|z - x'\|$ for all $x' \in M$. Let $y = z - x$. We will show that $y \in M^\perp$. For all $\lambda \in \mathbb{C}$ and $x' \in M$ we have

$$\begin{aligned}\|y\|^2 = \|z - x\|^2 &\le \|z - x - \lambda x'\|^2 = \|y - \lambda x'\|^2 \\ &= \|y\|^2 - \overline{\lambda}\langle y, x'\rangle - \lambda\langle x', y\rangle + |\lambda|^2\langle x', x'\rangle.\end{aligned}$$

Choosing $x' \in M$ with $\|x'\| = 1$ and $\lambda = \langle y, x'\rangle$ we obtain

$$\|y\|^2 \le \|y\|^2 - |\lambda|^2.$$

This implies $0 = \lambda = \langle y, x'\rangle$ for all $x' \in M$, and thus $y \in M^\perp$. Since $z \in X$ was arbitrary and since we have that it can be represented in the form $z = x + y$ where $x \in M$ and $y \in M^\perp$, the proof is complete. □

Bibliography

[Axl20] Sheldon Axler. *Measure, integration & real analysis*, volume 282 of *Graduate Texts in Mathematics*. Springer, Cham, [2020] ©2020.

[Enf73] Per Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Math.*, 130:309–317, 1973.

[Jam51] Robert C. James. A non-reflexive Banach space isometric with its second conjugate space. *Proc. Nat. Acad. Sci. U.S.A.*, 37:174–177, 1951.

[Kre89] Erwin Kreyszig. *Introductory functional analysis with applications.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1989.

[NS82] Arch W. Naylor and George R. Sell. *Linear Operator Theory in Engineering and Science.* Springer-Verlag, New York, 1982.

[Rud87] Walter Rudin. *Real and complex analysis.* Mathematics series. McGraw-Hill, 1987.

[TL80] Angus E. Taylor and David C. Lay. *Introduction to Functional Analysis.* John Wiley & Sons, Ltd., 1980.

Index