Lecture Notes for Functional Analysis

Emil Wiedemann
Universität Ulm, WS 2021/22

Preface

These are lecture notes for a first course in functional analysis, offered for advanced bachelor's and master's students of mathematics at Ulm University in the Winter Term of 2021/22. Moreover, the first half (Chapters 1 & 2) of these notes cover the newly established course *Functional Analysis for Data Science* for first year master's students in the Mathematical Data Science Programme. The full course is taught in approximately 32 sessions of 90 minutes each.

Functional analysis is one of the most elegant and beautiful parts of mathematics. Developed from the early twentieth century by mathematicians such as Banach and his colleagues from Lwów (such as Steinhaus, Mazur, Ulam and others), the Riesz brothers, and by mathematical physicists such as von Neumann and Weyl, it has become a fundamental theory indispensable for anyone who wants to go deeper into any kind of (even remotely) applied mathematics. Functional analytic notions are crucial in such diverse fields as partial differential equations, quantum mechanics, stochastic analysis, statistics, numerics, etc. But parts of functional analysis have also developed into very 'pure' directions, such as the study of operator algebras, giving rise to noncommutative geometry.

So what is functional analysis? A short answer would be: It is the study of infinite-dimensional vector spaces. One can, at least formally, imagine a vector $x = (x_1, x_2, \dots)$ with infinitely many entries, and a linear transformation encoded by an $(\infty \times \infty)$-matrix with entries (a_{ij}), where $i, j \in \mathbb{N}$; then the i-th component of the transformed vector would be given by the infinite series

$$\sum_{j=1}^{\infty} a_{ij} x_j.$$

Does this series converge? This depends crucially on the metric structure of the vector space in question, which leads us from linear algebra to analysis.

The elements of an infinite-dimensional vector space are often functions; for instance, the set $C(\mathbb{R})$ of continuous functions $\mathbb{R} \to \mathbb{R}$ forms a real vector space under pointwise addition and scalar multiplication. Another example would be the Lebesgue space $L^p(\Omega)$, where $1 \leq p \leq \infty$ and Ω is a measure space. By framing such function spaces within the context of linear algebra, one may thus apply geometric intuitions and concepts, like orthogonality, angles, projections, etc., to function spaces. Thus, functional analysis can be said to bring together the three classical branches of mathematics – algebra, geometry, and analysis. This certainly is one important reason why many mathematicians find functional analysis so appealing.

This course requires a solid knowledge of linear algebra, analysis, and measure theory, as should have been acquired in the first two years of an undergraduate mathematics programme. These notes have been inspired by the book of Werner [**5**], lecture notes of my colleagues Anna Dall'Acqua, Markus Kunze, and Rico Zacher, and the bachelor's thesis of Michael Staněk on Reproducing Kernel Hilbert Spaces. I would like to thank them for sharing their materials with me. I also wish to thank Dennis Gallenmüller for preparing and teaching the examples classes.

Contents

CHAPTER 1

Normed Spaces and Linear Operators

Throughout the entire course, we will denote by $\mathbb{K}$ the field of real ($\mathbb{K} = \mathbb{R}$) or complex ($\mathbb{K} = \mathbb{C}$) numbers.

1.1. Definitions and Examples

Recall the notion of a *normed space*:

DEFINITION 1.1. A $\mathbb{K}$-vector space X together with a map $\|\cdot\| \to \mathbb{R}$ is called a *normed space* if the following is true:

(1) For all $x \in X$, $\|x\| \geq 0$, and $\|x\| = 0$ only if $x = 0$;

(2) For all $x \in X$ and $\alpha \in \mathbb{K}$,

$$\|\alpha x\| = |\alpha|\|x\|;$$

(3) For all $x, y \in X$,

$$\|x + y\| \leq \|x\| + \|y\|.$$

The last property is called the *triangle inequality* (why?). By abuse of notation, we will frequently denote a normed space by X, suppressing the norm $\|\cdot\|$. Of course, in cases of possible ambiguity, one should explicitly mention the norm.

A sequence $(x_n)_{n\in\mathbb{N}} \subset X$ is said to *converge* to $x \in X$ if, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for every $n \geq N$, $\|x_n - x\| < \epsilon$. We use usual notation like $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ in this case.

A sequence $(x_n)_{n\in\mathbb{N}} \subset X$ is called *Cauchy* if, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that, for all $n, m \geq N$, $\|x_n - x_m\| < \epsilon$.

You know from analysis that every convergent sequence is Cauchy. Conversely, does every Cauchy sequence converge? In $\mathbb{K}$ or $\mathbb{K}^d$ this is the case, as these spaces are *complete.* In $\mathbb{Q}$ for instance, there are Cauchy sequences that do not converge (take for instance a rational approximation of $\sqrt{2}$). Generally, we define:

DEFINITION 1.2. A normed space X is *complete* if for every Cauchy sequence $(x_n)_{n\in\mathbb{N}} \subset X$ there exists $x \in X$ such that $x_n \to x$. A complete normed space is called *Banach.*

More briefly: A normed space is complete if every Cauchy sequence converges.

EXAMPLE 1.3. (1) Of course, $\mathbb{K}^d$ is Banach.

(2) If Ω is a measure space, then the spaces $L^p(\Omega)$, $1 \leq p \leq \infty$, are complete, as you saw in measure theory. Recall that $L^p(\Omega)$ consists of equivalence classes of measurable functions $\Omega \to \mathbb{R}$ that agree up to a set of measure zero and for which the norm

$$\|f\|_p := \left(\int_\Omega |f|^p dx\right)^{1/p} \quad (p < \infty), \qquad \|f\|_\infty := \operatorname{esssup}_{x\in\Omega} |f(x)|.$$

is finite[1].

[1] Recall that the essential supremum of $|f|$ over Ω is the infimum of all numbers $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for almost every $x \in \Omega$, where the meaning of 'almost every' depends on the underlying measure.

(3) Let l^∞ be the space of $\mathbb{K}$-valued sequences $x = (x_n)_{n\in\mathbb{N}}$ that are bounded, that is: $\|x\|_\infty := \sup_{n\in\mathbb{N}} |x_n| < \infty$. One can check (which you are encouraged to do) that $\|\cdot\|_\infty$ is a norm on l^∞. We want to show that l^∞ with this norm is in fact Banach, i.e., complete.

Indeed, let $(x^k)^{k\in\mathbb{N}}$ be a Cauchy sequence in l^∞, that is: For every $\epsilon > 0$, there is $K \in \mathbb{N}$ such that for all $k, l \geq K$,

$$\sup_{n\in\mathbb{N}} |x_n^k - x_n^l| < \epsilon.$$

In particular, for each (fixed) $n \in \mathbb{N}$, the sequence $(x_n^k)^{k\in\mathbb{N}}$ is Cauchy in $\mathbb{K}$. But since $\mathbb{K}$ is complete, we obtain for each $n \in \mathbb{N}$ a limit x_n, that is, $\lim_{k\to\infty} x_n^k = x_n$.

We need to show that $x := (x_n)_{n\in\mathbb{N}} \in l^\infty$, and that $x^k \to x$ in l^∞.

To show this, let $\epsilon > 0$. By the Cauchy assumption, there exists $N \in \mathbb{N}$ such that

$$\sup_{n\in\mathbb{N}} |x_n^k - x_n^l| < \frac{\epsilon}{2}$$

if $k, l \geq N$.

On the other hand, given $n \in \mathbb{N}$, there exists $K = K(n) \in \mathbb{N}$ such that

$$|x_n^K - x_n| < \frac{\epsilon}{2};$$

without loss of generality, we may assume $K(n) \geq N$ for all $n \in \mathbb{N}$.

Therefore, for $n \in \mathbb{N}$, we estimate for $k \geq N$:

$$|x_n^k - x_n| \leq |x_n^k - x_n^{K(n)}| + |x_n^{K(n)} - x_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

and since N does not depend on n, we even deduce

$$\sup_{n\in\mathbb{N}} |x_n^k - x_n| < \epsilon \tag{1.1}$$

for $k \geq N$; this implies, first, that

$$\sup_{n\in\mathbb{N}} |x_n| \leq \sup_{n\in\mathbb{N}} |x_n^N| + \sup_{n\in\mathbb{N}} |x_n - x_n^N| < \|x^N\|_\infty + \epsilon < \infty,$$

so that $x \in l^\infty$; and the convergence $x^k \to x$ in l^∞ then follows precisely from (1.1).

(4) For $1 \leq p < \infty$, the space of $\mathbb{K}$-valued sequences $x = (x_n)_{n\in\mathbb{N}}$ such that

$$\|x\|_p := \left(\sum_{n=1}^\infty |x_n|^p\right)^{1/p} < \infty$$

can be shown to be Banach (see [**5**, p. 12f.]). This space is called l^p.

(5) The real vector space $C([0,1])$ of continuous functions $[0,1] \to \mathbb{R}$ can be equipped with several different norms, for example:

$$\|f\|_\infty := \sup_{x\in[0,1]} |f(x)|$$

or

$$\|f\|_1 := \int_0^1 |f(x)| dx.$$

We will show below that $C([0,1])$ is Banach with respect to $\|\cdot\|_\infty$, and it is a worthwhile exercise to show it is *not* Banach with respect to $\|\cdot\|_1$. Hence, the completeness property depends crucially on the choice of norm.

Let X be a normed space; recall that a subset $U \subset X$ is called *closed* in X if the following holds: If $(x_n)_{n\in\mathbb{N}} \subset U$ is a sequence with $\lim_{n\in\mathbb{N}} x_n = x \in X$, then $x \in U$. In other words: U is closed if it contains all its limit points.

LEMMA 1.4. *If X is Banach and $U \subset X$ is a closed subspace, then also U is Banach.*

PROOF. It follows from the definition of a normed space that a subspace of a normed space is again a normed space[2]. So it suffices to show completeness of U.

Let $(x_n)_{n\in\mathbb{N}}$ be Cauchy in U and therefore also in X. As X is complete, there exists a limit $X \ni x = \lim_{n\to\infty} x_n$, but since U is closed, $x \in U$, so U is complete.

□

A paradigmatic application of this lemma is given as follows: We show that $C([0,1])$ is Banach with respect to the supremum norm $\|\cdot\|_\infty$. To this end, we set $X := L^\infty(0,1)$, which we have already seen to be Banach. Set $U = C([0,1])$, then clearly U is a subspace of X (as linear combinations of continuous functions are still continuous), and U is even closed in X: Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of continuous functions that converges with respect to $\|\cdot\|_\infty$ to some $f \in X$, that is:

$$\sup_{x\in[0,1]} |f_n(x) - f(x)| \to 0 \quad \text{as } n \to \infty.$$

This means that the convergence is uniform. But uniform limits of continuous functions are themselves continuous (Analysis I), so $f \in U$ and hence U is closed. By Lemma 1.4, we conclude that $(C([0,1]), \|\cdot\|_\infty)$ is Banach, as claimed.

1.2. Separability and the Theorem of Stone-Weierstraß

1.2.1. Separability.

DEFINITION 1.5. A normed space is *separable* if it contains a countable dense subset.

Recall that a subset A of a normed space X is called *dense* if for every $\epsilon > 0$ and every $x \in X$ there exists $a \in A$ such that $\|x - a\| < \epsilon$. For instance, $\mathbb{R}$ (with the standard norm given by the absolute value $|\cdot|$) is separable, because the countable set $\mathbb{Q}$ of rationals is dense in $\mathbb{R}$.

LEMMA 1.6. *Let X be a normed space. The following are equivalent:*

(1) *X is separable;*
(2) *There is a countable subset $A \subset X$ such that $X = \overline{\operatorname{span} A}$.*

Here, $\operatorname{span} A$ denotes the linear span of A, and the overline denotes the (topological) closure with respect to the given norm[3].

PROOF. If X is separable, then it suffices to choose A as a countable dense subset.

Conversely, assume (2) for some set A. Suppose for the moment that $\mathbb{K} = \mathbb{R}$. We will show that the set of linear combinations with rational coefficients

$$B := \{\sum_{j=1}^{n} \lambda_j x_j : \quad n \in \mathbb{N}, \quad \lambda_j \in \mathbb{Q}, \quad x_j \in A\}$$

is countable and dense. In fact, countability is obvious, as countable unions and Cartesian products of countable sets are still countable.

As far as the density of B is concerned, we know by assumption that for every $x \in X$ and $\epsilon > 0$, there exists a linear combination $\sum_{j=1}^{n} \lambda_j x_j$ with $x_j \in A$ and $\lambda_j \in \mathbb{R}$ such that

$$\left\| x - \sum_{j=1}^{n} \lambda_j x_j \right\| < \frac{\epsilon}{2}.$$

[2]To be pedantic, if $(X, \|\cdot\|)$ is the original normed space and $U \subset X$ is a subspace, then $(U, \|\cdot\| \restriction_U)$ is also a normed space.

[3]Please look up these notions from your linear algebra and analysis courses if you feel uncomfortable with them.

Now, for every $j \in \mathbb{N}$ there exists some $\lambda_j' \in \mathbb{Q}$ such that

$$|\lambda_j - \lambda_j'| < \frac{\epsilon}{2^{j+1}\|x_j\|}.$$

Certainly $\sum_{j=1}^n \lambda_j' x_j \in B$ and

$$\begin{aligned}\left\| x - \sum_{j=1}^n \lambda_j' x_j \right\| &\le \left\| x - \sum_{j=1}^n \lambda_j x_j \right\| + \left\| x - \sum_{j=1}^n (\lambda_j - \lambda_j') x_j \right\| \\ &< \frac{\epsilon}{2} + \sum_{j=1}^n |\lambda_j - \lambda_j'| \|x_j\| < \epsilon,\end{aligned}$$

where we used $\sum_{j=1}^\infty 2^{-(j+1)} = \frac{1}{2}$ in the last step. This proves the density of B in X.

If $\mathbb{K} = \mathbb{C}$, then simply replace $\lambda_j \in \mathbb{Q}$ by $\lambda_j \in \mathbb{Q} + i\mathbb{Q}$ in the definition of B. □

Example 1.7.
(1) Recall from Example 1.3 the space l^p ($1 \le p < \infty$) with norm $\|x\|_p = \left(\sum_{j=1}^\infty |x_j|^p\right)^{1/p}$. We show separability of l^p. To this end, let e_j be the sequence whose j-th element is one and all other elements zero, and set

$A := \{e_j : j \in \mathbb{N}\}$.

Let $x = (x_j)_{j\in\mathbb{N}} \in l^p$ and $\epsilon > 0$. If $N \in \mathbb{N}$ is sufficiently large, then $\sum_{j=N+1}^\infty |x_j|^p < \epsilon^p$, and therefore

$$\left\| x - \sum_{j=1}^N x_j e_j \right\|_p = \left\| \sum_{j=N+1}^\infty x_j e_j \right\|_p = \left(\sum_{j=N+1}^\infty |x_j|^p \right)^{1/p} < \epsilon,$$

which means that the span of A is dense. Separability of l^p now follows from Lemma 1.6.

(2) As in Example 1.3, we consider l^∞, the Banach space of bounded sequences, and show that it is *not* separable. For let $S \subset \mathbb{N}$ any subset, and define $x_S \in l^\infty$ by

$$(x_S)_j = \begin{cases} 1 & \text{if } j \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Then, whenever $S_1 \ne S_2$, we will have $\|x_{S_1} - x_{S_2}\|_\infty = 1$. Therefore, if $A \subset l^\infty$ is any countable subset, then for any $x \in A$, the ball[4] $B_{\frac{1}{2}}(x)$ will contain at most one x_S.

Hence, there are at most countably many x_S for which there exists $x \in A$ such that $x_S \in B_{\frac{1}{2}}(x)$. However, as there are uncountably many subsets of $\mathbb{N}$, there are also uncountably many x_S, and therefore there exists an x_S (in fact infinitely many) that has distance at least $1/2$ from any element of A. We conclude that A cannot be dense in l^∞.

(3) One can show that L^p spaces behave similarly: On a measure space Ω, the space $L^p(\Omega)$ is separable for $1 \le p < \infty$, and not separable for $p = \infty$.

1.2.2. The Theorem of Stone-Weierstraß. Having talked about dense subsets of normed spaces, we will make an excursion into the related field of *approximation theory* for the rest of this section.

Recall[5] the notion of a *compact metric space*: A set X together with a map $d : X \times X \to \mathbb{R}$ is a *metric space* if

(1) For all $x, y \in X$, $d(x, y) \ge 0$, and equality holds if and only if $x = y$;
(2) For all $x, y \in X$, $d(x, y) = d(y, x)$;

[4] As usual we write $B_r(x) := \{y \in X : \|y - x\| < r\}$.

[5] The concept of a metric space and of open, closed, and compact sets should be known from Analysis II. If you do not have this background, let me know so we can do something about it.

(3) For all $x,y,z \in X$, $d(x,z) \le d(x,y) + d(y,z)$.

Every normed space is, in particular, a metric space with metric given as $d(x,y) := \|x-y\|$. In metric spaces one may define balls $B_r(x) := \{y \in X : d(x,y) < r\}$, and a subset $U \subset X$ is called *open* if for every $x \in U$ there exists $r > 0$ such that $B_r(x) \subset U$.

A metric space (X,d) is called *compact* if the following is true: If $X = \bigcup_{j\in I} U_j$ is a cover of X by (possibly uncountably many) open sets U_j, then there exists a finite subcover $X = \bigcup_{n=1}^N U_{j_n}$.

Let (X,d) be a compact metric space. As in the example after Lemma 1.4, we see that the space $C(X)$ of continuous functions $X \to \mathbb{K}$ is Banach when equipped with the norm given as $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

DEFINITION 1.8. A subspace $A \subset C(X)$ is called a *subalgebra* if it is closed under multiplication, that is: If $f,g \in A$, then $fg \in A$.

For instance, the following spaces are subalgebras of $C([0,1])$: the space of polynomial functions on $[0,1]$; the space $C^1([0,1])$ of continuously differentiable functions (Leibniz rule!); the subspace generated by functions of the form $\exp(ik\cdot)$ $(k \in \mathbb{Z})$, i.e., the trigonometric polynomials. On the other hand, for fixed $N \in \mathbb{N}$, the space of polynomials of degree at most N is *not* a subalgebra, because the product of two polynomials of degree N can have degree larger than N.

A subalgebra A is said to *separate the points* of X if, for every pair $x,y \in X$ with $x \neq y$, there exists $f \in A$ such that $f(x) \neq f(y)$. For instance, the algebra of polynomial functions separates the points of $[0,1]$ (simply choose $f(x) = x$), but the algebra of even functions[6] on $[-1,1]$ does not separate points, as two points $\pm x$ are always assigned the same value.

If A is a subalgebra of $C(X)$, then we denote by $\overline{A}$ its closure (with respect to $\|\cdot\|_\infty$); that is, $\overline{A}$ is the set of all uniform limits of sequences in A. It is easy to check that $\overline{A}$ is a subalgebra itself.

If $\mathbb{K} = \mathbb{C}$, then a subalgebra is *stable under conjugation* if for every $f \in A$ also the complex conjugate $\bar{f}$ is in A. For example, the space of trigonometric polynomials is stable under conjugation, as

$$\overline{\exp(i\cdot)} = \exp(-i\cdot).$$

LEMMA 1.9. *Let X be a compact metric space and A a subalgebra of $C(X;\mathbb{R})$ that contains the constant function 1. If $f,g \in A$, then also $\min\{f,g\} \in \bar{A}$, $\max\{f,g\} \in \bar{A}$, and $|f| \in \bar{A}$.*

PROOF. Let $f,g \in A$. It is easy to see

$$\min\{f,g\} = \frac{1}{2}\left(f+g-|f-g|\right), \quad \max\{f,g\} = \frac{1}{2}\left(f+g+|f-g|\right).$$

Since A is a vector space, we can immediately deduce from $|f|,|g| \in \bar{A}$ the assertions for minimum and maximum. Thus it suffices to show $|f| \in \bar{A}$.

If f is identically zero, then the statement is trivial. So suppose $\|f\|_\infty > 0$. It is possible to construct a family $(p_n)_{n\in\mathbb{N}}$ of polynomial functions that converge uniformly on $[-1,1]$ to the absolute value function $|\cdot|$ (exercise). Hence, for $\epsilon > 0$, there is $N \in \mathbb{N}$ such that

$$||t| - p_N(t)| < \frac{\epsilon}{\|f\|_\infty} \quad \text{for all } t \in [-1,1].$$

For each $x \in X$, the choice $t = \frac{f(x)}{\|f\|_\infty}$ then yields

$$\left|\left|\frac{f(x)}{\|f\|_\infty}\right| - p_N\left(\frac{f(x)}{\|f\|_\infty}\right)\right| < \frac{\epsilon}{\|f\|_\infty}$$

[6] A function $[-1,1] \to \mathbb{K}$ is even if $f(-x) = f(x)$ for all $x \in [-1,1]$.

and after multiplication with $\|f\|_\infty$:

$$\left| |f(x)| - \|f\|_\infty p_N\left(\frac{f(x)}{\|f\|_\infty}\right)\right| < \epsilon.$$

As A contains the constant 1, and thus for any $f \in A$ it contains $p_n \circ f$, we find that $\|f\|_\infty p_N\left(\frac{f}{\|f\|_\infty}\right) \in A$. In any neighbourhood of $|f|$, therefore, there exists an element of A. It follows that $|f| \in \overline{A}$, as claimed.

□

THEOREM 1.10 (Stone-Weierstraß). *Let X be a compact metric space and A a subalgebra of $C(X;\mathbb{K})$ which*

- *contains the constant function 1,*
- *separated the points of X, and*
- *is stable under conjugation in the case $\mathbb{K} = \mathbb{C}$.*

Then A is dense in $C(X;\mathbb{K})$, i.e., for every $f \in C(X,\mathbb{K})$ und every $\epsilon > 0$ there exists $a \in A$ such $\|f - a\|_\infty < \epsilon$.

PROOF. The first proof steps deal with the case $\mathbb{K} = \mathbb{R}$.

Step 1. Let $f \in C(X;\mathbb{R})$. We show: For any $y, z \in X$ there is a function $h_{y,z} \in A$ such that

$$h_{y,z}(y) = f(y), \quad h_{y,z}(z) = f(z). \tag{1.2}$$

To this end, we may assume $y \neq z$ (for $y = z$ simply choose the constant function $h_{y,z} = f(y)$). By assumption there exists some $g \in A$ separating the points y and z, i.e., $g(y) \neq g(z)$. Therefore, the following function is well-defined and in A:

$$h_{y,z}(x) = f(y) + \frac{f(z) - f(y)}{g(z) - g(y)}(g(x) - g(y)).$$

Clearly, this function satisfies (1.2).

Step 2. For $y, z \in X$, set

$$U_{y,z} := \{x \in X : h_{y,z}(x) < f(x) + \epsilon\}, \quad V_{y,z} := \{x \in X : h_{y,z}(x) > f(x) - \epsilon\},$$

so that $U_{y,z}$ and $V_{y,z}$ are the preimages, respectively, of the open sets $(-\infty, \epsilon) \subset \mathbb{R}$ and $(-\epsilon, \infty) \subset \mathbb{R}$ under the continuous function $h_{y,z} - f$. Since preimages of open sets under continuous functions are again open, we find that $U_{y,z}$ and $V_{y,z}$ are open subsets of X. Thanks to (1.2), we also have $y \in U_{y,z}$ and $z \in V_{y,z}$.

For fixed $z \in X$, therefore, $X = \bigcup_{y \in X} U_{y,z}$ is an open cover, and by compactness of X there is a finite subcover $X = \bigcup_{j=0}^m U_{y_j,z}$.

Now set

$$h_z := \min_{j=1,\dots,m} h_{y_j,z}.$$

By Lemma 1.9, $h_z \in \overline{A}$, and for all $x \in X$

$$h_z(x) < f(x) + \epsilon, \tag{1.3}$$

because there exists some j with $x \in U_{y_j,z}$, and by definition $h_z \le h_{y_j,z}$.

Step 3. For $z \in X$ we set

$$V_z := \bigcap_{j=1}^m V_{y_j,z}.$$

As a finite intersection of open sets, V_z is itself open, and $z \in V_{y,z}$ for all $y \in X$ entails $z \in V_z$. Hence $X = \bigcup_{z \in X} V_z$ is an open cover, which by compactness admits a finite subcover $X = \bigcup_{l=1}^r V_{z_l}$.

Finally set

$$h := \max_{l=1,\dots,r} h_{z_l}.$$

By Lemma 1.9, $h \in \overline{A}$, und by (1.3) we have $h < f + \epsilon$. Note also that for any $x \in X$ there is an index l with $x \in V_{z_l}$, so that by definition of V_{z_l} we infer $x \in V_{y_j,z_l}$ for all $j = 1,\dots,m$. Hence (by definition of $V_{y,z}$) $h_{y_j,z_l}(x) > f(x) - \epsilon$, therefore $h_{z_l}(x) > f(x) - \epsilon$, and thereby even $h(x) > f(x) - \epsilon$. In summary we have proved

$$\|f - h\|_\infty < \epsilon.$$

As $h \in \overline{A}$, there is $a \in A$ with $\|h - a\|_\infty < \epsilon$, and thus $\|f - a\|_\infty < 2\epsilon$. This shows the Theorem for the real-valued case.

Step 4. We reduce the complex case to the real one. So let now A be a subalgebra of $C(X;\mathbb{C})$ satisfying the assumptions of the Theorem. Let $A' \subset A$ be the set of functions in A that take only real values. It is a subalgebra of $C(X;\mathbb{R})$, and we have $A = A' + iA' := \{a + ib : a, b \in A'\}$: Indeed, as A is stable under conjugation, then for every $a \in A$ also

$$\Re a = \frac{a + \bar{a}}{2} \in A, \quad \Im a = \frac{a - \bar{a}}{2i} \in A,$$

and hence also $\Re a, \Im a \in A'$. It follows that $A \subset A' + iA'$, since for $a \in A$ we have $a = \Re a + i\Im a \in A' + iA'$. On the other hand, $A \supset A' + iA'$ because of $A' \subset A$ and A being a $\mathbb{C}$-vector space.

A' contains the constant 1 and separates the points of X: Indeed, if $x \neq y$, then by assumption there exists $a \in A$ with $a(x) \neq a(y)$, and hence $\Re a(x) \neq \Re a(y)$ or $\Im a(x) \neq \Im a(y)$. So $\Re a$ or $\Im a$ is an element of A' that separates x and y.

According to the already established real-valued version of Stone-Weierstraß, for every $\epsilon > 0$ and $f \in C(X;\mathbb{C})$ there exist $a, b \in A'$ such that

$$\|\Re f - a\|_\infty < \frac{\epsilon}{2}, \quad \|\Im f - b\|_\infty < \frac{\epsilon}{2},$$

and it follows that $\|f - (a + ib)\|_\infty < \epsilon$. As $a + ib \in A$, the Theorem is proved. □

1.2.3. Some Consequences of Stone-Weierstraß.

1.2.3.1. *Polynomial Approximation.* As a first corollary of the Stone-Weierstraß Theorem, we obtain the classical approximation theorem of Weierstraß from 1885:

COROLLARY 1.11 (Weierstraß Approximation Theorem). *Let $[a,b] \subset \mathbb{R}$ be a compact interval ($a < b$) and $f : [a,b] \to \mathbb{R}$ continuous. Then for any $\epsilon > 0$ there exists a polynomial function*[7] *$p : [a,b] \to \mathbb{R}$ such that*

$$\|f - p\|_\infty = \sup_{x \in [a,b]} |f(x) - p(x)| < \epsilon.$$

PROOF. It suffices to show that the set of real polynomial functions satisfies the conditions of the (real version of the) Stone-Weierstraß Theorem. Since linear combinations and products of polynomials are again polynomial, and since polynomial functions are continuous, they form a subalgebra of $C([a,b])$. This algebra contains constant functions and separates points (just take $p(x) = x$ as a separating function for any two distinct points of $[a,b]$). □

[7] A polynomial function, of course, is a function of the form $p(x) = \sum_{k=0}^{N} a_k x^k$ for coefficients $a_k \in \mathbb{K}$.

1.2.3.2. *Trigonometric Polynomials.* In the theory of Fourier series, to be developed further in the next chapter, *trigonometric polynomials* play a crucial role. We prefer here to work with functions of the form e^{ikx}. Often, trigonometric polynomials instead are written in terms of sine and cosine functions; such a formalism can easily be obtained from the framework presented here by virtue of the famous formula $e^{ix} = \cos(x) + i\sin(x)$.

DEFINITION 1.12. A *trigonometric polynomial function* is a function $\mathbb{R} \to \mathbb{C}$ of the form

$$x \mapsto \sum_{k=-N}^{N} c_k e^{ikx}$$

with coefficients $c_k \in \mathbb{C}$.

Obviously, every trigonometric polynomial function is 2π-periodic[8]. Let us denote by $C_{per}(\mathbb{R};\mathbb{C})$ the space of 2π-periodic continuous functions. We want to use the Stone-Weierstraß Theorem to show that every such function can be approximated by trigonometric polynomials in a suitable sense. To this end, it is useful to identify a 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$ with a function $\tilde{f} : \mathbb{S}^1 \to \mathbb{C}$, where $\mathbb{S}^1$ denotes the one-dimensional unit circle, i.e., the subset $\{z \in \mathbb{C} : |z| = 1\}$; simply set

$$\tilde{f}\left(e^{ix}\right) = f(x)$$

(this is well-defined, as $x \mapsto e^{ix}$ is a bijection between $[0, 2\pi)$ and $\mathbb{S}^1$).

Consider first the algebra $A_{\mathbb{S}^1}$ of continuous functions $\mathbb{S}^1 \to \mathbb{C}$ of the form

$$\sum_{k=-N}^{N} c_k z^k \tag{1.4}$$

with $c_k \in \mathbb{C}$, $k = -N, \dots, N$. It is a subalgebra of $C(\mathbb{S}^1;\mathbb{C})$.

LEMMA 1.13. *The subalgebra $A_{\mathbb{S}^1}$ is dense in $C(\mathbb{S}^1;\mathbb{C})$ with respect to the supremum norm, i.e., for every $f \in C(\mathbb{S}^1;\mathbb{C})$ and $\epsilon > 0$ there exists $a \in A_{\mathbb{S}^1}$ such that $\|f - a\|_\infty < \epsilon$.*

PROOF. As $\mathbb{S}^1 \subset \mathbb{C}$ is closed and bounded, it is compact by the Theorem of Heine-Borel. Therefore, $\mathbb{S}^1$ becomes a compact metric space with the metric $d(x,y) = |x - y|$ induced by $\mathbb{C}$.

Setting $c_0 = 1$ and all other coefficients zero in (1.4), we see that $1 \in A_{\mathbb{S}^1}$. Let $x, y \in \mathbb{S}^1$ be two distinct points, then they are separated by the identity $\mathrm{id} : z \mapsto z$, and $\mathrm{id} \in A_{\mathbb{S}^1}$ (set $c_1 = 1$ and all other coefficients zero in (1.4)). Finally, $A_{\mathbb{S}^1}$ is stable under conjugation, because for every $z \in \mathbb{S}^1$ and $k \in \mathbb{Z}$, $\overline{z^k} = z^{-k}$ (recall $\overline{e^{ikx}} = e^{-ikx}$).

Hence all conditions of (the complex version of) the Theorem of Stone-Weierstraß are satisfied, and we obtain density of $A_{\mathbb{S}^1}$ in $C(\mathbb{S}^1;\mathbb{C})$. □

Let $C_{per}(\mathbb{R};\mathbb{C})$ be the algebra of 2π-periodic continuous functions. We obtain

THEOREM 1.14. *Trigonometric polynomial functions are dense in $C_{per}(\mathbb{R};\mathbb{C})$ with respect to the supremum norm.*

PROOF. Let $f \in C_{per}(\mathbb{R};\mathbb{C})$ and define $\tilde{f} \in C(\mathbb{S}^1;\mathbb{C})$ by $\tilde{f}\left(e^{ix}\right) = f(x)$. Let $\epsilon > 0$. By Lemma 1.13 there exists a function $a : \mathbb{S}^1 \to \mathbb{C}$ of the form $a(z) = \sum_{k=-N}^{N} c_k z^k$ such that

$$\|\tilde{f} - a\|_\infty < \epsilon.$$

Thus for every $x \in [0, 2\pi)$,

$$|f(x) - a\left(e^{ix}\right)| = |\tilde{f}\left(e^{ix}\right) - a\left(e^{ix}\right)| < \epsilon. \tag{1.5}$$

But $x \mapsto a\left(e^{ix}\right) = \sum_{k=-N}^{N} c_k e^{ikx}$ is a trigonometric polynomial, and periodicity implies (1.5) for all $x \in \mathbb{R}$. □

[8] A function $f : \mathbb{R} \to \mathbb{C}$ is called l-periodic if $f(x + l) = f(x)$ for all x.

1.2.3.3. *Neural Networks.* We sketch here the idea of Hornik-Stinchcombe-White [**2**] to show that neural networks can approximately represent any continuous function.

A *squashing function* is a measurable non-decreasing function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\lim_{s\to-\infty} \phi(s) = 0$ and $\lim_{s\to\infty} \phi(s) = 1$ (it 'squashes' $\mathbb{R}$ into $[0,1]$). We consider affine maps $a : \mathbb{R}^d \to \mathbb{R}$ of the form

$$a(x) = w \cdot x + b,$$

where $w \in \mathbb{R}^d$ is interpreted as the vector of network weights, $b \in \mathbb{R}$ as a bias, and the argument x as the network input. The set of such affine maps will be denoted $\mathbb{A}^d$.

Given a measurable function $\phi : \mathbb{R} \to \mathbb{R}$ and a dimension $d \in \mathbb{N}$, we define the set of *single hidden layer feedforward networks* as

$$\Sigma^d(\phi) = \left\{ f : \mathbb{R}^d \to \mathbb{R} : \quad f(x) = \sum_{j=1}^{N} \alpha_j \phi(a_j(x)), \quad N \in \mathbb{N}, \alpha_j \in \mathbb{R}, a_j \in \mathbb{A}^d \right\}.$$

Theorem 1.15 (Universal Approximation Theorem, version of [**2**]). *Let ϕ be a squashing function, $K \subset \mathbb{R}^d$ compact, $F : K \to \mathbb{R}$ continuous, and $\epsilon > 0$. Then there exists an $f \in \Sigma^d(\phi)$ such that*

$$\sup_{x\in K} |F(x) - f(x)| < \epsilon.$$

Sketch of proof. *Step 1.* Let $K \subset \mathbb{R}^d$ be compact. We observe that, on K, the set $\Sigma^d(\cos)$ is a subalgebra of $C(K;\mathbb{R})$ that contains constants and separates points: First, it is clear that $\Sigma^d(\cos)$ is a vector space of continuous functions. But it is even an algebra, as can be seen by the trigonometric identity

$$\cos(s)\cos(t) = \frac{1}{2}(\cos(s+t) + \cos(s-t)).$$

The constant function 1 is obtained by setting $N = 1$, $\alpha_1 = 1$, $a_1 = 0$ in the definition of $\Sigma(\phi)$. To see the separation property, let $x, y \in K$ with $x \neq y$ and pick two arbitrary real numbers α, β such that $\cos(\alpha) \neq \cos(\beta)$; then, if $a \in \mathbb{A}^1$ is an affine function such that $a(x) = \alpha$ and $a(y) = \beta$, the function $\cos \circ a \in \Sigma^d(\cos)$ will separate x and y.

Therefore, by Stone-Weierstraß, we conclude the density of $\Sigma^d(\cos)$ in $C(K)$.

Step 2. It is easy to check, using the basic properties of sine, that the following defines a continuous squashing function:

$$\mathrm{cs}(s) = \begin{cases} 0 & s < -\frac{\pi}{2}, \\ \frac{1}{2}(1 + \sin(s)) & -\frac{\pi}{2} \le s \le \frac{\pi}{2}, \\ 1 & s > \frac{\pi}{2}. \end{cases}$$

Moreover, for given $M > 0$, it is a little combinatorial exercise to see that there is a function in $\Sigma^1(\mathrm{cs})$ that agrees with cos on $[-M, M]$.

Step 3. Next we show that for any $\epsilon > 0$, there is a function $g \in \Sigma^1(\phi)$ such that $|\mathrm{cs}(s) - g(s)| < \epsilon$ for all $s \in \mathbb{R}$.

For this, let $N \in \mathbb{N}$ be so large that $\frac{1}{N} < \frac{\epsilon}{2}$, and $M \in \mathbb{R}$ so large that

$$\phi(-M) < \frac{\epsilon}{2N}, \quad \phi(M) > 1 - \frac{\epsilon}{2N},$$

which is possible since ϕ is a squashing function. For $j = 1, \dots, N-1$ we denote by s_j the unique real number such that $\mathrm{cs}(s) = \frac{j}{N}$, and by s_N the unique real number such that $\mathrm{cs}(s) = 1 - \frac{1}{2N}$.

Next, for $r < s$ let $a_{r,s} : \mathbb{R} \to \mathbb{R}$ denote the unique affine function such that $a_{r,s}(r) = -M$ and $a_{r,s}(s) = M$. One can then check that $g \in \Sigma^1(\phi)$ given by

$$g(s) = \frac{1}{N} \sum_{j=1}^{N-1} \phi(a_{s_j,s_{j+1}}(s))$$

indeed has the property $|\operatorname{cs}(s) - g(s)| < \epsilon$ for all $s \in \mathbb{R}$.

Steps 2 and 3 together, thus, imply for each $M > 0$ and $\epsilon > 0$ the existence of an $h \in \Sigma^1(\phi)$ such that

$$|\cos(s) - h(s)| < \epsilon$$

for all $s \in [-M, M]$.

Step 4. Let $F \in C(K)$ and $\epsilon > 0$. By Step 1, there exists a function of the form $\sum_{j=1}^N \alpha_j \cos(a_j(x))$ such that for all $x \in K$,

$$\left| F(x) - \sum_{j=1}^{N} \alpha_j \cos(a_j(x)) \right| < \frac{\epsilon}{2}. \tag{1.6}$$

Choose $M > 0$ so large that $a_j(K) \subset [-M, M]$ for all $j = 1, \dots, N$. For this M, we have (cf. end of Step 3) a function $h \in \Sigma^1(\phi)$ approximating the cosine, so that for all $x \in K$

$$\left| \sum_{j=1}^{N} \alpha_j (\cos(a_j(x)) - h(a_j(x))) \right| < \frac{\epsilon}{2}. \tag{1.7}$$

Since the composition of affine functions is affine, we have that

$$f := \sum_{j=1}^{N} \alpha_j h \circ a_j \in \Sigma^d(\phi),$$

and (1.6) and (1.7) together yield, for all $x \in K$,

$$|F(x) - f(x)| < \epsilon,$$

as claimed. □

1.3. Linear Operators and Dual Spaces

Maps between normed spaces are often called *operators*. We use the terms *map* and *operator* synonymously. Of special interest to us are *linear* operators $T : X \to Y$, where X and Y are normed spaces. This means of course that $T(\alpha x_1 + \beta x_2) = \alpha T x_1 + \beta T x_2$ for all $\alpha, \beta \in \mathbb{K}$ and $x_1, x_2 \in X$. We have used the common notation Tx instead of $T(x)$.

Definition 1.16. Let X, Y be normed spaces over $\mathbb{K}$.

(1) A linear operator is called *bounded* if

$$\|T\| := \sup_{\|x\| \le 1} \|Tx\| < \infty.$$

The thus defined number $\|T\|$ is called the *operator norm* of T. The set of bounded linear operators $X \to Y$ is denoted $L(X, Y)$.

(2) If $Y = \mathbb{K}$, then a linear operator $X \to Y$ is called a linear *functional*. The set of bounded linear functionals $X \to \mathbb{K}$ is called the *dual space* of X and is denoted $X' = L(X, \mathbb{K})$.

Of course, $L(X, Y)$ is itself a vector space (think it over!) and becomes a normed space using the operator norm from the above definition. Indeed, if $\|T\| = 0$ then $Tx = 0$ for all $x \in \overline{B_1(0)}$, and thus by linearity $Tx = 0$ for all $x \in X$. If $\alpha \in \mathbb{K}$, then clearly $\|\alpha T\| = |\alpha| \|T\|$, and for $T, S \in L(X, Y)$

$$\|T + S\| = \sup_{\|x\| \le 1} \|(T + S)x\| \le \sup_{\|x\| \le 1} (\|Tx\| + \|Sx\|) \le \sup_{\|x\| \le 1} \|Tx\| + \sup_{\|x\| \le 1} \|Sx\| = \|T\| + \|S\|,$$

so that the axioms of a normed space are all satisfied. In particular, the dual space of a normed space is itself a normed space.

We can say even more:

PROPOSITION 1.17. *If X is a normed space and Y is Banach, then $L(X,Y)$ is Banach. In particular, the dual space of a normed space is always Banach.*

PROOF. Let $(T_n)_{n\in\mathbb{N}}$ be Cauchy in $L(X,Y)$, which means that for $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for $n,m \geq N$ and for all $x \in \overline{B_1(0)}$,

$$\|T_n x - T_m x\| < \epsilon.$$

In particular, for every $x \in X$, $(T_n x)_{n\in\mathbb{N}}$ is Cauchy in Y, hence by assumption it converges in Y to some element that we denote by Tx.

It remains to show that the map T thus defined is linear and bounded, and that $T_n \to T$ in the operator norm. If $x_1, x_2 \in X$, then

$$T(x_1 + x_2) = \lim_{n\to\infty} T_n(x_1 + x_2) = \lim_{n\to\infty} T_n x_1 + \lim_{n\to\infty} T_n x_2 = Tx_1 + Tx_2.$$

Similarly one sees $T(\alpha x) = \alpha T x$ for $\alpha \in \mathbb{K}$ and $x \in X$. To see boundedness, note that a Cauchy sequence in a normed space is bounded (think it over!) and therefore $\limsup_{n\to\infty} \|T_n\| < \infty$. Therefore, for all $x \in \overline{B_1(0)}$,

$$\|Tx\| = \|\lim_{n\to\infty} T_n x\| = \lim_{n\to\infty} \|T_n x\| \leq \limsup_{n\to\infty} \|T_n\| \|x\| \leq \limsup_{n\to\infty} \|T_n\|,$$

and since this is independent of x, we conclude boundedness. Note we have used the continuity of the norm, as established on the first exercise sheet.

To see that $T_n \to T$ in the operator norm, Let $x \in \overline{B_1(0)}$. Choose N (independent of x) so large that $\|T_n - T_m\| < \frac{\epsilon}{2}$ for $n,m \geq N$ (using the Cauchy assumption), and $M = M(x)$ so large that $\|T_M x - Tx\| < \frac{\epsilon}{2}$. Without loss of generality, $M \geq N$. Then, for all $n \geq N$,

$$\|T_n x - Tx\| \leq \|T_n x - T_M x\| + \|T_M x - Tx\| < \epsilon,$$

and as N was independent of x, we conclude $T_n \to T$ in the operator norm. □

PROPOSITION 1.18. *Let $T : X \to Y$ be a linear operator between normed spaces X and Y. Then the following are equivalent:*

(1) *T is bounded;*
(2) *T is continuous;*
(3) *T is continuous at $x = 0$.*

PROOF. (1)⇒(2): Let T be bounded and $x_n \to x$ in X, that is, $\|x_n - x\| \to 0$. Therefore

$$\|Tx_n - Tx\| \leq \|T\| \|x_n - x\| \to 0,$$

so that T is continuous.

(2)⇒(3) is trivial.

(3)⇒(1): Suppose T were unbounded, so there existed a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ such that $\|x_n\| \leq 1$ but $\|Tx_n\| > n$. Then $\left(\frac{x_n}{n}\right)_{n\in\mathbb{N}}$ is a sequence in X converging to zero, but

$$\left\|T\left(\frac{x_n}{n}\right)\right\| = \frac{1}{n}\|Tx_n\| > 1$$

for all $n \in \mathbb{N}$, in contradiction to the continuity at zero. □

EXAMPLE 1.19. (1) Any $(m \times n)$-matrix gives rise to a linear operator $\mathbb{K}^n \to \mathbb{K}^m$, and in this finite-dimensional situation, linear operators are always bounded (exercise).

(2) Consider the Banach space $C([0,1];\mathbb{R})$ with the supremum norm. A linear functional on this space is given by

$$Tf := \int_0^1 f(x)dx.$$

It is bounded because $|Tf| \le \|f\|_\infty$, whence one also sees that $\|T\| \le 1$. But since $T1 = 1$, it is even true that $\|T\| = 1$.

For each $x \in [0,1]$, another bounded linear functional is defined by the point evaluation:

$$S_x f := f(x).$$

It also has norm 1.

(3) Now let $X = C^1([0,1])$ the space of continuously differentiable functions and $Y = C([0,1])$, both equipped with the supremum norm. Then the differential operator $D : X \to Y$, $Df = f'$, is linear but unbounded: For instance, the sequence $(\sin(n\cdot))_{n\in\mathbb{N}} \subset X$ is bounded (as $\|\sin(n\cdot)\|_\infty = 1$ for every $n \in \mathbb{N}$), but the sequence of derivatives $(n\cos(n\cdot))_{n\in\mathbb{N}}$ is unbounded.

However, if we replace on X the supremum norm by the norm

$$\|f\|_{C^1} := \|f\|_\infty + \|f'\|_\infty,$$

then it is easy to check that the differential operator is bounded. We learn from this that the boundedness of a linear operator depends crucially on the choice of norms.

(4) Let $1 \le p \le \infty$. On l^p, define the *left shift* operator $L : l^p \to l^p$ by

$$L(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots),$$

and correspondingly the *right shift operator* $R : l^p \to l^p$ by

$$R(x_1, x_2, x_3, \dots) = (0, x_1, x_2, x_3, \dots).$$

It is easy to check that L and R are bounded continuous operators of norm 1. Note however the following phenomenon: LR is the identity on l^p, but RL is not. This is in contrast to the finite-dimensional case, where the existence of a right inverse implies the existence of a left inverse, which is always equal to the right inverse.

(5) Let $1 < p < \infty$ and q the dual exponent defined by the property $\frac{1}{p} + \frac{1}{q} = 1$. Let Ω be a measure space with a σ-finite measure μ. Let $g \in L^q(\Omega;\mathbb{R})$. On $L^p(\Omega;\mathbb{R})$, a linear functional is defined by

$$T_g f := \int_\Omega f(x)g(x)d\mu(x),$$

and by Hölder's inequality this is well-defined and bounded, as $|T_g f| \le \|f\|_{L^p}\|g\|_{L^q}$. This shows that $\|T_g\| \le \|g\|_{L^q}$. In fact, the reverse inequality is also true: Assuming $1 < p < \infty$ and thus also $1 < q < \infty$ for simplicity, we set

$$f = \frac{g}{|g|}\left(\frac{|g|}{\|g\|_{L^q}}\right)^{\frac{q}{p}}$$

and find that $\|f\|_{L^p} = 1$ and $\|T_g f\| = \|g\|_{L^q}$, which implies $\|T_g\| \ge \|g\|_{L^q}$. In summary, $\|T_g\| = \|g\|_{L^q}$.

It turns out[9] that *all* linear functionals on $L^p(\Omega)$ are of the form T_g for some $g \in L^q(\Omega)$, so that the dual space of $L^p(\Omega)$ can be identified with $L^q(\Omega)$ under the isometric isomorphism[10] $L^p(\Omega)' \to L^q(\Omega)$, $T_g \mapsto g$.

Similarly, one can show that $L^1(\Omega)'$ is isometrically isomorphic to $L^\infty(\Omega)$; however, the dual space of $L^\infty(\Omega)$ is *not* isomorphic to $L^1(\Omega)$.

Note in particular that for $p = 2$, also $q = 2$, hence the dual space of $L^2(\Omega)$ can be identified with itself. This is a general feature of Hilbert spaces, as we shall see in the next chapter.

(6) We shall prove that $(l^1)'$ is (isometrically) isomorphic to l^∞. First of all, it is quite clear that for any $y \in l^\infty$,

$$T_y x := \sum_{j=1}^{\infty} x_j y_j$$

defines a bounded linear functional on l^1 (i.e., an element of $(l^1)'$). Its norm is at most $\|y\|_\infty$, because

$$|T_y x| = \left| \sum_{j=1}^{\infty} x_j y_j \right| \le \|x\|_1 \|y\|_\infty.$$

Let $\epsilon > 0$, then there exists $n \in \mathbb{N}$ such that $|y_n| > \|y\|_\infty - \epsilon$ (by definition of the supremum). Choosing $x = e_n \in l^1$, we discover

$$|T_y e_n| = |y_n| > \|y\|_\infty - \epsilon,$$

and since ϵ was arbitrary, we conclude $\|T_y\| = \|y\|_\infty$. Hence the linear map $l^\infty \to (l^1)'$, $y \mapsto T_y$ preserves norms and therefore is bounded and injective. It remains to prove surjectivity.

To this end, let $T \in (l^1)'$ and set $y := \sum_{j=1}^\infty T e_j e_j$. Since $|Te_j| \le \|T\|$ for all $j \in \mathbb{N}$, we have $y \in l^\infty$. Moreover, for any $x \in l^1$ and $N \in \mathbb{N}$,

$$\sum_{j=1}^{N} x_j y_j = \sum_{j=1}^{N} x_j T e_j = \sum_{j=1}^{N} T(x_j e_j) = T\left(\sum_{j=1}^{N} x_j e_j \right).$$

On the right hand side, the argument of T converges in l^1 to x, as $N \to \infty$, and as T is continuous, the right hand side converges to Tx. The left hand side converges to $\sum_{j=1}^\infty x_j y_j$, and we conclude $T = T_y$ and hence the map $y \mapsto T_y$ is surjective onto $(l^1)'$, as claimed.

1.4. The Theorem of Hahn-Banach

The Theorem of Hahn-Banach comes in two versions, pertaining to the extension of linear functionals and the separation of convex sets, respectively. For simplicity, we shall assume $\mathbb{K} = \mathbb{R}$ in this entire section. We note however that a suitable formulation of the Hahn-Banach Theorem still holds in complex vector spaces (in both versions).

But first of all, we need to recall a fundamental result from the foundations of mathematics, Zorn's Lemma.

---

[9] We omit the proof, as it requires the Radon-Nikodým Theorem from measure theory, which some of you might not know. See [**5**, Satz II.2.4].

[10] An isometric isomorphism between two normed spaces X and Y is a bounded linear bijection $\iota : X \to Y$ that preserves norms, i.e., $\|\iota(x)\|_Y = \|x\|_X$ for all $x \in X$.

1.4.1. Zorn's Lemma. Let M be a set and $\lesssim$ a partial order, that is, a relation on M that is reflexive, antisymmetric, and transitive:

(1) $\forall x \in M : x \lesssim x$,
(2) $\forall x, y \in M : x \lesssim y \wedge y \lesssim x \Rightarrow x = y$,
(3) $\forall x, y, z \in M : x \lesssim y \wedge y \lesssim z \Rightarrow x \lesssim z$.

A *chain* in M is a totally ordered subset of M, that is, a set $C \subset M$ such that

$$\forall x, y \in C : x \lesssim y \vee y \lesssim x.$$

An element $x \in M$ is said to be an *upper bound* for a subset $C \subset M$ if

$$\forall y \in C : y \lesssim x.$$

An element $m \in M$ is called *maximal* if

$$\forall x \in M : m \lesssim x \Rightarrow m = x.$$

THEOREM 1.20 (Zorn's Lemma). *Let $\lesssim$ be a partial order on the nonempty set M. If every chain in M has an upper bound in M, then there exists a maximal element $m \in M$.*

We will not prove Zorn's Lemma (see for instance [**3**, Chapter 7] for an accessible proof), but just remark that it is equivalent to the axiom of choice. If you don't want to prove Zorn's Lemma, therefore, you may just take it as an axiom (and then prove from it the axiom of choice, if necessary).

1.4.2. Extension of Linear Functionals. If X is a vector space, then we call a functional $p : X \to \mathbb{R}$ *sublinear* if

(1) $p(\lambda x) = \lambda p(x)$ for all $\lambda \ge 0$ and $x \in X$,
(2) $p(x + y) \le p(x) + p(y)$ for all $x, y \in X$.

Obviously, every linear function is sublinear. Another example of a sublinear functional is given by any norm on X. An important class of sublinear functionals is formed by the *Minkowski functionals* that we will discuss shortly.

The following lemma is purely algebraic:

LEMMA 1.21. *Let X be a (real) vector space and $U \subset X$ a subspace. Let $p : X \to \mathbb{R}$ be sublinear and $l : U \to X$ linear such that*

$$l(x) \le p(x) \quad \forall x \in U.$$

Then there exists a linear extension $L : X \to \mathbb{R}$ (that is, $L\restriction_U = l$) such that

$$L(x) \le p(x) \quad \forall x \in X.$$

PROOF. *Step 1.* We first prove the extension result for the special case that the codimension of U in X is one; that is, there exists $x_0 \notin U$ such that $X = \operatorname{span}\{U, x_0\}$. In this case, every $x \in X$ has a unique representation as $x = u + \lambda x_0$ with $u \in U$ and $\lambda \in \mathbb{R}$.

Then, L is a linear extension of l from U to X if and only if there exists $r \in \mathbb{R}$ such that

$$L(x) = l(u) + \lambda r$$

for all $x \in X$ with the decomposition $x = u + \lambda x_0$ given above. It remains to choose $r \in \mathbb{R}$ in such a way that $L \le p$.

If $\lambda = 0$, then $L(x) \le p(x)$ is satisfied by assumption. For $\lambda > 0$, we have $L(u + \lambda x_0) \le p(u + \lambda x_0)$ for all $u \in U$ if and only if

$$r \le p\left(\frac{u}{\lambda} + x_0\right) - l\left(\frac{u}{\lambda}\right)$$

for all $u \in U$, which in turn is equivalent to

$$r \le \inf_{v \in U} (p(v + x_0) - l(v)).$$

For $\lambda < 0$, an analogous calculation gives the requirement

$$r \geq \sup_{v \in U}(l(v) - p(v - x_0)).$$

Therefore, there exists a suitable r if and only if, for all $v, w \in U$,

$$l(w) - p(w - x_0) \leq p(v + x_0) - l(v). \tag{1.8}$$

But, by assumption, for $v, w \in U$ we have

$$l(v + w) \leq p(v + w) \leq p(v + x_0) + p(w - x_0),$$

where in the last step we used the triangle inequality for p. This establishes (1.8) and thus the statement of the Lemma in the case of codimension one.

Step 2. In this step, we apply Zorn's Lemma to the set M of pairs (V, L_V) where $V \supset U$ is a subspace of X and L_V is a linear map $V \to \mathbb{R}$ such that $L_V \restriction_U = l$ and $L_V \leq p \restriction_V$.

On this set, we consider the partial ordering defined by

$$(V_1, L_{V_1}) \lesssim (V_2, L_{V_2}) \quad \text{if and only if} \quad V_1 \subset V_2 \text{ and } L_{V_2} \restriction_{V_1} = L_{V_1}.$$

M is nonempty because, by assumption, $(U, l) \in M$. Also, if $(V_i, L_{V_i})_{i \in I}$ is a totally ordered subset (a chain) in M, then an upper bound is given by

$$V = \bigcup_{i \in I} V_i, \qquad L_V(x) = L_{V_i}(x) \quad \text{if } x \in V_i;$$

indeed, since the (V_i) are totally ordered, V thus defined is a subspace of X, and L_V is a well-defined linear map.

We are therefore in a position to apply Zorn's Lemma, and obtain a maximal element $(X_0, L_{X_0}) \in M$. If we can show $X_0 = X$, we are done. But if $X_0 \neq X$, then there would exist $x_0 \in X \setminus X_0$, and by Step 1 we could extend L_{X_0} to a linear map L_{X_1} on the space $X_1 = \operatorname{span}\{X_0, x_0\}$ in such a way that $L_{X_1} \leq p \restriction_{X_1}$, in contradiction with the maximality of (X_0, L_{X_0}). □

THEOREM 1.22 (Hahn-Banach Extension Theorem). *Let X be a normed space and $U \subset X$ a subspace. For any bounded linear functional $u' \in U'$ there exists a bounded linear functional $x' \in X'$ such that $x' \restriction U = u'$ and $\|x'\| = \|u'\|$ (where $\|\cdot\|$ denotes the operator norm in X' and U', respectively).*

PROOF. For given $u' \in U'$, we apply Lemma 1.21 with the sublinear functional $p(x) = \|u'\|\|x\|$. Clearly, for $v \in U$ we have $u'(v) \leq \|u'\|\|v\| = p(v)$ by definition of the operator norm, so that the Lemma gives us an $x' \in X'$ that extends u' and has $x'(x) \leq p(x)$ for all $x \in X$. Plugging in $-x$ instead of x, we also have $-x'(x) \leq p(x)$, so that in fact $|x'(x)| \leq p(x) = \|u'\|\|x\|$ for all $x \in X$, which implies $\|x'\| \leq \|u'\|$. On the other hand,

$$\|u'\| = \sup_{v \in U, \|v\| \leq 1} |u'(v)| = \sup_{v \in U, \|v\| \leq 1} |x'(v)| \leq \sup_{v \in X, \|v\| \leq 1} |x'(v)| = \|x'\|,$$

so that $\|x'\| = \|u'\|$, and the Theorem is proved. □

The Hahn-Banach Theorem has several important consequences, in particular for the relation between a normed space and its dual. For instance, the following result characterises the norm in nice duality with the definition of the operator norm:

COROLLARY 1.23. *Let X be a normed space, then $\|x\| = \sup_{\|x'\| \leq 1} |x'(x)|$ for all $x \in X$.*

PROOF. For all $x' \in X'$ with $\|x'\| \leq 1$, by definition of the operator norm $|x'(x)| \leq \|x\|$. Conversely, For $x \in X$ define a linear functional $u' : \operatorname{span}\{x\} \to \mathbb{R}$ by $\lambda x \mapsto \lambda\|x\|$, which has norm 1, and extend it by Hahn-Banach to a functional $x' \in X'$ of norm 1. Then $x'(x) = u'(x) = \|x\|$. □

From this, another consequence is immediate:

COROLLARY 1.24. *Let X be a normed space and $x \in X$. If $x'(x) = 0$ for all $x' \in X'$, then $x = 0$.*

Setting $x = x_1 - x_2$, we see that X' separates the points of X: If $x_1 \neq x_2$, then there exists $x' \in X'$ such that $x'(x_1) \neq x'(x_2)$.

1.4.3. Separation of Convex Sets. Before we state the Hahn-Banach Separation Theorem, let us introduce a specific class of sublinear functionals:

DEFINITION 1.25. Let X be a vector space and $A \subset X$ a subset. The *Minkowski functional* $p_A : X \to \mathbb{R}$ corresponding to A is defined by

$$p_A(x) = \inf\left\{\lambda > 0 : \frac{x}{\lambda} \in A\right\}.$$

A is called *absorbing* if $p_A(x) < \infty$ for all $x \in X$.

For example, if $A = B_1(0)$ with respect to a norm $\|\cdot\|$, then $p_A = \|\cdot\|$.
Recall that a subset U of a vector space is called *convex* if, for all $x, y \in U$ and $\lambda \in [0,1]$, also $\lambda x + (1-\lambda)y \in U$.

LEMMA 1.26. *Let X be a normed space and $U \subset X$ a convex subset whose interior*[11] *contains 0. Then,*

(1) $p_U(x) \leq \frac{1}{\epsilon}\|x\|$ *for all $x \in X$ and for any $\epsilon > 0$ such that $B_\epsilon(0) \subset U$. In particular, U is absorbing;*
(2) p_U *is sublinear;*
(3) *If U is open, then $U = p_u^{-1}([0,1))$.*

PROOF. (1) The estimate follows from the fact that $p_U \leq p_V$ whenever $V \subset U$, and $p_{B_\epsilon(0)}(x) = \frac{1}{\epsilon}\|x\|$. Since 0 is contained in the interior of U, there exists $\epsilon > 0$ such that $B_\epsilon(0) \subset U$, and then the estimate implies that p_U is finite, so U is absorbing.

(2) It follows immediately from the definition that $p_U(\lambda x) = \lambda p_U(x)$ for $\lambda \geq 0$. For the triangle inequality, let $x, y \in X$ and $\epsilon > 0$. Pick $\lambda, \mu > 0$ such that $\frac{x}{\lambda} \in U$, $\frac{y}{\mu} \in U$, and $\lambda \leq p_U(x) + \epsilon$ as well as $\mu \leq p_U(y) + \epsilon$. By convexity of U,

$$\frac{\lambda}{\lambda+\mu}\frac{x}{\lambda} + \frac{\mu}{\lambda+\mu}\frac{y}{\mu} = \frac{x+y}{\lambda+\mu} \in U,$$

whence $p_U(x+y) \leq \lambda + \mu \leq p_U(x) + p_U(y) + 2\epsilon$. Since ϵ was arbitrary, the sublinearity of p_U follows.

(3) If $p_U(x) < 1$, then there exists $0 < \lambda < 1$ such that $\frac{x}{\lambda} \in U$, and as $0 \in U$, we have by convexity $x = \lambda\frac{x}{\lambda} + (1-\lambda)\cdot 0 \in U$.

Conversely, if $p_U(x) \geq 1$, then $\frac{x}{\lambda} \notin U$ for all $\lambda < 1$. But as the complement U^c of U is closed, we have that

$$x = \lim_{\lambda \nearrow 1} \frac{x}{\lambda} \in U^c.$$

□

We introduce the notation $A \pm B := \{a \pm b : a \in A, b \in B\}$ for two subsets A, B of a vector space X. Now we are ready to prove the Separation Theorem:

THEOREM 1.27 (Hahn-Banach Separation Theorem). *Let X be a normed space, $V_1, V_2 \subset X$ disjoint convex subsets, and V_1 open. Then there exists $x' \in X'$ such that*

$$x'(v_1) < x'(v_2) \quad \text{for all } v_1 \in V_1, v_2 \in V_2.$$

[11]Recall that the interior of a set U is the union of all open subsets of U.

Proof. We may assume V_1 and V_2 to be nonempty, since otherwise the statement is trivial (and rather meaningless).

Step 1. In the first step, we show that if $V \subset X$ is nonempty, convex, and open with $0 \notin V$, then there exists $x' \in X'$ such that $x' \restriction_V < 0$.

To this end, let $x_0 \in V$, set $y_0 := -x_0$ and $U = V - \{x_0\}$. U is still open and convex, and we have $0 \in U$ but $y_0 \notin U$.

By Lemma 1.26, the corresponding Minkowski functional p_U is sublinear, and $p_U(y_0) \geq 1$. Let $Y := \operatorname{span}\{y_0\}$ and define on Y the linear functional

$$y'(\lambda y_0) := \lambda p_U(y_0), \quad \lambda \in \mathbb{R}.$$

Then $y' \leq p_U \restriction_Y$, because for $\lambda \geq 0$ equality holds, and for $\lambda < 0$ we have $y'(\lambda y_0) \leq 0$. Hence, by Lemma 1.21, we obtain a linear extension x' of y' such that $x' \leq p_U$. In fact, x' is bounded, because by Lemma 1.26, for any $x \in X$,

$$|x'(x)| = \max\{x'(x), -x'(x)\} \leq \max\{p_U(x), p_U(-x)\} \leq \frac{1}{\epsilon}\|x\|,$$

where $\epsilon > 0$ is such that $B_\epsilon(0) \subset U$ (such ϵ exists because $0 \in U$ and U is open).

If $x \in V$, then $x = u - y_0$ for some $u \in U$, and so

$$x'(x) = x'(u) - x'(y_0) \leq p_U(u) - p_U(y_0),$$

where we used $x' \leq p_U$ and $x'(y_0) = u'(y_0) = p_U(y_0) \geq 1$. But as by Lemma 1.26 $p_U(u) < 1$, we get $x'(x) < 0$, so that $x' \in X'$ is as desired.

Step 2. Let now V_1, V_2 as in the statement of the Theorem. Set $V := V_1 - V_2$. Then V is convex, because if $v_1, w_1 \in V_1$ and $v_2, w_2 \in V_2$, then for $\lambda \in [0,1]$,

$$\lambda(v_1 - v_2) + (1-\lambda)(w_1 - w_2) = [\lambda v_1 + (1-\lambda)w_1] - [\lambda v_2 + (1-\lambda)w_2] \in V_1 - V_2.$$

Moreover, $V = \bigcup_{v \in V_2}(V_1 - \{v\})$ is open as a union of open sets. Also, $0 \notin V$ since V_1 and V_2 are disjoint.

The application of Step 1 now yields a functional $x' \in X'$ such that $x'(v) < 0$ for all $v \in V$, which means that for all $v_1 \in V_1$ and $v_2 \in V_2$ we have $0 > x'(v_1 - v_2) = x'(v_1) - x'(v_2)$, as claimed. □

1.5. Reflexivity and Weak Convergence

1.5.1. Reflexivity. Let X be a normed space, then X'', i.e., the dual of the dual space of X, is called the *bidual* space of X. Consider the canonical inclusion $\iota : X \to X''$,

$$\iota[x] : X' \to \mathbb{K}, \quad \iota[x](x') = x'(x).$$

This may seem confusing at first glance. The map ι sends an element $x \to X$ to an element $\iota[x] \in X''$, that is, $\iota[x]$ is a bounded linear map from X' to $\mathbb{K}$, and it assigns to $x' \in X'$ the value $x'(x)$.

Since x' is linear, then so is ι. For boundedness, observe

$$\|\iota[x]\| = \sup_{\|x'\| \leq 1} |\iota[x](x')| = \sup_{\|x'\| \leq 1} |x'(x)| = \|x\|,$$

where we used Corollary 1.23 in the last step. Therefore, ι is an isometric (and hence injective) linear map from X into X''.

Definition 1.28. A Banach space X is called *reflexive* if the canonical inclusion $\iota : X \to X''$ is surjective.

Note that an incomplete normed space can never have this property, because by Proposition 1.17, the bidual space is always complete, and there can be no isometric isomorphism between an incomplete and a complete normed space.

EXAMPLE 1.29. (1) $\mathbb{K}^d$ is reflexive: It is known from Linear Algebra that every element $x' \in (\mathbb{K}^d)'$ has the form $x \mapsto j(x') \cdot x$ with $j(x') \in \mathbb{K}^d$, and $j : X' \to \mathbb{K}^d$ is an isometric isomorphism. Likewise, every element $x'' \in (\mathbb{K}^d)''$ takes the form $x' \mapsto i(x'') \cdot j(x')$ for some $i(x'') \in \mathbb{K}^d$, and $i : X'' \to \mathbb{K}^d$ is an isometric isomorphism.

Let now $x'' \in X''$, then for all $x' \in X'$,

$$\iota[i(x'')](x') = x'(i(x'')) = j(x') \cdot i(x'') = x''(x'),$$

so that $\iota[i(x'')] = x''$, and thus ι is surjective and $\mathbb{K}^d$ is reflexive.

(2) Let Ω be a measure space. For $1 < p < \infty$, $L^p(\Omega)$ is reflexive: We have seen that every element of $(L^p)'$ can be identified with an element of L^q, where $\frac{1}{p} + \frac{1}{q} = 1$. In turn, every element of $(L^q)' \simeq (L^p)''$ can be identified with an element of L^p. Under these identifications (which can be expressed as isometric isomorphisms as in the previous example), for any $f \in L^p$ and $g \in (L^p)' \simeq L^q$, we have

$$\iota[f](g) = g(f) = \int_\Omega fg d\mu = f(g),$$

where, by abuse of notation[12], we used f to denote both an element of L^p and the corresponding element of the bidual, and similarly for g.

(3) We saw in Example 1.19(6) that $(l^1)' \simeq l^\infty$, where an element $y \in l^\infty$ can be identified with the functional $x \mapsto \sum_{j=1}^\infty x_j y_j$. The canonical map $l^1 \to (l^\infty)'$ is therefore given by

$$\iota[x](y) = y(x) = \sum_{j=1}^\infty x_j y_j.$$

We show that this map is not surjective. To this end, let $c \subset l^\infty$ be the subspace of convergent sequences. Then $\lim : c \to \mathbb{K}$, $x \mapsto \lim_{n\to\infty} x_n$ is a bounded linear functional on c, which by the Theorem of Hahn-Banach can be extended to a bounded linear operator $L \in (l^\infty)'$. Suppose this L were in the image of ι, that is, there existed an $l \in l^1$ such that $L(y) = \sum_{j=1}^\infty l_j y_j$ for all $y \in l^\infty$. But then, for each $k \in \mathbb{N}$,

$$l_k = L(e_k) = \lim_{n\to\infty} (e_k)_n = 0,$$

and so $L = 0$, in contradiction to L being an extension of the limit operator (which is of course not identically zero). This shows that l^1 is not reflexive.

(4) Generally, it is known that l^p is reflexive for $1 < p < \infty$ but not for $p = 1$ or $p = \infty$, and the same is true for $L^p(\Omega)$.

1.5.2. Weak Convergence.

DEFINITION 1.30. Let X be a normed space. A sequence $(x_n)_{n\in\mathbb{N}} \subset X$ *converges weakly* to $x \in X$ if

$$x'(x_n) \to x'(x) \quad \text{as } n \to \infty$$

for every $x' \in X'$. In this case, we write $x_n \rightharpoonup x$ as $n \to \infty$.

The usual notion of convergence (with respect to the norm) is sometimes called *strong convergence* in order to distinguish it from weak convergence.

The weak limit, if it exists, is unique, because X' separates the points of X (Corollary 1.24).

[12]The argument can be made cleaner, but more complicated, by explicitly spelling out the respective isomorphisms as in the previous example.

EXAMPLE 1.31. (1) Every (strongly) convergent sequence is weakly convergent (with the same limit). Indeed, let $x_n \to x$ and $x' \in X'$. Then,

$$|x'(x_n) - x'(x)| = |x'(x_n - x)| \leq \|x'\| \|x_n - x\| \to 0.$$

(2) Let $1 < p < \infty$. We know that $(l^p)' \simeq l^q$ where $\frac{1}{p} + \frac{1}{q} = 1$. Consider the sequence $(e_n)_{n\in\mathbb{N}} \subset l^p$. Then for each $n \in \mathbb{N}$, we have $\|e_n\|_p = 1$. On the other hand, let $y \in l^q$ be given. Since $y \in l^q$ implies $\lim_{n\to\infty} y_n = 0$, we have

$$y(e_n) = \sum_{j=1}^{\infty} y_j (e_n)_j = y_n \to 0,$$

so that $e_n \rightharpoonup 0$ although the norms of e_n do not converge to zero. In fact, (e_n) cannot converge strongly: If it did, then the strong and weak limits would be the same (i.e., zero), but then also the norms would converge to zero, which they don't.

(3) For $1 \leq p < \infty$, consider the space $L^p(0;1)$ and set $u_n = \sin(n\cdot)$. Then $(u_n)_{n\in\mathbb{N}}$ is a sequence in $L^p(0,1)$. The dual space is (isomorphic to) $L^q(0,1)$. Let $g \in L^q(0,1)$, then by the Riemann-Lebesgue Lemma (which you might or might not know),

$$\lim_{n\to\infty} \int_0^1 u_n(x) g(x) dx = 0,$$

so that $u_n \rightharpoonup 0$. However, as is easy to check, $\|u_n\|_p$ does not converge to zero, so that (u_n) is not strongly convergent. This example shows that the gap between weak and strong convergence is often caused by *high frequency oscillations.*

We have seen in the last two examples that the norm may 'drop' after passing to the weak limit – e.g., $\|e_n\|_p = 1$ but the weak limit is zero (and thus has zero norm, of course). This means, in particular, that the norm is not continuous with respect to weak convergence. In general, we have the following result:

PROPOSITION 1.32 (weak lower semicontinuity of the norm). *Let X be a normed space and $x_n \rightharpoonup x$. Then,*

$$\|x\| \leq \liminf_{n\to\infty} \|x_n\|.$$

PROOF. The proof is a one-liner, once we recall Corollary 1.23:

$$\|x\| = \sup_{\|x'\|\leq 1} x'(x) = \sup_{\|x'\|\leq 1} \lim_{n\to\infty} x'(x_n) \leq \sup_{\|x'\|\leq 1} \liminf_{n\to\infty} \|x_n\| \|x'\| = \liminf_{n\to\infty} \|x_n\|.$$

□

A main reason why weak convergence is useful is that it restores some of the compactness properties that got lost in infinite dimensions. For instance, $(e_n)_{n\in\mathbb{N}}$ has no strongly convergent subsequence although it is bounded in the norm $\|\cdot\|_p$, in contrast to the finite dimensional case (Bolzano-Weierstraß). A weak limit, however, does exist.

A prototypical weak compactness result reads as follows:

THEOREM 1.33 (Theorem of Banach-Alaoglu (reflexive case)). *Let X be a reflexive Banach space with separable dual space. Then every bounded*[13] *sequence $(x_n)_{n\in\mathbb{N}}$ has a weakly convergent subsequence.*

REMARK 1.34. The assumption of separability can be removed (see [**5**, Theorem III.3.7]): A bounded sequence in a reflexive space always has a weakly convergent subspace.

[13] Of course, a sequence $(x_n)_{n\in\mathbb{N}}$ in a normed space is called bounded if $\sup_{n\in\mathbb{N}} \|x_n\| < \infty$.

PROOF. As X' is separable, there is a countable dense (in the operator norm) set $(x'_n)_{n\in\mathbb{N}} \subset X'$. Let $(x_k)_{k\in\mathbb{N}}$ be a bounded sequence in X, then for each $n \in \mathbb{N}$, the sequence $(x'_n(x_k))_{k\in\mathbb{N}}$ is bounded in $\mathbb{K}$.

Step 1. By Bolzano-Weierstraß, we find a subsequence $(x_j^1)_{j\in\mathbb{N}}$ of $(x_k)_{k\in\mathbb{N}}$ such that $x'_1(x_j^1)$ converges, as $j \to \infty$. From this, we may select a further subsequence $(x_j^2)_{j\in\mathbb{N}}$ such that $x'_2(x_j^2)$ converges; and since (x_j^2) is a subsequence of (x_j^1), also $x'_1(x_j^2)$ converges. Continuing in this way, for each $n \in \mathbb{N}$ we find a subsequence $(x_j^n)_{j\in\mathbb{N}}$ of the original sequence $(x_k)_{k\in\mathbb{N}}$ such that $(x'_m(x_j^n))_{j\in\mathbb{N}}$ converges for all $m = 1, \ldots, n$. Upon setting

$$y_j := x_j^j,$$

we find that $(y_j)_{j\in\mathbb{N}}$ is a subsequence of $(x_k)_{k\in\mathbb{N}}$ such that $(x'_n(y_j))_{j\in\mathbb{N}}$ is convergent for every $n \in \mathbb{N}$. (The type of argument used so far is known as a *diagonal argument* and is quite standard in analysis.)

Step 2. Let now $x' \in X'$ be arbitrary and $\epsilon > 0$. Denote $M := \sup_{k\in\mathbb{N}} \|x_k\| < \infty$. Pick $n \in \mathbb{N}$ such that

$$\|x' - x'_n\| < \frac{\epsilon}{4M},$$

which is possible by density of $(x'_n)_{n\in\mathbb{N}}$ in X'. Let also $N \in \mathbb{N}$ be so large that

$$|x'_n(y_k) - x'_n(y_j)| < \frac{\epsilon}{2}$$

for $k, j \geq N$. Then, for such k, j, we have

$$\begin{aligned} |x'(y_k) - x'(y_j)| &\leq |x'(y_k) - x'_n(y_k)| + |x'_n(y_k) - x'_n(y_j)| + |x'_n(y_j) - x'(y_j)| \\ &< M\|x' - x'_n\| + \frac{\epsilon}{2} < \epsilon, \end{aligned}$$

since $\|y_k\| \leq M$ for all $k \in \mathbb{N}$. Therefore, $(x'(y_j))_{j\in\mathbb{N}}$ is Cauchy and thus convergent.

Step 3. Finally we identify the weak limit of (y_j), for which we will use reflexivity. Consider the linear functional $l : X' \to \mathbb{K}$ defined as

$$l(x') := \lim_{j\to\infty} x'(y_j).$$

It is bounded, because

$$|l(x')| = \lim_{j\to\infty} |x'(y_j)| \leq \liminf_{j\to\infty} \|y_j\|\|x'\| \leq M\|x'\|,$$

so that $l \in X''$. As X is assumed reflexive, there exists $x \in X$ such that $\iota(x) = l$, and thus

$$x'(x) = l(x') = \lim_{j\to\infty} x'(y_j)$$

for all $x' \in X'$, which means that $y_j \rightharpoonup x$.

□

Note in particular that Theorem 1.33 applies to the spaces L^p and l^p: A bounded sequence in L^p (or l^p) has a weakly convergent subsequence.

We conclude this section by noting the following interesting connection between weak convergence and convexity. We saw in Example 1.31(2) that the unit vectors e_n, though all contained in the closed set $\{x \in X : \|x\| = 1\}$, weakly converge to zero, which is no longer in this closed set. One might say that *(strongly) closed sets are in general not weakly closed.* However, a closed set remains weakly closed if it is convex:

THEOREM 1.35. *Let X be a normed space and $U \subset X$ a closed and convex subset. Suppose $(x_n)_{n\in\mathbb{N}} \subset U$ converges weakly to x. Then $x \in U$.*

PROOF. We prove the statement only for $\mathbb{K} = \mathbb{R}$ (as we didn't bother to prove the complex version of the Hahn-Banach Theorem). Suppose $x \notin U$. Since U is closed, its complement U^c is open, so that there exists an open ball $B_\epsilon(x)$ that is disjoint from U. By the Hahn-Banach Separation Theorem, therefore, there exists $x' \in X' \setminus \{0\}$ such that $x'(x+y) < x'(u)$ for all $y \in B_\epsilon(0)$ and $u \in U$. As $\sup_{\|y\| \le \epsilon} x'(y) = \epsilon \|x'\|$ by definition of the operator norm, we get

$$x'(x) + \epsilon \|x'\| \le x'(x_n)$$

for all $n \in \mathbb{N}$, in contradiction with $x'(x) = \lim_{n\to\infty} x'(x_n)$. □

1.6. The Fréchet Derivative

We take a very brief excursion into nonlinear functional analysis, i.e., the analysis of *nonlinear* operators between two normed spaces. As in finite-dimensional analysis, the question arises how best to approximate a nonlinear operator by a linear one near a given point, and the answer is given by a sort of differential. In analogy with the finite-dimensional case, therefore, we define:

DEFINITION 1.36 (Fréchet derivative). Let X, Y be normed spaces, $U \subset X$ open, and $f : U \to Y$ a (not necessarily linear) map. Then f is said to be *Fréchet differentiable* at a point $x_0 \in U$ if there exists $T \in L(X,Y)$ such that

$$\lim_{h\to 0} \frac{f(x_0 + hv) - f(x_0)}{h} = Tv$$

uniformly in $v \in \overline{B_1(0)}$. In this case, we write $T =: Df(x_0)$.

The map f is called Fréchet differentiable on U if it is so at each $x_0 \in U$. The function $Df : U \to L(X,Y)$ is then called the *Fréchet derivative* of f.

It may seem confusing that the derivative of a map $X \to Y$ at a point x_0 is an *operator*, i.e., an element of $L(X,Y)$. Note however that, for a map $\mathbb{R}^n \to \mathbb{R}^m$, the total differential (or Jacobian, or whatever you like to call it) is as well a linear map $\mathbb{R}^n \to \mathbb{R}^m$ represented by an $(m \times n)$-matrix, so $Df(x_0) \in L(X,Y)$ does make a lot of sense.

The idea (well-known from Analysis I & II) that the derivative should be viewed as a linear approximation is made precise by the following result.

THEOREM 1.37 (Fréchet derivative as a linear approximation). *A map $f : U \to Y$ is Fréchet differentiable at $x_0 \in U$ if and only if there exists $T \in L(X,Y)$ and $r : U - \{x_0\} \to Y$ such that*

$$f(x_0 + u) = f(x_0) + Tu + r(u)$$

and[14]

$$\lim_{\|u\|\to 0} \frac{r(u)}{\|u\|} = 0. \tag{1.9}$$

In this case, $T = Df(x_0)$.

PROOF. Suppose that f is Fréchet differentiable at x_0. Setting

$$r(u) := f(x_0 + u) - f(x_0) - Df(x_0)u,$$

we estimate

$$\frac{r(u)}{\|u\|} = \frac{f\left(x_0 + \|u\| \frac{u}{\|u\|}\right) - f(x_0)}{\|u\|} - Df(x_0)\left(\frac{u}{\|u\|}\right) \to 0$$

[14]To be precise, the limit condition means: For all $\epsilon > 0$ there exists $\delta > 0$ such that $\|u\| < \delta$ implies $\frac{r(u)}{\|u\|} < \epsilon$.

as $\|u\| \to 0$. Note we have used the *uniform* convergence of the difference quotient, as the direction $\frac{u}{\|u\|}$ can change as $\|u\| \to 0$.

Conversely, assume (1.9) for some T and r as required, and let $v \in \overline{B_1(0)}$. Then,

$$\left\| \frac{f(x_0 + hv) - f(x_0)}{h} - Tv \right\| = \|v\| \left\| \frac{f(x_0 + hv) - f(x_0) - T(hv)}{h\|v\|} \right\| = \|v\| \frac{\|r(hv)\|}{\|hv\|} \to 0$$

uniformly in $v \in \overline{B_1(0)}$ as $h \to 0$. □

It is easy to see that the Fréchet derivative is linear: If two maps $f, g : U \to Y$ are Fréchet-differentiable, then so is $\alpha f + \beta g$ ($\alpha, \beta \in \mathbb{K}$), and $D(\alpha f + \beta g) = \alpha Df + \beta Dg$. In fact, several familiar properties of derivatives, like the chain rule, a version of the Mean Value Theorem, or the Implicit Function Theorem, remain valid in the Fréchet sense. But instead of developing the theory further[15], let us just give give a simple example.

EXAMPLE 1.38. On $C([0,1];\mathbb{R})$ with the supremum norm, define $f : X \to X$ by $f(x) = x^2$ (note x is a continuous function and the square is taken pointwise). For any $x, v \in X$,

$$\frac{f(x + hv) - f(x)}{h} = \frac{2hxv + h^2v^2}{h} = 2xv + hv^2 \to 2xv$$

uniformly, as $h \to 0$, in $v \in \overline{B_1(0)}$. It follows that $Df(x) = 2x$, that is: the Fréchet derivative $Df(x) \in L(X, X)$ is the multiplication operator with $2x$, so that for $v \in X$, $Df(x)v = 2xv$.

[15]More on the topic can be learned in Prof. Dall'Acqua's course *Nonlinear Functional Analysis* next term.

CHAPTER 2

Hilbert Spaces

Hilbert spaces are a special class of Banach spaces whose norm stems from an inner product. Such spaces are in certain ways very similar to (finite-dimensional) Euclidean space and therefore admit geometric intuition. For instance, a Hilbert space is dual to itself, has an orthonormal basis, allows for orthogonal projections, etc. The notion of Hilbert space is fundamental to quantum mechanics.

2.1. Definition and Fundamental Properties

DEFINITION 2.1 (inner product space). A $\mathbb{K}$-vector space X is an *inner product space* if it is equipped with an *inner product* $X \times X \to \mathbb{K}$ satisfying the following:

(1) For all $x \in X$: $\mathbb{R} \ni (x,x) \geq 0$ with equality if and only if $x = 0$;
(2) For all $x, y, z \in X$ and $\alpha, \beta \in \mathbb{K}$:
$$(\alpha x + \beta y, z) = \alpha(x,z) + \beta(y,z);$$
(3) For all $x, y \in X$: $(x,y) = \overline{(y,x)}$.

When $\mathbb{K} = \mathbb{R}$, then the complex conjugation in the last axiom does nothing, and we just have $(x,y) = (y,x)$ (which also entails linearity in the second argument). In this case, the inner product can be characterised as a *positive definite symmetric bilinear form*, and an inner product space over $\mathbb{R}$ is often called a *Euclidean* space.

If $\mathbb{K} = \mathbb{C}$, the inner product is not linear in the second argument, but only linear up to complex conjugation[1]; in this case, one says that the inner product is a *positive definite conjugate symmetric sesquilinear*[2] *form*, and an inner product space over $\mathbb{C}$ is also called a *unitary* space.

Obviously, $\mathbb{K}^d$ is an inner product space with the usual inner product $(x,y) = \sum_{j=1}^d x_j \overline{y_j}$.

If X is an inner product space, then
$$\|\cdot\| : X \to \mathbb{R}, \quad \|x\| := \sqrt{(x,x)}$$
is well-defined, because $(x,x) \geq 0$ according to (1). We will show shortly that, as the notation suggests, $\|\cdot\|$ is a norm on X, but as a preparation, we first prove the Cauchy-Schwarz inequality[3]:

PROPOSITION 2.2 (Cauchy-Schwarz inequality). *Let X be an inner product space and $x, y \in X$. Then,*
$$|(x,y)| \leq \|x\|\|y\|.$$

PROOF. If $y = 0$, then both sides are zero. So let $y \neq 0$ and therefore $\|y\| > 0$. If $\lambda \in \mathbb{K}$, then from sesquilinearity,
$$0 \leq (x - \lambda y, x - \lambda y) = \|x\|^2 + \lambda\bar{\lambda}\|y\|^2 - \bar{\lambda}(x,y) - \lambda(y,x).$$

[1] More precisely: $(x, \alpha y + \beta z) = \bar{\alpha}(x,y) + \bar{\beta}(x,z)$.

[2] from latin "sesqui": one and a half.

[3] Please make sure to always spell the Schwarz (Hermann Amandus, 1843-1921) from the Cauchy-Schwarz inequality correctly. In fact, there is another Schwartz (Laurent, 1915-2002) spelt differently. The latter is known for the theory of distributions and appears, e.g., in *Schwartz space*.

The choice $\lambda = \frac{(x,y)}{\|y\|^2}$ yields

$$0 \le \|x\|^2 + \frac{|(x,y)|^2}{\|y\|^4}\|y\|^2 - 2\frac{|(x,y)|^2}{\|y\|^2} = \|x\|^2 - \frac{|(x,y)|^2}{\|y\|^2},$$

which implies the statement. □

PROPOSITION 2.3. *Let X be an inner product space. Then the map $\|\cdot\|$ is a norm on X.*

This norm is called the norm *induced* by the inner product.

PROOF. As $(x,x) \ge 0$, the map $\|\cdot\|$ is real-valued and non-negative. We have $\|x\| = 0$ if and only if $(x,x) = 0$, i.e., if and only if $x = 0$.

For $\lambda \in \mathbb{K}$ and $x \in X$,

$$\|\lambda x\| = \sqrt{(\lambda x, \lambda x)} = \sqrt{|\lambda|^2 (x,x)} = |\lambda|\|x\|$$

thanks to sesquilinearity.

For $x, y \in X$, finally,

$$\begin{aligned}\|x+y\|^2 &= (x+y, x+y) = (x,x) + (x,y) + (y,x) + (y,y)\\ &= \|x\|^2 + 2\Re(x,y) + \|y\|^2\\ &\le \|x\|^2 + 2|(x,y)| + \|y\|^2\\ &\le \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,\end{aligned}$$

where we used the Cauchy-Schwarz inequality. □

DEFINITION 2.4 (Hilbert space). A Hilbert space is an inner product space which is complete with respect to the norm induced by its inner product.

So every Hilbert space is, in particular, a Banach space.

EXAMPLE 2.5.
(1) $\mathbb{R}^d$ and $\mathbb{C}^d$ are (respectively real and complex) Hilbert spaces, as already noted.
(2) If Ω is a measure space, then it is easy to check that

$$(f,g) := \int_\Omega f(x)\overline{g(x)}dx \tag{2.1}$$

defines an inner product on the vector space $L^2(\Omega;\mathbb{K})$, and this inner product induces the norm of L^2. As L^2 is complete, we conclude that it is a Hilbert space.
(3) Taking $\Omega = \mathbb{N}$ together with the counting measure, then $L^2(\Omega) = l^2$, so that in particular l^2 is Hilbert.
(4) We can as well equip the vector space $C([0,1];\mathbb{K})$ with the same inner product (2.1), but this inner product space will not be complete (its completion is precisely $L^2(0,1)$ – think it over!), so it is not Hilbert.

For most of this section, we will not make use of completeness, so that all results except Corollary 2.12 are true in (possibly incomplete) inner product spaces.

Two vectors $x, y \in X$ are *orthogonal* if $(x,y) = 0$. In particular, the zero vector is orthogonal to any vector in X. One occasionally writes $x \perp y$ to express orthogonality. If $U \subset X$ is a subspace, then

$$U^\perp := \{x \in X : x \perp u \quad \forall u \in U\}$$

is called the *orthogonal complement* of U in X. The orthogonal complement is always a closed subspace, even when U itself is not closed (exercise).

A (not necessarily finite) family $(e_k)_{k\in I}$ indexed by a set I is called an *orthonormal system* in X if

$$(e_k, e_l) = \begin{cases} 0 & \text{if } k \neq l, \\ 1 & \text{if } k = l. \end{cases}$$

PROPOSITION 2.6 (Theorem of Pythagoras). *Let $(e_k)_{k\in I}$ be an orthonormal system and $I' \subset I$ a finite set of indices. Then, any linear combination $x = \sum_{k\in I'} c_k e_k$ ($c_k \in \mathbb{K}$) satisfies:*

$$\|x\|^2 = \sum_{k\in I'} |c_k|^2.$$

PROOF. Orthonormality and sesquilinearity yield

$$\|x\|^2 = \left(\sum_{k\in I'} c_k e_k, \sum_{j\in I'} c_j e_j \right) = \sum_{j,k\in I'} c_k \bar{c}_j (e_k, e_j) = \sum_{k\in I'} |c_k|^2.$$

□

If $I' \subset I$ is a finite set of indices and $x \in X$, then

$$P_{I'} v := \sum_{k\in I'} (x, e_k) e_k$$

the *orthogonal projection* of x onto the subspace spanned by $(e_k)_{k\in I'}$. By Pythagoras, $\|P_{I'}x\|^2 = \sum_{k\in I'} |(x, e_k)|^2$.

PROPOSITION 2.7 (properties of the projection). *Let $x \in X$, $I' \subset I$ finite, and $(c_k)_{k\in I'} \subset \mathbb{K}$. For the orthogonal projection $P_{I'}x$ we have:*

(1) $P_{I'} \left(\sum_{k\in I'} c_k e_k\right) = \sum_{k\in I'} c_k e_k$, *in particular:* $P_{I'}^2 = P_{I'}$;
(2) $\|P_{I'}x\| \leq \|x\|$.

PROOF. (1) From orthonormality,

$$P_{I'} \left(\sum_{k\in I'} c_k e_k \right) = \sum_{j\in I'} \left(\sum_{k\in I'} c_k e_k, e_j \right) e_j = \sum_{j,k\in I'} c_k (e_k, e_j) e_j = \sum_{k\in I'} c_k e_k.$$

(2) It holds that

$$\begin{aligned} \|x - P_{I'}x\|^2 &= \|x\|^2 - 2\Re(x, P_{I'}x) + \|P_{I'}x\|^2 \\ &= \|x\|^2 - 2\Re\left(x, \sum_{k\in I'} (x, e_k) e_k \right) + \|P_{I'}x\|^2 \\ &= \|x\|^2 - 2 \sum_{k\in I'} |(x, e_k)|^2 + \|P_{I'}x\|^2 \\ &= \|x\|^2 - \|P_{I'}x\|^2, \end{aligned}$$

where we used Pythagoras in the last step.

□

An orthonormal system $(e_k)_{k\in I}$ is called a *Hilbert basis* if for every $x \in X$ and $\epsilon > 0$ there exists a finite linear combination such that

$$\left\| x - \sum_{k\in I'} c_k e_k \right\| < \epsilon \quad (I' \subset I \text{ finite}, \quad c_k \in \mathbb{K}).$$

EXAMPLE 2.8. (1) For the Hilbert space l^2, the unit vectors $(e_n)_{n\in\mathbb{N}}$ form a Hilbert basis.

(2) We want to find a Hilbert basis for $L^2(0,1)$. To this end, note first that the polynomials are dense in $C([0,1])$ with respect to the supremum norm, and *a fortiori* with respect to the L^2 norm. Moreover, $C([0,1])$ is dense in $L^2(0,1)$, hence $\operatorname{span}\{x \mapsto x^n : n \in \mathbb{N} \cup \{0\}\}$ is dense. However, the monomials $x \mapsto x^n$ are not orthonormal. But they can be orthonormalised via the Gram-Schmidt procedure, which you hopefully know from Linear Algebra (if not, please give me a shout!).
(3) In the next section we shall see a Hilbert basis for $L^2_{per}(\mathbb{R}, \mathbb{C})$, the space of 2π-periodic locally L^2 functions $\mathbb{R} \to \mathbb{C}$.

REMARK 2.9. A basis in the sense of Linear Algebra is then called *algebraic basis* or *Hamel basis* in order to distinguish it from a Hilbert basis. If $(e_k)_{k\in K}$ is an algebraic basis, then for each $x \in X$ there is a finite set $I' \subset I$ and coefficients $(c_k)_{k\in I'} \subset \mathbb{K}$ such that $x = \sum_{k\in I'} c_k e_k$. In contrast, a Hilbert basis generally only allows to *approximate*, but not represent exactly, the vector x by a finite linear combination.

You proved in the exercises, using Zorn's Lemma, that every vector space has a basis. In a similar way one can show that every inner product space has a Hilbert basis. The decisive advantage of a Hilbert basis over an algebraic one is that, in many cases, a Hilbert basis can be explicitly constructed, whereas usually (in infinite dimensions) nothing is known about an algebraic basis save its mere existence.

THEOREM 2.10. *Assume X has a countable Hilbert basis $(e_k)_{k\in\mathbb{Z}}$. Then for every $x \in X$,*

$$x = \sum_{k=-\infty}^{\infty} (x, e_k)e_k := \lim_{N\to\infty} \sum_{k=-N}^{N} (x, e_k)e_k,$$

and for $x, y \in X$ we have the identity

$$(x, y) = \sum_{k=-\infty}^{\infty} (x, e_k)(e_k, y).$$

In particular, Parseval's identity holds:

$$\|x\|^2 = \sum_{k=-\infty}^{\infty} |(x, e_k)|^2.$$

PROOF. Let $\epsilon > 0$. By assumption there exists $N \in \mathbb{N}$ and coefficients $(c_k)_{k=-N,\dots,N} \subset \mathbb{K}$ such that

$$\left\| x - \sum_{k=-N}^{N} c_k e_k \right\| < \frac{\epsilon}{2}.$$

Denote $P_N : X \to X$ the orthogonal projection onto the subspace spanned by $(e_k)_{k=-N,\dots,N}$, then by Proposition 2.7,

$$\frac{\epsilon}{2} > \left\| P_N \left(x - \sum_{k=-N}^{N} c_k e_k \right) \right\| = \left\| \sum_{k=-N}^{N} (x, e_k)e_k - \sum_{k=-N}^{N} c_k e_k \right\|,$$

hence in total

$$\left\| x - \sum_{k=-N}^{N} (x, e_k)e_k \right\| \le \left\| x - \sum_{k=-N}^{N} c_k e_k \right\| + \left\| \sum_{k=-N}^{N} (x, e_k)e_k - \sum_{k=-N}^{N} c_k e_k \right\| < \epsilon.$$

So we have $\lim_{N\to\infty} P_N x = x$ and $\lim_{N\to\infty} P_N y = y$, and also

$$\sum_{k=-\infty}^{\infty} |(x, e_k)|^2 \le \|x\|^2$$

because the sum on the left hand side equals $\lim_{N\to\infty} \|P_N x\|^2$. Therefore,

$$\sum_{k=-\infty}^{\infty} (x, e_k)(e_k, y) = \lim_{N\to\infty} \sum_{k=-N}^{N} (x, e_k)(e_k, y) = \lim_{N\to\infty} (P_N x, P_N y) = (x, y),$$

where the last equality follows from continuity of the scalar product (think it over). □

Of course we could as well have indexed our Hilbert basis with $\mathbb{N}$ instead of $\mathbb{Z}$, but taking $\mathbb{Z}$ as an index set facilitates notation in the next section.

The following Proposition tells us when a Hilbert space actually has a countable[4] Hilbert basis:

PROPOSITION 2.11. *A Hilbert space H has a countable Hilbert basis if and only if it is separable.*

PROOF. If H has a countable Hilbert basis $(e_n)_{n\in\mathbb{N}}$, then

$$\left\{\sum_{n=1}^{N} \alpha_n e_n : \alpha_n \in \mathbb{Q}, N \in \mathbb{N}\right\}$$

is countable and dense in H (if $\mathbb{K} = \mathbb{C}$, replace $\mathbb{Q}$ by $\mathbb{Q} + i\mathbb{Q}$).

Conversely, let $(x_n)_{n\in\mathbb{N}}$ be a countable dense subset of H. We may assume the family $(x_n)_{n\in\mathbb{N}}$ to be linearly independent (otherwise check for every $n \in \mathbb{N}$ whether x_n is a finite linear combination of the other vectors, and delete it if necessary). Then, the following recursion (known as Gram-Schmidt orthonormalisation) turns $(x_n)_{n\in\mathbb{N}}$ into a Hilbert basis:

$$e_1 := \frac{x_1}{\|x_1\|}, \qquad e_n := \frac{x_n - \sum_{k=1}^{n-1}(x_n, e_k)e_k}{\|x_n - \sum_{k=1}^{n-1}(x_n, e_k)e_k\|} \quad (n > 1).$$

□

Let H_1 and H_2 both be Hilbert. We call a bijective linear operator $U : H_1 \to H_2$ *unitary* if it preserves inner products, that is, if $(Ux, Uy)_{H_2} = (x, y)_{H_1}$ for all $x, y \in H_1$. Unitary operators are of course bounded with operator norm one. If there exists a unitary operator between two Hilbert spaces H_1 and H_2, then these spaces are called isometrically isomorphic (in particular, they are also isometrically isomorphic as normed spaces).

COROLLARY 2.12. *Every infinite-dimensional separable Hilbert space is isometrically isomorphic with l^2.*

PROOF. Let H be an infinite-dimensional separable Hilbert space, which has a countably infinite Hilbert basis $(e_n)_{n\in\mathbb{N}}$ by Proposition 2.11. Define $U : H \to l^2$ by

$$Ux = ((x, e_n))_{n\in\mathbb{N}}.$$

Then indeed $U(x) \in l^2$ by Parseval's identity, and it is an isometry, because for $x, y \in H$,

$$(Ux, Uy) = \sum_{k=1}^{\infty}(x, e_k)(e_k, y) = (x, y),$$

where we used Theorem 2.10. As an isometry, it is automatically injective. But U is also surjective: For any $x = (x_n)_{n\in\mathbb{N}} \in l^2$, we have

$$x = U\left(\sum_{k=1}^{\infty} x_n e_n\right).$$

Notice that $\sum_{k=1}^{\infty} x_n e_n \in H$ because H is complete and the sequence $\left(\sum_{k=1}^{N} x_n e_n\right)_{N\in\mathbb{N}}$ is Cauchy, because, by Pythagoras,

$$\left\|\sum_{k=N+1}^{M} x_n e_n\right\|^2 = \sum_{k=N+1}^{M} |x_n|^2 \le \sum_{k=N+1}^{\infty} |x_n|^2 \to 0$$

as $N \to \infty$, since $(x_n)_{n\in\mathbb{N}} \in l^2$. □

[4]Here, by 'countable' we mean 'finite or countably infinite'.

2.2. Convergence of Fourier Series

We exemplify the findings of the preceding section as follows. Let $L^2_{per}(\mathbb{R};\mathbb{C})$ be the space of measurable functions $\mathbb{R} \to \mathbb{C}$ that are 2π-periodic (i.e., $f(x+2\pi) = f(x)$ for almost every $x \in \mathbb{R}$) and such that

$$\|f\|_2^2 := \int_0^{2\pi} |f(x)|^2 dx$$

is finite. This norm is induced by the scalar product

$$(f,g) := \int_0^{2\pi} f(x)\overline{g(x)}dx. \tag{2.2}$$

THEOREM 2.13. *For $k \in \mathbb{Z}$ set*

$$e_k := \frac{1}{\sqrt{2\pi}} e^{ik\cdot}.$$

Then $(e_k)_{k\in\mathbb{Z}}$ is a Hilbert basis of $L^2_{per}(\mathbb{R};\mathbb{C})$.

PROOF. Obviously $e_k \in L^2_{per}$ for all $k \in \mathbb{Z}$. Using $\overline{\exp(ikx)} = \exp(-ikx)$ and the functional equation of exp, we find that

$$\begin{aligned}
(e_k, e_l) &= \frac{1}{2\pi}\int_0^{2\pi} e^{ikx}\overline{e^{ilx}}dx = \frac{1}{2\pi}\int_0^{2\pi} e^{ikx}e^{-ilx}dx \\
&= \frac{1}{2\pi}\int_0^{2\pi} e^{i(k-l)x}dx \\
&= \begin{cases} 1 & \text{if } k = l, \\ \frac{1}{2\pi i(k-l)} e^{i(k-l)x}|_0^{2\pi} = 0 & \text{if } k \neq l. \end{cases}
\end{aligned}$$

Therefore, $(e_k)_{k\in\mathbb{Z}}$ is an orthonormal system.

To see completeness, let $\epsilon > 0$ and $f \in C_{per}(\mathbb{R};\mathbb{C})$; by Theorem 1.14 there exists a linear combination $\sum_{k=-N}^{N} c_k e^{ikx}$ such that

$$\sup_{x\in[0,2\pi]} \left| f(x) - \sum_{k=-N}^{N} c_k e^{ikx} \right| < \epsilon.$$

On the bounded interval $[0,2\pi)$, the supremum norm controls the L^2 norm, more precisely:

$$\begin{aligned}
\left\| f - \sum_{k=-N}^{N} c_k e^{ik\cdot} \right\|_2^2 &= \int_0^{2\pi} \left| f(x) - \sum_{k=-N}^{N} c_k e^{ikx} \right|^2 dx \\
&\leq 2\pi \sup_{x\in[0,2\pi]} \left| f(x) - \sum_{k=-N}^{N} c_k e^{ikx} \right|^2 < 2\pi\epsilon^2,
\end{aligned}$$

and we also know that $C_{per}(\mathbb{R};\mathbb{C})$ is dense in $L^2_{per}(\mathbb{R};\mathbb{C})$ (in the norm $\|\cdot\|_2$), so that finite linear combinations of $(e_k)_{k\in\mathbb{Z}}$ are dense in $L^2_{per}(\mathbb{R};\mathbb{C})$. □

Let $f \in L^2_{per}(\mathbb{R};\mathbb{C})$. We write $\hat{f}(k)$ for the inner product (f, e_k) and call it the k-th *Fourier coefficient* of f. Writing it out, we have

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x)e^{-ikx}dx.$$

Theorem 2.10 then tells us:

THEOREM 2.14 (L^2-Convergence of Fourier Series). *Let $f \in L^2_{per}(\mathbb{R}, \mathbb{C})$, then*

$$f = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \hat{f}(k) e^{ik\cdot}, \tag{2.3}$$

and we have Parseval's identity:

$$\int_0^{2\pi} |f(x)|^2 dx = \sum_{k=-\infty}^{\infty} |\hat{f}(k)|^2.$$

The series $(\hat{f}(k))_{k\in\mathbb{Z}}$ is called *Fourier series* of f.

As remarked in Corollary 2.12, the Fourier series thus gives an isometric isomorphism between $L^2_{per}(\mathbb{R}; \mathbb{C})$ and l^2.

2.3. Orthogonal Projection and the Theorem of Riesz-Fréchet

Orthogonal projections can be defined not just onto a subspace, but onto any closed convex (nonempty) set. We will show that such a projection is well-defined, and that it has geometrically intuitive properties. After that, as a by-product, so to speak, we characterise the dual space of a Hilbert space. It turns out that every Hilbert space is essentially *self-dual*, that is, isometrically (anti-)isomorphic to itself. This statement is known as the Riesz-Fréchet Theorem.

2.3.1. Orthogonal Projection onto a Closed Convex Subset. We have already made sense of the orthonormal projection onto a finite-dimensional subspace. However, the concept extends way beyond such subspaces.

Let H be Hilbert and $C \subset H$ a nonempty closed convex subset. If $x \in H$, then we call an element $\bar{x} \in C$ a *proximum* of x in C if

$$\|x - \bar{x}\| \le \|x - c\| \quad \forall c \in C.$$

In other words, the proximum (if it exists) minimises the distance $\|x - c\|$ among all $c \in C$.

THEOREM 2.15 (existence and uniqueness of the proximum). *Let H be Hilbert and $\emptyset \ne C \subset H$ closed and convex. Then, for every $x \in H$ there exists a unique proximum in C.*

PROOF. Let $x \in H$. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ such that

$$\|x - x_n\| \to d := \inf\{\|x - c\| : c \in C\} < \infty$$

as $n \to \infty$.

We will use the following auxiliary identity: For $u, v \in H$,

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2, \tag{2.4}$$

which is immediate to verify. Apply this with $u := x - x_n$ and $v := x_m - x$ to get

$$\begin{aligned} \|x_n - x_m\|^2 &= 2\|x - x_n\|^2 + 2\|x - x_m\|^2 - \|2x - x_n - x_m\|^2 \\ &= 2\|x - x_n\|^2 + 2\|x - x_m\|^2 - 4\left\|x - \frac{1}{2}(x_n - x_m)\right\|^2 \\ &\le 2\|x - x_n\|^2 + 2\|x - x_m\|^2 - 4d^2 \\ &\to 2d^2 + 2d^2 - 4d^2 = 0 \end{aligned}$$

as $m, n \to \infty$, where we used $\frac{1}{2}(x_n + x_m) \in C$ by convexity. Therefore, $(x_n)_{n\in\mathbb{N}}$ is Cauchy. Since H is complete and C is closed, we infer the existence of a limit $\bar{x} \in C$. By continuity of the norm,

$$\|x - \bar{x}\| = \lim_{n\to\infty} \|x - x_n\| = d,$$

so that $\bar{x}$ is a proximum.

It remains to show uniqueness. Suppose $\bar{x}$ and $\hat{x}$ are two proxima. Invoking once more (2.4) with $u = x - \bar{x}$ and $v = x - \hat{x}$,

$$d^2 = \frac{1}{2}\|x-\bar{x}\|^2 + \frac{1}{2}\|x-\hat{x}\|^2 = \left\|x - \frac{1}{2}(\bar{x}+\hat{x})\right\|^2 + \frac{1}{4}\|\bar{x}-\hat{x}\|^2 \geq d^2 + \frac{1}{4}\|\bar{x}-\hat{x}\|^2,$$

whence $\|\bar{x}-\hat{x}\| = 0$ and so $\bar{x} = \hat{x}$. □

DEFINITION 2.16. Let C be nonempty, closed and convex. The map $P : H \to H$ that maps x to its proximum in C is called the *orthogonal projection* onto C.

Note that P is, in general, not linear.

LEMMA 2.17. *Let C be nonempty, convex, and closed. Then $\bar{x} \in C$ is the proximum of $x \in H$ in C if and only if*

$$\Re(x-\bar{x}, y-\bar{x}) \leq 0 \quad \forall y \in C. \tag{2.5}$$

PROOF. Suppose $\bar{x}$ is the proximum of x in C, and let $y \in C$. By convexity,

$$\lambda y + (1-\lambda)\bar{x} = \bar{x} + \lambda(y-\bar{x}) \in C$$

for all $\lambda \in [0,1]$. Since $\bar{x}$ is the proximum, for every $\lambda \in [0,1]$,

$$\begin{aligned}\|x-\bar{x}\|^2 &\leq \|x-(\bar{x}+\lambda(y-\bar{x})))\|^2\\ &= \|x-\bar{x}\|^2 - 2\lambda\Re(x-\bar{x},y-\bar{x}) + \lambda^2\|y-\bar{x}\|^2,\end{aligned}$$

whence

$$\Re(x-\bar{x},y-\bar{x}) \leq \frac{\lambda}{2}\|y-\bar{x}\|^2$$

for $\lambda \in (0,1]$. Since this holds for arbitrarily small positive λ, (2.5) follows.

Conversely, assume (2.5) and let $y \in C$. Then,

$$\|x-y\|^2 = \|(x-\bar{x})+(\bar{x}-y)\|^2 = \|x-\bar{x}\|^2 - 2\Re(x-\bar{x},y-\bar{x}) + \|y-\bar{x}\|^2 \geq \|x-\bar{x}\|^2,$$

so that $\bar{x}$ is indeed the proximum. □

LEMMA 2.18. *Let H Hilbert and $C \subset H$ nonempty, closed and convex. Then the orthogonal projection $P : H \to H$ is Lipschitz continuous with Lipschitz constant one, that is,*

$$\|Px-Py\| \leq \|x-y\| \quad \forall x,y \in H.$$

PROOF. Let $x, y \in H$. By the previous lemma, we have

$$\Re(x-Px, z-Px) \leq 0, \quad \Re(y-Py, z-Py) \leq 0 \quad \forall z \in C.$$

Pick $z = Py$ in the first and $z = Px$ in the second of these inequalities, we get

$$\Re(x-Px, Py-Px) \leq 0, \quad -\Re(y-Py, Py-Px) \leq 0.$$

Adding both yields

$$0 \geq \Re(x-Px+Py-y, Py-Px) = \|Py-Px\|^2 + \Re(x-y, Py-Px),$$

so that rearranging and the Cauchy-Schwarz inequality result in

$$\|Px-Py\|^2 \leq \|x-y\|\|Px-Py\|.$$

□

The following result is about the special case of a closed subspace.

LEMMA 2.19. *Let H be Hilbert and $H_0 \subset H$ a closed subspace. Then the orthogonal projection $P : H \to H$ onto H_0 is a bounded linear operator whose norm, unless $H_0 = \{0\}$, is one. Also, $z = Px$ if and only if $x - z \in H_0^\perp$.*

Proof. Let $x \in H$. From Lemma 2.17 we know that the orthogonal projection is characterised by $\Re(x - Px, y - Px) \le 0$ for all $y \in H_0$. But since H_0 is a subspace, $y \in H_0$ if and only if $y + Px \in H_0$, and the same is true for $-y + Px$ as well as $\pm iy + Px$. It follows that the orthogonal projection is characterised by $(x - Px, w) = 0$ for all $w \in H_0$, or equivalently, $x - Px \in H_0^\perp$.

Let $\alpha, \beta \in \mathbb{K}$ and $x, y \in H$. Then, as H_0 is a subspace, $\alpha Px + \beta Py \in H_0$, and we have for each $w \in H_0$:

$$(\alpha x + \beta y - \alpha Px - \beta Py, w) = \alpha(x - Px, w) + \beta(y - Py, w) = 0,$$

whence $P(\alpha x + \beta y) = \alpha Px + \beta Py$, so P is linear. By Lemma 2.18 (setting $y = 0$), $\|Px\| \le x$ for all $x \in H$, so that $\|P\| \le 1$; but also $\|P\| \ge 1$ because restricted to H_0, P is the identity. In total we see $\|P\| = 1$. □

2.3.2. Characterisation of the Dual of a Hilbert Space.

Theorem 2.20 (Riesz-Fréchet). *Let H be Hilbert and $x' \in H'$. Then there exists exactly one $y \in H$ such that*

$$x' = (\cdot, y). \tag{2.6}$$

Proof. As x' is linear and bounded, $H_0 := \ker x'$ is a closed subspace of H. If $H_0^\perp = \{0\}$, then $H_0 = H$: Indeed, If there were $x \in H \setminus H_0$, then $0 \ne x - Px$ and $x - Px \in H_0^\perp$ by Lemma 2.19, contradicting $H_0^\perp = \{0\}$. So $x' = 0$, and then the choice $y = 0$ uniquely satisfies (2.6).

Otherwise, we may choose $y_1 \in H_0^\perp$ such that $x'(y_1) = 1$. Then, for any $x \in H$, $x - x'(x)y_1 \in H_0$, hence

$$0 = (x - x'(x)y_1, y_1) = (x, y_1) - x'(x)\|y_1\|^2.$$

Thus, $y := \frac{y_1}{\|y_1\|^2}$ satisfies (2.6).

For uniqueness, Suppose z is another vector fulfilling (2.6). Then,

$$\|y - z\|^2 = (y - z, y) - (y - z, z) = x'(y - z) - x'(y - z) = 0,$$

so $z = y$. □

Owing to this theorem, there exists a bijective map $T : H \to H'$, $y \mapsto T[y] \in H'$, such that $(x, y) = T[y](x)$ for all $x \in H$. This map is antilinear, that is, $T[\alpha y_1 + \beta y_2] = \bar{\alpha} T[y_1] + \bar{\beta} T[y_2]$. It is norm-preserving, because

$$\|T[y]\| = \sup_{\|x\| \le 1} |T[y](x)| = \sup_{\|x\| \le 1} |(x, y)| = \|y\|$$

(simply set $x = \frac{y}{\|y\|}$).

Hence, the Riesz-Fréchet Theorem tells us that a *real* Hilbert space is isometrically isomorphic to its dual[5], and a complex Hilbert space is 'isometrically anti-isomorphic' to its dual.

2.4. The Spectral Theorem for Compact Self-Adjoint Operators

In Linear Algebra you learn that a real symmetric matrix A is diagonalisable, meaning that there exists an orthogonal matrix O such that OAO^t is diagonal. Equivalently, there exists an orthonormal basis of eigenvectors of A. In a Hilbert space H, it turns out that for *compact* operators in $L(H, H)$, the situation is very analogous: If the operator is *self-adjoint* (which is the analogue of symmetry) and compact, then there exists a Hilbert basis of eigenvectors of the operator. The assumption of compactness is crucial as it puts us

[5]In fact we have only seen that T preserves norms, but we have not considered inner products. However preservation of the latter follows from preservation of norms by the polarisation identity $(x, y) = \frac{1}{4}\|x + y\|^2 - \frac{1}{4}\|x - y\|^2$.

into a similar situation as in the finite-dimensional case. The theory of diagonalisation, in a broad sense, is called *spectral theory*.

For most of this section we will be occupied with a clarification of the notions of self-adjointness, compact operators, and spectrum. The actual proof of the spectral theorem will then not be particularly difficult.

2.4.1. The Adjoint Operator.

Definition 2.21 (adjoint). Let H be Hilbert and $T \in L(H,H)$. Then an operator $T^* \in L(H,H)$ is called its *adjoint* if

$$(Tx,y) = (x,T^*y) \quad \forall x,y \in H.$$

For instance, if $H = \mathbb{R}^d$ and an operator is represented by a $(d \times d)$-matrix A, then A^t represents the unique adjoint, because for $x,y \in \mathbb{R}^d$,

$$Ax \cdot y = \sum_{k,l=}^{d} A_{kl} x_l y_k = \sum_{k,l=}^{d} A^t_{lk} x_l y_k = x \cdot A^t y.$$

Likewise, the unique adjoint of a complex matrix $A \in \mathbb{C}^{d\times d}$ is given as $A^* := \overline{A^t}$.

Theorem 2.22 (existence & uniqueness of the adjoint). *Let H be Hilbert and $T \in L(H,H)$. Then there exists exactly one adjoint operator $T^* \in L(H,H)$ for T, and $\|T^*\| = \|T\|$.*

Proof. This is a straightforward consequence of the Riesz-Fréchet Theorem: For $y \in H$, define the linear functional $x \mapsto (Tx,y)$, which is bounded by $\|T\|\|y\|$. By Riesz-Fréchet, there exists a unique $T^*y \in H$ such that $(Tx,y) = (x,T^*y)$. It remains to show that $T^* : y \mapsto T^*y$ is linear and bounded with norm $\|T\|$.

For linearity, note that for $\lambda_1,\lambda_2 \in \mathbb{K}$ and $y_1,y_2 \in \mathbb{K}$,

$$\begin{aligned}(x,\lambda_1 T^*y_1 + \lambda_2 T^*y_2) &= \overline{\lambda_1}(x,T^*y_1) + \overline{\lambda_2}(x,T^*y_2)\\ &= \overline{\lambda_1}(Tx,y_1) + \overline{\lambda_2}(Tx,y_2) = (Tx,\lambda_1 y_1 + \lambda_2 y_2),\end{aligned}$$

whence $T^*(\lambda_1 y_1 + \lambda_2 y_2) = \lambda T^*y_1 + \lambda_2 T^* y_2$ follows by virtue of uniqueness in the Riesz-Fréchet Theorem.

For boundedness, observe

$$\|T\| = \sup_{\|x\|\le 1} \|Tx\| = \sup_{\|x\|\le 1}\sup_{\|y\|\le 1} |(Tx,y)| = \sup_{\|x\|\le 1}\sup_{\|y\|\le 1} |(x,T^*y)| = \|T^*\|,$$

where we used the theorems of Hahn-Banach and Riesz-Fréchet to write $\|x\| = \sup_{\|y\|\le 1} |(x,y)|$. □

Definition 2.23 (self-adjoint operators). An operator $T \in L(H,H)$ is called *self-adjoint* if $T = T^*$.

For instance, in $\mathbb{R}^d$ a matrix is self-adjoint if it is symmetric, and in $\mathbb{C}^d$ it is self-adjoint if it is Hermitian, that is, $A^t = \overline{A}$.

As another example, the orthogonal projection onto a closed subspace of a Hilbert space is self-adjoint, as you already saw in the exercises.

Proposition 2.24. *Let H be Hilbert and $T \in L(H,H)$.*

(1) *Let $K = \mathbb{C}$. Then T is self-adjoint if and only if $(Tx,x) \in \mathbb{R}$ for all $x \in H$.*
(2) *If T is self-adjoint, then $\|T\| = \sup_{\|x\|\le 1} |(Tx,x)|$.*

Proof. (1) Let T be self-adjoint. Then, for any $x \in H$,

$$(Tx,x) = (x,Tx) = \overline{(Tx,x)},$$

so $(Tx,x) \in \mathbb{R}$.

Conversely, suppose $(Tx,x) \in \mathbb{R}$ for all $x \in H$. Let $x, y \in H$ and $\lambda \in \mathbb{C}$, then

$$(T(x+\lambda y), x+\lambda y) = (Tx,x) + \overline{\lambda}(Tx,y) + \lambda(Ty,x) + |\lambda|^2(Ty,y).$$

By assumption, complex conjugation of this equality only affects the middle terms on the right hand side, so that

$$(T(x+\lambda y), x+\lambda y) = (Tx,x) + \lambda(y,Tx) + \overline{\lambda}(x,Ty) + |\lambda|^2(Ty,y).$$

Subtract both equalities to obtain

$$\overline{\lambda}(Tx,y) + \lambda(Ty,x) = \lambda(y,Tx) + \overline{\lambda}(x,Ty).$$

Set first $\lambda = 1$ and then $\lambda = i$ in this equality to obtain

$$\begin{aligned} (Tx,y) + (Ty,x) &= (y,Tx) + (x,Ty), \\ -i(Tx,y) + i(Ty,x) &= i(y,Tx) - i(x,Ty). \end{aligned}$$

Dividing the second line by i and then adding both equalities, we arrive at $(Ty,x) = (y,Tx)$, so that T is self-adjoint.

(2) Denote $C := \sup_{\|x\|\le 1} |(Tx,x)|$. Clearly, $C \le \|T\|$. For $x, y \in H$, we have

$$(T(x+y),x+y)-(T(x-y),x-y) = 2(Tx,y)+2(Ty,x) = 2(Tx,y)+2\overline{(Tx,y)} = 4\Re(Tx,y),$$

where we used the self-adjointness of T. From the definition of C and equality (2.4) it follows that

$$4\Re(Tx,y) \le C(\|x+y\|^2 + \|x-y\|^2) = 2C(\|x\|^2 + \|y\|^2).$$

In particular, if $\|x\| = \|y\| = 1$, then $\Re(Tx,y) \le C$. Replacing x by $e^{is}x$, where $s \in \mathbb{R}$ is chosen such that

$$e^{is}(Tx,y) = |(Tx,y)|,$$

we conclude $|(Tx,y)| \le C$, which proves the statement as $\|T\| = \sup_{\|x\|=\|y\|=1} |(Tx,y)|$. □

2.4.2. Compact Operators. The content of this and the next section is not specific to Hilbert spaces, so we work more generally in normed spaces again.

An operator $T : X \to Y$ is called *compact* if it maps bounded sets to precompact ones, that is:

DEFINITION 2.25 (compact operators). Let X, Y be normed spaces and $T : X \to Y$ a linear map. It is *compact* if, for every bounded sequence $(x_n)_{n\in\mathbb{N}}$, the sequence $(Tx_n)_{n\in\mathbb{N}}$ has a convergent subsequence.

Compact operators are always bounded: Otherwise, there would be a bounded sequence $(x_n)_{n\in\mathbb{N}}$ such that $\|Tx_n\| \ge n$ for all $n \in \mathbb{N}$, and this cannot have a convergent subsequence as convergent sequences are bounded.

EXAMPLE 2.26. A bounded operator $X \to Y$ whose range has finite dimension is compact, by the Theorem of Bolzano-Weierstraß. Such operators are called *finite rank operators*. For instance, orthogonal projections from a Hilbert space onto a finite-dimensional subspace are finite-rank operators. One can show that limits (in the operator norm) of sequences of finite rank operators are still compact. The converse question – whether every compact operator is a limit of finite-rank operators – was open for decades and only solved in 1972 by Per Enflo (the answer is 'no').

2.4.3. The Spectrum of an Operator. The following concepts (resolvent and spectrum) are known from Linear Algebra, at least in the finite-dimensional case. The idea is that the *spectrum* of a square matrix A is the set of its eigenvalues, which coincides with the set of $\lambda \in \mathbb{K}$ for which the matrix[6] $\lambda I - A$ is singular (i.e., λ is a root of the characteristic polynomial of A). Trivially reformulated, we can also say that the spectrum is the complement in $\mathbb{K}$ of the set of λ for which $\lambda I - A$ is invertible.

Analogously, we define:

DEFINITION 2.27 (resolvent and spectrum). Let X be a normed space and $T \in L(X, X)$. The *resolvent set* $\rho(T) \subset \mathbb{K}$ is the set of all $\lambda \in \mathbb{K}$ for which $\lambda I - T$ is bijective and $(\lambda I - T)^{-1} \in L(X, X)$. For $\lambda \in \rho(T)$, we call

$$R(\lambda, T) := (\lambda I - T)^{-1}$$

the *resolvent* of T in λ.

The complement $\sigma(T) := \mathbb{K} \setminus \rho(T)$ is called the *spectrum* of T.

One is tempted to identify $\sigma(T)$ with the set of eigenvalues of T, i.e., the set of $\lambda \in \mathbb{K}$ such that there exists $X \ni x \neq 0$ satisfying $Tx = \lambda x$. In finite dimensions, this is certainly true. However, we have the following counterexample: On l^2, consider the right shift operator R as in Example 1.19:

$$R(x_1, x_2, x_3, \ldots) = (0, x_1, x_2, \ldots).$$

Obviously, R is not surjective, hence $0 \notin \rho(R)$ and so $0 \in \sigma(R)$. However, 0 is not an eigenvalue, as $Rx = 0$ implies $x = 0$. The converse is true: Every eigenvalue belongs to the spectrum. The set of eigenvalues of an operator is called its *point spectrum*.

The following looks like the summation formula for a geometric series:

LEMMA 2.28 (Neumann Series). *Let X be Banach and $T \in L(X, X)$ with $\|T\| < 1$. Then, $I - T$ is bijective, $(I - T)^{-1} \in L(X, X)$, and*

$$(I - T)^{-1} = \sum_{n=0}^{\infty} T^n.$$

PROOF. Since X is complete, then so is $L(X, X)$ by Proposition 1.17. From the definition of the operator norm, one sees that $\|PS\| \leq \|P\|\|S\|$ for linear operators $P, S \in L(X, X)$. Therefore, $\|T^n\| \leq \|T\|^n$ for any $n \in \mathbb{N}$, and so

$$\sum_{n=0}^{\infty} \|T^n\| \leq \sum_{n=0}^{\infty} \|T\|^n = \frac{1}{1 - \|T\|} < \infty,$$

and $\sum_{n=0}^{\infty} T^n$ is absolutely convergent in $L(X, X)$. But in Banach spaces, absolute convergence implies convergence, hence $\sum_{n=0}^{\infty} T^n =: S \in L(X, X)$, and we have

$$(I - T)S = \lim_{N \to \infty} (I - T) \sum_{n=0}^{N} T^n = \lim_{N \to \infty} (I - T^{N+1}) = I,$$

and similarly $S(I - T) = I$. Therefore, $S = (1 - T)^{-1}$, as claimed. □

THEOREM 2.29 (compactness of the spectrum). *Let X be Banach and $T \in L(X, X)$. Then $\sigma(T) \subset \mathbb{K}$ is compact with $|\lambda| \leq \|T\|$ for all $\lambda \in \sigma(T)$.*

PROOF. If $|\lambda| > \|T\|$, then $\|\lambda^{-1}T\| < 1$, and by Lemma 2.28, $I - \lambda^{-1}T$ is invertible with bounded inverse, and so is $\lambda I - T = \lambda(I - \lambda^{-1}T)$. This shows $|\lambda| \leq \|T\|$ for all $\lambda \in \sigma(T)$.

It remains to show that $\sigma(T)$ is closed, or equivalently, $\rho(T)$ is open. To this end, let $\lambda_0 \in \rho(T)$ and $\epsilon := \|R(\lambda_0, T)\|^{-1} > 0$. If $\lambda \in B_\epsilon(\lambda_0)$, then

$$\lambda I - T = \lambda_0 I - T - (\lambda_0 - \lambda)I = (\lambda_0 I - T)\left[I - (\lambda_0 - \lambda)(\lambda_0 I - T)^{-1}\right].$$

[6]Here and in the following, we denote by I the identity matrix, or the identity operator.

Set $R := (\lambda_0 - \lambda)(\lambda_0 I - T)^{-1}$. The computation shows that $\lambda I - T$ will be invertible, with continuous inverse, if $I - R$ is (as $\lambda_0 I - T$ is invertible with continuous inverse). But, by choice of ϵ,

$$\|R\| \le |\lambda_0 - \lambda| \|R(\lambda_0, T)\| < \epsilon \|R(\lambda_0, T)\| = 1,$$

so the claim that $\lambda \in \rho(T)$ follows from Lemma 2.28. Therefore, the resolvent set is open and the spectrum is compact. □

LEMMA 2.30 (spectrum of the adjoint). *Let H be Hilbert and $T \in L(H,H)$. Then, $\sigma(T^*) = \overline{\sigma(T)}$, where the overline denotes complex conjugation. In particular, the spectrum of a self-adjoint operator is real.*

PROOF. Note first that, for any two operators $S, T \in L(H,H)$, $(ST)^* = T^* S^*$, because for $x, y \in H$ we have $(STx, y) = (Tx, S^* y) = (x, T^* S^* y)$.

Now let $\lambda \in \rho(T)$. By definition of the resolvent,

$$I = (\lambda I - T) R(\lambda, T) = R(\lambda, T)(\lambda I - T),$$

and therefore

$$I = I^* = [(\lambda I - T) R(\lambda, T)]^* = R(\lambda, T)^* (\overline{\lambda} - T^*)$$

and similarly

$$I = I^* = [R(\lambda, T)(\lambda I - T)]^* = (\overline{\lambda} - T^*) R(\lambda, T)^*.$$

This shows $R(\lambda, T)^* = R(\overline{\lambda}, T^*)$, so that the resolvent set of T^* is the conjugate of the resolvent set of T; here one uses that an operator is invertible with bounded inverse if and only if this is the case for its adjoint (note $(S^*)^{-1} = (S^{-1})^*$). Therefore, the same is true for the spectrum: $\sigma(T^*) = \overline{\sigma(T)}$. □

2.4.4. The Spectral Theorem.

THEOREM 2.31 (Spectral Theorem for compact self-adjoint operators). *Let H be an infinite-dimensional Hilbert space and $T \in L(H,H)$ compact and self-adjoint. Then there is an orthonormal system $(e_n)_{n\in\mathbb{N}}$ and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of real numbers with $|\lambda_n| \searrow 0$ such that, for each $x \in H$,*

$$Tx = \sum_{n=1}^{\infty} \lambda_n (x, e_n) e_n.$$

In particular, $Te_n = \lambda_n e_n$ for every $n \in \mathbb{N}$, so that the λ_n are eigenvalues of T.

Represented in the orthonormal system $(e_n)_{n\in\mathbb{N}}$ (extended by a Hilbert basis of $\ker T$ if necessary), T is thus in a sense an $\mathbb{N} \times \mathbb{N}$ diagonal matrix with diagonal entries λ_n. This is analogous to the diagonalisation of a symmetric (or Hermitian) finite-dimensional square matrix in Linear Algebra.

PROOF. We may assume $T \neq 0$, as otherwise the statement is trivial.

Step 1. We show first $\sigma(T) \subset [-\|T\|, \|T\|]$, and $\|T\|$ or $-\|T\|$ is an eigenvalue of T. The first assertion follows immediately from Theorem 2.29 and Lemma 2.30.

From Proposition 2.24, we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ with $\|x_n\| \le 1$ such that $|(Tx_n, x_n)| \to \|T\| =: \lambda > 0$ as $n \to \infty$. We assume $\lim_{n\to\infty} (Tx_n, x_n) > 0$ (otherwise set $\lambda = -\|T\|$ in the following).

By compactness of T, we may assume $(Tx_n)_{n\in\mathbb{N}}$ to be convergent. Next, observe that

$$\limsup_{n\to\infty} \|Tx_n - \lambda x_n\|^2 \le \limsup_{n\to\infty} (\|Tx_n\|^2 + \lambda^2 \|x_n\|^2 - 2\lambda (Tx_n, x_n)) \le \lambda^2 + \lambda^2 - 2\lambda^2 = 0,$$

where we used the self-adjointness of T. This shows that, as Tx_n converges, so does λx_n and (since $\lambda \neq 0$) therefore x_n. Set $x := \lim_{n\to\infty} x_n \neq 0$. As T is continuous,

$$\lambda x = \lim_{n\to\infty}(\lambda x_n) = \lim_{n\to\infty} Tx_n = Tx,$$

so that $\lambda = \|T\|$ is an eigenvalue of T.

Step 2. By Step 1, $\lambda_1 = \pm\|T\|$ is an eigenvalue of T. Let e_1 be a corresponding unit eigenvector. Set $R_1 := \operatorname{span}\{e_1\}^\perp$, which is a closed subspace of H. Note $TR_1 \subset R_1$, because if $x \perp e_1$ then

$$(Tx, e_1) = (x, Te_1) = \overline{\lambda_1}(x, e_1) = 0.$$

Also, the restriction $T_1 = T\restriction_{R_1}$ is of course still compact and self-adjoint.

We may therefore iterate Step 1 to obtain an orthonormal system $(e_n)_{n\in\mathbb{N}}$ of eigenvectors of T with eigenvalue $\lambda_n = \pm\|T_n\|$, where T_n is the restriction of T to $\operatorname{span}\{e_1,\dots,e_n\}^\perp$. Clearly, $|\lambda_n| = \|T_n\|$ is non-increasing in n.

Step 3. It remains to show that $(e_n)_{n\in\mathbb{N}}$ is a Hilbert basis for the range of T. To this end, set

$$H_\infty := \overline{\operatorname{span}\{e_n : n \in \mathbb{N}\}},$$

which is a closed subspace of H. Write $R_\infty := H_\infty^\perp$. Again, $TR_\infty \subset R_\infty$. Denote $T_\infty := T\restriction_{R_\infty}$ and P_∞ the orthogonal projection onto R_∞.

The orthogonal projection of $x \in H$ onto H_∞ is given by $\sum_{n\in\mathbb{N}}(x, e_n)e_n$, so that x can be written as

$$x = \sum_{n\in\mathbb{N}}(x, e_n)e_n + P_\infty x.$$

Applying T, therefore, yields,

$$Tx = \sum_{n\in\mathbb{N}}\lambda_n(x, e_n)e_n + T_\infty P_\infty x,$$

so it suffices to prove $T_\infty = 0$. By construction, $\|T_\infty\| \le \|T_n\| = |\lambda_n|$ for all $n \in \mathbb{N}$. We shall show $|\lambda_n| \to 0$.

Indeed, suppose there were a subsequence $(\lambda_{n_k})_{k\in\mathbb{N}}$ converging to $\mu \neq 0$. By compactness of T, $Te_{n_k} = \lambda_{n_k}e_{n_k}$ would also have a convergent subsequence, and therefore (due to the assumption $\lambda_{n_k} \to \mu \neq 0$) also $(e_{n_k})_{k\in\mathbb{N}}$ would have a convergent subsequence, which is doesn't (because $\|e_k - e_l\| = \sqrt{2}$ for $k \neq l$). This shows $|\lambda_n| \to 0$ and thus completes the proof. □

EXAMPLE 2.32. (1) Let $(e_n)_{n=1,\dots,N}$ be a finite orthonormal system in an infinite-dimensional Hilbert space H, and denote $H_0 = \operatorname{span}\{e_n : n = 1,\dots,N\}$. Then the orthogonal projection onto H_0 is represented as

$$Px = \sum_{n=1}^{N}(x, e_n)e_n,$$

so P has eigenvalues 1 and 0.

(2) This example will be a bit sketchy, but very instructive nevertheless.

Consider functions in $L^2_{per}(\mathbb{R};\mathbb{C})$ with $\hat{f}(0) = \frac{1}{\sqrt{2\pi}}\int_0^{2\pi} f(x)dx = 0$. As this integral condition is inherited by L^2 limits, such functions form a closed subspace of $L^2_{per}(\mathbb{R};\mathbb{C})$ and thus a Hilbert space H in its own right. Define the antiderivative operator $I : H \to H$ by

$$If(x) = \int_0^x f(t)dt + \frac{1}{2\pi}\int_0^{2\pi} tf(t)dt.$$

Indeed, $(If)' = f$ for almost every $x \in (0, 2\pi)$, and I does map into H as If is continuous for $f \in H$ (so *a fortiori* L^2 in $[0, 2\pi)$), periodic (because $If(2\pi) =$

$If(0) = \frac{1}{2\pi}\int_0^{2\pi} tf(t)dt)$), and $\int_0^{2\pi} If(x)dx = 0$, as can be seen by an elementary integration exercise. Obviously, I is linear.

We want to show compactness of I. Recall the Arzelà-Ascoli Theorem: A sequence of continuous functions $(F_n)_{n\in\mathbb{N}} : [0, 2\pi] \to \mathbb{C}$ has a uniformly convergent subsequence if it is uniformly bounded and is equicontinuous (meaning that for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $n \in \mathbb{N}$, $|x-y| < \delta$ implies $|F_n(x)-F_n(y)| < \epsilon$).

Let $(f_n)_{n\in\mathbb{N}}$ be bounded in H (hence in L^2). Then $(If_n)_{n\in\mathbb{N}}$ are bounded pointwise by

$$\int_0^{2\pi} |f_n(t)|dt + \frac{1}{2\pi}\int_0^{2\pi} t|f_n(t)|dt \le \left(\sqrt{2\pi} + \frac{\sqrt{2\pi}}{\sqrt{3}}\right)\|f_n\|_{L^2},$$

where we used Hölder's inequality with $p = q = 2$. But this is bounded independently of n.

For equicontinuity, we find

$$|If_n(x) - If_n(y)| = \left|\int_y^x f_n(t)dt\right| \le \int_y^x |f_n(t)|dt \le \sqrt{|x-y|}\|f_n\|_{L^2},$$

again using Hölder's inequality. Since the right hand side becomes arbitrary small uniformly in n, we infer equicontinuity. The Arzelà-Ascoli Theorem thus gives a uniformly convergent subsequence of $(If_n)_n$, which *a fortiori* converges in L^2. This shows that $I : H \to H$ is compact.

Since I is compact, then so is I^2 (think it over: the composition of compact operators is again compact). And I^2 is self-adjoint: If $F = I^2 f$, then $F'' = f$ almost everywhere and $F \in H$. In fact F is uniquely determined by these properties (why?). Integration by parts therefore gives, for $f, g \in H$,

$$\begin{aligned}(I^2 f, g) &= (F, (I^2 g)'') = -(F', (I^2 g)') + [F, Ig]_0^{2\pi} \\ &= (F'', I^2 g) - \left[F', I^2 g\right]_0^{2\pi} + [F, Ig]_0^{2\pi} = (f, I^2 g),\end{aligned}$$

where the boundary terms vanished thanks to periodicity. Hence, the Spectral Theorem gives us an orthonormal system and a family of eigenvalues such that

$$I^2 f = \sum_{n\in\mathbb{N}} \lambda_n (f, e_n) e_n.$$

In fact, the Hilbert basis $(e_n)_{n\in\mathbb{Z}\setminus\{0\}}$ for H from Section 2.3 does the job: By explicit computation of the integrals, it turns out that for $e_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$,

$$I^2 e_n = -\frac{1}{n^2} e_n,$$

so $\lambda_n = -\frac{1}{n^2}$ for $n \in \mathbb{Z} \setminus \{0\}$.

The Fourier coefficients of $I^2 f$ are therefore given as $-\frac{\hat{f}(n)}{n^2}$. Conversely, if $F \in H$ is twice differentiable, then the second derivative operator is represented in terms of the Fourier series being multiplied by $-n^2$. One therefore says that *the Fourier transform $H \to l^2(\mathbb{C})$ diagonalises differential operators.*

2.5. Reproducing Kernel Hilbert Spaces

The theory of Reproducing Kernel Hilbert Spaces has become an important concept in machine learning, where the 'kernel trick' is used for nonlinear classification problems in Support Vector Machines, for example. We discuss some basic theory and take a glimpse into applications. For further study of the topic, we refer to [4].

2.5.1. Basic Concepts and Examples. Let X be a set, then the set $\mathbb{K}^X$ of maps $X \to \mathbb{K}$ forms a vector space under pointwise addition and scalar multiplication.

DEFINITION 2.33 (RKHS). Let X be a set. A subset $H \subset \mathbb{K}^X$ is called a *Reproducing Kernel Hilbert Space (RKHS)* if H is a Hilbert space with some inner product $(\cdot,\cdot)$, and if for each $x \in X$ the evaluation functional

$$E_x : H \to \mathbb{K}, \quad E_x(f) = f(x)$$

is bounded.

Note that the main condition (boundedness of E_x) is not satisfied in quite typical cases. For instance, L^2 is *not* RKHS: L^2 does not consist of functions, but of equivalence classes of functions, so that the evaluation map is not even well-defined. We will see examples of RKHS soon.

Let H be an RKHS. By the Riesz-Fréchet Theorem, for any $x \in X$ there exists a unique $k_x \in H$ such that

$$E_x(f) = (f, k_x) \quad \forall f \in H.$$

DEFINITION 2.34 (reproducing kernel). In the situation as described, the *reproducing kernel* of H is the function $K : X \times X \to \mathbb{K}$ defined by

$$K(x,y) = k_y(x) = (k_y, k_x).$$

EXAMPLE 2.35.

- $\mathbb{K}^d$, viewed as the set of maps $\{1,2,\dots,d\} \to \mathbb{K}$, is an RKHS, because in $\mathbb{K}^d$ in fact *every* linear functional is bounded. The evaluation functional $E_k : \mathbb{K}^d \to \mathbb{K}$ is simply given as $E_k(x) = x_k$ ($k = 1,\dots,d$), and the corresponding representation from Riesz-Fréchet is $k_k = e_k$, the k-th standard unit vector. The reproducing kernel, therefore, is given by

  $$K(k,l) = (e_l, e_k) = \delta_{lk}.$$

- Let H be a real Hilbert space and H' its dual, which is isomorphic to H by the Theorem of Riesz-Fréchet via an isometric isomorphism $T : H \to H'$. In particular, H' is a Hilbert space in its own right, with inner product $(x', y') = (T^{-1}x', T^{-1}y')$. We claim that $H' \subset \mathbb{K}^H$ is RKHS. Let $x \in H$, then

  $$E_x(y') = y'(x) = (x, T^{-1}y') = (Tx, y'),$$

  which of course is bounded with norm $\|x\|$ because $\|T^{-1}\| = 1$. Moreover, $k_x = Tx$ and therefore

  $$K(x,y) = (k_y, k_x) = (Ty, Tx) = (x,y),$$

  so that the reproducing kernel is simply the inner product in H.
- Let $X = (-1,1) \subset \mathbb{R}$ and define

  $$H = \left\{ f : X \to \mathbb{R} : \quad \exists a \in l^2(\mathbb{R}) \quad \forall x \in X : \quad f(x) = \sum_{k=0}^{\infty} a_k x^k \right\}.$$

  Estimating by Cauchy-Schwarz and geometric series, we find

  $$\left| \sum_{k=0}^{\infty} a_k x^k \right| \le \sum_{k=0}^{\infty} |a_k||x^k| \le \|a\|_{l^2} \left(\sum_{k=0}^{\infty} x^{2k} \right)^{1/2} \le \|a\|_{l^2} \frac{1}{\sqrt{1-x^2}}, \tag{2.7}$$

  So that the series converges absolutely on X and thus $H \subset \mathbb{R}^X$ is well-defined. Of course, H is a vector space as l^2 is. For ease of notation, let us denote $f = \sum_{k=0}^{\infty} a_k x^k$ instead of $f = \sum_{k=0}^{\infty} a_k (\cdot)^k$. Define a linear map

  $$\phi : H \to l^2(\mathbb{R}), \quad \phi\left(\sum_{k=0}^{\infty} a_k x^k \right) = a$$

and set

$$(f,g) := (\phi(f),\phi(g))$$

for all $f,g \in H$. It is easy to check that this defines an inner product on H. Also, completeness of H under this inner product follows from completeness of l^2, so H is a Hilbert space.

For the evaluation functional E_x $(x \in X)$, notice

$$|E_x(f)| = |f(x)| \le \|a\|_{l^2} \frac{1}{\sqrt{1-x^2}} = \|f\| \frac{1}{\sqrt{1-x^2}},$$

so that for each $x \in X$, E_x is bounded (although the bound is not uniform in X). Thus, H is an RKHS.

Let us compute the reproducing kernel. Let $y \in X$, then $k_y = \sum_{k=0}^{\infty} y^k x^k$, because

$$\left(\sum_{k=0}^{\infty} a_k x^k, k_y\right) = \sum_{k=0}^{\infty} a_k y^k = E_y\left(\sum_{k=0}^{\infty} a_k x^k\right).$$

Therefore,

$$K(x,y) = (k_x,k_y) = \sum_{k=0}^{\infty} x^k y^k = \frac{1}{1-xy}.$$

2.5.2. The Moore-Aronszajn Theorem. The reproducing kernel of an RKHS turns out to be positive semidefinite:

Proposition 2.36. *Let H be RKHS and $K : X \times X \to \mathbb{K}$ its reproducing kernel. Then K is positive semidefinite, i.e., for any finite set $\{x_1,\dots,x_N\} \subset X$ of pairwise distinct elements, the matrix $K(x_k,x_l)_{k,l=1,\dots,N}$ is positive semidefinite.*

Proof. Let $\alpha_1,\dots,a_N \in \mathbb{K}$ and $x_1,\dots,x_N \in X$, then

$$\sum_{k,l=1}^{N} \alpha_l \overline{\alpha_k} K(x_k,x_l) = \sum_{k,l=1}^{N} (\alpha_l k_{x_l}, \alpha_k k_{x_k}) = \|\sum_{k=1}^{N} \alpha_k k_{x_k}\|^2 \ge 0.$$

□

Obviously, K is also antisymmetric, as $K(x,y) = (k_y,k_x) = \overline{(k_x,k_y)} = \overline{K(y,x)}$.

The converse is of course more interesting:

Theorem 2.37 (Moore-Aronszajn). *Let X be a set and $K : X \times X \to \mathbb{K}$ antisymmetric and positive semidefinite. Then there exists an RKHS H whose reproducing kernel is K.*

Proof. Define

$$H_0 := \operatorname{span}\{K(\cdot,y) : y \in X\}.$$

Then, H_0 is certainly a vector space contained in $\mathbb{K}^X$. On H_0, we define the inner product

$$\left(\sum_{k=1}^{n} \alpha_k K(\cdot,y_k), \sum_{l=1}^{m} \beta_l K(\cdot,z_l)\right) := \sum_{k=1}^{n}\sum_{l=1}^{m} \alpha_k \overline{\beta_l} K(z_l,y_k).$$

Is this map even well-defined? After all, there might be several ways to represent the same element of H_0 as a linear combination of maps $K(\cdot,y)$. So assume $\sum_{k=1}^{n} \alpha_k K(\cdot,y_k) =$

$\sum_{j=1}^{r} \gamma_j K(\cdot, x_j)$, then

$$\begin{aligned}\left(\sum_{k=1}^{n} \alpha_k K(\cdot, y_k), \sum_{l=1}^{m} \beta_k K(\cdot, z_l)\right) &= \sum_{k=1}^{n}\sum_{l=1}^{m} \alpha_k \overline{\beta_l} K(z_l, y_k) \\ &= \sum_{j=1}^{r}\sum_{l=1}^{m} \gamma_j \overline{\beta_l} K(z_l, x_j) \\ &= \left(\sum_{j=1}^{r} \gamma_j K(\cdot, x_j), \sum_{l=1}^{m} \beta_k K(\cdot, z_l)\right),\end{aligned}$$

and similarly one checks the independence of the map of the representation of the second argument (for this, we would use antisymmetry of K).

Let us check this is really an inner product. Clearly, it is sesquilinear and antisymmetric, as K is antisymmetric. It is also positive semidefinite, because K is positive semidefinite.

For positivity, let $f \in H_0$ be such that $(f, f) = 0$, and let $g \in H_0$. We first want to show $(f, g) = 0$. For $\lambda \in \mathbb{K}$, note

$$0 \le (f + \lambda g, f + \lambda g) = (f, f) + |\lambda|^2 (g, g) + 2\Re(\lambda(f, g)) = |\lambda|^2 (g, g) + 2\Re(\lambda(f, g)). \quad (2.8)$$

If $(g, g) = 0$, then $(f, g) = 0$ follows by choosing $\lambda = \pm(g, f)$. If, on the other hand, $(g, g) \ne 0$, then set

$$z := \frac{(f, g)}{(g, g)},$$

so that (2.8) becomes (using $(g, g) > 0$)

$$\Re(\lambda z) \ge -\frac{|\lambda|^2}{2} \quad \forall \lambda \in \mathbb{K}.$$

Replacing λ by $e^{is}\lambda$ if necessary, we get $\pm|\lambda||z| \ge -\frac{|\lambda|^2}{2}$ for all λ, which is possible only when $z = 0$.

But as $(f, g) = 0$ for all $g \in H_0$, we can take in particular $g_y := K(\cdot, y)$ for any $y \in X$, which gives

$$0 = (f, g_y) = f(y) \quad \forall y \in X,$$

whence $f = 0$. So we have defined an inner product on H_0.

Let now $\hat{H}$ be the completion of H_0 under the given inner product; that is, $\hat{H}$ is the space of equivalence classes of Cauchy sequences in H_0, where two Cauchy sequences in H_0 are considered equivalent if their difference converges to zero[7]. We equip $\hat{H}$ with the inner product

$$(\hat{f}, \hat{g}) := \lim_{n\to\infty} (f_n, g_n),$$

whenever the equivalence class $\hat{f}$ contains the Cauchy sequence $(f_n)_{n\in\mathbb{N}} \subset H_0$, and similar for $\hat{g}$. It is readily checked that this is indeed a well-defined scalar product (note the limit exists thanks to completeness of $\mathbb{K}$). One can also identify H_0 as a dense subspace of $\hat{H}$.

Let $\Phi : \hat{H} \to \mathbb{K}^X$ be defined through

$$\Phi(\hat{f})(x) := (\hat{f}, K(\cdot, x)) \quad \forall x \in X.$$

Clearly, Φ is linear. It is injective, because $\Phi(\hat{f}) = 0$ implies $(\hat{f}, g) = 0$ for any $g \in H_0 \subset \hat{H}$; but since H_0 is dense in $\hat{H}$, it follows already that $\hat{f} = 0$. Also, if $\hat{f} \in H_0 \subset \mathbb{K}^X$, then $\Phi(\hat{f})(x) = \hat{f}(x)$, so that $\Phi \restriction_{H_0}$ is the identity.

[7]Cf. the construction of $\mathbb{R}$ as completion of $\mathbb{Q}$.

This allows us to identify $\hat{H}$ with the subset $H := \Phi(\hat{H}) \subset \mathbb{K}^X$. As an injective linear image of a Hilbert space, H is itself Hilbert with the inner product

$$(f,g)_H := (\Phi^{-1}(f), \Phi^{-1}(g))_{\hat{H}}.$$

Finally, we show that H is RKHS with reproducing kernel K. Indeed, for the evaluation functional and $f = \Phi(\hat{f})$ we have

$$\begin{aligned} |E_x(f)| &= |f(x)| = |\Phi(\hat{f})(x)| = |(\hat{f}, K(\cdot,x))_{\hat{H}}| \\ &= |(f, K(\cdot,x))_H| \le \|f\|_H \|K(\cdot,x)\|_H = \|f\|_H \sqrt{K(x,x)}, \end{aligned} \tag{2.9}$$

hence E_x is bounded with norm at most $\sqrt{K(x,x)}$. Also, K is the reproducing kernel, because $E_x(f) = (f, k_x)_H$ with $k_x := K(\cdot,x)$, as is apparent from (2.9), and $K(x,y) = (k_y, k_x)$ for $x, y \in X$. □

In fact, given an antisymmetric positive semidefinite map K, the Hilbert space with reproducing kernel K is uniquely determined. This follows from the following Lemma:

LEMMA 2.38. *Let H be an RKHS over X with reproducing kernel K. Set $k_x := K(\cdot,x)$ for each $x \in X$. Then,*

$$V := \operatorname{span}\{k_x : x \in X\}$$

is dense in H.

PROOF. If $f \in H$ and $f \perp k_x$ for all $x \in X$, then $f(x) = (f, k_x) = 0$ for all $x \in X$ and therefore $f = 0$. But a subspace of a Hilbert space whose orthogonal complement is zero must be dense (see the beginning of the proof of Riesz-Fréchet). □

Therefore, the strategy of proof of the Moore-Aronszajn Theorem was the only possible: If there is an RKHS whose kernel is K, then it *must* be the completion of the span of the functions k_x. We summarise:

DEFINITION 2.39. Let X be a set and $K : X \times X \to \mathbb{K}$ antisymmetric and positive semidefinite. The unique $\mathbb{K}$-RKHS over X whose reproducing kernel is K is denoted $H(K)$.

The kernel function K thus contains the entire information about the associated RKHS $H(K)$. As we shall see for the case of regression, many problems in high- or even infinite-dimensional Hilbert spaces can be reduced to computations involving only the kernel. This is the main reason why reproducing kernels play such an important role in data science.

2.5.3. Applications. We use the theory to investigate regression problems as they occur frequently in statistics and machine learning. It should be said in advance that linear or affine regression (as you have probably seen in Numerical Linear Algebra) can be easily solved by classical means of Linear Algebra and does not require the theory of RKHS. However, RKHS theory allows for vast generalisations to *nonlinear* classification problems that can no longer be handled by Linear Algebra alone. Due to time constraints, however, we have to stick to the simplest available examples.

Prior to the description of the regression problem, let us state an easy observation:

LEMMA 2.40. *Let X be a set, K a kernel (i.e., an antisymmetric positive semidefinite map $X \times X \to \mathbb{K}$), and for a finite subset $\{x_1, \dots, x_n\}$ of pairwise distinct elements of X denote $A := (K(x_k, x_l))_{k,l=1,\dots,n}$. If $w = (\alpha_1, \dots, \alpha_n) \in \mathbb{K}^n$ has $Aw = 0$, then $H(K) \ni f := \sum_{j=1}^n \alpha_j k_{x_j} = 0$.*

PROOF. This is just a one-line calculation:

$$\|f\|^2 = \sum_{k,l=1}^n \overline{\alpha_k} \alpha_l (k_{x_l}, k_{x_k}) = \sum_{k,l=1}^n \overline{\alpha_k} \alpha_l K(x_k, x_l) = \sum_{k,l=1}^n \overline{\alpha_k} \alpha_l A_{kl} = (A\alpha, \alpha) = 0.$$

□

Let us now consider the following problem: We are given n distinct 'data points' $\{x_1, x_2, \dots x_n\}$ from a set X that are already classified, i.e., they have known values $\lambda_1, \dots, \lambda_n \in \mathbb{K}$. Out of some class of 'nice' functions, we wish to find one that predicts the (unknown) values of all $x \in X$ in the best possible accordance with the known data; ideally, we would want $f(x_j) = \lambda_j$ $(j = 1, \dots n)$ *exactly.* However, such a function may not be available in the given class, so that we contend ourselves with an f that minimises the quadratic error, given as

$$Q[f] := \sum_{j=1}^{n} |f(x_j) - \lambda_j|^2.$$

In general, there may be many such minimisers. However, in case the chosen class of functions is an RKHS, there exists a minimiser that is unique under the additional condition that its norm be minimal:

THEOREM 2.41 (regression in RKHS). *Let $H = H(K)$ be an RKHS over X and $\{x_1, \dots, x_n\} \subset X$ pairwise distinct. Let moreover $v = (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n$ and*

$$A = (K(x_j, x_k))_{j,k=1,\dots,n}.$$

Then there exists a vector $w = (\alpha_1, \dots, \alpha_n) \in \mathbb{K}^n$ such that $v - Aw \in \ker(A)$, and the unique minimiser of the quadratic error $Q[f]$ with minimal norm of all functions $f \in H$ is given as

$$f = \sum_{j=1}^{n} \alpha_j k_{x_j}.$$

PROOF. Let $V := \operatorname{span}\{k_{x_1}, \dots, k_{x_n}\} \subset H$ and $f \in H$. Denoting by P_V the orthogonal projection onto V, we have that $f - P_V f \perp V$ and in particular

$$0 = (f - P_V f, k_{x_j}) = f(x_j) - P_V f(x_j) \quad \forall j = 1, \dots, n,$$

so that $f(x_j) = P_V(x_j)$ for all j. Let $P_V f = \sum_{k=1}^{n} \alpha_k k_{x_k}$ and write $w = (\alpha_1, \dots, \alpha_n)$, then it follows for all $k = 1, \dots, n$ that

$$(Aw)_k = \sum_{j=1}^{n} K(x_k, x_j)\alpha_j = \sum_{j=1}^{n} (\alpha_j k_{x_j}, k_{x_k}) = P_V f(x_k) = f(x_k),$$

and therefore $Q[f] = \sum_{k=1}^{n} |f(x_k) - \lambda_k|^2 = \|Aw - v\|^2$.

From Linear Algebra it is known (and from geometric intuition it should be clear) that $w \in \mathbb{C}^n$ is a minimiser of $\|Aw - v\|^2$ if and only if $Aw = P_{\operatorname{ran}(A)} v$. Therefore,

$$f = \sum_{j=1}^{n} \alpha_j k_{x_j}$$

is indeed a minimiser with the choice of $w = (\alpha_1, \dots, \alpha_n)$ as described.

Next, observe that among all minimisers in H of the error Q, those with smallest norm (if any) must be in V, because $f \in H$ coincides with $P_V f$ on $\{x_1, \dots, x_n\}$ and $\|P_V f\| < \|f\|$ if $f \notin V$.

Finally, suppose $f = \sum_{j=1}^{n} \alpha_j k_{x_j}$ and $g = \sum_{j=1}^{n} \beta_j k_{x_j}$ are both minimisers of Q, then for $w = (\alpha_j)$ and $\tilde{w} = (\beta_j)$ we must have $Aw = v = A\tilde{w}$, hence $A(w - \tilde{w}) = 0$. By Lemma 2.40, then, $f = g$. Hence f is the unique minimiser of Q in V, and therefore the unique minimiser of Q of minimal norm.

□

EXAMPLE 2.42 (linear regression). Let X be a real Hilbert space and $H = X'$ its dual. Setting the regression problem in this framework amounts to *linear* regression, as the space H of admissible functions, by which we wish to approximate the given data, consists precisely of the set of (bounded) linear functionals on X. (Think of $X = \mathbb{R}$ for the simplest possible situation.) We have already seen in Example 2.35 that H is RKHS and $K(x,y) = (x,y)$ for all $x, y \in H$. Using the notation of the preceding theorem, the minimiser of smallest norm is then given by

$$f(x) = \left(x, \sum_{j=1}^{n} \alpha_j x_j\right) \quad \forall x \in X,$$

where $w = (\alpha_1, \ldots, \alpha_n)$ is chosen such that $Aw = P_{\mathrm{ran}(A)} v$ and $A_{jl} = (x_j, x_l)$.

In many situations, linear functions form an arguably too small set of functions. We show next how to extend linear regression to the slightly larger space of *affine* functions. The method uses a *feature map* that maps the original data space to a higher dimensional space, where the problem becomes linear again:

EXAMPLE 2.43 (affine regression). Consider the set $\mathbb{A}^d$ of affine maps $\mathbb{R}^d \to \mathbb{R}$, that is,

$$\mathbb{A}^d = \left\{ f_{v,c} : \mathbb{R}^d \to \mathbb{R} : \quad f_{v,c}(x) = (v,x) + c \quad \forall x \in \mathbb{R}^d \right\}.$$

Define the 'feature map' $\phi : \mathbb{R}^d \to \mathbb{R}^{d+1}$ by

$$\phi(x_1, \ldots, x_d) := (x_1, \ldots, x_d, 1)$$

and observe there is a bijection between $(\mathbb{R}^{d+1})'$ and $\mathbb{A}^d$ given by

$$(\mathbb{R}^{d+1})' \ni g \mapsto g \circ \phi \in \mathbb{A}^d;$$

indeed, every $f_{v,c} \in \mathbb{A}^d$ can be written as $g \circ \phi$ with

$$g(x_1, \ldots, x_d, x_{d+1}) = \sum_{j=1}^{d} x_j v_j + c x_{d+1},$$

and clearly $g \mapsto g \circ \phi$ is injective.

Let again $\{x_1, \ldots, x_d\}$ be data points with given values $v = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$. We want to minimise the quadratic error among all functions in $\mathbb{A}^d$. To this end, we minimise the quadratic error for the linear regression problem with data points $\{\phi(x_1), \ldots, \phi(x_d)\}$ and the same values v by means of Example 2.42. This gives a minimiser of

$$\sum_{j=1}^{d} |g(\phi(x_j)) - \lambda_j|^2$$

among all $g \in (\mathbb{R}^{d+1})'$, and hence a minimiser of $\sum_{j=1}^{d} |f(x_j) - \lambda_j|^2$ in $\mathbb{A}^d$ by setting $f = g \circ \phi$.

Let us discuss this last example a bit further: We characterised the affine maps, which we were originally interested in, as the set

$$\{g \circ \phi : g \in (\mathbb{R}^{d+1})'\},$$

where $\phi : \mathbb{R}^d \to \mathbb{R}^{d+1}$ was a feature map transforming the problem from $\mathbb{R}^d$ into the (slightly) higher-dimensional space $\mathbb{R}^{d+1}$.

It turns out that the composition with ϕ leaves the property of being RKHS untouched; indeed, let us state without proof the Pullback Theorem:

THEOREM 2.44 (pullback of RKHS). *Let X and Y be sets and $\phi : Y \to X$. If $K : X \times X \to \mathbb{K}$ is positive semidefinite and antisymmetric, then so is $K \circ \phi$ (where we write $(K \circ \phi)(x,y) = K(\phi(x), \phi(y))$), and*

$$H(K \circ \phi) = H(K) \circ \phi := \{f \circ \phi : f \in H(K)\}.$$

Hence the regression problem over $H(K \circ \phi)$ can be reduced to one in $H(K)$. This becomes particularly advantageous if the space of interest has a kernel of the form $K(x, y) = (\phi(x), \phi(y))$, in which case the problem can be transformed into a linear regression problem as in Example 2.42.

Let us give another example to conclude this chapter.

EXAMPLE 2.45. Let $X = \mathbb{R}^2$ and

$$M = \left\{ f : \mathbb{R}^2 \to \mathbb{R} : f(x_1, x_2) = a_1 x_1^2 + a_2 x_1 x_2 + a_3 x_2^2 + b_1 x_1 + b_2 x_2 + c : a_1, a_2, a_3, b_1, b_2, c \in \mathbb{R} \right\}$$

the space of polynomials of order at most 2. Define the map

$$\phi : \mathbb{R}^2 \to \mathbb{R}^6, \quad (x_1, x_2) \mapsto (x_1^2, x_1 x_2, x_2^2, x_1, x_2, 1).$$

Then, for a polynomial f as in the definition of M, we observe

$$\begin{aligned} f(x_1, x_2) &= a_1 x_1^2 + a_2 x_1 x_2 + a_3 x_2^2 + b_1 x_1 + b_2 x_2 + c \\ &= (\phi(x_1, x_2), (a_1, a_2, a_3, b_1, b_2, c)), \end{aligned}$$

so that $M = \{g \circ \phi : g \in (\mathbb{R}^6)'\}$, and thus by the Pullback Theorem, $M = H(K \circ \phi)$, where K is simply the standard scalar product in $\mathbb{R}^6$, and so the kernel of M is given by $K(x, y) = (\phi(x), \phi(y))$.

Therefore, the solution of the *linear* regression problem with data points $\{\phi(x_1), \ldots, \phi(x_n)\}$ in $\mathbb{R}^6$ yields the solution of the *nonlinear* regression problem for M with data points $\{x_1, \ldots, x_n\}$ in $\mathbb{R}^2$.

CHAPTER 3

The Baire Category Theorem and its Consequences

3.1. The Baire Category Theorem

Let (X, d) be a metric space and $M \subset X$. Recall the *interior* $\operatorname{int} M \subset M$ is defined as the largest open subset of M, or equivalently, the set of all $x \in M$ such that there exists $\epsilon > 0$ with $B_\epsilon(x) \subset M$. Similarly, the *closure* $\overline{M}$ is the smallest closed set containing M, or equivalently, the union of M with the set of its limit points. Of course, these notions are dependent on the metric d.

Definition 3.1 (nowhere dense sets). A subset $M \subset X$ is called *nowhere dense* if $\operatorname{int} \overline{M} = 0$.

A nowhere dense set can be considered very 'small'. In fact, nowhere dense sets are so small that their countable union can never be the whole space:

Theorem 3.2 (Baire Category Theorem). *Let (X, d) be a nonempty complete metric space. Then X is not the countable union of nowhere dense sets.*

Proof. Let $(X_n)_{n\in\mathbb{N}}$ be a family of nowhere dense sets. We construct sequences $(x_n)_{n\in\mathbb{N}} \subset X$ and $(\epsilon_n)_{n\in\mathbb{N}} \subset \mathbb{R}^+$ such that

- $\epsilon_n \searrow 0$ as $n \to \infty$;
- $\overline{B_{\epsilon_n}(x_n)} \cap \overline{X_n} = \varnothing$ for all $n \in \mathbb{N}$;
- $B_{\epsilon_{n+1}}(x_{n+1}) \subset B_{\epsilon_n}(x_n)$ for all $n \in \mathbb{N}$.

Suppose there is such a sequence, then by the last property, if $m > n \geq N$, then

$$x_m \in B_{\epsilon_{m-1}}(x_{m-1}) \subset \ldots \subset B_{\epsilon_n}(x_n) \subset B_{\epsilon_N}(x_n), \tag{3.1}$$

so that $d(x_n, x_m) < \epsilon_N$ and so $(x_n)_{n\in\mathbb{N}}$ is Cauchy. As X is complete, there is a limit $x = \lim_{n\to\infty} x_n$, and we claim that $x \notin \overline{X_n}$ for all $n \in \mathbb{N}$. Indeed, similarly as in (3.1) we have $x \in \overline{B_{\epsilon_n}(x_n)}$ for any $n \in \mathbb{N}$, and by the second property it follows $x \notin \overline{X_n}$.

Now for the construction of the sequence. As $\overline{X_1}$ is closed and has empty interior, its complement is dense in X (so in particular nonempty, as X is nonempty) and open; we may therefore find $\epsilon_1 > 0$ and $x_1 \in X$ such that $\overline{B_1(0)} \cap \overline{X_1} = \varnothing$.

Suppose x_n and ϵ_n have already been constructed and satisfy the stated properties. Again, $\overline{X_{n+1}}$ has open and dense complement, so there exists $x_{n+1} \in B_{\epsilon_n}(x_n)$ and $\epsilon_{n+1} < \frac{1}{2}\epsilon_n$ such that $B_{\epsilon_{n+1}}(x_{n+1}) \subset B_{\epsilon_n}(x_n)$ and $\overline{B_{\epsilon_n}(x_n)} \cap \overline{X_n} = \varnothing$, as desired. □

Remark 3.3.
(1) In fact, the proof shows something more: If X is a complete metric space and $(X_n)_{n\in\mathbb{N}}$ is a countable family of nowhere dense sets, then the complement of $\bigcup_{n=1}^\infty X_n$ is dense in X. This can be seen from the above proof, as x_1 can be chosen from a dense set in X and ϵ_1 can be chosen arbitrarily small.
(2) An alternative formulation of the Theorem is: If $(O_n)_{n\in\mathbb{N}}$ is a countable family of dense open sets, then $\bigcap_{n=1}^\infty O_n$ is nonempty (in fact it is even dense).
(3) Some terminology: If $M \subset X$ is a countable union of nowhere dense sets, then it is called *meagre* (or *of first category*). Sets that are not meagre are called *of second category*. Complements of meagre sets are called *comeagre* or (in a disturbingly misleading terminology) *residual* or, colloquially, *fat*. In this terminology, the

Baire Category Theorem can be stated as: A nonempty complete metric space is of second category/not meagre (with respect to itself).

EXAMPLE 3.4 (continuous but nowhere differentiable functions). A direct application of the Baire Category Theorem is the existence of functions that are continuous but nowhere differentiable. The first example of such a function was explicitly given by Weierstraß in 1872 and sparked some controversy among mathematicians. Here, we use the Baire Theorem to show the existence of such 'monsters' in a rather slick topological way.

Consider $C([0,1];\mathbb{R})$ with the supremum norm, as usual. We have seen that this space is complete. We want to show that the set of functions in this space that are differentiable at at least one point is meagre. To this end, define for $n \in \mathbb{N}$

$$X_n := \left\{ f \in C([0,1]) : \quad \exists x \in [0,1] \quad \sup_{0<|h|\le 1/n} \frac{|f(x+h)-f(x)|}{|h|} \le n \right\},$$

where we can extend f to the left by $f(0)$ and to the right by $f(1)$ to ensure well-definedness. Clearly, if $f \notin X_n$ for all $n \in \mathbb{N}$, then f is nowhere differentiable, because its difference quotients are not bounded (let alone convergent) in any neighbourhood of any point. By the Baire Category Theorem, then, it suffices to show that each X_n is nowhere dense, i.e., $\operatorname{int}\overline{X_n} = \varnothing$.

First, we show that $\overline{X_n} = X_n$ (that is, X_n is closed). Fix $n \in \mathbb{N}$. Let $(f_k)_{k\in\mathbb{N}} \subset X_n$ converge uniformly to $f \in C([0,1])$, then we need to show $f \in X_n$. Since $f_k \in X_n$, there exists $x_k \in [0,1]$ such that

$$\sup_{0<|h|\le 1/n} \frac{|f_k(x_k+h)-f_k(x_k)|}{|h|} \le n.$$

By Bolzano-Weierstraß, There exists a convergent subsequence (not relabelled), so that $x_k \to x \in [0,1]$ as $k \to \infty$. Let $0 < |h| \le 1/n$, then

$$\frac{|f(x+h)-f(x)|}{|h|} = \lim_{k\to\infty} \frac{|f_k(x_k+h)-f_k(x_k)|}{|h|} \le n,$$

because $f_k(x_k) \to f(x)$ thanks to uniform convergence (think it over and cook up an example where this fails for just pointwise convergence). So indeed, $f \in X_n$ and X_n is closed.

It thus remains to show that X_n has empty interior, or equivalently, for any $f \in X_n$ and $\epsilon > 0$ there exists $\tilde{f} \in C([0,1]) \setminus X_n$ such that $\|f - \tilde{f}\|_\infty < \epsilon$. By the Weierstraß Approximation Theorem, there exists a polynomial function $p \in C([0,1])$ such that $\|f - p\|_\infty < \frac{\epsilon}{2}$. In particular, p is differentiable and p' is bounded in $[0,1]$. Let $s_\epsilon \in C([0,1])$ be a sawtooth function with $\|s_\epsilon\|_\infty < \frac{\epsilon}{2}$ and $|s_\epsilon'| = M$ almost everywhere, with M yet to be determined. Then, for $x \in [0,1]$, $\tilde{f} := p + s_\epsilon \in B_\epsilon(f)$ satisfies

$$\begin{aligned}\sup_{0<|h|\le 1/n} \frac{|\tilde{f}(x+h)-\tilde{f}(x)|}{|h|} \ge& \sup_{0<|h|\le 1/n} \frac{|s_\epsilon(x+h)-s_\epsilon(x)|}{|h|} - \sup_{0<|h|\le 1/n} \frac{|p(x+h)-p(x)|}{|h|} \\ \ge& M - \|p'\|_\infty > n\end{aligned}$$

provided M is chosen sufficiently large (depending on f but not on x). Note we used the Mean Value Theorem for the estimate of the difference quotient of p. This shows $\tilde{f} \notin X_n$.

We conclude that the set of nowhere differentable functions is fat, and in particular dense, in $C([0,1])$ (and, very much in particular, there exists such a function).

3.2. The Uniform Boundedness Principle

3.2.1. The Theorem and Some Corollaries. Recall from your first year how sacrilegious it is to confuse the order of existential and universal quantifiers: It is strictly prohibited to conclude $\exists x \forall y$ from $\forall y \exists x$. Thus the following theorem is quite surprising.

THEOREM 3.5 (Banach-Steinhaus/Uniform Boundedness Principle). *Let X be Banach and Y a normed space, and assume $(T_j)_{j\in J} \subset L(X,Y)$ is a (possibly uncountable) family of bounded linear operators such that*

$$\sup_{j\in J} \|T_j x\| < \infty \quad \forall x \in X.$$

Then,

$$\sup_{j\in J} \|T_j\| < \infty.$$

PROOF. For $n \in \mathbb{N}$, set

$$X_n := \{x \in X : \sup_{j\in J} \|T_j x\| \le n\}.$$

Note X_n is closed, because it can be written as

$$X_n = \bigcap_{j\in J} \|T_j(\cdot)\|^{-1}([0,n])$$

and thus as an intersection of closed sets (of course, $x \mapsto \|T_j x\|$ is continuous). By assumption, $X = \bigcup_{n\in\mathbb{N}} X_n$, so by the Baire Category Theorem, there exists $N \in \mathbb{N}$ such that X_N has non-empty interior. Let therefore $x_0 \in X_N$ and $\epsilon > 0$ such that $B_\epsilon(x_0) \subset X_N$. As X_N is symmetric in the sense that $x \in X_N$ if and only if $-x \in X_N$, then also $B_\epsilon(-x_0) \subset X_N$. Even better, X_N is convex by virtue of the triangle inequality for the norm $\|\cdot\|$ in Y: Indeed, if $\|T_j x_1\| \le N$ and $\|T_j x_2\| \le N$ for all $j \in J$, and if $\lambda_1, \lambda_2 \in [0,1]$ sum up to 1, then

$$\|\lambda_1 T_j x_1 + \lambda_2 T_j x_2\| \le \lambda_1 \|T x_1\| + \lambda_2 \|T x_2\| \le N.$$

As a consequence, if $\|y\| < \epsilon$, then

$$y = \frac{1}{2}(x_0 + y) + \frac{1}{2}(-x_0 + y) \in X_N,$$

so that if $x \in B_1(0)$, then

$$\|T_j x\| = \frac{1}{\epsilon}\|T_j(\epsilon x)\| \le \frac{N}{\epsilon} \quad \forall j \in J,$$

whence $\sup_{j\in J} \|T_j\| \le \frac{N}{\epsilon} < \infty$. □

The Uniform Boundedness Principle has a number of beautiful corollaries:

COROLLARY 3.6. *Let X be a normed space and $M \subset X$. Then M is bounded if and only if for all $x' \in X'$, the set $x'(M) \subset \mathbb{K}$ is bounded.*

PROOF. If M is bounded by, say, $S > 0$, then for all $x \in M$, $|x'(x)| \le \|x'\|\|x\| \le S\|x'\|$, whence boundedness of $x'(M)$ already follows.

Conversely, assume $x'(M)$ is bounded for all $x' \in X'$. Consider the canonical embedding $\iota : X \to X''$. Then, for every $x' \in X'$,

$$\infty > \sup_{x\in M} |x'(x)| = \sup_{x\in M} |\iota[x](x')|,$$

and since X' is Banach, we may apply the Uniform Boundedness Principle to get

$$\infty > \sup_{x\in M} \|\iota[x]\| = \sup_{x\in M} \|x\|.$$

□

COROLLARY 3.7 (weakly convergent sequences are bounded). *Weakly convergent sequences are bounded.*

PROOF. Let X be a normed space and $(x_n)_{n\in\mathbb{N}} \subset X$ weakly convergent. Then for any $x' \in X'$, the convergent sequence $(x'(x_n))_{n\in\mathbb{N}} \subset \mathbb{K}$ is bounded, hence the set $\{x_n : n \in \mathbb{N}\}$ is bounded by virtue of the preceding corollary. □

COROLLARY 3.8 (dual version of Corollary 3.6). *Let X be Banach and $M \subset X'$. Then M is bounded if and only if for all $x \in X$, the set $\{x'(x) : x' \in M\} \subset \mathbb{K}$ is bounded.*

PROOF. If M is bounded, then the boundesness of $\{x'(x) : x' \in M\}$ follows from $|x'(x)| \le \|x'\|\|x\|$ and boundedness of M. The converse is the special case $Y = \mathbb{K}$ of the Uniform Boundedness Principle. □

COROLLARY 3.9 (Pointwise limits of continuous linear operators are continuous (!)). *Let X Banach and Y a normed space. Let $(T_n)_{n\in\mathbb{N}} \subset L(X,Y)$ converge pointwise to T, that is, $\lim_{n\to\infty} T_n x = Tx$ for all $x \in X$. Then $T \in L(X,Y)$.*

PROOF. The linearity of T is standard, see the proof of Proposition 1.17. So we only show continuity. As $(T_n x)_{n\in\mathbb{N}}$ is convergent, it is bounded, and so $\sup_{n\in\mathbb{N}} T_n x < \infty$ for each $x \in X$. By the Uniform Boundedness Principle, $\sup_{n\in\mathbb{N}} \|T_n\| =: M < \infty$. Therefore,

$$\|Tx\| = \lim_{n\to\infty} \|T_n x\| \le M\|x\| \quad \forall x \in X.$$

□

3.2.2. An Application to Fourier Series. As an application, we wish to show that the Fourier series of a periodic continuous function on $\mathbb{R}$ does not necessarily converge pointwise, although, as we have seen, it converges in $L^2_{per}(\mathbb{R};\mathbb{C})$.

Recall from Section 2.2 that a function $f \in L^2_{per}(\mathbb{R};\mathbb{C})$ can be written as

$$f = \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} \hat{f}(k) e^{ik\cdot}, \tag{3.2}$$

where

$$\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-ikx} dx, \quad k \in \mathbb{Z}. \tag{3.3}$$

Note carefully that the convergence implied in the infinite sum in (3.2) is understood in the L^2 sense. While it is well-known that each L^2-convergent sequence has a subsequence that converges *almost everywhere*[1], it makes no sense to ask about *everywhere* convergence, as an L^2 function is only defined up to a nullset. However, if f is additionally assumed to be continuous (and thus everywhere defined), it is a very reasonable conjecture that its Fourier series would converge at every $x \in \mathbb{R}$. Owing to the Uniform Boundedness Principle, we can show that this is *not* the case in general.

As a preparation, let us represent the partial sums of the Fourier series in a useful new way. Using (3.3), we compute for $n \in \mathbb{N}$:

$$\frac{1}{\sqrt{2\pi}} \sum_{k=-n}^{n} \hat{f}(k) e^{ikx} = \frac{1}{2\pi} \sum_{k=-n}^{n} \int_0^{2\pi} f(y) e^{-iky} dy e^{ikx} = \frac{1}{2\pi} \int_0^{2\pi} f(y) \sum_{k=-n}^{n} e^{ik(x-y)} dy.$$

Observe that for any $z \in \mathbb{R}$, we may write (thanks to the summation formula for finite geometric sums)

$$\sum_{k=-n}^{n} e^{ikz} = \sum_{k=0}^{2n} e^{i(k-n)z} = e^{-inz} \frac{1 - e^{i(2n+1)z}}{1 - e^{iz}} = \frac{e^{-i(n+\frac{1}{2})z} - e^{i(n+\frac{1}{2})z}}{e^{-i\frac{z}{2}} - e^{i\frac{z}{2}}} = \frac{\sin\left((n+\frac{1}{2})z\right)}{\sin\left(\frac{z}{2}\right)},$$

so in total we get

$$\frac{1}{\sqrt{2\pi}} \sum_{k=-n}^{n} \hat{f}(k) e^{ikx} = \int_0^{2\pi} f(y) D_n(x-y) dy.$$

[1]Therefore, a subsequence of the Fourier series of $f \in L^2_{per}$ will converge almost everywhere to f. In fact, one does not even need to pass to a subsequence: This is a deep result of Carleson from 1966.

Here,

$$D_n(z) := \frac{1}{2\pi} \frac{\sin\left((n+\frac{1}{2})z\right)}{\sin\left(\frac{z}{2}\right)}$$

is called the *Dirichlet kernel* ($n \in \mathbb{N}$). Note the Dirichlet kernel is 2π-periodic and continuous (even at zero – think it over).

Recall the space $C_{per}(\mathbb{R};\mathbb{C})$ of 2π-periodic continuous functions, which as usual we equip with the supremum norm. Setting

$$s_n(f,x) := \frac{1}{\sqrt{2\pi}} \sum_{k=-n}^{n} \hat{f}(k) e^{ikx}$$

for the n-th partial sum of the Fourier series ($n \in \mathbb{N}$), and choosing $x = 0$, we can view $s_n(\cdot, 0)$ as an element of the dual space C'_{per}, because

$$|s_n(f,0)| = \left| \int_0^{2\pi} f(y) D_n(y) dy \right| \le \|f\|_\infty \int_0^{2\pi} |D_n(y)| dy,$$

where the integral on the right hand side is finite (for each fixed $n \in \mathbb{N}$) as $D_n \in C_{per}$. Note we used $D_n(y) = D_n(-y)$ for all $y \in \mathbb{R}$.

In fact, the operator norm of $s_n(f,0)$ is precisely given by $\int_0^{2\pi} |D_n(y)| dy$, as can be seen by setting

$$f_\epsilon := \frac{\overline{D_n}}{|D_n| + \epsilon}$$

and letting $\epsilon \searrow 0$ (you are encouraged to work out the details yourself).

THEOREM 3.10 (failure of everywhere convergence of the Fourier series). *There exists a function $f \in C_{per}(\mathbb{R};\mathbb{C})$ whose Fourier series does not converge at $x = 0$.*

PROOF. Suppose the Fourier series of every $f \in C_{per}$ converged at $x = 0$, so in particular the sequences $(s_n(f,0))_{n\in\mathbb{N}}$ would be bounded for all $f \in C_{per}$. Since C_{per} is Banach, we can apply the Uniform Boundedness Principle to deduce that

$$\sup_{n\in\mathbb{N}} \|s_n(\cdot,0)\| = \sup_{n\in\mathbb{N}} \int_0^{2\pi} |D_n(y)| dy < \infty.$$

We will reach the desired contradiction once we show that the integrals of $|D_n|$ go off to infinity. This is shown by an elementary estimate:

$$\begin{aligned} 2\pi \int_0^{2\pi} |D_n(y)| dy &= \int_0^{2\pi} \frac{\left|\sin\left(\left(n+\frac{1}{2}\right)y\right)\right|}{\left|\sin\left(\frac{y}{2}\right)\right|} dy \ge 2 \int_0^{2\pi} \frac{\left|\sin\left(\left(n+\frac{1}{2}\right)y\right)\right|}{y} dy \\ &= 2 \int_0^{(2n+1)\pi} \frac{|\sin(x)|}{x} dx \ge 2 \sum_{k=1}^{2n} \int_{(k-1)\pi}^{k\pi} \frac{|\sin(x)|}{x} dx \\ &\ge 2 \sum_{k=1}^{2n} \frac{1}{k\pi} \int_{(k-1)\pi}^{k\pi} |\sin(x)| dx = \frac{2}{\pi} I \sum_{k=1}^{2n} \frac{1}{k} \to \infty \end{aligned}$$

as $n \to \infty$, where we set $I := \int_0^\pi |\sin(x)| dx > 0$. We used the variable transformation $x = (n+\frac{1}{2})y$ in passing from the first to the second line, and the π-periodicity of $|\sin|$ as well as the divergence of the harmonic series in the last step. □

REMARK 3.11. The question of convergence of Fourier series has been of great interest in the mathematical area of harmonic analysis (and is well-suited for a bachelor's or master's thesis). The above proof of non-convergence via Uniform Boundedness is very pretty, but also non-constructive. Du Bois-Reymond gave an explicit construction of such a function in 1876 (long before the Uniform Boundedness Principle was available), thus disproving the everywhere convergence conjecture that had been held by the leading mathematicians of the time. The *almost everywhere* convergence of the Fourier series of an L^2_{per} function, however, was proved by Carleson in 1966, as mentioned.

One may ask whether additional assumptions on f ensure everywhere convergence. Indeed, if f is continuously differentiable, then the Fourier series converges everywhere (see [**5**, Satz IV.2.9]) and even uniformly.

We know from the Stone-Weierstraß Theorem that every function in C_{per} can be uniformly approximated by trigonometric polynomials. Theorem 3.10 shows that the Fourier series does *not* give such an approximation. Can we somehow use the Fourier series to give an explicit sequence of trigonometric polynomials that converges uniformly to f? The answer is *yes*: By taking the so-called *Cesàro mean* of the Fourier series, we obtain a sequence of trigonometric polynomials converging uniformly to f. More precisely, if $f \in C_{per}$ and

$$c_N(f,x) := \frac{1}{N}\sum_{n=1}^{N} s_n(f,x),$$

then $c_N(f,\cdot)$ converges uniformly to f as $N \to \infty$ (see [**5**, IV.2.11]). The reason is that c_N can be represented by arithmetic means of the Dirichlet kernels D_n (these arithmetic means are called *Fejér* kernels), and the Fejér kernels behave much better than the Dirichlet kernels, as the former are non-negative and have integral 1 for all $n \in \mathbb{N}$.

3.3. The Open Mapping Theorem

Definition 3.12 (open mappings). A map between two metric spaces is called *open* if it maps open sets to open sets.

Compare this with the topological definition of continuity: A map between two metric (or even topological) spaces is continuous if the *preimages* of open sets are open. Open maps have this property in the other direction, so to speak. In particular, if an open map has an inverse, then the inverse will be continuous.

Lemma 3.13. *Let X, Y be normed spaces and $T : X \to Y$ linear. Then the following are equivalent:*

(1) *T is open;*
(2) *For every $r > 0$, the image $T(B_r(0)) \subset Y$ is a neighbourhood*[2] *of zero;*
(3) *$T(B_1(0)) \subset Y$ is a neighbourhood of zero.*

Proof. (1)⇒ (2): By linearity, $T(0) = 0$. Since $B_r(0)$ is open, then so is $T(B_r(0))$ as T is open. Hence $T(B_r(0))$ is an open neighbourhood of zero.

(2)⇒(1): Let $O \subset X$ be open and $Tx \in T(O)$ for some $x \in O$. As O is open, there exists $r > 0$ such that $B_r(x) \subset O$, and then (by linearity) $T(B_r(x)) = Tx + T(B_r(0)) \subset T(O)$. By (2), $T(B_r(0))$ is a neighbourhood of zero, which means there exists $\epsilon > 0$ such that $B_\epsilon(0) \subset T(B_r(0))$, which implies $B_\epsilon(Tx) \subset T(O)$. This shows that $T(O)$ is open.

(2)⇔(3): This is obvious from the linearity of T. □

Clearly, any open linear map is surjective, because it maps onto a neighbourhood of the origin. The converse, however, is anything but obvious. Yet, for maps between Banach spaces, it is true:

Theorem 3.14 (Open Mapping Theorem, Banach). *Let X, Y be Banach and $T \in L(X, Y)$ surjective. Then T is open.*

[2]Recall: In a metric space, a *neighbourhood* of x is any set N such that $B_\epsilon(x) \subset N$ for some $\epsilon > 0$.

PROOF. By the preceding Lemma, it suffices to show that $T(B_1(0))$ is a neighbourhood of zero.

Step 1. First we show there exists $\epsilon_0 > 0$ such that $B_{\epsilon_0}(0) \subset \overline{T(B_1(0))}$. To this end, note that by assumption of surjectivity,

$$Y = \bigcup_{n\in\mathbb{N}} T(B_n(0)),$$

so that the Baire Category Theorem gives us an $N \in \mathbb{N}$ for which $\overline{T(B_N(0))}$ has nonempty interior, so there is $y_0 \in Y$ and $\epsilon > 0$ with $B_\epsilon(y_0) \subset \overline{T(B_N(0))}$.

The proof now proceeds just like the proof of the Uniform Boundedness Principle: $\overline{T(B_N(0))}$ is symmetric (i.e., it contains z if and only if it contains $-z$) and convex, as $T(B_N(0))$ has these properties, which are preserved by the closure. Therefore, if $y \in B_\epsilon(0)$,

$$y = \frac{1}{2}(x_0 + y) + \frac{1}{2}(-x_0 + y) \in \overline{T(B_N(0))},$$

which completes the first step of the proof upon choosing $\epsilon_0 := \frac{\epsilon}{N}$.

Step 2. We shall improve Step 1 to the effect that $B_{\epsilon_0}(0) \subset T(B_1(0))$, thereby completing the proof of the Open Mapping Theorem. So let $y \in B_{\epsilon_0}(0)$, then there exists $0 < \epsilon < \epsilon_0$ such that even $y \in B_\epsilon(0)$. Set $\bar{y} := \frac{\epsilon_0}{\epsilon} y$. Let $0 < \alpha < 1$ to be chosen later. We recursively construct a sequence $(x_n)_{n\in\mathbb{N}} \subset B_1(0)$ in X such that

$$\left\| \bar{y} - T\left(\sum_{k=0}^{n-1} \alpha^k x_{k+1} \right) \right\| < \alpha^n \epsilon_0, \quad n \in \mathbb{N}, \tag{3.4}$$

in the following way: As $\bar{y} \in B_{\epsilon_0}(0)$, by Step 1, $\bar{y} \in \overline{T(B_1(0))}$, so there exists $y_1 = Tx_1$ with $x_1 \in B_1(0)$ and $\|\bar{y} - y_1\| < \alpha\epsilon_0$.

If $x_1, \ldots, x_n$ have been constructed for some $n \in \mathbb{N}$, observe that

$$\frac{\bar{y} - T\left(\sum_{k=0}^{n-1} \alpha^k x_{k+1}\right)}{\alpha^n} \in B_{\epsilon_0}(0),$$

so by another application of Step 1 we find $x_{n+1} \in B_1(0)$ such that

$$\left\| \frac{\bar{y} - T\left(\sum_{k=0}^{n-1} \alpha^k x_{k+1}\right)}{\alpha^n} - Tx_{n+1} \right\| < \alpha\epsilon_0.$$

It follows that (3.4) is still satisfied with n replaced by $n+1$.

Since $\sum_{k=0}^{\infty} \|\alpha^k x_{k+1}\| \le \frac{1}{1-\alpha} < \infty$ (recall $\|x_k\| < 1$), the sum is absolutely convergent and thus, as X is Banach, also convergent to some $\bar{x} \in X$. By (3.4) and continuity of T, we have $\bar{y} = T\bar{x}$.

Finally set $x := \frac{\epsilon}{\epsilon_0}$, then $Tx = y$ and

$$\|x\| = \frac{\epsilon}{\epsilon_0}\|\bar{x}\| \le \frac{\epsilon}{\epsilon_0} \sum_{k=0}^{\infty} \|\alpha^k x_{k+1}\| \le \frac{\epsilon}{\epsilon_0} \frac{1}{1-\alpha} < 1$$

if $\alpha > 0$ is chosen sufficiently small. This shows $y \in T(B_1(0))$, and the Theorem is proved. □

COROLLARY 3.15. *Let X and Y be Banach and $T \in L(X,Y)$ bijective, then $T^{-1} \in L(Y,X)$.*

PROOF. Linearity of T^{-1} is clear. As the images under T of open sets are open, then the preimages under T^{-1} of open sets are open, so T^{-1} is continuous. □

COROLLARY 3.16. *Let X be a vector space and $\|\cdot\|_1, \|\cdot\|_2 : X \to \mathbb{R}$ two norms with respect to both of which X is Banach. If there exists $M > 0$ such that*

$$\|\cdot\|_1 \le M\|\cdot\|_2,$$

then the two norms are equivalent.

This means we get the other inequality $\|\cdot\|_2 \leq M'\|\cdot\|_1$ 'for free'.

PROOF. By assumption, the identity map $(X, \|\cdot\|_2) \to (X, \|\cdot\|_1)$, $x \mapsto x$ is bounded. By Corollary 3.15, then also the inverse map $(X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$, $x \mapsto x$, is bounded, that is, $\|\cdot\|_2 \leq M'\|\cdot\|_1$ for some $M' > 0$. □

3.4. The Closed Graph Theorem

Let X and Y be normed spaces and $D \subset X$. Recall the *graph* of a map $f : D \to Y$ is given as

$$\Gamma_f := \{(x, y) \in X \times Y : x \in D, y = f(x)\}.$$

In fact, according to the usual set-theoretic definition, a map simply *is* its graph[3]. Anyway, note that if f is linear, then Γ_f is a subspace of $X \times Y$ (in particular, $D \subset X$ must then be a subspace).

DEFINITION 3.17 (maps with closed graph). Let X and Y be normed spaces, $D \subset X$ a subspace and $T : D \to Y$ a linear map. We say that T has *closed graph* if its graph is closed in $X \times Y$, that is: Whenever $(x_n)_{n\in\mathbb{N}} \subset D$ with $(x_n, Tx_n) \to (x, y) \in X \times Y$, then $x \in D$ and $y = Tx$.

In the case that D is a *closed* subspace, note that having a closed graph is weaker that being continuous: If T is continuous, then $x_n \to x$ implies the convergence of $(Tx_n)_{n\in\mathbb{N}}$, namely to Tx. However, if T only has closed graph, then we need to *assume* convergence of $(Tx_n)_{n\in\mathbb{N}}$ in order to conclude that the limit is Tx.

EXAMPLE 3.18.
(1) Let $X = Y = C([-1, 1])$ with the supremum norm and $D = C^1([-1, 1]) \subset X$. Then, the differential operator $D \to Y$, $f \mapsto f'$, has closed graph: For let $(f_n)_{n\in\mathbb{N}}$ be a sequence of C^1 functions converging uniformly to f and assume $f' \to g$ uniformly, then it is known from Analysis (see for instance [**1**, §21, Satz 5]) that $f \in C^1$ and $g = f'$. Hence the differential operator has closed graph although it is discontinuous, see Example 1.19(3).
(2) Changing the norms in the preceding example, we exhibit an operator whose graph is not closed: Let $X = Y = L^2(-1, 1)$, $D = C^1([-1, 1])$, and consider again the differential operator $D \to Y$, $f \mapsto f'$. Choosing

$$f_n(x) = \left(x^2 + \frac{1}{n}\right)^{1/2}, \quad f(x) = |x|$$

and

$$g(x) = \begin{cases} -1 & x < 0, \\ 1 & x > 0, \end{cases}$$

we find that $D \ni f_n \to f$ in L^2 (even uniformly!) and $f_n' \to g$ in L^2 (but not uniformly); however, $f \notin D$ and so the differential operator does not have closed graph with respect to the chosen spaces.

On the domain D of definition of a linear operator, we may define the *graph norm* by $\|x\|_\Gamma := \|x\| + \|Tx\|$. It is clear that this is really a norm.

LEMMA 3.19. *Let X, Y be Banach, $D \subset X$ a subspace, and assume $T : D \to Y$ has closed graph. Then, D together with the graph norm is Banach, and T is continuous from $(D, \|\cdot\|_\Gamma)$ to Y.*

[3] This is a manifestation of the extensionality principle.

Proof. Let $(x_n)_{n\in\mathbb{N}} \subset D$ be Cauchy with respect to the graph norm, so that $(x_n)_{n\in\mathbb{N}}$ and $(Tx_n)_{n\in\mathbb{N}}$ are Cauchy in X and Y, respectively. As X and Y are Banach, we have $x_n \to x \in X$ and $Tx_n \to y \in Y$. But since T is closed, we can deduce $x \in D$ and $y = Tx$, thereby implying $x_n \to x$ in the graph norm. This shows the first assertion.

The second claim is obvious because $\|Tx\|_Y \leq \|x\|_X + \|Tx\|_Y = \|x\|_\Gamma$ for every $x \in D$. □

With this observation we can generalise the Open Mapping Theorem in the following way:

Theorem 3.20 (Surjective operators with closed graph are open). *Let X, Y be Banach and $D \subset X$ a subspace. If $T : D \to Y$ has closed graph and is surjective, then T is open. In particular, if in addition T is injective, then the inverse $T^{-1} : Y \to D$ is continuous.*

Proof. By the previous Lemma and the Open Mapping Theorem, T is open as a map $(D, \|\cdot\|_\Gamma) \to Y$, meaning that the image of an open set in X with respect to the graph norm is open in Y. But any open set with respect norm induced by X is also open with respect to the graph norm: Indeed, let $O \subset D$ be open with respect to $\|\cdot\|_X$, then for any $x \in O$ there is $\epsilon > 0$ such that $\|x - y\|_X < \epsilon$ implies $y \in O$. So if $\|x - y\|_X \leq \|x - y\|_\Gamma < \epsilon$, then $y \in O$. Continuity of T^{-1} now follows just as in Corollary 3.15. □

The proof of the main theorem in this section is now easy.

Theorem 3.21 (Closed Graph Theorem). *Let X, Y be Banach and $T : X \to Y$ linear. If T has closed graph, then T is continuous.*

Proof. By Lemma 3.19, T is continuous with respect to the graph norm. But by Corollary 3.16, the graph norm and the norm of X are equivalent. Therefore, T is even continuous with respect to the norm of X. □

3.5. Projections in Banach Spaces

In Hilbert spaces, we know how to project (orthogonally) onto nonempty closed and convex sets, in particular onto closed subspaces. Let's take a brief look into the situation for Banach spaces.

Definition 3.22 (projections). If X is a vector space, then any linear map $P : X \to X$ with $P^2 = P$ is called a (linear) *projection.*

If X is a vector space and $U, V \subset X$, then we say X is the *direct sum* of U and V and write $X = U \oplus V$ if $X = \{u + v : u \in U, v \in V\}$ and $U \cap V = \{0\}$; in this case, the representation of each vector in X as a sum of vectors in U and V is unique.

Proposition 3.23 (properties of projections). *Let X be a normed space and $P : X \to X$ a continuous linear projection. Then,*

(1) $P = 0$ *or* $\|P\| \geq 1$.
(2) *The kernel and the range of P are closed.*
(3) $X = \ker P \oplus P(X)$.

Proof. (1) This follows immediately from $\|P\| = \|P^2\| \leq \|P\|^2$.

(2) Since $\ker P = P^{-1}(\{0\})$ and P is continuous, the kernel is closed as the preimage of a closed set under a continuous map. Similarly, if $I : X \to X$ is the identity, then $I - P$ is also a continuous linear projection (as $(I - P)^2 = I^2 - 2P + P^2 = I - P$), so its kernel is closed; therefore,

$$P(X) = \ker(I - P)$$

is also closed.

(3) Clearly, for each $x \in X$, we can write

$$x = (x - Px) + Px,$$

where $x - Px \in \ker P$ and $Px \in P(X)$. Also, suppose $x \in \ker P \cap P(X)$: Then $x = Py$ for some $y \in X$ but also $0 = Px = P^2 y = Py = x$, so the sum is indeed direct. □

As an example, consider the map $P : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ given by $Pf = \chi_{(0,1)} f$, where $\chi_{(0,1)}$ denotes the indicator function of $(0,1)$. As $\chi^2_{(0,1)} = \chi_{0,1}$, this is a projection. It is linear and bounded, as $\|Pf\|_p \le \|f\|_p$. Its kernel is given by the set of L^p functions whose essential support is disjoint from $(0,1)$, and its image is the set of functions whose essential support is contained in $[0,1]$. Accordingly, we can uniquely write any L^p function f as the sum $(1 - \chi_{(0,1)})f + \chi_{(0,1)} f$.

Does every closed subspace of a Banach space admit a continuous projection? The answer, maybe surprisingly, is no: It can be shown (see [**5**, Satz IV.6.5]) that the closed subspace $c_0 \subset l^\infty$ of nullsequences does not have a complement, that is, there does not exist any closed subspace $U \subset l^\infty$ such that $l^\infty = U \oplus c_0$, which by Proposition 3.23 is a necessary condition for the existence of a continuous projection.

For *finite-dimensional* subspaces, however, there is always a continuous projection:

THEOREM 3.24 (projection onto finite-dimensional subspaces). *Let X be a normed space and $U \subset X$ a finite-dimensional subspace, then there exists a bounded linear projection P onto U such that $\|P\| \le \dim U$.*

PROOF. Let $\{u_1, \dots, u_n\}$ be a so-called *Auerbach basis* for U, that is, a basis for which there exists a dual basis $\{u'_1, \dots, u'_n\}$ of U' such that $\|u_k\| = 1$ and $\|u'_k\| = 1$ for $k = 1, \dots, n$, and $u'_j(u_k) = \delta_{jk}$ for $j, k = 1, \dots, n$. The proof of existence of an Auerbach basis for any finite-dimensional normed space can be found, e.g., in [**5**, II.2.6].

By the Hahn-Banach Theorem, for every $k = 1, \dots, n$, we can extend the bounded linear functionals u'_k from U' to $x'_k \in X'$ preserving the norms, so that still $\|x'_k\| = 1$. Set

$$Px := \sum_{k=1}^{n} x'_k(x) u_k,$$

then P is the desired projection. □

As mentioned, we see from Proposition 3.23 that a necessary condition for a closed subspace U of a Banach space X to have a bounded linear projection is the existence of a *complementary space* for U, that is, another closed subspace $V \subset X$ such that $X = U \oplus V$. The next theorem states that this is even sufficient:

THEOREM 3.25 (criterion for existence of a continuous linear projection). *Let X be Banach and $U \subset X$ a closed subspace. Assume there exists another closed subspace $V \subset X$ such that $X = U \oplus V$[4]. Then the norm defined as $\|x\|_1 := \|u\| + \|v\|$ is equivalent to the norm of X (where $u \in U$ and $v \in V$ are the unique vectors such that $x = u + v$), and there exists a continuous linear projection onto U.*

PROOF. By the triangle inequality, $\|x\| \le \|u\| + \|v\|$ whenever $x \in X$ and $x = u + v$ with $u \in U$, $v \in V$. An application of Corollary 3.16 will yield equivalence of the norms once we show that X is also Banach with respect to $\|\cdot\|_1$. So let $(x_n)_{n\in\mathbb{N}}$ be Cauchy in $\|\cdot\|_1$ with corresponding decomposition $x_n = u_n + v_n$, so that for all $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for $m, n \ge N$,

$$\epsilon > \|x_n - x_m\|_1 = \|u_n - u_m\|_X + \|v_n - v_m\|_X,$$

so that $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are Cauchy and therefore convergent to u and v, respectively, since X with its original norm is assumed Banach. But then $x_n \to x = u + v$ in the norm $\|\cdot\|_1$, so that equivalence of both norms is proved.

[4]The complementary space will typically not be unique: Consider for instance the span of the unit vector e_1 as a closed subset of $\mathbb{R}^2$, then the spans of e_2 and of $e_1 + e_2$ are both complementary to $\operatorname{span}\{e_1\}$.

As for the existence of a bounded projection, by the equivalence of the two norms we have the 'reverse triangle inequality' $\|u\| + \|v\| \le M\|u+v\|$ for some $M > 0$ and all $u \in U$, $v \in V$, so that the linear map $u + v \to u$ is bounded, and thus this map is in fact the desired projection. □

Note that in the special case that X is Hilbert, we can always take V to be the orthogonal complement of U; this gives an alternative proof of the existence of the projection onto any closed subspace of a Hilbert space. As mentioned, in Banach spaces a complementary space need not always exist.

CHAPTER 4

Spectral Theory

We have already seen the simplest spectral theorem for compact self-adjoint operators on Hilbert spaces in Section 2.4. We will now generalise this theory in two directions: to compact operators on *Banach* spaces, and to bounded but possibly non-compact operators on Hilbert spaces.

4.1. The Adjoint Operator

DEFINITION 4.1 (The adjoint in normed spaces). Let X, Y be normed spaces and $T \in L(X,Y)$. The *adjoint operator* $T' \in L(Y', X')$ is defined by

$$T'y'(x) = y'(Tx).$$

It is clear that this defines a bounded linear operator $Y' \to X'$ (because $|y'(Tx)| \le \|y'\|\|T\|\|x\|$). Also, note if $X = Y$ is Hilbert, then this definition is related to our previous Definition 2.21 thanks to the Riesz-Fréchet Theorem. Indeed, if $X = Y = H$ is Hilbert and $x' = (\cdot, y)$ for some $y \in H$, then the adjoint T' of $T \in L(H,H)$ in the sense of Definition 4.1 is represented as

$$T'x' = (\cdot, T^*y),$$

where $T^* \in L(H,H)$ is the Hilbert space adjoint as in Definition 2.21.

EXAMPLE 4.2.
(1) For $1 \le p < \infty$, consider $X = Y = L^p(0,1)$. Given a function $h \in L^\infty(0,1)$, we may define the multiplication operator $T : X \to Y$, $f \mapsto hf$ (note $hf \in L^p$ if $f \in L^p$). Let L^q be the dual of L^p (meaning $\frac{1}{p} + \frac{1}{q} = 1$); then for $f \in L^p$ and $g \in L^q$ we have

$$g(Tf) = g(hf) = \int_0^1 g(x)h(x)f(x)dx = (gh)(f),$$

so that $T' : Y' \to X'$ is given by $T'g = gh$, so T' is itself the multiplication operator with h.

(2) Let X be a normed space, then we have seen the injection operator $\iota : X \to X''$, $\iota[x](x') = x'(x)$. Then the adjoint $\iota' : X''' \to X'$ is defined as

$$\iota'[x'''](x) = x'''(\iota(x)),$$

so $\iota'[x'''] = x''' \circ \iota$.

THEOREM 4.3. *Let X, Y, Z be normed spaces. The map $L(X,Y) \to L(Y', X')$, $T \mapsto T'$ is a linear isometry, and for $T \in L(X,Y)$ and $S \in L(Y,Z)$, we have $(ST)' = T'S'$.*

PROOF. Linearity is clear by definition, and we note

$$\|T\| = \sup_{\|x\|\le 1} \|Tx\| = \sup_{\|y'\|\le 1, \|x\|\le 1} |y'(Tx)| = \sup_{\|y'\|\le 1, \|x\|\le 1} |T'y'(x)| = \|T'\|,$$

where we made use of Corollary 1.23.

For the second statement, let $x \in X$ and $z' \in Z'$, then

$$(ST)'z'(x) = z'(STx) = S'z'(Tx) = T'S'z'(x).$$

□

As an isometry, $T \mapsto T'$ is injective. Note it may not be surjective: If, for instance, Y is not Banach, then $L(X,Y)$ need not be Banach, but $L(Y',X')$ is Banach in any case (because X' is).

In the case of Hilbert spaces, $T^{**} = T$. In general, we have the following:

LEMMA 4.4. *Let X,Y be normed spaces and $T \in L(X,Y)$, then*

$$T'' \circ \iota_X = \iota_Y \circ T,$$

where $\iota_X : X \to X''$ and $\iota_Y : Y \to Y''$ denote the canonical injections.

PROOF. Note the operators on both sides should map X to Y''. So let $x \in X$ and $y' \in Y'$, then

$$T''(\iota_X[x])(y') = \iota_X[x](T'y') = T'y'(x) = y'(Tx) = \iota_Y[Tx](y').$$

□

THEOREM 4.5 (Schauder). *Let X and Y be normed spaces and $T \in L(X,Y)$.*

(1) *If T is compact, then so is T'.*
(2) *If Y is Banach and T' is compact, then so is T.*

PROOF. (1) Suppose T is compact. Let $(y'_n)_{n\in\mathbb{N}}$ be a bounded sequence in Y'. By compactness of T, the set $K := \overline{T(B_1(0))} \subset Y$ is compact. The restrictions $y'_n \restriction_K$ form a sequence of continuous functions $K \to \mathbb{K}$, uniformly bounded by $\|T\| \sup_{n\in\mathbb{N}} \|y'_n\| < \infty$. The sequence is also equicontinuous, because for all $n \in \mathbb{N}$,

$$|y'_n(y_1) - y'_n(y_2)| \le \sup_{k\in\mathbb{N}} \|y'_k\| \|y_1 - y_2\|.$$

Invoking now the Theorem of Arzelà-Ascoli, we find a uniformly convergent subsequence $(y'_{n_k} \restriction_K)_{k\in\mathbb{N}}$. Therefore,

$$\|T'y'_{n_k} - T'y'_{n_l}\| = \sup_{x\in B_1(0)} \|y'_{n_k}(Tx) - y'_{n_l}(Tx)\| = \|y'_{n_k} \restriction_K - y'_{n_l} \restriction_K \|_\infty,$$

where the last expression can be made arbitrarily small choosing k,l large enough. Hence, $(T'y'_{n_k})_{k\in\mathbb{N}}$ is Cauchy and hence convergent, so indeed T' is compact.

(2) Assume in addition Y to be Banach, and let T' be compact. By the first part of the proof, T'' is then also compact, and so is $T'' \circ \iota_X$ as the composition of a compact and a bounded operator. By Lemma 4.4, $T'' \circ \iota_X = \iota_Y \circ T$, so the latter is compact. Therefore, if $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence in X, then there is a convergent subsequence $(\iota_Y(Tx_{n_k}))_{k\in\mathbb{N}}$ in Y''. In particular, the sequence is Cauchy, and so is the sequence $(Tx_{n_k})_{k\in\mathbb{N}}$, because ι_Y is an isometry. Convergence of $(Tx_{n_k})_{k\in\mathbb{N}}$ now follows from the completeness of Y. □

THEOREM 4.6 (A linear operator is invertible iff its adjoint is). *Let X,Y be Banach. Then $T \in L(X,Y)$ is bijective if and only if T' is, and in this case $(T')^{-1} = (T^{-1})'$.*

PROOF. Let T be bijective, then it is invertible, and its inverse $T^{-1} \in L(Y,X)$ by virtue of the Open Mapping Theorem (Corollary 3.15). Therefore, $(T^{-1})' \in L(X',Y')$ is well-defined, and for $y' \in Y'$ and $y \in Y$ we have

$$(T^{-1})'[T'y'](y) = T'y'(T^{-1}y) = y'(TT^{-1}y) = y'(y),$$

so that $(T^{-1})'T'$ is the identity on Y'. Likewise, for $x' \in X'$ and $x \in X$,

$$T'[(T^{-1})'x'](x) = (T^{-1})'x'(Tx) = x'(T^{-1}Tx) = x'(x),$$

so that $T'(T^{-1})'$ is the identity on X'. It follows that T' is invertible with $(T')^{-1} = (T^{-1})'$.

Conversely, assume T' is invertible, and denote $S := (T')^{-1}$. T is injective: Suppose $Tx = 0$ and let $x' \in X'$. As T' is surjective, there exists $y' \in Y'$ such that $T'y' = x'$, and then we have

$$x'(x) = T'y'(x) = y'(Tx) = 0,$$

so $x = 0$ follows from Corollary 1.24.

For surjectivity, we show first that $T(X) \subset Y$ is closed. So let $(Tx_n)_{n\in\mathbb{N}}$ be a sequence in $T(X)$ converging to $y \in Y$. Then the sequence (Tx_n) is Cauchy, and by Corollary 1.23,

$$\begin{aligned}\|x_n - x_m\| &= \sup_{\|x'\|\le 1} |x'(x_n - x_m)| = \sup_{\|x'\|\le 1} |T'Sx'(x_n - x_m)| \\ &= \sup_{\|x'\|\le 1} |Sx'(Tx_n - Tx_m)| \le \|S\| \|Tx_n - Tx_m\|,\end{aligned}$$

which shows that $(x_n)_{n\in\mathbb{N}}$ is Cauchy and thus convergent to $x \in X$. But then $Tx = y$ by continuity of T, and so $y \in T(X)$.

Suppose now T were not surjective, so there existed $y_0 \in Y \setminus T(X)$. By the Hahn-Banach Theorem, taking into account the closedness of $T(X)$, we can find a functional $y' \in Y'$ such that $y' \restriction_{T(X)} = 0$ but $y'(y_0) = 1$. Indeed, let $\omega : Y \to Y/T(X)$, $y \mapsto y + T(X)$ be the canonical quotient map. Then $\omega(Tx) = 0$ for all $x \in X$ and $\omega(y_0) \neq 0$. By Hahn-Banach (applied to the span of $\omega(y_0)$ as a subspace of $Y/T(X)$), there exists a functional $l \in (Y/T(X))'$ such that $l(\omega(y_0)) = 1$. The functional $y' := l \circ \omega$ is then as desired. But then

$$0 = y'(Tx) = T'y'(x)$$

for every $x \in X$, whence from injectivity of T' we have $y' = 0$, in contradiction with $y'(y_0) = 1$.

The formula $S = (T^{-1})'$ follows by the same computation as in the beginning of the proof.

□

4.2. Basic Concepts of Spectral Theory for Bounded Operators

Recall from Subsection 2.4.3 some basic notions of spectral theory: If X is Banach and $T \in L(X, X)$, then the resolvent set $\rho(T)$ is the set of $\lambda \in \mathbb{K}$ such that $\lambda I - T$ is invertible (meanwhile we have learned from the Open Mapping Theorem that then the inverse is automatically bounded). For $\lambda \in \rho(T)$, we denoted the resolvent by $R(\lambda, T) = (\lambda I - T)^{-1}$. The complement $\sigma(T) := \mathbb{K} \setminus \rho(T)$ was called the spectrum.

We made the observation that not every element of $\sigma(T)$ needs to be an eigenvalue. We therefore distinguish three parts of the spectrum:

- The *point spectrum* $\sigma_p(T)$ is the set of eigenvalues of T, i.e., the set of $\lambda \in \mathbb{K}$ such that $\lambda I - T$ is not injective;
- the *continuous spectrum* $\sigma_c(T)$ is the set of $\lambda \in \mathbb{K}$ such that $\lambda I - T$ is injective but not surjective, and the range of $\lambda I - T$ is dense in X;
- the *residual spectrum* $\sigma_r(T)$ is the set of $\lambda \in \mathbb{K}$ such that $\lambda I - T$ is injective but not surjective, and the range of $\lambda I - T$ is not dense in X.

It is clear from these definitions that $\mathbb{K} = \rho(T) \dot\cup \sigma_p(T) \dot\cup \sigma_c(T) \dot\cup \sigma_r(T)$, that is, each $\lambda \in \mathbb{K}$ is an element of exactly one of these four sets. The terminology will make more sense later on.

Proposition 4.7 (Spectra of an operator and its adjoint are the same). *If X is a Banach space and $T \in L(X, X)$, then $\sigma(T') = \sigma(T)$.*

Proof. By Theorem 4.6, the map $\lambda I - T$ is invertible if and only its adjoint is. But the adjoint is precisely $\lambda I - T'$, because for any $x' \in X'$ and $x \in X$,

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

□

In the case that X is Hilbert, note carefully that the spectrum of the Hilbert space adjoint T^* is the *complex conjugate* of the spectrum of T (Lemma 2.30). This annoying discrepancy between the Banach and Hilbert adjoints is due to the fact that the identification of a Hilbert space with its dual via Riesz-Fréchet is only an anti-isomorphism (i.e., an isomorphism up to complex conjugation), cf. the discussion after the Riesz-Fréchet Theorem.

EXAMPLE 4.8. (1) We choose $X = C([0,1])$ with the usual supremum norm, so that X is Banach. Set

$$(Tx)(s) := \int_0^s x(t)dt,$$

which defines a continuous linear operator. We wish to show $\sigma(T) = \sigma_r(T) = \{0\}$.

First, observe that for $\lambda \neq 0$, the operator $\lambda I - T$ is invertible: Indeed, let $y \in C^1([0,1])$ for the moment, then $\lambda x - Tx = y$ is equivalent with the initial value problem

$$\dot{x} - \frac{1}{\lambda}x = \frac{1}{\lambda}\dot{y}, \quad x(0) = \frac{y(0)}{\lambda},$$

which has a unique solution explicitly given by

$$x(t) = e^{t/\lambda}\left(\frac{1}{\lambda}\int_0^t e^{-s/\lambda}\dot{y}(s)ds + \frac{y(0)}{\lambda}\right) = \frac{1}{\lambda^2}\int_0^t e^{(t-s)/\lambda}y(s)ds + \frac{1}{\lambda}y(t),$$

where we used integration by parts in the last step. In fact, this formula still yields a solution of $\lambda x - Tx = y$ for general $y \in C([0,1])$. This solution is unique because $\lambda x - Tx = 0$ is uniquely solved by zero, so if $\lambda x_1 - Tx_1 = y = \lambda x_2 - Tx_2$, then $x_1 - x_2 = 0$. Therefore, $\lambda I - T$ is bijective.

Let us consider the case $\lambda = 0$. If $Tx = 0$ then $x = 0$, so T is injective. On the other hand, the image $T(X)$ is not dense in X, because $(Tx)(0) = 0$ for all $x \in X$. Hence $0 \in \sigma_r(T)$.

(2) Consider the same operator T, but now on the smaller space $X = \{x \in C([0,1]) : x(0) = 0\}$, still with the supremum norm. Exactly as in the previous situation, one shows $\lambda \in \rho(T)$ for $\lambda \neq 0$, and T is injective. But now $T(X)$ is dense in X, because $T(X)$ is the space of C^1 functions such that $y(0) = \dot{y}(0) = 0$, and it is an analysis exercise to check that this space is dense in X.

THEOREM 4.9. *Let X be Banach and $T \in L(X,X)$.*

(1) *The resolvent map $\rho(T) \to L(X,X)$, $\lambda \mapsto R(\lambda,T)$ is analytic, that is: It can be locally expressed as a power series in λ with coefficients in $L(X,X)$.*
(2) *When $\mathbb{K} = \mathbb{C}$ then $\sigma(T) \neq \varnothing$.*

PROOF. (1) This follows from the Neumann Series representation (Lemma 2.28):

$$\begin{aligned} R(\lambda,T) &= (\lambda I - T)^{-1} = (\lambda_0 I - T)^{-1}\left[I - (\lambda_0 - \lambda)(\lambda_0 I - T)^{-1}\right]^{-1} \\ &= \sum_{n=0}^{\infty}(\lambda_0 - \lambda)^n((\lambda_0 I - T)^{-1})^{n+1} \end{aligned}$$

for any $\lambda_0 \in \rho(T)$ and λ sufficiently close to λ_0.

(2) Suppose for a contradiction that $\sigma(T) = \varnothing$. Then $R(\lambda,T)$ is analytic on all of $\mathbb{C}$. Let $l' \in L(X,X)'$, then by the previous step we can write

$$l'(R(\lambda,T)) = \sum_{n=0}^{\infty}(-1)^n l'(R(\lambda_0,T)^{n+1})(\lambda - \lambda_0)^n, \tag{4.1}$$

so that $\lambda \mapsto l'(R(\lambda,T))$ is an analytic function $\mathbb{C} \to \mathbb{C}$.

This map is also bounded, because if $|\lambda| > 2\|T\|$, then again thanks to the Neumann Series,

$$|l'(R(\lambda,T))| \le \|l'\|\|(\lambda I - T)^{-1}\| = \|l'\|\frac{1}{\lambda}\sum_{n=0}^{\infty}\left\|\frac{T}{\lambda}\right\|^n \le \frac{\|l'\|}{\|T\|}$$

(without loss of generality, $T \neq 0$, because the spectrum of the zero operator is $\{0\}$); and on the compact set $\overline{B_{2\|T\|}(0)}$, the map is bounded anyway as it is continuous.

It follows from Liouville's Theorem (from complex analysis) that $\lambda \mapsto l'(R(\lambda,T))$ is in fact constant, which implies that all except the zero-th coefficient in (4.1) vanish. In particular, for the choice $\lambda_0 = 0$ (which is valid as, by assumption, $\rho(T) = \mathbb{C}$), we obtain for $n = 1$ that $0 = l'(R(0,T)^2) = l'(T^{-2})$, and it follows from Corollary 1.24 that $T^{-2} = 0$, which is the desired contradiction (as the zero operator is not the inverse of anything). □

From Theorem 2.29 we know that $|\lambda| \le \|T\|$ for all $\lambda \in \sigma(T)$. In order to improve this result, we need a technical lemma:

LEMMA 4.10. *Let $(a_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ be a real sequence such that $0 \le a_{n+m} \le a_n a_m$ for $n,m \in \mathbb{N}$. Then,*

$$\lim_{n\to\infty} a_n^{1/n} = \inf_{n\in\mathbb{N}} a_n^{1/n}.$$

PROOF. Let $\epsilon > 0$ and $N \in \mathbb{N}$ such that $a_N^{1/N} < \inf_{n\in\mathbb{N}} a_n^{1/n} + \epsilon$. Set

$$b := \max\{a_1, \dots, a_N\},$$

and note that b depends on ϵ. Let $n \in \mathbb{N}$ be written as $n = kN + r$ with $1 \le r \le N$. Then, denoting $a := \inf_{n\in\mathbb{N}} a_n^{1/n}$,

$$a_n^{1/n} = a_{kN+r}^{1/n} \le (a_N^k a_r)^{1/n} \le (a+\epsilon)^{kN/n} b^{1/n} = (a+\epsilon)(a+\epsilon)^{-r/n} b^{1/n} \le a + 2\epsilon$$

for n sufficiently large, which proves the claim. □

DEFINITION 4.11 (spectral radius). If X is Banach and $T \in L(X,X)$, then the *spectral radius* of T is defined as

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

Note that $a_n = \|T^n\|$ satisfies the requirement of Lemma 4.10, so that the spectral radius is well-defined. Since $\|T^n\| \le \|T\|^n$, it is clear that always $r(T) \le \|T\|$. The following result is therefore an improvement over the bound from Theorem 2.29, and justifies the terminology 'spectral radius':

THEOREM 4.12 (sharp bound for the spectrum). *Let X be Banach and $T \in L(X,X)$. Then, for every $\lambda \in \sigma(T)$, we have $|\lambda| \le r(T)$. Moreover, when $\mathbb{K} = \mathbb{C}$, then there exists $\lambda \in \sigma(T)$ with $|\lambda| = r(T)$.*

PROOF. Let $|\lambda| > r(T)$ and consider the series $\lambda^{-1}\sum_{n=0}^{\infty}(T/\lambda)^n$. This series is absolutely convergent by the Root Test, because

$$\limsup_{n\to\infty}\left\|\left(\frac{T}{\lambda}\right)^n\right\|^{1/n} = \frac{1}{|\lambda|}\lim_{n\to\infty}\|T^n\|^{1/n} = \frac{r(T)}{|\lambda|} < 1.$$

It is readily checked that $\lambda^{-1}\sum_{n=0}^{\infty}(T/\lambda)^n$ is the inverse of $\lambda I - T$ (cf. the proof of Lemma 2.28), which means $\lambda \in \rho(T)$. Therefore, whenever $\lambda \in \sigma(T)$, then $|\lambda| \le r(T)$.

For the sharpness assertion, set

$$r_0 := \max\{|\lambda| : \lambda \in \sigma(T)\},$$

which is well-defined by compactness of $\sigma(T)$ (Theorem 2.29). We have already seen $r_0 \le r(T)$. Let $\mu \in \mathbb{C}$ with $|\mu| > r_0$. If we can show $|\mu| \ge r(T)$, then the proof is complete.

To this end, let $l' \in L(X,X)'$, and consider the the function

$$f_{l'} : \mathbb{C} \smallsetminus \overline{B_{r_0}(0)} \to \mathbb{C}, \quad \lambda \mapsto l'(R(\lambda,T)),$$

which is well-defined by choice of r_0 and analytic by Theorem 4.9. As before, if $|\lambda| > r(T)$, then

$$f_{l'}(\lambda) = \sum_{n=0}^{\infty} l'(T^n)\lambda^{-n-1}.$$

It is known from complex analysis that this series converges on any open annulus on which $f_{l'}$ is analytic, so in particular it converges at μ, which implies

$$\lim_{n\to\infty} l'\left(\frac{T^n}{\mu^{n+1}}\right) = 0.$$

As $l' \in L(X,X)'$ was arbitrary, this means that $\left(\frac{T^n}{\mu^{n+1}}\right)_{n\in\mathbb{N}}$ converges weakly to zero in $L(X,X)$, so it is bounded by Corollary 3.7. So there exists $M > 0$ such that

$$\|T^n\|^{1/n} \le M^{1/n}|\mu|^{(n+1)/n} \to |\mu|,$$

which proves $|\mu| \ge r(T)$ as desired. □

We saw in the proof of the Spectral Theorem for compact self-adjoint operators on Hilbert spaces that, in that case, in fact $r(T) = \|T\|$. In general this is no longer true: The operator T from Example 4.8 has $\|T\| = 1$ but $\sigma(T) = \{0\}$ and therefore, thanks to Theorem 4.12, $r(T) = 0$.

4.3. Spectral Theory for Compact Operators

4.3.1. The Theorem of Riesz-Schauder. Throughout this subsection, let X be Banach, $T \in L(X,X)$ a compact operator, and $S := I - T$, where as usual I denotes the identity operator. For $n \in \mathbb{N} \cup \{0\}$, set $N_n := \ker S^n$ (in particular $N_0 = \{0\}$) and $R_n := S^n(X)$ (in particular $R_0 = X$). Clearly, $N_0 \subset N_1 \subset N_2 \subset \dots$ and $R_0 \supset R_1 \supset R_2 \supset \dots$.

We shall prove the following properties:

Lemma 4.13. (1) *There exists a smallest number $p \in \mathbb{N} \cup \{0\}$ such that $N_p = N_{p+1}$, and for this p we have $N_{p+r} = N_p$ for all $r \in \mathbb{N}$. Also, $N_p \cap R_p = \{0\}$.*

(2) *There exists a smallest number $q \in \mathbb{N} \cup \{0\}$ such that $R_q = R_{q+1}$, and for this q we have $R_{q+r} = R_q$ for all $r \in \mathbb{N}$. Also, $N_q + R_q = X$.*

(3) *$p = q$.*

Proof. (1) Note that all N_n are closed as preimages of the closed set $\{0\}$ under the continuous map S^n. If there were no p such that $N_p = N_{p+1}$, then we would have

$$N_0 \subsetneq N_1 \subsetneq N_2 \subsetneq \dots$$

Then, by the Lemma of Riesz (exercise), for every $n \in \mathbb{N}$ there exists $x_n \in N_n$ such that $\|x_n\| = 1$ and[1] $\operatorname{dist}(x_n, N_{n-1}) > \frac{1}{2}$. If $n > m \ge 1$, then

$$\|Tx_n - Tx_m\| = \|x_n - (Sx_n + x_m - Sx_m)\| > \frac{1}{2}$$

as $S(N_n) \subset N_{n-1}$ and therefore $Sx_n + x_m - Sx_m \in N_{n-1}$. But this implies that $(Tx_n)_{n\in\mathbb{N}}$ cannot have a convergent subsequence, in contradiction with compactness of T.

Fix p to be the smallest number with $N_p = N_{p+1}$. Next, let $r \in \mathbb{N}$. We need to show $N_{p+r} \subset N_p$. Let $x \in N_{p+r}$, then $S^{r-1}(x) \in N_{p+1} = N_p$ so $x \in N_{p+r-1}$. Iterate this argument to get $N_{p+r} \subset N_p$.

Let $x \in N_p \cap R_p$, then $S^p x = 0$ and there is $y \in X$ such that $x = S^p y$, so that $S^{2p}(y) = 0$ and thus $y \in N_{2p} = N_p$. Hence $0 = S^p y = x$.

[1]Here, $\operatorname{dist}(x_n, N_{n-1})$ means $\inf\{\|x_n - x\| : x \in N_{n-1}\}$.

(2) We argue as in the first step, only we need that all R_n are closed. Note that

$$S^n = (I-T)^n = I - \sum_{k=1}^{n} \binom{n}{k} (-1)^{k+1} T^k = I - \tilde{T}$$

with $\tilde{T}$ compact, as the compact operators form an algebra (in fact even an ideal) in $L(X,X)$. It will be shown below in Theorem 4.16 that operators of this form always have closed range, and the proof of this part of Theorem 4.16 will of course be independent of this Lemma.

Now let us proceed as in the first part: Suppose there did not exist $q \in \mathbb{N} \cup \{0\}$ such that $R_q = R_{q+1}$, then

$$R_0 \supsetneq R_1 \supsetneq R_2 \supsetneq \dots,$$

and again we could choose $x_n \in R_n$ with $\|x_n\| = 1$ and $\operatorname{dist}(x_n, R_{n+1}) > \frac{1}{2}$ by the Riesz Lemma. Then as before, for $m > n \geq 1$,

$$\|Tx_n - Tx_m\| = \|x_n - (Sx_n + x_m - Sx_m)\| > \frac{1}{2},$$

because $S(R_n) = R_{n+1}$ and thus $Sx_n + x_m - Sx_m \in R_{n+1}$. Again this contradicts compactness of T.

So let q be the smallest number with $R_q = R_{q+1}$. Let now $r \in \mathbb{N}$ and $x \in R_{q+r}$, then there is $y \in X$ with $x = S^{q+r}y$. As $S^q y \in R_q = R_{q+1}$, it follows that $x \in R_{q+r+1}$. It now follows by induction that $R_{q+r} = R_q$ for all $r \in \mathbb{N}$.

Let $x \in X$, then $S^q x \in R_q = R_{2q}$, so there is $y \in X$ such that $S^q x = S^{2q} y$. Therefore

$$x = (x - S^q y) + S^q y \in N_q + R_q.$$

(3) Suppose $p > q$, then by (2) $R_p = R_q$ but there would exist $x \in N_p \setminus N_q$. By (2) we can write $x = y + z \in N_q + R_q$, and $z = x - y \in N_p + N_q = N_p$. On the other hand, $z \in R_q = R_p$. By (1), $z = 0$, so that $x = y \in N_q$, which is a contradiction. Therefore $p \leq q$.

But suppose now $p < q$. Then by (1) $N_p = N_q$ but there is $x \in R_p \setminus R_q$. Again by (2), we can write $x = y + z \in N_q + R_q$, and so $y = x - z \in R_p + R_q = R_p$. But also $y \in N_q = N_p$, so again from (1) we deduce $y = 0$ and thus $x = z \in R_q$, contradiction.

□

Corollary 4.14. *There are closed subspaces $\hat{N}$ and $\hat{R}$ with the following properties:*

1. $\dim \hat{N} < \infty$;
2. $X = \hat{N} \oplus \hat{R}$
3. $S(\hat{N}) \subset \hat{N}$, $S(\hat{R}) \subset \hat{R}$;
4. $S \restriction_{\hat{R}}$ *is an isomorphism from* $\hat{R}$ *to* $\hat{R}$.

Proof. Choose $\hat{N} = N_p$ and $\hat{R} = R_q$, where N_p and R_q are the spaces from Lemma 4.13. Both spaces are closed, and $X = \hat{N} \oplus \hat{R}$ follows from $p = q$ and from $N_p \cap R_p = \{0\}$ and $X = N_p + R_p$. That $\hat{N}$ is finite-dimensional will follow independently from Theorem 4.16 below.

Note also

$$S(\hat{N}) = S(N_p) \subset N_{p-1} \subset N_p = \hat{N}$$

and

$$S(\hat{R}) = S(R_p) = R_{p+1} = R_p = \hat{R}.$$

In particular, $S \restriction_{\hat{R}}$ is surjective from $\hat{R}$ to $\hat{R}$. But it is also injective, because if $Sy = 0$ for some $y \in \hat{R}$, then there is $x \in X$ with $y = S^p x$, then $S^{p+1}x = 0$, thus $x \in N_{p+1} = N_p$, and then $y = S^p x = 0$. By the Open Mapping Theorem, the inverse of $S \restriction_{\hat{R}}$ is continuous, so indeed we have an isomorphism. □

Before we go on, let us recall some facts about quotient spaces (see also the corresponding exercise). If X is a normed space and $U \subset X$ a closed subspace, we can define an equivalence relation on X by $x \sim y$ if $x - y \in U$. The equivalence classes are then, for $x \in X$, sets of the form

$$x + U := \{x + u : u \in U\}.$$

The set of such equivalence classes becomes a vector space, denoted X/U, under the linear structure

$$(x + U) + (y + U) = (x + y) + U, \quad \alpha(x + U) = \alpha x + U,$$

which is easily seen to be well-defined. On the quotient space X/U one defines the norm

$$\|x + U\| := \inf\{\|x - y\| : y \in U\}.$$

If X is Banach, then so is X/U (exercise). If Y is another normed space and $R : X \to Y$ is a bounded linear operator, then the induced operator

$$\hat{R} : X/\ker R \to R(X), \quad x + U \mapsto Rx$$

is well-defined and bijective (think it over). It is also continuous, because

$$\|\hat{R}(x + \ker R)\| = \|Rx\| = \|R(x - y)\| \le \|R\|\|x - y\| \quad \forall y \in \ker R.$$

Lemma 4.15. *Let $R \in L(X, X)$ such that $R(X)$ is closed, then*

$$\ker(R)^\perp = R'(X'),$$

where $U^\perp := \{x' \in X' : x'(x) = 0 \quad \forall x \in U\}$ for a closed subspace $U \subset X$.

Proof. Let first $x' \in R'(X')$, so there is $y' \in X'$ such that $x' = R'y'$. Let $x \in \ker(R)$, then $x'(x) = R'y'(x) = y'(Rx) = 0$, so $x' \in \ker(R)^\perp$ as claimed.

For the converse inclusion, first observe there exists $K > 0$ such that for all $y \in R(X)$ there is $x \in X$ with $Rx = y$ and $\|x\| \le K\|y\|$; indeed, consider the induced operator $\hat{R} : X/\ker R \to R(X)$ defined by $\hat{R}(x + \ker R) := Rx$. Then, as mentioned, $\hat{R}$ is bijective and continuous from the Banach space $X/\ker R$ to the Banach space $R(X)$. By the Open Mapping Theorem, $\hat{R}^{-1}$ is continuous, and the existence of K follows from the definition of the norm in the quotient space.

Let now $x' \in \ker(R)^\perp$. Consider the map

$$z' : R(X) \to \mathbb{K}, \quad Rx \mapsto x'(x),$$

which is well-defined because $x' \in \ker(R)^\perp$. We claim that z' is continuous. So let K be the constant from the preceding argument, then for $y = Rx \in R(X)$ and $\|x\| \le K\|y\|$ it holds that

$$|z'(y)| = |x'(x)| \le \|x'\|\|x\| \le \|x'\|K\|y\|,$$

which shows continuity of z'. Let now $y' \in X'$ be a Hahn-Banach extension of z' to all of X. Then $x' = R'y'$ because

$$x'(x) = z'(Rx) = y'(Rx) = R'y'(x)$$

for every $x \in X$, and it follows $x' \in R'(X')$. □

After all this preparation we can state the following important result:

Theorem 4.16 (Riesz-Schauder). *Let X be Banach, $T \in L(X, X)$ a compact operator, and $S = I - T$. Then:*

(1) $\ker S$ *is finite-dimensional;*
(2) $S(X)$ *is closed and the quotient space $X/S(X)$ is finite-dimensional;*
(3) $\dim(X/S(X)) = \dim(\ker S) = \dim(X'/S'(X')) = \dim(\ker S')$.

PROOF. (1) Let $(x_n)_{n\in\mathbb{N}} \subset \ker S$ be a bounded sequence. By compactness, there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $(Tx_{n_k})_{k\in\mathbb{N}}$ converges. Since $0 = Sx_{n_k} = x_{n_k} - Tx_{n_k}$, also $(x_{n_k})_{k\in\mathbb{N}}$ converges to a limit $x \in \ker S$, because $\ker S$ is closed as the preimage of the closed set $\{0\}$ under a continuous map. Thus, every bounded sequence in $\ker S$ has a convergent subsequence, so $\ker S$ is finite-dimensional by problem 4 on exercise sheet 6.

(2) To show that $S(X)$ is closed, consider once more the induced operator $\hat{S} : X/\ker S \to S(X)$ defined by $\hat{S}(x+\ker S) := Sx$. Then, as we know, $\hat{S}$ is bijective and continuous. If we can prove that $\hat{S}^{-1}$ is also continuous, we have thus established an isomorphism between $X/\ker S$ and $S(X)$, and since the former is complete, then so is the latter, and therefore $S(X)$ will be closed. Note however that we can not use the Open Mapping Theorem for continuity of the inverse, because for this we would require $S(X)$ to be complete, which is precisely what we want to prove.

Suppose now $\hat{S}^{-1}$ were not continuous, then there would exist a sequence $(x_n)_{n\in\mathbb{N}}$ such that $Sx_n \to 0$ but $\|x_n + \ker S\| = 1$. Without loss of generality we may assume $\|x_n\| \le 2$ (otherwise replace x_n by $x_n - y_n$ for suitable $y_n \in \ker S$). Therefore, there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $(Tx_{n_k})_{k\in\mathbb{N}}$ is convergent, so also $x_{n_k} = Sx_{n_k} + Tx_{n_k}$ converges. Denote $x = \lim_{k\to\infty} x_{n_k}$. By continuity, $Sx = 0$, hence $\|x + \ker S\| = 0$, although by continuity of the norm in the quotient space we also have $\|x + \ker S\| = 1$. This is the desired contradiction, so we have shown closedness of $S(X)$.

For finite-dimensionality, we first observe that for a closed subspace $U \subset X$, we have the isomorphy $(X/U)' \simeq U^\perp$, where $U^\perp := \{x' \in X' : x'(x) = 0 \quad \forall x \in U\}$. Indeed, it is easy to check that an isomorphism between these spaces is given by $l \mapsto l \circ \omega$ for $l \in (X/U)'$, where $\omega : X \to X/U$ is the canonical quotient map given by $x \mapsto x + U$. Applying this observation to $U = S(X)$ and noting $S(X)^\perp = \ker S'$ (exercise sheet 11, problem 4), we arrive at

$$(X/S(X))' \simeq \ker S'$$

and therefore

$$\dim(X/S(X)) = \dim(X/S(X))' = \dim(\ker S') < \infty, \tag{4.2}$$

where we used Step 1 applied to S' and Schauder's Theorem, which shows that T' is compact; for the first equality we also used that a finite-dimensional space and its dual have the same dimension.

(3) One can check that, for a closed subspace $U \subset X$, we have the isomorphy $X'/U^\perp \simeq U'$ via the isomorphism $x' + U^\perp \mapsto x' \restriction_U$. Applying this with $U = \ker S$ and taking into consideration Lemma 4.15, we obtain

$$\dim(X'/S'(X')) = \dim(\ker S)' = \dim(\ker S).$$

Thanks to this and (4.2), it only remains to show $\dim(X/S(X)) = \dim(\ker S)$. We consider the decomposition $X = \hat{N} \oplus \hat{R}$ from Corollary 4.14. Set $\hat{S} := S \restriction_{\hat{N}}$.

First, we show that $X/S(X)$ is isomorphic (as a vector space) to $\hat{N}/S(\hat{N})$, so their dimensions are equal. To this end, consider the obvious map $\Phi : \hat{N}/S(\hat{N}) \to X/S(X)$, $x + \hat{S}(\hat{N}) \mapsto x + S(X)$, which is well-defined as $\hat{S}(\hat{N}) \subset S(X)$. Linearity is clear. For injectivity, let $x \in \hat{N}$ with $\Phi(x + \hat{S}(\hat{N})) = 0$, i.e., $x \in S(X)$. We need to show $x \in \hat{S}(\hat{N})$. Write $x = Sy$ for some $y \in X$, and $y = y_1 + y_2 \in \hat{N} + \hat{R}$. Then, $Sy_2 = Sy - Sy_1 = x - Sy_1 \in \hat{N}$. But note also $Sy_2 \in \hat{R}$, so that $Sy_2 = 0$. Since $S \restriction_{\hat{R}}$ is an isomorphism and $y_2 \in \hat{R}$, this implies $y_2 = 0$, so that $x = Sy_1 \in \hat{S}(\hat{N})$. For surjectivity, let $X \ni x = x_1 + x_2 \in \hat{N} \oplus \hat{R} = \hat{N} \oplus S(\hat{R})$. Then $x + S(X) = x_1 + S(X)$, so that $\Phi(x_1 + \hat{S}(\hat{N})) = x + S(X)$.

Secondly, as $\hat{N}$ is finite-dimensional, it is a well-known result from Linear Algebra that $\dim(\hat{N}/\hat{S}(\hat{N})) = \dim(\ker \hat{S})$.

Thirdly, we show $\ker \hat{S} = \ker S$. As $\hat{S}$ is the restriction of S to $\hat{N}$, it suffices to show that any $x \in X$ with $Sx = 0$ is contained in $\hat{N}$. Write such an x as $x = x_1 + x_2 \in \hat{N} \oplus \hat{R}$, then $Sx = 0$ implies $Sx_2 = -Sx_1$. As S leaves the spaces $\hat{N}$ and $\hat{R}$ invariant, this implies $Sx_2 = -Sx_1 \in \hat{N} \cap \hat{R}$, hence $Sx_1 = Sx_2 = 0$. Since $S\restriction_{\hat{R}}$ is an isomorphism, it follows that $x_2 = 0$, so that $x = x_1 \in \hat{N}$, as claimed.

Putting together these three observations, we have

$$\dim(X/S(X)) = \dim(\hat{N}/S(\hat{N})) = \dim(\ker \hat{S}) = \dim(\ker S).$$

□

4.3.2. Consequences of the Riesz-Schauder Theorem. We still assume X to be Banach and $T \in L(X, X)$ a compact operator. Recall from exercise sheet 11, problem 4, the definition

$$V_\perp := \{x \in X : \quad x'(x) = 0 \quad \forall x' \in V\}$$

for a subspace $V \subset X'$.

THEOREM 4.17 (Fredholm Alternative). *Let $0 \neq \lambda \in \mathbb{K}$. Then, either the homogeneous equation*

$$\lambda x - Tx = 0$$

has the unique solution $x = 0$, in which case also the inhomogeneous equation

$$\lambda x - Tx = y$$

has a unique solution for all $y \in X$; or there exist $n = \dim(\ker(\lambda I - T)) < \infty$ many linearly independent solutions of the homogeneous equation, and the same number of linearly independent solutions of the adjoint equation

$$\lambda x' - T'x' = 0,$$

in which case the inhomogeneous equation has a solution if and only if $y \in \ker(\lambda - T')_\perp$.

PROOF. Without loss of generality, take $\lambda = 1$ (because $\lambda I - T = \lambda\left(I - \frac{T}{\lambda}\right)$, and $\frac{T}{\lambda}$ is compact). If $x = 0$ is the only solution of $x = Tx$, then by Theorem 4.16

$$\dim(X/(I-T)(X)) = \dim(\ker(I-T)) = 0,$$

which implies that the range of $I - T$ is all of X. Therefore, $x - Tx = y$ has a solution for all $y \in X$, and uniqueness of the solution follows from injectivity of $I - T$.

If, however, $I - T$ is not injective, then by Theorem 4.16, the solution space of the homogeneous equation (i.e., the kernel of $I - T$) is finite-dimensional, and the solution space of the adjoint equation (i.e., the kernel of $I - T'$) has the same dimension. In this case, $x - Tx = y$ has a solution if and only if $y \in (I - T)(X) = \ker(I - T')_\perp$, where we used closedness of the range of $I - T$ (Theorem 4.16) and problem 4 on exercise sheet 11.

□

Let us present an application of the Fredholm Alternative in the theory of integral equations. Let $k \in C([0,1]^2)$ and

$$T : C([0,1]) \to C([0,1]), \quad Tx(s) = \int_0^s k(s,t)x(t)dt.$$

In the exercises you showed that T is compact. Consider the equation $\lambda x - Tx = 0$ with $\lambda \neq 0$, which reads as

$$\lambda x(s) = \int_0^s k(s,t)x(t)dt.$$

Without loss of generality, set $\lambda = 1$, because otherwise consider the compact operator $\frac{T}{\lambda}$. We show injectivity of $I - T$: If $x = Tx$, then

$$|x(s)| = |Tx(s)| \le \int_0^s |k(s,t)||x(t)|dt \le s\|k\|_\infty \|x\|_\infty.$$

Substituting this estimate into itself, so to speak, we obtain

$$|x(s)| = |Tx(s)| \le \int_0^s |k(s,t)|t\|k\|_\infty \|x\|_\infty dt \le \frac{s^2}{2}\|k\|_\infty^2 \|x\|_\infty,$$

and iteration yields

$$|x(s)| \le \frac{s^n \|k\|_\infty^n}{n!}\|x\|_\infty$$

for all $n \in \mathbb{N}$, which converges to zero as $n \to \infty$. Hence $x = 0$ and $I - T$ is injective. By the Fredholm Alternative, the equation $\lambda x - Tx = y$ has a unique solution for every $y \in C([0,1])$. In other words, for any continuous function y and $\lambda \ne 0$, there is a unique solution to

$$\lambda x(s) = y(s) + \int_0^s k(s,t)x(t)dt.$$

Another application of the Riesz-Schauder Theorem is the Spectral Theorem for compact operators:

THEOREM 4.18 (Spectral Theorem for compact operators). *Let $T \in L(X,X)$ be compact.*

(1) *If X has infinite dimension, then $0 \in \sigma(T)$.*
(2) *$\sigma(T)$ is at most countable.*
(3) *If $\lambda \in \sigma(T) \setminus \{0\}$, then λ is an eigenvalue with finite-dimensional eigenspace.*
(4) *$\sigma(T)$ has no accumulation point except possibly 0.*

PROOF. (1) If $0 \in \rho(T)$, then T is invertible with continuous inverse, and since the composition of a compact operator and a continuous operator is again compact (exercise), we have that $I = TT^{-1}$ is compact. This implies that X is finite-dimensional.

(3) Again, we may set $\lambda = 1$ without loss of generality. If $I - T$ is injective, then $\ker(I - T) = \{0\}$, so by the Riesz-Schauder Theorem, $X/(I-T)(X)$ has dimension zero, which implies $I - T$ is also surjective. But then $1 \in \rho(T)$. Therefore, if $1 \in \sigma(T)$, $I - T$ is not injective, meaning that 1 is an eigenvalue. Finite-dimensionality of the eigenspace $\ker(\lambda I - T)$ also follows from the Riesz-Schauder Theorem.

(2) and (4) are both implied by the following claim: For any $\epsilon > 0$, the set $\{\lambda \in \sigma(T) : |\lambda| \ge \epsilon\}$ is finite.

Let us prove this by contradiction. If the claim were not true, then we could find an $\epsilon > 0$ and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of pairwise distinct eigenvalues of T (here we use (3)) such that $|\lambda_n| \ge \epsilon$. Denote by $x_n \ne 0$ a corresponding eigenvector. Then $\{x_n : n \in \mathbb{N}\}$ is linearly independent; for otherwise there would exist $N \in \mathbb{N}$ and a linear combination

$$x_{N+1} = \sum_{j=1}^N \alpha_j x_j,$$

for linearly independent $\{x_1, \dots, x_N\}$, and therefore

$$Tx_{N+1} = \sum_{j=1}^N \alpha_j Tx_j = \sum_{j=1}^N \alpha_j \lambda_j x_j$$

but also

$$Tx_{N+1} = \lambda_{N+1}x_{N+1} = \sum_{j=1}^N \alpha_j \lambda_{N+1} x_j.$$

It would then follow from independence of $\{x_1, \dots, x_N\}$ that $\alpha_j \lambda_j = \alpha_j \lambda_{N+1}$ for all $j = 1, \dots, N$, and since not all α_j can be zero, we obtain some j such that $\lambda_j = \lambda_{N+1}$, in contradiction with the assumption that the λ's are pairwise distinct. This shows linear independence of $\{x_n : n \in \mathbb{N}\}$.

But now, if $E_n := \operatorname{span}\{x_1, \dots, x_n\}$, then

$$E_1 \subsetneq E_2 \subsetneq E_3 \subsetneq \dots.$$

By the Riesz Lemma, we may find $y_n = \sum_{j=1}^n \alpha_j^n x_j \in E_n$ with $\|y_n\| = 1$ and $\operatorname{dist}(y_n, E_{n-1}) > \frac{1}{2}$ for $n \geq 2$. If $n > m \geq 2$, therefore,

$$\|Ty_n - Ty_m\| = \|\lambda_n y_n - (Ty_m + \lambda_n y_n - Ty_n)\|. \tag{4.3}$$

Note that $Ty_m \subset E_m \subset E_{n-1}$ because y_m is a linear combination of $\{x_1, \dots, x_m\}$, which are eigenvectors of T. Note further that $\lambda_n y_n - Ty_n = \sum_{j=1}^n (\lambda_n - \lambda_j) \alpha_j^n x_j \in E_{n-1}$, so that from (4.3) we infer

$$\|Ty_n - Ty_m\| > \frac{1}{2}|\lambda_n| \geq \frac{\epsilon}{2},$$

and thus the desired contradiction with compactness of T.

□

4.4. Spectral Theory for Bounded Operators on Hilbert Spaces

We get back to the study of Hilbert spaces. In this entire section, H will denote a *complex* Hilbert space.

4.4.1. Continuous Functional Calculus. The Spectral Theorem from Linear Algebra tells us that a complex matrix $T \in \mathbb{C}^n$ that is Hermitian (i.e., $T^t = \overline{T}$) can be diagonalised, that is, there exists a unitary matrix $U \in \mathbb{C}^n$ (i.e., $U^* := \overline{U}^t = U^{-1}$) such that

$$UTU^{-1} = D,$$

where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix with *real* entries $\lambda_1, \dots, \lambda_n$, which are the eigenvalues of T. (You might have learned in Linear Algebra that the Spectral Theorem is true more generally for *normal* matrices, i.e., those such that $T^*T = TT^*$, but then the eigenvalues might no longer be real. The theory presented here for bounded linear operators on Hilbert spaces also transfers to normal operators, but we will stick to self-adjoint operators anyway.)

The Spectral Theorem, as you have seen in Ordinary Differential Equations, gives rise to the definition of the matrix exponential e^T, which can be defined as

$$e^T = U^{-1} \operatorname{diag}(e^{\lambda_1}, \dots, e^{\lambda_n}) U.$$

Of course there is nothing special here about the exponential function: We could have defined $f(T)$ for any function defined on the spectrum of T. This procedure of defining a function of a matrix is known as a *functional calculus*. Here and in the next section, we shall develop a functional calculus for bounded self-adjoint operators on a Hilbert space, and then see what this tells us about the spectrum of such an operator.

In the sequel, let us denote by 1 the constant function $t \mapsto 1$ and by t the identity function $t \mapsto t$. As before, for an operator $T \in L(H, H)$, we denote by $T^* \in L(H, H)$ its Hilbert space adjoint. The following Proposition has already been known to us in the case of *compact* self-adjoint operators, see the proof of Theorem 2.31.

Proposition 4.19 (spectral radius of self-adjoint operators). *Let $T \in L(H, H)$ be self-adjoint. Then*

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

PROOF. From Proposition 2.24, noting that T^2 is self-adjoint if T is, we have

$$\|T^2\| = \sup_{\|x\|\le 1} (x, T^2 x) = \sup_{\|x\|\le 1} (Tx, Tx) = \sup_{\|x\|\le 1} \|Tx\|^2 = \|T\|^2.$$

We infer $\|T^{2^k}\| = \|T\|^{2^k}$ for all $k \in \mathbb{N}$ and therefore

$$r(T) = \lim_{n\to\infty} \|T^n\|^{1/n} = \lim_{k\to\infty} \|T^{2^k}\|^{2^{-k}} = \|T\|.$$

The equality $r(T) = \max\{|\lambda| : \lambda \in \sigma(T)\}$ is precisely Theorem 4.12. □

THEOREM 4.20 (Continuous Functional Calculus). *Let $T \in L(H,H)$ be self-adjoint. There exists a unique map $\Phi : C(\sigma(T);\mathbb{C}) \to L(H,H)$ such that*

(1) *$\Phi(1) = I$ and $\Phi(t) = T$;*
(2) *Φ is a homomorphism of algebras and an involution, that is:*
  - *Φ is linear,*
  - *Φ is multiplicative, i.e., $\Phi(fg) = \Phi(f) \circ \Phi(g)$,*
  - *$\Phi(\overline{f}) = \Phi(f)^*$ (involution property);*

(3) *Φ is continuous.*

PROOF. Let's prove uniqueness first, as that is very simple. As Φ is linear and multiplicative, the values $\Phi(1)$ and $\Phi(t)$, which are prescribed, uniquely determine Φ on the space of polynomial functions. But the space of polynomial functions is dense in $C(\sigma(T))$ (by the Stone-Weierstraß Theorem, noting that $\sigma(T)$ is compact), and Φ is continuous, so Φ is already determined on all of $C(\sigma(T))$.

Now for existence: Obviously, if $f : t \mapsto \sum_{k=0}^n a_k t^k$ is polynomial, then Φ must return

$$\Phi(f) = \sum_{k=0}^n a_k T^k,$$

whereby (1) and (2) are satisfied on polynomial functions (for $\Phi(\overline{f}) = \Phi(f)^*$ recall that $\sigma(T) \subset \mathbb{R}$ as T is self-adjoint). If Φ, restricted to the polynomials, can be shown to be continuous, then there exists a continuous extension to all of $C(\sigma(T))$. We proceed in several steps.

Step 1. We claim: If f is a polynomial function on $\sigma(T)$, then

$$\sigma(\Phi(f)) = \{f(\lambda) : \lambda \in \sigma(T)\}. \tag{4.4}$$

To prove this, let first $\lambda \in \sigma(T)$, then we need to show $f(\lambda) \in \sigma(\Phi(f))$. As λ is a root of $f - f(\lambda)$, there exists a polynomial g such that $f(t) - f(\lambda) = (t-\lambda)g(t)$ for all $t \in \sigma(T)$. So by definition of Φ, we have $\Phi(f) - f(\lambda)I = (T - \lambda I)\Phi(g)$.

Now if $f(\lambda) \in \rho(\Phi(f))$, it would hold that

$$I = (T-\lambda I)\Phi(g)(\Phi(f) - f(\lambda)I)^{-1} = (\Phi(f) - f(\lambda)I)^{-1}\Phi(g)(T - \lambda I)$$

(because all the operators commute with each other) – but this implies $\lambda \in \rho(T)$, contradiction. This proves the inclusion $\{f(\lambda) : \lambda \in \sigma(T)\} \subset \sigma(\Phi(f))$.

Conversely, let $\mu \in \sigma(\Phi(f))$ and assume f is not constant (otherwise the claim is clearly satisfied), so there exists a factorisation

$$f - \mu = a(t-\lambda_1)\cdots(t-\lambda_n)$$

and therefore

$$\Phi(f) - \mu I = a(T - \lambda_1 I)\cdots(T - \lambda_n I).$$

Hence if $\lambda_k \in \rho(T)$ for all $k = 1,\dots,n$, then also $\mu \in \rho(\Phi(f))$, which is not the case, so that there is some $k = 1,\dots,n$ with $\lambda_k \in \sigma(T)$. But $f(\lambda_k) = \mu$, which completes the proof of the claim.

Step 2. In this step we show continuity of Φ on the space of polynomial functions. We compute

$$\begin{aligned} \|\Phi(f)\|^2 &= \|\Phi(f)^*\Phi(f)\| = \|\Phi(\overline{f}f)\| \\ &= \sup\{|\lambda| : \lambda \in \sigma(\Phi(\overline{f}f))\} \\ &= \sup\{|(\overline{f}f)(\lambda)| : \lambda \in \sigma(T)\} = \sup\{|f(\lambda)|^2 : \lambda \in \sigma(T)\}. \end{aligned} \tag{4.5}$$

Here, in the passage from the first to the second line we used Proposition 4.19; note also that $\Phi(\overline{f}f)$ is self-adjoint (check it yourself). For the passage from the second to the third line, we used Step 1.

The shows that $\|\Phi\| = \|f\|_\infty$, so indeed Φ is continuous with norm 1 on the space of polynomial functions on $\sigma(T)$.

Step 3. As the polynomials are dense in $C(\sigma(T))$, we can uniquely extend the continuous linear operator Φ to all of $C(\sigma(T))$ with the same norm[2]. It remains to show that this operator has all the required properties.

In fact, linearity and continuity are clear (see footnote), and $\Phi(1) = I$ as well as $\Phi(t) = T$. Multiplicativity and the involution property follow by a simple approximation argument; we give it only for the involution property: Let $f \in C(\sigma(T))$ and $(f_n)_{n\in\mathbb{N}}$ a sequence of polynomials that converges uniformly to f. Then,

$$\begin{aligned} \Phi(\overline{f}) &= \Phi(\overline{\lim_{n\to\infty} f_n}) = \Phi(\lim_{n\to\infty} \overline{f_n}) = \lim_{n\to\infty} \Phi(\overline{f_n}) \\ &= \lim_{n\to\infty} \Phi(f_n)^* = \Big(\lim_{n\to\infty} \Phi(f_n)\Big)^* = \Phi(\lim_{n\to\infty}(f_n))^* = \Phi(f)^*, \end{aligned}$$

where we used the continuity of $f \mapsto \overline{f}$, of $S \mapsto S^*$, and the involution property of Φ restricted to polynomials. ☐

Instead of $\Phi(f)$, we usually write $f(T)$ (just like we write e^T for the matrix exponential). Let us collect some properties of the functional calculus. For a linear operator $R : X \to X$, we write $R \geq 0$ if $(x, Rx) \geq 0$ for all $x \in H$.

Theorem 4.21 (properties of the continuous functional calculus). *Let $T \in L(H,H)$ self-adjoint and $f \in C(\sigma(T))$.*

1. *$\|f(T)\| = \|f\|_\infty$;*
2. *When $f \geq 0$, then also $f(T) \geq 0$;*
3. *If $Tx = \lambda x$, then also $f(T)x = f(\lambda)x$;*
4. *$\sigma(f(T)) = f(\sigma(T))$;*
5. *The set $\{f(T) : f \in C(\sigma(T))\}$ forms a commutative algebra of operators[3]. The operator $f(T)$ is self-adjoint if and only if f is real-valued.*

Proof. (1) was already showed in (4.5) when f is a polynomial; but since polynomials are dense in $C(\sigma(T))$ and $f \mapsto f(T)$ is continuous, the equality holds for general $f \in C(\sigma(T))$.

(2) If $C(\sigma(T)) \ni f \geq 0$, then there is $0 \leq g \in C(\sigma(T))$ with $g^2 = f$. But then, for all $x \in H$,

$$(x, f(T)x) = (g(T)^*x, g(T)x) = (\overline{g}(T)x, g(T)x) = (g(T)x, g(T)x) = \|g(T)x\|^2 \geq 0.$$

(3) If $f = t^n$, then clearly $Tx = \lambda x$ implies $f(T)x = T^n x = \lambda^n x = f(\lambda)x$, and by linearity this remains true for all polynomial functions. The general statement then follows again from density of polynomials in $\sigma(T)$.

[2]Indeed, if $f \in C(\sigma(T))$, pick a sequence of polynomials such that $f_n \to f$ and define $\Phi(f) := \lim_{n\to\infty} \Phi(f_n)$. It is not difficult to show that the limit exists and is independent of the choice of approximating polynomials, and that Φ thus defined is itself linear and continuous.

[3]This means that the set forms a vector space and additionally is closed under multiplication (here: composition of operators), which is commutative.

(4) For the inclusion $\sigma(f(T)) \subset f(\sigma(T))$, suppose $\mu \notin f(\sigma(T))$, then we need to show $\mu \in \rho(f(T))$. By choice of μ, the function $g := (f-\mu)^{-1} \in C(\sigma(T))$ is well-defined with

$$g(f-\mu) = (f-\mu)g = 1.$$

Applying the functional calculus,

$$g(T)(f(T)-\mu I) = (f(T)-\mu I)g(T) = I,$$

hence $f(T)-\mu I$ is invertible and thus $\mu \in \rho(f(T))$ as claimed.

For the converse inclusion, note that $\sigma(f(T)) = f(\sigma(T))$ when f is a polynomial, see (4.4). Let $\mu = f(\lambda)$ for some $\lambda \in \sigma(T)$, then we need to show $\mu \in \sigma(f(T))$. For $n \in \mathbb{N}$, let g_n be a polynomial function such that $\|f-g_n\|_\infty < \frac{1}{n}$. Then, in particular, $|f(\lambda)-g_n(\lambda)| < \frac{1}{n}$ and also, as the functional calculus is continuous of norm one (see (4.5)), $\|f(T)-g_n(T)\| < \frac{1}{n}$.

Therefore, $\mu I - f(T)$ is the limit (in the operator norm) of the sequence $(g_n(\lambda)I - g_n(T))$, which consists of non-invertible operators (because $\sigma(g_n(T)) = g_n(\sigma(T))$ as g_n are polynomials). Now the set of non-invertible operators is closed in $L(H,H)$ (think it over – hint: Neumann Series), so that also $\mu I - f(T)$ is non-invertible and thus $\mu \in \sigma(f(T))$ as claimed.

(5) The property of being an algebra follows immediately from linearity and multiplicativity of the functional calculus. Commutativity is clear for polynomials (because powers of T commute with each other), and for the general case one proceeds by approximation.

If f is real-valued and polynomial, then clearly $f(T)$ is self-adjoint because T is. The non-polynomial case follows again by approximation. On the other hand, if f is not real-valued, then by (4) $\sigma(f(T))$ contains a number that is not real; but self-adjoint operators have real spectrum (Lemma 2.30), so $f(T)$ cannot be self-adjoint. □

Let us describe the link between the continuous functional calculus and the spectrum in the case of a *compact* self-adjoint operator $T \in L(H,H)$, for which we already have a spectral theorem (Theorem 2.31) at our disposal. We saw that we can write

$$Tx = \sum_{n=1}^{\infty} \lambda_n (x, e_n) e_n \tag{4.6}$$

for every $x \in H$, where $(e_n)_{n\in\mathbb{N}}$ is a certain orthonormal system and $\lambda_n \neq 0$ a corresponding eigenvalue. We can extend $(e_n)_{n\in\mathbb{N}}$ to an orthonormal basis by adding an orthonormal basis of $\ker T$, if necessary. Denote $P_n : x \mapsto (x, e_n)e_n$, which is the orthogonal projection onto the span of e_n, and P_0 the orthogonal projection onto $\ker T$, so that $T = \sum_{n=0}^{\infty} \lambda_n P_n$ (set $\lambda_0 = 0$).

For $f \in C(\sigma(T))$, then, define

$$f(T) = \sum_{n=0}^{\infty} f(\lambda_n) P_n.$$

We claim $T \mapsto f(T) = \Phi(f)$ is the functional calculus. For this, it suffices to check the properties of Theorem 4.20, as these uniquely determine Φ. First, from (4.6) it is clear that $1(T) = I$ and $t(T) = T$. Secondly, $f \mapsto f(T)$ is linear, but also multiplicative, as

$$\begin{aligned}(fg)(T)x &= f(0)g(0)P_0x + \sum_{n=1}^{\infty} f(\lambda_n)g(\lambda_n)(x,e_n)e_n \\ &= f(0)P_0 g(T)x + \sum_{n=1}^{\infty} f(\lambda_n)\left(g(T)x, e_n\right) e_n = f(T)g(T)x.\end{aligned}$$

It is also involutive, because

$$\begin{aligned}(\overline{f}(T)x,y) &= \left(\overline{f(0)}P_0x + \sum_{n=1}^{\infty}\overline{f(\lambda_n)}(x,e_n)e_n, P_0y + \sum_{k=1}^{\infty}(y,e_k)e_k\right)\\ &= \overline{f(0)}(P_0x,P_0y) + \sum_{n=1}^{\infty}\overline{f(\lambda_n)}(x,e_n)\overline{(y,e_n)}\\ &= (P_0x, f(0)P_0y) + \sum_{n=1}^{\infty}(x,e_n)\overline{f(\lambda_n)(y,e_n)} = (x, f(T)y).\end{aligned}$$

Finally, it is continuous as a map $C(\sigma(T)) \to L(H,H)$, because

$$\begin{aligned}\|f(T)x\|^2 &= |f(0)|^2\|P_0x\|^2 + \sum_{n=1}^{\infty}|f(\lambda_n)|^2|(x,e_n)|^2\\ &\le \|f\|_\infty^2\left(\|P_0x\|^2 + \sum_{n=1}^{\infty}|(x,e_n)|^2\right) = \|f\|_\infty^2\|x\|^2.\end{aligned}$$

This shows, for compact operators, the relation between the functional calculus and spectral theory. As we have seen, given the spectral theorem, we could construct the functional calculus. But it would also work the other way round: Given $f \mapsto f(T)$, we can reconstruct the projections P_n onto the eigenspaces via

$$P_n = f_n(T), \quad f_n(\lambda_j) = \delta_{nj} \quad (n \ge 1, j \ge 0) \tag{4.7}$$

where we set $\lambda_0 = 0$ and thus define a continuous (!) function f_n on the set $\sigma(T) = \{\lambda_n : n \in \mathbb{N}\} \cup \{0\}$.

For bounded (but not necessarily compact) operators, we do not yet have a spectral theorem at our disposal, but the functional calculus is available from Theorem 4.20. We will therefore follow the strategy just outlined to gain insight into the eigenspaces from the functional calculus. A technical problem is that, for bounded operators, the spectrum need no longer be discrete[4], so the functions as in (4.7) would not be continuous on $\sigma(T)$. This means we have to extend our calculus from continuous to *measurable* functions.

4.4.2. Measurable Functional Calculus. For a compact set $K \subset \mathbb{C}$, let $\mathcal{L}^\infty(K)$ be the Banach space (!) of bounded measurable functions[5] $M \to \mathbb{C}$, together with the supremum norm.

LEMMA 4.22. *$\mathcal{L}^\infty(K)$ is the smallest function space containing $C(K)$ that is closed under pointwise limits of uniformly bounded functions.*

More precisely: Assume $C(K) \subset U \subset \mathcal{L}^\infty(K)$ has the property that every $(f_n)_{n\in\mathbb{N}} \subset U$ with $\sup_{n\in\mathbb{N}}\|f_n\|_\infty < \infty$ and $f_n \to f$ pointwise satisfies $f \in U$. Then $U = \mathcal{L}^\infty(K)$.

PROOF. Let V be the intersection over all sets S with the properties described, i.e., $S \supset C(K)$ is closed with respect to pointwise limits of uniformly bounded functions (such sets exists because $\mathcal{L}^\infty$ is one of them). By definition, $V \supset C(K)$.

Let us show first that V is a vector space. So let $f \in C(K)$ and set $V_f := \{g \in \mathcal{L}^\infty(K) : f + g \in V\}$. It is easy to see $V_f \supset C(K)$ and V_f is closed with respect to pointwise limits of uniformly bounded functions. This implies $V_f \supset V$, or in other words: If $f \in C(K)$ and $g \in V$, then $f + g \in V$.

Let now $h \in V$ and consider $V_h := \{g \in \mathcal{L}^\infty(K) : h + g \in V\}$. By the argument we just put forward, $V_h \supset C(K)$. Also, V_h is closed with respect to pointwise limits of uniformly bounded functions. It follows that $V \subset V_h$, whereby we have showed: If $h \in V$ and $g \in V$, then $h + g \in V$ (as $g \in V$ implies $g \in V_h$).

[4]This is why the sum in (4.6) will have to be replaced by an integral.

[5]Note this is *not* the space $L^\infty(K)$, because the latter consists of (equivalence classes of) functions that are bounded only up to a nullset.

If $\alpha \in \mathbb{C}$ and $g \in V$ then a similar argument (where one would show that $V_\alpha := \{g \in \mathcal{L}^\infty : \alpha g \in V\}$ contains $C(K)$ and is closed with respect to pointwise limits of uniformly bounded functions) yields $\alpha g \in V$, so we have showed that V is a vector space. Also, V is closed with respect to the supremum norm, because if $g_n \to g$ uniformly and $g_n \in V$ for all $n \in \mathbb{N}$, then g_n are uniformly bounded and pointwise convergent, so g is contained in any space closed with respect to pointwise limits of uniformly bounded functions, and thus in V.

Next we show that step functions[6] are contained in V. Since (as you might remember from measure theory) step functions are dense in $\mathcal{L}^\infty$ with respect to uniform convergence, it will then already follow $V = \mathcal{L}^\infty(K)$.

In fact, since V is a vector space, it suffices to show $\chi_E \in V$ for any measurable $E \subset K$. We do this by a standard measure theory argument: We show that $\Delta := \{E \in \Sigma : \chi_E \in V\}$ is a Dynkin system containing a $\cap$-stable generator of Σ, where Σ denotes the Borel σ-algebra of K. Then $\Delta = \Sigma$ by a well-known theorem in measure theory.

As a $\cap$-stable generator of Σ, consider the open sets in K. They are contained in Δ: Indeed, when $E \subset K$ is open, then there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset C(K)$ with $0 \le f_n \le 1$ for all $n \in \mathbb{N}$ and $f_n \to \chi_E$ pointwise. So $\chi_E \in V$.

It remains to prove Δ is Dynkin. There are two defining properties of Dynkin spaces: First, if $E, F \in \Delta$ and $F \subset E$, then also $E \setminus F \in \Delta$. This is true because $\chi_{E\setminus F} = \chi_E - \chi_F$ and V is a vector space. Secondly, if $(E_n)_{n\in\mathbb{N}} \subset \Delta$ are pairwise disjoint, then $E := \bigcup_{n\in\mathbb{N}} E_n$ is also in Δ. This is also true, as $\chi_E = \lim_{N\to\infty} \sum_{n=1}^N \chi_{E_n} \in V$ as V is closed under pointwise convergence of uniformly bounded functions.

□

As a preparation for the construction of the measurable functional calculus, let me state without proof the *Riesz Representation Theorem* (not to be confused with the Riesz-Fréchet Representation Theorem): If $K \subset \mathbb{C}$ is compact and $l \in C(K)'$ is a bounded linear functional, then there exists a complex-valued measure $\mu : \Sigma \to \mathbb{C}$, where Σ is again the Borel σ-algebra on M, such that

$$l(f) = \int_K f(z)d\mu(z) \quad \forall f \in C(K).$$

Moreover, $\|l\| = \|\mu\|$, where $\|\mu\|$ is the *total variation norm* of μ explained below.

The notion of complex-valued measure requires some explanation, as you probably have only seen measures with non-negative real values so far. A map $\Sigma \to \mathbb{C}$ is called a (complex) measure if it is σ-additive, that is, if

$$\mu\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \sum_{n=1}^\infty \mu(E_n) \in \mathbb{C} \quad \text{whenever } E_n \in \Sigma \text{ are pairwise disjoint.}$$

We will not go into the theory of complex measures here, but it can be shown that a complex measure μ can be decomposed as $\mu = \mu_{\Re}^+ - \mu_{\Re}^- + i(\mu_{\Im}^+ - \mu_{\Im}^-)$, where all four measures are finite measures with values in $\mathbb{R}_0^+$, and then the integral of $f \in C(K;\mathbb{C})$ with respect to μ is defined as

$$\begin{aligned}\int_K f(z)d\mu(z) = &\int_K f(z)d\mu_{\Re}^+(z) - \int_K f(z)d\mu_{\Re}^-(z) \\ &+ i\left(\int_K f(z)d\mu_{\Im}^+(z) - \int_K f(z)d\mu_{\Im}^-(z)\right).\end{aligned}$$

Accordingly, many properties of usual measures transfer to complex measures, like the Dominated Convergence Theorem.

[6]Recall from measure theory that step functions take the form $\sum_{k=1}^n \alpha_k \chi_{E_k}$, where $\alpha_k \in \mathbb{C}$ and E_k are measurable, i.e., they are contained in the Borel σ-algebra of K. The function χ_{E_k} is then the indicator function of E_k.

The vector space (!) of complex measures on K can be equipped with a norm, the *total variation norm*, given as

$$\|\mu\|^2 := (\mu_{\mathfrak{R}}^+(K) + \mu_{\mathfrak{R}}^-(K))^2 + (\mu_{\mathfrak{J}}^+(K) + \mu_{\mathfrak{J}}^-(K))^2 .$$

One then has the estimate

$$\left| \int_K f(z) d\mu(z) \right| \le \|f\|_\infty \|\mu\| \quad \forall f \in \mathcal{L}^\infty(K).$$

THEOREM 4.23 (Measurable Functional Calculus). *Let $T \in L(H,H)$ be self-adjoint. There exists a unique map $\Phi : \mathcal{L}^\infty(\sigma(T)) \to L(H,H)$ such that*

1. *$\Phi(1) = I$ and $\Phi(t) = T$;*
2. *Φ is a homomorphism of algebras and an involution;*
3. *Φ is continuous.*
4. *If $f_n \in \mathcal{L}^\infty(\sigma(T))$ for all $n \in \mathbb{N}$, $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$ and $f_n \to f$ pointwise, then also $(\Phi(f_n)x, y) \to (\Phi(f)x, y)$ for all $x, y \in H$.*

PROOF. Again let us prove uniqueness first. We have seen that (1), (2), and (3) already determine Φ on $C(\sigma(T))$. By Lemma 4.22, we can approximate each $f \in \mathcal{L}^\infty$ as pointwise limit of uniformly bounded continuous functions (think it over). Property (4) then determines $\Phi(f)$ from its values along the approximation sequence.

For existence, let $f \in \mathcal{L}^\infty(\sigma(T))$ and $x, y \in H$. Consider $l_{x,y} : C(\sigma(T)) \to \mathbb{C}$ given as

$$l_{x,y}(g) = (g(T)x, y),$$

which is linear and bounded, because

$$|l_{x,y}(g)| \le \|g(T)\| \|x\| \|y\| = \|g\|_\infty \|x\| \|y\|. \tag{4.8}$$

By the Riesz Representation Theorem, there is a complex measure $\mu_{x,y}$ with $\|\mu_{x,y}\| \le \|x\| \|y\|$ such that

$$l_{x,y}(g) = \int_{\sigma(T)} g(z) d\mu_{x,y}(z) \quad \forall g \in C(\sigma(T)).$$

Consider now the sesquilinear map

$$(x, y) \mapsto \int_{\sigma(T)} f(z) d\mu_{x,y}(z),$$

which is bounded since

$$\left| \int_{\sigma(T)} f(z) d\mu_{x,y}(z) \right| \le \|f\|_\infty \|\mu_{x,y}\| \le \|f\|_\infty \|x\| \|y\|.$$

Owing to the Lax-Milgram Theorem (exercise sheet 13, problem 4) there is $\Phi(f) \in L(H,H)$ such that

$$(\Phi(f)x, y) = \int_{\sigma(T)} f(z) d\mu_{x,y}(z) \quad \forall x, y \in H.$$

Note, by definition, that Φ coincides with the continuous functional calculus on $C(\sigma(T))$, so that (1) is already clear, and so is (3).

For (4), we use the Dominated Convergence Theorem for the complex measure $\mu_{x,y}$ to find

$$(\Phi(f_n)x, y) = \int_{\sigma(T)} f_n(z) d\mu_{x,y}(z) \to \int_{\sigma(T)} f(z) d\mu_{x,y}(z) = (\Phi(f)x, y).$$

It remains to prove (2). We only show multiplicativity, using Lemma 4.22. Note first that multiplicativity holds on $C(\sigma(T))$, because there Φ is simply the continuous functional calculus.

Fix now $g \in C(\sigma(T))$ and define

$$U = \{f \in \mathcal{L}^\infty(\sigma(T)) : \Phi(fg) = \Phi(f) \circ \Phi(g)\}.$$

Then $C(\sigma(T)) \subset U$. To invoke Lemma 4.22, let $(f_n)_{n\in\mathbb{N}} \subset U$ be uniformly bounded and pointwise convergent to $f \in \mathcal{L}^\infty(\sigma(T))$. By property (4), we have

$$(\Phi(f_n)(\Phi(g)x), y) \to (\Phi(f)(\Phi(g)x), y) \tag{4.9}$$

as $n \to \infty$, but also, as $f_n \in U$,

$$(\Phi(f_n)(\Phi(g)x), y) = (\Phi(f_n g)x, y) \to (\Phi(fg)x, y)\,. \tag{4.10}$$

It follows that $\Phi(f) \circ \Phi(g) = \Phi(fg)$, hence $f \in U$. Lemma 4.22 then yields $U = \mathcal{L}^\infty(\sigma(T))$.

Now let $f \in \mathcal{L}^\infty(\sigma(T))$ and set

$$V = \{g \in \mathcal{L}^\infty(\sigma(T)) : \Phi(fg) = \Phi(f) \circ \Phi(g)\}.$$

By the previous argument, $V \supset C(\sigma(T))$, and with the same computations as in (4.9), (4.10) one finds by Lemma 4.22 that $V = \mathcal{L}^\infty(\sigma(T))$, whereby Φ is multiplicative, as claimed. □

Again we shall write $f(T)$ instead of $\Phi(f)$. Note that $f(T)$ and $g(T)$ commute, since $f(T)\circ g(T) = (fg)(T) = (gf(T)) = g(T)\circ f(T)$. If f is real-valued, then $f(T)$ is self-adjoint by virtue of the involution property.

COROLLARY 4.24. *If $f_n \in \mathcal{L}^\infty(\sigma(T))$ for all $n \in \mathbb{N}$, $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$ and $f_n \to f$ pointwise, then also $f_n(T)x \to f(T)x$ for all $x \in H$.*

PROOF. By part (4) of Theorem 4.23, under the given assumptions we have *weak* convergence of $f_n(T)x$ to $f(T)x$ for every $x \in H$. As weak convergence and convergence of the norms imply strong convergence (why?), it suffices to show $\|f_n(T)x\| \to \|f(T)x\|$. But this is true since

$$\begin{aligned}\|f_n(T)x\|^2 &= (f_n(T)x, f_n(T)x) = (f_n(T)^* f_n(T)x, x) = ((\overline{f_n} f_n)(T)x, x)\\ &\to ((\overline{f} f)(T)x, x) = \|f(T)x\|^2,\end{aligned}$$

where we applied property (4) of Theorem 4.23 to $\overline{f_n} f_n$. □

4.4.3. Spectral Measures. Within our measurable functional calculus, functions of the form $f = \chi_A \in \mathcal{L}^\infty(\sigma(T))$ with measurable $A \subset \sigma(T)$ are of particular interest. As before, we denote by $\Sigma \in \mathcal{P}(\sigma(T))$ the Borel σ-Algebra generated by the (relatively) open subsets of $\sigma(T)$. (Note Σ can equivalently be characterised as the family of sets of the form $B \cap \sigma(T)$, where B is a measurable subset of $\mathbb{R}$.)

LEMMA 4.25. *Let $A \in \Sigma$ and $E_A := \chi_A(T)$. Then E_A is an orthogonal projection, that is, $E_A^2 = E_A$ and $\ker(E_A) \perp E_A(H)$.*

PROOF. E_A is a projection because $E_A^2 = \chi_A(T)^2 = \chi_A^2(T) = \chi_A(T) = E_A$. Also, as χ_A is real-valued, E_A is self-adjoint, so for $x \in \ker E_A$ and $y \in H$ we have $(x, E_A y) = (E_A x, y) = 0$. □

LEMMA 4.26. *Let $T \in L(H,H)$ be self-adjoint. Then,*

1. $\chi_\varnothing(T) = 0$, $\chi_{\sigma(T)}(T) = I$;
2. *for pairwise disjoint $A_1, A_2, \ldots \in \Sigma$ and $x \in H$,*
$$\sum_{k=1}^\infty \chi_{A_k}(T)x = \chi_{\bigcup_{k=1}^\infty A_k}(T)x;$$
3. $\chi_A(T)\chi_B(T) = \chi_{A\cap B}(T)$ *for $A, B \in \Sigma$.*

PROOF. (1) The function $\chi_\varnothing$ is nothing but the zero function, so $\chi_\varnothing(T)$ is the zero operator. Likewise, $\chi_{\sigma(T)}$ is constantly 1 on $\sigma(T)$, so $\chi_{\sigma(T)}(T) = I$.

(2) For each $N \in \mathbb{N}$, note that $\chi_{\bigcup_{k=1}^N A_k} = \sum_{k=1}^N \chi_{A_k}$, as the sets A_k are pairwise disjoint. Thus, for finite N the assertion follows from linearity of the functional calculus. In the limit $N \to \infty$, we use Corollary 4.24 to conclude.

(3) This follows from $\chi_A \chi_B = \chi_{A\cap B}$ and multiplicativity of the functional calculus. □

DEFINITION 4.27 (spectral measure). Let Σ be the Borel σ-algebra on $\mathbb{R}$. A map $E : \Sigma \to L(H,H)$, $A \mapsto E_A$ is called *spectral measure* if each E_A is an orthogonal projection, and

(1) $E_\varnothing = 0$, $E_\mathbb{R} = I$;
(2) for pairwise disjoint $A_1, A_2, \ldots \in \Sigma$ and $x \in H$,

$$\sum_{k=1}^{\infty} E_{A_k} x = E_{\bigcup_{k=1}^\infty A_k} x.$$

We say that a spectral measure E is compactly supported if there is a compact $K \subset \mathbb{R}$ with $E_K = I$.

The previous two lemmata tell us that, for a self-adjoint bounded operator T, $E_A = \chi_A(T)$ defines a spectral measure compactly supported on $\sigma(T)$. In general, a spectral measure is really a measure (because it assigns the value zero to the empty set and is σ-additive), only with values in $L(H,H)$. One therefore often talks about *operator-valued measures* in this context.

If E is a spectral measure, then E_A is self-adjoint for all $A \in \Sigma$, because orthogonal projections are always self-adjoint. In case a spectral measure arises from the functional calculus of a bounded self-adjoint operator T, then $E_A E_B = E_B E_A = E_{A\cap B}$, because

$$E_A E_B = \chi_A(T)\chi_B(T) = (\chi_A \chi_B)(T) = \chi_{A\cap B}(T) = E_{A\cap B} = E_B E_A.$$

One can show that this is true in general for spectral measures:

PROPOSITION 4.28. *Let E be a spectral measure. Then for $A, B \in \Sigma$,*

$$E_A E_B = E_B E_A = E_{A\cap B}.$$

In particular, the values of a spectral measure always commute with each other.

PROOF. Let us assume for the moment that A and B are disjoint. Then, by additivity,

$$\begin{aligned} E_A + E_B &= E_{A\cup B} = E_{A\cup B}^2 = (E_A + E_B)^2 \\ &= E_A^2 + E_A E_B + E_B E_A + E_B^2 = E_A + E_A E_B + E_B E_A + E_B, \end{aligned}$$

whence $E_A E_B = -E_B E_A$. Using this,

$$E_A E_B = E_A E_B^2 = -E_B E_A E_B = E_B^2 E_A = E_B E_A,$$

so that $E_A E_B = \pm E_B E_A$, which is only possible when $E_A E_B = E_B E_A = 0 = E_\varnothing = E_{A\cap B}$.

In the general situation, we can use this to compute

$$\begin{aligned} E_A E_B &= (E_{A\cap B} + E_{A\setminus B})(E_{A\cap B} + E_{B\setminus A}) \\ &= E_{A\cap B}^2 + E_{A\cap B} E_{B\setminus A} + E_{A\setminus B} E_{A\cap B} + E_{A\setminus B} E_{B\setminus A} = E_{A\cap B}, \end{aligned}$$

and of course also $E_B E_A = E_{A\cap B}$ by interchanging A and B. □

The purpose of a measure's life is to be integrated against (discuss). This is no different for spectral measures. Let us outline how one can integrate a function $f \in \mathcal{L}^\infty(K)$ with respect to a spectral measure supported on K.

First of all, if $f = \sum_{k=1}^n \alpha_k \chi_{A_k}$ is an elementary function (where $\alpha_k \in \mathbb{C}$ for $k = 1, \dots, n$ and $(A_k)_{k=1,\dots,n} \subset \Sigma$ are pairwise disjoint), then of course we set

$$\int_K f(\lambda) dE_\lambda := \sum_{k=1}^n \alpha_k E_{A_k}.$$

It is easy to check (just as in measure theory) that the value of the integral does not depend on the specific way the elementary function is written.

Next, from measure theory it is known that for general $f \in \mathcal{L}^\infty(K)$ there exists a sequence of elementary functions f_n that converge to f uniformly, so we set

$$\int_K f(\lambda) dE_\lambda := \lim_{n\to\infty} \int_K f_n(\lambda) dE_\lambda.$$

Again one can show without great difficulties that the limit exists and is independent of the choice of approximating sequence $(f_n)_{n\in\mathbb{N}}$ (see [**5**, p. 323 f.]). Note however that the value of the integral is an element of $L(H, H)$, so the limit is taken in the operator norm.

Sometimes we write $\int f dE$ instead of $\int_K f(\lambda) dE_\lambda$. We condense the discussion into the following theorem:

THEOREM 4.29 (integration w.r.t. a spectral measure). *Let a spectral measure E be supported on a compact set $K \subset \mathbb{R}$. There exists a linear and continuous map*

$$\mathcal{L}^\infty(K) \to L(H, H), \quad f \mapsto \int_K f(\lambda) dE_\lambda,$$

and $\left\| \int f dE \right\| \le \|f\|_\infty$. If f is real-valued, then $\int f dE$ is self-adjoint.

PROOF. The only thing left to prove is the estimate $\left\| \int f dE \right\| \le \|f\|_\infty$. (The self-adjoint property is clear for elementary functions and then follows by approximation in the general case.) It suffices to prove the estimate for elementary functions, as the general case follows by approximation. First note that $E_A x \perp E_B x$ if $A \cap B = \emptyset$, because $(E_A x, E_B x) = (x, E_A E_B x) = (x, E_\emptyset x) = 0$.

So let $f = \sum_{k=1}^n \alpha_k \chi_{A_k}$ with measurable and pairwise disjoint $A_k \subset K$ and $x \in H$, then we use Pythagoras' Theorem to compute

$$\begin{aligned} \left\| \int f dE x \right\|^2 &= \left\| \sum_{k=1}^n \alpha_k E_{A_k} x \right\|^2 = \sum_{k=1}^n \|\alpha_k E_{A_k} x\|^2 = \sum_{k=1}^n |\alpha_k|^2 \|E_{A_k} x\|^2 \\ &\le \sup_{k=1,\dots,n} |\alpha_k|^2 \sum_{k=1}^n \|E_{A_k} x\|^2 = \|f\|_\infty^2 \left\| \sum_{k=1}^n E_{A_k} x \right\|^2 = \|f\|_\infty^2 \left\| E_{\bigcup_{k=1}^n A_k} x \right\|^2 \\ &\le \|f\|_\infty^2 \|x\|^2, \end{aligned}$$

where in the last step we used that $E_{\bigcup_{k=1}^n A_k}$ is an orthogonal projection, so its norm is one. □

4.4.4. The Spectral Theorem for Bounded Self-Adjoint Operators.

THEOREM 4.30. *Let E be a spectral measure of compact support and $T := \int_\mathbb{R} \lambda dE_\lambda$. Then the measurable functional calculus of T is given by*

$$f \mapsto f(T) := \int_{\sigma(T)} f dE.$$

PROOF. Note that T is self-adjoint as the identity function t is real. If $f \in \mathcal{L}^\infty(\sigma(T))$, then extend f to all of $\mathbb{R}$ by zero. Then, the operator

$$\Psi : \mathcal{L}^\infty(\sigma(T)) \to L(H, H), \quad \Psi(f) := \int_\mathbb{R} f dE$$

is linear and continuous by Theorem 4.29.

Multiplicativity is true for indicator functions, because $E_A E_B = E_{A\cap B}$ and therefore

$$\int_{\mathbb{R}} \chi_A \chi_B dE = \int_{\mathbb{R}} \chi_{A\cap B} dE = E_{A\cap B} = E_A E_B = \Big(\int_{\mathbb{R}} \chi_A dE\Big)\Big(\int_{\mathbb{R}} \chi_B dE\Big).$$

It then extends to $\mathcal{L}^\infty(\sigma(T))$ by approximation. Similarly for the involution property: If $f = \alpha\chi_A$ with $\alpha \in \mathbb{C}$, then for $x, y \in H$,

$$(x, \overline{f}(T)y) = (x, \overline{\alpha}E_A y) = (\alpha E_A x, y) = (f(T)x, y),$$

and the general statement follows by approximation.

Property (4) from Theorem 4.23 is seen as follows: If Σ is again the Borel σ-algebra on $\sigma(T)$, then for $x, y \in H$ the map $\mu_{x,y} : \Sigma \to \mathbb{C}$, $A \mapsto (E_A x, y)$ is a complex measure and

$$(\Psi(f)x, y) = \int_{\mathbb{R}} f d\mu_{x,y}$$

(prove this for step functions and approximate). So when $f_n \to f$ pointwise where $(f_n)_{n\in\mathbb{N}} \subset \mathcal{L}^\infty(\sigma(T))$ is uniformly bounded, then by Dominated Convergence

$$(\Psi(f_n)x, y) = \int_{\mathbb{R}} f_n d\mu_{x,y} \to \int_{\mathbb{R}} f d\mu_{x,y} = (\Psi(f)x, y),$$

which establishes (4).

It remains to prove $\Psi(1) = I$ and $\Psi(t) = T$. Since the extension by zero of 1 to all of $\mathbb{R}$ is $\chi_{\sigma(T)}$, then $\Psi(1) = \int_{\sigma(T)} 1 dE = E_{\sigma(T)}$, so we will prove $E_{\sigma(T)} = I$. From this it will also follow that E is supported on $\sigma(T)$ and thus $\Psi(t) = \int_{\mathbb{R}} \lambda dE_\lambda = T$ by definition of T.

In order to show $E_{\sigma(T)} = I$, let $(a, b] \subset \mathbb{R}$ be an interval that contains the (compact!) support of E, so that $E_{(a,b]} = I$. Let $\mu \in \rho(T)$. We claim that $E_U = 0$ for some neighbourhood $U \ni \mu$. Indeed, since $\mu I - T$ is invertible, an easy argument involving Neumann Series shows that there exists $\delta > 0$ such that $\|S - (\mu I - T)\| \le \delta$ implies S is invertible with

$$\|S^{-1}\| \le C := \|(\mu I - T)^{-1}\| + 1.$$

Without loss of generality, set $\delta := \frac{b-a}{N}$ for sufficiently large $N \in \mathbb{N}$ and assume $\delta < \frac{1}{C}$. Let $a_k := a + k\delta$ define the corresponding equidistant partition of $(a, b]$, where $k = 0, \dots, N$, and consider the elementary function

$$f = \sum_{k=1}^{N} a_k \chi_{(a_{k-1}, a_k]},$$

which uniformly approximates the identity t on $(a, b]$. The estimate $\left\|\int g dE\right\| \le \|g\|_\infty$ from Theorem (4.29) implies

$$\left\|T - \int f dE\right\| = \left\|\int_{(a,b]} (\lambda - f) dE_\lambda\right\| \le \|t - f\|_\infty = \delta. \tag{4.11}$$

Set $E_k := E_{(a_{k-1}, a_k]}$ for $k = 1, \dots, N$ and note $\int_{(a,b]} f dE = \sum_{k=1}^N a_k E_k$ as well as $\sum_{k=1}^N E_k = I$ thanks to σ-additivity. Estimate (4.11) thus implies

$$\left\|(\mu I - T) - \sum_{k=1}^{N} (\mu - a_k) E_k\right\| \le \delta.$$

By choice of δ, the operator $\sum_{k=1}^N (\mu - a_k) E_k$ is invertible, and the norm of its inverse is at most C. On the other hand, the norm of the inverse is seen to be at least[7]

$$\sup\{|\mu - a_k|^{-1} : E_k \neq 0\},$$

[7] Indeed, if $E_k \neq 0$, pick a unit vector $\tilde{y} \in E_k(H)$ and set $y := \frac{\overline{\mu - a_k}}{|\mu - a_k|^2}\tilde{y}$, then $\|y\| = |\mu - a_k|^{-1}$ and $\|\sum_{k=1}^N (\mu - a_k) E_k y\| = \|\tilde{y}\| = 1$. Note the interval $(a, b]$ and the fineness δ can always be chosen such that $\mu - a_k \neq 0$ for all $k = 1, \dots, N$.

implying that $E_k = 0$ whenever $|\mu - a_k| < \frac{1}{C}$. Since $\delta < \frac{1}{C}$, there exists at least one k such that $|\mu - a_k| < \frac{1}{C}$, $\mu \in (a_{k-1}, a_k)$, and the corresponding $E_k = E_{(a_{k-1},a_k]}$ is zero. This shows (with $U = (a_{k-1}, a_k)$) that there exists a neighbourhood $U \ni \mu$ such that $E_U = 0$.

Finally, let $K \subset \rho(T)$ be compact. As we just showed, for each $\mu \in K$, there is a neighbourhood $U_\mu \ni \mu$ such that $E_{U_\mu} = 0$. As K is compact, there exist finitely many μ_j ($j = 1, \dots n$) such that $K \subset \bigcup_{j=1}^n U_{\mu_j}$, and by σ-additivity, $E_K = 0$. For $x \in H$, the finite measure $A \mapsto (E_A x, x)$ is non-negative (why?) and therefore inner regular (cf. [**5**, Satz I.2.14]), meaning that the measure of any open set is the supremum of the measures of compact sets contained in the given open set. In particular, we have $(E_{\rho(T)}x, x) = 0$ and therefore $E_{\sigma(T)} = I$.

□

THEOREM 4.31 (Spectral Theorem). *Let $T \in L(H, H)$ be self-adjoint. Then there exists a unique spectral measure E supported on $\sigma(T)$ such that*

$$T = \int_{\sigma(T)} \lambda dE_\lambda. \tag{4.12}$$

The map

$$\mathcal{L}^\infty(\sigma(T)) \to L(H, H), \quad f \mapsto f(T) = \int f dE \tag{4.13}$$

is precisely the measurable functional calculus of T.

PROOF. Let E be the spectral measure associated with T through the functional calculus, i.e., $E_A = \chi_A(T)$. Then E is supported on $\sigma(T)$, because $\chi_{\sigma(T)}(T) = I$. We shall show (4.12).

To this end, set $S := \int_{\sigma(T)} \lambda dE_\lambda$, then we need to show $S = T$. For $\epsilon > 0$, let $f_\epsilon = \sum_{k=1}^n \alpha_k \chi_{A_k}$ be an elementary function such that $\|t - f_\epsilon\|_\infty < \epsilon$ on $\sigma(T)$, then

$$\|T - S\| \le \|T - f_\epsilon(T)\| + \|f_\epsilon(T) - f_\epsilon(S)\| + \|f_\epsilon(S) - S\|.$$

We treat the three terms individually:

- $\|T - f_\epsilon(T)\| \le \|t - f_\epsilon\|_\infty < \epsilon$, where we used that the functional calculus has norm one;
- $\|S - f_\epsilon(S)\| = \left\|\int_{\sigma(T)} (\lambda - f_\epsilon(\lambda)) dE_\lambda\right\| \le \|t - f_\epsilon\|_\infty < \epsilon$, where we used Theorem 4.30 to find $f_\epsilon(S) = \int_{\sigma(T)} f_\epsilon dE$ and Theorem 4.29 for the estimate.
- $f_\epsilon(T) - f_\epsilon(S) = \sum_{k=1}^n \alpha_k \chi_{A_k}(T) - \sum_{k=1}^n \alpha_k E_{A_k} = 0$.

Together, we have $\|T - S\| < 2\epsilon$, and since ϵ was arbitrary, $T = S$, so (4.12) is proved. The fact that the functional calculus is then given as (4.13) is the content of Theorem 4.30. Uniqueness then follows by taking $f = \chi_A$ in (4.13), which determines E_A uniquely for any $A \in \Sigma$. □

We have thus obtained a one-to-one correspondence between bounded self-adjoint operators and spectral measures of compact support, and we have completely characterised the functional calculus of bounded self-adjoint operators.

Let us close this course with a few examples.

EXAMPLE 4.32. (1) If T is compact, then we know there exists a countable orthonormal system $(e_n)_{n \in \mathbb{N}}$ of eigenvectors of T with corresponding eigenvalues λ_n, and $T = \sum_{n=1}^\infty \lambda_n P_n$, where P_n denotes the orthogonal projection onto $\operatorname{span}\{e_n\}$. Hence, the spectral measure of T is given as

$$E = \sum_{n=0}^{\infty} \delta_{\lambda_n} P_n,$$

where δ_{λ_n} denotes the (real-valued) Dirac measure centred at λ_n and as before we set $\lambda_0 = 0$ and P_0 the orthogonal projection onto $\ker T$. In other words, for $A \in \Sigma$, E_A is the orthogonal projection of the subspace spanned by all e_n such that $\lambda_n \in A$.

The operator T is then written as

$$Tx = \int_{\mathbb{R}} \lambda dE_\lambda x = \sum_{n=1}^{\infty} \lambda_n P_n x = \sum_{n=1}^{\infty} \lambda_n (x, e_n) e_n \quad (x \in H)$$

in accordance with Theorem 2.31. Also, for a bounded function f defined on $\{\lambda_n : n \in \mathbb{N} \cup \{0\}\}$ and $x \in H$,

$$f(T)x = f(0)P_0 x + \sum_{n=1}^{\infty} f(\lambda_n)(x, e_n)e_n.$$

(2) On $H = L^2(0,1;\mathbb{C})$, consider a multiplication operator of the form $T : f \mapsto \phi f$, where $\phi \in L^\infty(0,1)$ is given and real-valued (so that T is self-adjoint). Then, for a measurable set $A \subset \mathbb{R}$, E_A has the form

$$E_A f = \chi_{\phi^{-1}(A)} f.$$

Let us check that E thus given is really a spectral measure: Clearly E_A is a projection, because $\chi^2_{\phi^{-1}(A)} = \chi_{\phi^{-1}(A)}$. It is also orthogonal, because the kernel of E_A is the set of L^2 functions essentially supported outside $\phi^{-1}(A)$, its range is the set of functions essentially supported in $\phi^{-1}(A)$, and both spaces are obviously orthogonal.

For $A = \varnothing$, $E_A f = \chi_\varnothing f = 0$ and $E_{\mathbb{R}} f = \chi_{[0,1]} f = f$. Also, E is supported on the compact set $\{x \in \mathbb{R} : |x| \le \|\phi\|_\infty\}$.

If $(A_n)_{n\in\mathbb{N}} \subset \Sigma$ are pairwise disjoint, then so are the sets $\phi^{-1}(A_n)$, and therefore

$$\chi_{\bigcup_{n=1}^\infty \phi^{-1}(A_n)} = \sum_{n=1}^{\infty} \chi_{\phi^{-1}(A_n)},$$

whence follows the σ-additivity of $E : \Sigma \to L(H,H)$.

Finally, we show that indeed $T = \int \lambda dE$: For $\epsilon > 0$ let $\Psi_\epsilon : [-\|\phi\|_\infty, \|\phi\|_\infty] \to \mathbb{R}$ be an elementary function such that $\|t - \Psi_\epsilon\|_\infty < \epsilon$. We shall write $\Psi_\epsilon = \sum_{n=1}^N \alpha_n \chi_{A_n}$ and we may assume in addition that $\bigcup_{n=1}^N A_n \supset \phi([0,1])$. Since $\|t - \Psi_\epsilon\|_\infty < \epsilon$, we know that $|\alpha_n - \lambda| < \epsilon$ whenever $\lambda \in A_n$.

Observe that then, for any $x \in [0,1]$, there is exactly one $n \in \{1, \dots, N\}$ such that $\phi(x) \in A_n$, so that

$$|\sum_{k=1}^{N} \alpha_k \chi_{\phi^{-1}(A_k)}(x) - \phi(x)| = |\alpha_n - \phi(x)| < \epsilon. \tag{4.14}$$

Therefore,

$$\left\| \int_{\mathbb{R}} \Psi_\epsilon(\lambda) dE_\lambda - \phi \right\| = \left\| \sum_{n=1}^{N} \alpha_n E_{A_n} - \phi \right\| = \left\| \sum_{n=1}^{N} \alpha_n \chi_{\phi^{-1}(A_n)} - \phi \right\| < \epsilon, \tag{4.15}$$

and it follows that indeed

$$T = \int_{\mathbb{R}} \lambda dE_\lambda.$$

Note carefully that the last norm in (4.15) is the operator norm of the multiplication operator with the function $\sum_{n=1}^N \alpha_n \chi_{\phi^{-1}(A_n)} - \phi$, which equals the supremum norm of the function itself (which in turn is less than ϵ by (4.14)).

(3) Let $f \in L^2_{per}(0,2\pi)$ and $\phi \in L^\infty_{per}(0,2\pi;\mathbb{R})$ such that $\phi(x) = \phi(-x)$ for all $x \in \mathbb{R}$, then the *convolution* of these two functions is defined as

$$f * \phi(x) := \int_0^{2\pi} f(y)\phi(x-y)dy.$$

We claim $f * \phi$ is again a periodic L^2 function[8], and for fixed ϕ the linear operator $Tf := f * \phi$ is bounded with norm at most $2\pi\|\phi\|_\infty$. Indeed,

$$\begin{aligned}\|f * \phi\|_{L^2}^2 &= \int_0^{2\pi} \left|\int_0^{2\pi} f(y)\phi(x-y)dy\right|^2 dx \\ &\le \int_0^{2\pi} \left(\int_0^{2\pi} |f(y)||\phi(x-y)|dy\right)^2 dx \\ &\le \|\phi\|_\infty^2 (2\pi)^2 \int_0^{2\pi}\int_0^{2\pi} |f(y)|^2 \frac{dy}{2\pi} dx = (2\pi\|\phi\|_\infty)^2 \|f\|_{L^2}^2,\end{aligned}$$

where for the last estimate we used Jensen's inequality for the convex function $|\cdot|^2$ and the probability measure $\frac{dx}{2\pi}$.

Moreover, T is self-adjoint, because by Fubini's Theorem and the property of ϕ of being real-valued and even,

$$(Tf,g) = \int_0^{2\pi}\int_0^{2\pi} f(y)\phi(x-y)\overline{g(x)}dydx = (f,Tg) \quad \forall f,g \in L^2_{per}(0,2\pi).$$

For $n \in \mathbb{Z}$, let $e_n(x) = \frac{1}{\sqrt{2\pi}}e^{inx}$ as in Section 2.2. These form an orthonormal basis of eigenfunctions of T with eigenvalue $\sqrt{2\pi}\hat{\phi}(n)$ (recall the definition of Fourier coefficients in Section 2.2), because

$$\begin{aligned}Te_n(x) &= \int_0^{2\pi} \frac{1}{\sqrt{2\pi}} e^{iny}\phi(x-y)dy = \int_0^{2\pi} \frac{1}{\sqrt{2\pi}} e^{inx}\overline{e^{in(x-y)}}\phi(x-y)dy \\ &= e_n(x)\int_0^{2\pi} e^{-iny}\phi(y)dy = \sqrt{2\pi}\hat{\phi}(n)e_n(x).\end{aligned}$$

We thus have the spectral decomposition

$$Tf = \sqrt{2\pi}\sum_{n=-\infty}^{\infty} \hat{\phi}(n)P_n f = \sqrt{2\pi}\sum_{n=-\infty}^{\infty} \hat{\phi}(n)\hat{f}(n)e_n,$$

where again P_n denotes the orthogonal projection onto $\operatorname{span}\{e_n\}$. The spectral measure is given as

$$E = \sum_{n=-\infty}^{\infty} \delta_{\sqrt{2\pi}\hat{\phi}(n)} P_n.$$

The situation is therefore similar to the first example – in fact it is not difficult to show that T is compact.

[8]In fact, when $f \in L^2$, then by *Young's inequality for convolutions*, $f * \phi \in L^2$ is even true when ϕ is only in L^1, and $\|f * \phi\|_{L^2} \le \|f\|_{L^2}\|\phi\|_{L^1}$.

Bibliography

[1] O. Forster. Analysis 1. *Vieweg, Braunschweig/Wiesbaden,* sixth ed., 2001.

[2] K. Hornik, M. Stinchcombe, H. White. Multilayer feedforward networks are universal approximators. *Neural Networks* **2**(1989), 359–366.

[3] J. Naas, W. Tutschke. Große Sätze und schöne Beweise der Mathematik. *Verlag Harri Deutsch, Frankfurt a. M.,* second ed., 1997.

[4] V. I. Paulsen, M. Raghupathi. An introduction to the theory of reproducing kernel Hilbert spaces. *Cambridge University Press, Cambridge (UK),* 2016.

[5] D. Werner. Funktionalanalysis. *Springer-Verlag, Berlin,* third ed., 2000.