MATH 421/510

Real Analysis II — Functional Analysis

January – April 2023

Sven Bachmann

Version of April 5, 2023

Contents

What is functional analysis?

- Study of topological spaces and of functional relations between them
- Study of spaces of functions
- Language of PDE, calculus of variations, integral equations
- Language of quantum mechanics

Functional analysis arose in the 19th century, was developed in the first part of the 20th century, in a paradigmatic shift from the study of (the properties of) a single function/solution to the study of (the properties of) sets of functions/solutions and the relations between them. It is the language of much of modern mathematics, encompassing (linear) algebra, analysis and stochastic analysis.

CHAPTER I

Topological spaces

1. Point set topology

Understanding limits and convergence is central to functional analysis. This ultimately has to do with the notions of open sets and neighbourhoods of a point. If the set is equipped with a distance, this can be done with open balls. In the more general setting of topological spaces, these concepts are introduced by the notion of a *topology*.

DEFINITION I.1.1. A *topological space* $(S, \mathcal{T})$ is a nonempty set S with a family of subsets $\mathcal{T}$ such that

- $\emptyset \in \mathcal{T}, S \in \mathcal{T}$
- $\mathcal{T}$ is closed under finite intersections:

$$A_1, \dots A_n \in \mathcal{T} \Rightarrow \bigcap_{j=1}^{n} A_j \in \mathcal{T}$$

- $\mathcal{T}$ is closed under arbitrary unions:

$$\{A_\alpha : \alpha \in I\} \subset \mathcal{T} \Rightarrow \bigcup_{\alpha \in I} A_\alpha \in \mathcal{T}$$

  where I is an arbitrary index set.

The elements of $\mathcal{T}$ are called the *open sets* of S.

EXAMPLE 1. (i) The discrete topology: $\mathcal{T} = \mathcal{P}(S)$ the power set of S, containing all subsets of S
(ii) The indiscrete topology: $\mathcal{T} = \{\emptyset, S\}$
(iii) Let $S = \mathbb{R}^n$ with the elementary notion of open sets, namely $X \in \mathcal{T}$ iff $\forall x \in X, \exists r > 0$ s.t. $\{y \in S : d(y, x) < r\} \subset X$, where $d(\cdot, \cdot)$ is the Euclidean distance.

A *metric space* is a set M equipped with a function $d : M \times M \to [0, \infty)$ such that (i) $d(x, y) = 0$ iff $x = y$, (ii) $d(x, y) = d(y, x)$, and (iii) $d(x, z) \leq d(x, y) + d(y, z)$, the triangle inequality.

The metric defines a topology as in the third example above. Since any metric on S gives rise to a topology, one may wonder whether every topology arises from a metric and the answer is, not surprisingly, no. If it is the case, $\mathcal{T}$ is called *metrizable*.

Topologies on a space S can be ordered in a set-theoretic fashion: $\mathcal{T}_1 \prec \mathcal{T}_2$ iff $\mathcal{T}_1 \subset \mathcal{T}_2$ and $\mathcal{T}_1$ is called *weaker* than $\mathcal{T}_2$.

Given a family $\mathcal{E} \subset \mathcal{P}(S)$, the unique weakest topology $\mathcal{T}(\mathcal{E})$ on S containing $\mathcal{E}$ is called the topology *generated by* $\mathcal{E}$. It can be shown that $\mathcal{T}(\mathcal{E})$ consists of $\emptyset, S$ and all unions and all finite intersections of elements of $\mathcal{E}$.

Definition I.1.2. A *base* of $\mathcal{T}$ is a family $\mathcal{B} \subset \mathcal{T}$ such that for any nonempty $O \in \mathcal{T}$, there is a family $\{B_\alpha : \alpha \in I\} \subset \mathcal{B}$ and $O = \cup_{\alpha \in I} B_\alpha$.

If $(S, \mathcal{T})$ is a topological space, and $X \subset S$, then $\mathcal{T}_X := \{O \cap X : O \in \mathcal{T}\}$ defines a topology on X called the *relative topology*.

The following concepts, familiar in $\mathbb{R}^n$, extend to general topological spaces. Let $X \subset S$.

- X is *closed* if there is $Y \in \mathcal{T}$ such that $X = Y^c$
- The *interior* X^o of X is the largest open set contained in X
- The *closure* $\overline{X}$ of X is the smallest closed set containing X
- The *boundary* ∂X of X is $\partial X = \overline{X} \setminus X^o$
- X is called *dense* in S if $\overline{X} = S$

A *neighbourhood* of $x \in S$ is a set $N_x \subset S$ such that $x \in N_x^o$. Note that a neighbourhood is not required to be open. A family $\mathcal{N}_x$ of subsets of S is a *neighbourhood base at* x if each $N \in \mathcal{N}_x$ is a neighbourhood of x and if for any neighbourhood M_x of x, there is an $N \in \mathcal{N}_x$ such that $N \subset M_x$.

There are two major classifications of topological spaces. The first one is about how well open sets separate points. While the classification has five classes denoted $T_0, \ldots, T_4$, we only introduce the following, which plays an important role in the discussion of compactness.

Definition I.1.3. A topological space $(S, \mathcal{T})$ is called *Hausdorff*, or T_2, if for all pairs $x, y \in S, x \neq y$, $\exists O_x, O_y \in \mathcal{T}$ with $O_x \cap O_y = \emptyset$, such that $x \in O_x, y \in O_y$.

The second classification is about countability and it is particularly relevant in discussing questions of convergence (and consequently its relation to compactness).

Definition I.1.4. A topological space $(S, \mathcal{T})$ is called

- *separable* if it has a countable dense set
- *first countable* if each $x \in S$ has a countable neighbourhood base
- *second countable* if S has a countable base

Proposition I.1.5. *(i) Second countable $\Rightarrow$ First countable*
(ii) Second countable $\Rightarrow$ Separable

Proof. Let $\mathcal{B}$ be a countable base of $\mathcal{T}$.
(i) For any $x \in S$, the family $\mathcal{N}_x := \{N \in \mathcal{B} : x \in N\}$ is a countable neighbourhood base at x. Indeed, N are all open by definition of a base, hence $x \in N^o$ and so $N \in \mathcal{N}_x$ is a neighbourhood of x. Moreover, if M_x is a neighbourhood of x, then M_x^o is an open set and since $\mathcal{B}$ is a base, there are $\{N_j \in \mathcal{B}\}$ such that $\cup_j N_j = M_x^o$. Hence there is j_0 such that $x \in N_{j_0} \subset M_x$, and $N_{j_0} \in \mathcal{N}_x$.
(ii) For each $B \in \mathcal{B}$, let $x_B \in B$. Then the set $D := \{x_B : B \in \mathcal{B}\}$ is countable. But $\overline{D}^c$ is open by construction it does not include any $B \in \mathcal{B}$. It follows from the definition of a base that $\overline{D}^c = \emptyset$, namely, D is dense. $\square$

Note that there are separable spaces that are not second countable.

Example 2. Consider $\mathbb{R}^n$ equipped with the usual topology. Then the family of all open balls (any centre, any radius) is a base. For any $x \in \mathbb{R}^n$ the family $\{\overline{B_{p/q}(x)} : p, q \in \mathbb{N}\}$ of closed balls for rational radii is a neighbourhood base. Hence $\mathbb{R}^n$ is first countable.

This again generalizes to general metric spaces. A metric space is first countable. Moreover, a metric space is second countable iff it is separable.
We are now ready to turn to the general notion of convergence.

Definition I.1.6. A sequence $(x_n)_{n\in\mathbb{N}}$ in a topological space $(S, \mathcal{T})$ is *convergent* if there is $x \in S$ such that for every neighbourhood N_x of x, there is n_0 such that $x_n \in N_x$ for all $n \geq n_0$.

Here is a first result that is valid only in first countable spaces, namely that the closure of a subset is given by the set of limit points of sequences.

Proposition I.1.7. *Let $(S, \mathcal{T})$ be a first countable topological space and $X \subset S$. Then $x \in \overline{X}$ iff x is the limit of a convergent sequence $(x_n)_{n\in\mathbb{N}}$ in X.*

Proof. Let $\mathcal{N}_x := \{O_n : n \in \mathbb{N}\}$ be a countable neighbourhood base of x such that $O_n \subset O_{n-1}$ for all $n \in \mathbb{N}$. If $x \in \overline{X}$, then for any $n \in \mathbb{N}$, $O_n \cap X \neq \emptyset$ (since otherwise $x \notin (O_n^o)^c$ would be a closed set containing X, but $x \in \overline{X} \subset (O_n^o)^c$ is a contradiction) and we can pick $x_n \in O_n \cap X$. This is a convergent sequence such that $\lim_{n\to\infty} x_n = x$: For any neighbourhood M of x, there is $n_0 \in \mathbb{N}$ such that $O_{n_0} \subset M$, and hence $O_n \subset M$ for all $n \geq n_0$; therefore, $x_n \in M$ for all $n \geq n_0$. Reciprocally, assume that $x \in (\overline{X})^c$. For any sequence $(y_n)_{n\in\mathbb{N}}$ in X, the open neighbourhood $(\overline{X})^c$ contains no point of the sequence, and hence $(y_n)_{n\in\mathbb{N}}$ does not converge to x. $\square$

Note that if $\mathcal{M}_x := \{U_n : n \in \mathbb{N}\}$ is any a countable neighbourhood base at x, the sets $O_j = \cap_{n=1}^j U_j$ form a 'decreasing' neighbourhood base as used in the proof.
If $(S, \mathcal{T})$ is not first countable, this criterion is not sufficient. The closure is given by limit points of *nets*, which are generalizations of sequences of the form $(x_\alpha)_{\alpha\in I}$ where I is not necessarily countable and only partially ordered.

Proposition I.1.8. *Let $(S, \mathcal{T})$ be a Hausdorff space. Let $(x_n)_{n\in\mathbb{N}}$ be a convergent sequence in S. Then the limit $x = \lim_{n\to\infty} x_n$ is unique.*

Proof. Let $x = \lim_{n\to\infty} x_n$ and let $y \neq x$. There exist disjoint $O_x, O_y \in \mathcal{T}$ with $x \in O_x, y \in O_y$. But $x_n \to x$ implies that there is n_0 such that $x_n \in O_x$ for all $n \geq n_0$, and in particular $x_n \notin O_y, n \geq n_0$. It follows that $(x_n)_{n\in\mathbb{N}}$ does not converge to y. $\square$

2. Compactness

In a topological space $(S, \mathcal{T})$, an *open cover* is a family $\mathcal{C} \subset \mathcal{T}$ such that $S = \cup_{O\in\mathcal{C}} O$. A *subcover* is a subset of $\mathcal{C}$ that is a cover.

Definition I.2.1. A topological space $(S, \mathcal{T})$ is *compact* if any open cover has a finite subcover.

A subset $X \subset S$ is a compact set if it is compact in the relative topology. It is called *precompact* if its closure is compact. Note that if a family of open sets $\mathcal{C} = \{O_\alpha \in \mathcal{T} : \alpha \in I\}$ is such that $X \subset \cup_{\alpha\in I} O_\alpha$, then $\mathcal{C}_X = \{O_\alpha \cap X \in \mathcal{T} : \alpha \in I\}$ is an open cover of X.

DEFINITION I.2.2. A topological space is *locally compact* if every $x \in S$ has a compact neighbourhood.

Compactness can also be formulated in terms of closed sets. $(S, \mathcal{T})$ is said to have the *finite intersection property* if any family $\mathcal{C}$ of closed set such that $\cap_{j=1}^n C_j \neq \emptyset$ for any finite subfamily $\{C_1, \ldots C_n\} \subset \mathcal{F}$ satisfies $\cap_{C \in \mathcal{C}} C \neq \emptyset$. We then have the following result, the proof of which is an exercise in Boole-Morgan's laws: S is compact iff S has the finite intersection property.

PROPOSITION I.2.3. *Let $X \subset S$ be a subset of a compact topological space $(S, \mathcal{T})$. If X is closed, then it is compact.*

PROOF. Let $\mathcal{C}$ be an open cover of X. By the definition of the relative topology, any $C \in \mathcal{C}$ is of the form $O_C \cap X$ with $O_C \in \mathcal{T}$. If $\mathcal{O}$ is the set of these O_C's, then $\mathcal{O} \cup \{X^c\}$ is an open cover of S since X is closed. S being compact, there is a finite subcover $\tilde{\mathcal{O}}$, which yields, by intersecting with X, a finite open cover $\tilde{\mathcal{C}}$ of X. $\square$

PROPOSITION I.2.4. *Let $(S, \mathcal{T})$ be a Hausdorff space and let K be a compact subset of S. Then K is closed.*

PROOF. For any $x \in K^c$, let $U_x \ni x$ be the open set given by the lemma below. Clearly $K^c \subset \cup_{x \in K^c} U_x$. Moreover, $K \cap U_x = \emptyset$ for all $x \in K^c$ implies that $\cup_{x \in K^c} U_x \subset K^c$. Hence $K^c = \cup_{x \in K^c} U_x$, which is open, and so K is closed. $\square$

LEMMA I.2.5. *Let $(S, \mathcal{T})$ be a Hausdorff space and let K be a compact subset of S. For any $x \in K^c$, there are disjoint open sets U, V such that $x \in U$, $K \subset V$.*

PROOF. Let $x \in K^c, y \in K$. There are disjoint open U_y, O_y such that $x \in U_y, y \in O_y$. Using the open cover $\{O_y : y \in K\}$, there are $\{y_1, \ldots, y_N\}$ in K such that

$$K \subset \cup_{j=1}^N O_{y_j} = V.$$

Moreover, the set $U = \cap_{j=1}^N U_{y_j}$ contains x and is disjoint from V. $\square$

It is worth pointing out that the Bolzano-Weierstrass theorem of real analysis does not hold in a general topological space. In fact, one must consider nets instead of sequences in the general case. However it does in a second countable space:

THEOREM I.2.6. *Let $(S, \mathcal{T})$ be a second countable topological space. Then S is compact iff every sequence has a convergent subsequence.*

We will need the following auxiliary lemma.

LEMMA I.2.7. *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a first countable topological space $(S, \mathcal{T})$. A point $x \in S$ is a cluster point of $(x_n)_{n\in\mathbb{N}}$ iff there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ that converges to x.*

By definition, a *cluster point* of $(x_n)_{n\in\mathbb{N}}$ is a $x \in S$ such that for every neighbourhood N_x of x, $x_n \in N_x$ for infinitely many n.

PROOF. Let x be a cluster point, and let $\mathcal{N}_x$ be a countable neighbourhood base of x, such that $N_j \subset N_{j-1}$. For each j, let $x_{n_j} \in N_j$. Then $(x_{n_j})_{j\in\mathbb{N}}$ converges to x. Indeed, let M_x be a neighbourhood of x and let $N_k \subset M_x$. Then $x_{n_j} \in N_j \subset N_k$ for all $j \geq k$. Reciprocally, if $(x_{n_j})_{j\in\mathbb{N}}$ converges to x, then for any neighbourhood N_x of x, $x_{n_j} \in N_x$ for all $j \geq j_0$. Hence x is a cluster point since $\{j \geq j_0\}$ is infinite. $\square$

PROOF OF THEOREM I.2.6. Assume that S is compact, let $(z_n)_{n\in\mathbb{N}}$ be a sequence in S that does not have a convergent subsequence. Since S is second countable, it is first countable, so that $(z_n)_{n\in\mathbb{N}}$ does not have a cluster point. Hence, for any $x \in S$, there is an open set $O_x \ni x$ such that $z_n \in O_x$ for only finitely many n's. In particular, there is $n_x \in \mathbb{N}$ such that $z_n \notin O_x$ for all $n \geq n_x$. Extracting a finite cover $\{O_{x_i} : 1 \leq i \leq N\}$ from $\{O_x : x \in S\}$, and letting $n_0 = \max\{n_{x_i} : 1 \leq i \leq N\}$, we have that $z_n \notin \cup_{i=1}^N O_{x_i} = S$ for all $n \geq n_0$, a contradiction.

Reciprocally, assume that every sequence has a convergent subsequence. Since S is second countable, it has a countable open base $\mathcal{C} = \{O_j : j \in \mathbb{N}\}$, which is also an open cover. Assume that there is no finite subcover of $\mathcal{C}$. Then for any $n \in \mathbb{N}$, there is $x_n \notin \cup_{j=1}^n O_j$. Let $(x_{n_k})_{k\in\mathbb{N}}$ be a convergent subsequence and let x be its limit. Since $\mathcal{C}$ is a cover, there is j_0 such that $x \in O_{j_0}$, and hence there is k_0 such that $x_{n_k} \in O_{j_0}$ for all $k \geq k_0$. This is contradiction with $x_{n_k} \notin \cup_{j=1}^{n_k} O_j$ for any $n_k > j_0$. Finally, since $U = \cup_{j_U} O_{j_U}$ for any open U, any open cover has a countable subcover, and hence a finite subcover by the above argument. $\square$

The property that every sequence has a convergent subsequence is called *sequential compactness*. The first part of the theorem shows that compactness implies sequential compactness in a first countable space (a fortiori in a second countable space and in a metric space).

3. Continuity

DEFINITION I.3.1. Let $(S_1, \mathcal{T}_1), (S_2, \mathcal{T}_2)$ be topological spaces. A function $f : S_1 \to S_2$ is *continuous* if $f^{-1}(O) \in \mathcal{T}_1$ for any $O \in \mathcal{T}_2$.

In other words, the preimage of any open set is open. This should not be confused with the following:

DEFINITION I.3.2. Let $(S_1, \mathcal{T}_1), (S_2, \mathcal{T}_2)$ be topological spaces. A function $f : S_1 \to S_2$ is *open* if $f(O) \in \mathcal{T}_2$ for any $O \in \mathcal{T}_1$.

An invertible function that is both open and continuous is a *homeomorphism*.
While continuity is defined in terms of two topologies, one can reciprocally use continuity to define topologies. Let S_1 be a set (not yet equipped with a topology) and let $(S_2, \mathcal{T}_2)$ be a topological space. Let $\mathcal{F}$ be a family of functions from S_1 to S_2. Then the topology on S_1 generated by $\{f^{-1}(O) : O \in \mathcal{T}_2\}$ is called the $\mathcal{F}$-weak topology. By definition, all functions $f \in \mathcal{F}$ are continuous with respect to this topology on S_1.

EXAMPLE 3. Let $S_1 = C([a, b]; \mathbb{R})$ be the set of continuous functions, and let $S_2 = \mathbb{R}$ with the usual metric topology. Let $E_x : S_1 \to S_2, E_x(f) = f(x)$ be the evaluation functions and let $\mathcal{F} = \{E_x : x \in [a, b]\}$. The $\mathcal{F}$-weak topology on $C([a, b]; \mathbb{R})$ is the topology of pointwise convergence.

Another useful result is that compactness is pushed forward by continuous functions. It in particular generalizes the well-known fact that a continuous, real-valued function defined on a compact interval reaches it maximum and minimum values.

PROPOSITION I.3.3. *Let $(S_1, \mathcal{T}_1), (S_2, \mathcal{T}_2)$ be topological spaces, and let $f : S_1 \to S_2$ be a continuous function. If S_1 is compact, then $f(S_1) \subset S_2$ is compact.*

PROOF. Let $\mathcal{C} = \{C_\alpha : \alpha \in I\}$ be an open cover of $f(S_1) \subset S_2$ in the relative topology. There are open sets $\{O_\alpha : \alpha \in I\}$ in S_2 such that $C_\alpha = O_\alpha \cap f(S_1)$, and $f^{-1}(O_\alpha)$ is open in

S_1 by continuity. Therefore, $\{f^{-1}(O_\alpha) : \alpha \in I\}$ is an open cover of S_1, from which one can extract a finite subcover $\{f^{-1}(O_n) : 1 \le n \le N\}$. But then $\{C_n = O_n \cap f(S_1) : 1 \le n \le N\}$ is a finite subcover of $\mathcal{C}$. $\square$

THEOREM I.3.4. *Let $(S_1, \mathcal{T}_1), (S_2, \mathcal{T}_2)$ be two compact Hausdorff spaces and let $f : S_1 \to S_2$ be a continuous bijection. Then f is a homeomorphism.*

PROOF. Let $C \subset S_1$ be closed. As a subset of the compact S_1, it is compact by Proposition I.2.3 and so $f(C)$ is compact by Proposition I.3.3. Since S_2 is Hausdorff, Proposition I.2.4 implies that $f(C)$ is closed. Hence, if O is open, then $f(O^c)$ is closed and since $f(O^c) = f(O)^c$ by injectivity, we conclude that f^{-1} is continuous. $\square$

4. Stone-Weierstrass theorems

First of all, we recall the 'classical' Weierstrass theorem:

PROPOSITION I.4.1. *If f is a continuous real-valued function on $[a, b]$, then there exists a sequence of polynomials $(P_n)_{n\in\mathbb{N}}$ such that*

$$\lim_{n\to\infty} P_n = f$$

uniformly on $[a, b]$.

In other words, the polynomials are dense in the set $C_{\mathbb{R}}([a, b])$ of continuous real-valued functions on the compact interval $[a, b]$. The Stone-Weierstrass theorem generalizes the result to an arbitrary compact Hausdorff space.

Let X be a compact Hausdorff space. We first note that $C_{\mathbb{R}}(X)$, the real-valued continuous functions on X equipped with the multiplication $(fg)(x) = f(x)g(x)$ is an algebra. It is a metric space with metric

$$d_\infty(f, g) = \sup\{|f(x) - g(x)| : x \in X\}.$$

We say that a subalgebra $\mathcal{A}$ of $C_{\mathbb{R}}(X)$ *separates points* if $x, y \in X$ such that $x \neq y$ implies $\exists f \in \mathcal{A}$ such that $f(x) \neq f(y)$.

THEOREM I.4.2. *Let X be a compact Hausdorff space. Let $\mathcal{A}$ be a closed subalgebra of $C_{\mathbb{R}}(X)$ that separates points. Then either $\mathcal{A} = C_{\mathbb{R}}(X)$ or $\exists x_0 \in X$ such that $\mathcal{A} = \{f \in C_{\mathbb{R}}(X) : f(x_0) = 0\}$.*

In particular, if $1 \in \mathcal{A}$, then the second case is excluded; there is no proper closed unital subalgebra of $C_{\mathbb{R}}(X)$ that separates points. We prove the theorem in this slightly easier case. Note that if $\mathcal{A}$ is not closed, the theorem applies to $\overline{\mathcal{A}}$ in which case it can be stated as: Any unital subalgebra $\mathcal{A}$ that separates points is dense in $C_{\mathbb{R}}(X)$ in the uniform topology.
We note that Hausdorffness is not used in the proof. However, it is a necessary condition for the existence of an algebra separating points. Indeed, if there is $f \in \mathcal{A}$ such that $f(x) \neq f(y)$, then $f(x), f(y)$ have disjoint open neihbourhoods (since $\mathbb{R}$ is Hausdorff) and their preimages must be disjoint open neighbourhoods of x, repsectively y.
The proof uses the concept of a lattice: A subset $\mathcal{F} \subset C_{\mathbb{R}}(X)$ is called a *lattice* if for all $f, g \in \mathcal{F}$, the functions $f \wedge g := \min\{f, g\}$ and $f \vee g := \max\{f, g\}$ are in $\mathcal{F}$.

LEMMA I.4.3. *Any closed unital subalgebra $\mathcal{A}$ of $C_{\mathbb{R}}(X)$ is a lattice.*

PROOF. Since

$$f \vee g = \frac{1}{2}|f-g| + \frac{1}{2}(f+g), \quad f \wedge g = -\big((-f) \vee (-g)\big),$$

it suffices to prove that $f \in \mathcal{A}$ implies $|f| \in \mathcal{A}$. Since there is nothing to prove is $f = 0$, we assume that $f \neq 0$. Since f is continuous on a compact X, it is bounded, namely $\|f\|_\infty = \sup_{x \in X} |f(x)| < \infty$. By the classical Weierstrass theorem, there is a sequence of polynomials such that $\big|P_n(x) - |x|\big| < n^{-1}$ for all $x \in [-1, 1]$. Hence

$$\|P_n \circ h - |h|\|_\infty < \frac{1}{n},$$

where $h = f/\|f\|_\infty$, namely $P_n(h) \to |h|$ uniformly. Since $\mathcal{A}$ is a unital algebra, $f \in \mathcal{A}$ implies $P_n(h) \in \mathcal{A}$, and the convergence just proved concludes the proof by Proposition I.1.7 since $\mathcal{A}$ is closed w.r.t. a metric (hence first countable) topology. □

The final part of the proof goes by the name of Kakutani-Krein theorem.

PROPOSITION I.4.4. *Let $\mathcal{L} \subset C_{\mathbb{R}}(X)$ be a closed lattice that contains 1 and that separates points. Then $\mathcal{L} = C_{\mathbb{R}}(X)$.*

Proof. Let $g \in C_{\mathbb{R}}(X)$. Let $x \neq y$ and let $\epsilon > 0$. The map $\mathcal{L} \ni h \mapsto (h(x), h(y)) \in \mathbb{R}^2$ is linear, and range contains $(1,1)$ since $1 \in \mathcal{L}$ as well as one element of the form (a,b) with $a \neq b$ since $\mathcal{L}$ separates points. Hence its range is all of $\mathbb{R}^2$, so that there is $f_{xy} \in \mathcal{L}$ such that $f_{xy}(x) = g(x), f_{xy}(y) = g(y)$.

By continuity of f_{xy}, g, there is a neighbourhood N_y of y such that $f_{xy}(z) + \epsilon > g(z)$ for all $z \in N_y$. By compactness, there is a finite set $\{y_1, \ldots, y_n\}$ such that $\{N_{y_j} : 1 \leq j \leq n\}$ is a subcover of X. The function $f_x := f_{xy_1} \vee \cdots \vee f_{xy_n}$, is such that $f_x(x) = g(x)$ and $f_x(z) + \epsilon > g(z)$ for all $z \in X$.

By continuity of f_{xy}, g, there is a neighbourhood $M_x \ni x$ such that $f_x(z) - \epsilon < g(z)$ for all $z \in M_x$. Extracting a finite subcover indexed by $\{x_1, \ldots, x_m\}$ and letting $f := f_{x_1} \wedge \cdots \wedge f_{x_m}$, we conclude that $f(z) - \epsilon < g(z)$ for all $z \in X$. By the previous part $f(z) + \epsilon > g(z)$, so that we have constructed $f \in \mathcal{L}$ such that $\|f - g\|_\infty < \epsilon$. Since ϵ is arbitrary, this shows that $\mathcal{L}$ is dense and hence equal to $C_{\mathbb{R}}(X)$ because it is closed. □

The Stone-Weierstrass extends is two directions. First of all, it extend to complex-valued functions, provided the subalgebra $\mathcal{A}$ is closed under complex conjugation, namely $f \in \mathcal{A}$ implies $\bar{f} \in \mathcal{A}$ (and indeed, the result is in general false). Indeed, any $f \in C_{\mathbb{C}}(X)$ can be written as $f = (f + \bar{f})/2 - \mathrm{i}(f - \bar{f})/2$, where both terms are in $\mathcal{A} \cap C_{\mathbb{R}}(X)$. The complex Stone-Weierstrass theorem follows from an application of the real one to the real and imaginary parts of f.

Secondly, it extends to locally compact Hausdorff (LCH) spaces. In that case, the relevant algebra is the set of functions that vanish at infinity, namely those $f \in C_{\mathbb{R}}(S)$ such that $\forall \epsilon > 0$, the set $\{x \in S : |f(x)| \geq \epsilon\}$ is compact. Indeed, it suffices to apply the above to the one-point compactification $X = S \cup \{\infty\}$ of S, noting that every continuous function on S vanishing at infinity has a continuous extension to X.

5. Urysohn's lemma

We conclude this chapter with Urysohn's lemma. It is again about separating sets, but now using continuous functions. Both the lemma and the following proposition upon which its proof lies can be phrased very explicitly in the context of metric spaces. Here, we present the proofs for a more general locally compact Hausdorff space. First of all, PROPOSITION I.5.1. *Let S be a LCH space. Let $K \subset U \subset S$, where K is compact and U is open. There is an open set O with compact closure such that*

$$K \subset O \subset \overline{O} \subset U.$$

PROOF. Since S is LCH, every point of K has an open neighbourhood with compact closure. Since K is compact, there is finite subcover of such neighbourhoods. Hence K is a subset of their union V which has a compact closure (indeed, $\overline{V}$ is the finite union of the compact closures of the neighbourhoods). If $U = S$, then $O = V$ satisfies the conclusion of the theorem. Otherwise, the complement U^c is nonempty. By the Hausdorff property, for any $x \in U^c \subset K^c$, there is an open set O_x such that $K \subset O_x$ and $x \notin \overline{O_x}$, see Lemma I.2.5. It follows that

$$\bigcap_{x \in U^c} U^c \cap \overline{V} \cap \overline{O_x} = \emptyset,$$

where each $U^c \cap \overline{V} \cap \overline{O_x}$ is a compact subset of $\overline{V}$, hence closed. By the finite intersection property, there are finitely many $\{x_1, \ldots, x_n\}$ such that

$$U^c \cap \overline{V} \cap \overline{O_{x_1}} \cap \cdots \cap \overline{O_{x_n}} = \emptyset$$

and the set $O = V \cap O_{x_1} \cap \cdots \cap O_{x_n} \supset K$ satisfies the conclusions of the theorem since $\overline{O} \subset \overline{V} \cap \overline{O_{x_1}} \cap \cdots \cap \overline{O_{x_n}} \subset U$ and $\overline{O}$ is compact as a closed subset of a compact set. $\square$

We recall that the *support* of a complex-valued function f is given by

$$\operatorname{supp}(f) = \overline{\{x \in S : f(x) \neq 0\}}.$$

We denote by $C_c(S)$ the set of compactly supported continuous functions on S. With these definitions, we denote

$$K \prec f$$

for a compact set K and a $f \in C_c(S)$ such that $0 \leq f(x) \leq 1$ for all $x \in S$ and that $f(x) = 1$ for all $x \in K$. We further denote

$$f \prec U$$

for an open set U and a $f \in C_c(S)$ such that $0 \leq f(x) \leq 1$ for all $x \in S$ and $\operatorname{supp}(f) \subset U$. In these notations, Urysohn's Lemma reads:

LEMMA I.5.2. *Let S be a LCH space, $K \subset U \subset S$ be respectively compact and open. There exists a $f \in C_c(S)$ such that*

$$K \prec f \prec U.$$

PROOF. A inductive application of Proposition I.5.1 yields a family of open set $\{O_r : r \in \mathbb{Q} \cap [0,1]\}$ with compact closures such that

$$K \subset O_1, \qquad \overline{O_0} \subset U$$

and

$$\overline{O_s} \subset O_r \quad \text{whenever} \quad s > r.$$

Let

$$f_r(x) = \begin{cases} r & \text{if } x \in O_r \\ 0 & \text{otherwise} \end{cases} \qquad g_s(x) = \begin{cases} 1 & \text{if } x \in \overline{O_s} \\ s & \text{otherwise} \end{cases}$$

namely $f_r = r\chi_{O_r}$ and $g_s = s + (1-s)\chi_{\overline{O_s}}$, and

$$f(x) = \sup\{f_r(x) : r \in \mathbb{Q} \cap [0,1]\}, \qquad g(x) = \inf\{g_s(x) : s \in \mathbb{Q} \cap [0,1]\}.$$

Since f_r is proportional to the characteristic function of the open set O_r, it is lower semicontinuous and f being the supremum thereof, it is again lower semicontinuous (namely $\{x : f(x) > a\}$ is open for all $a \in \mathbb{R}$). Similarly g is upper semicontinuous (namely $\{x : g(x) < a\}$ is open for all $a \in \mathbb{R}$). Moreover, $0 \leq f \leq 1$, $f(x) = 1$ for all $x \in K \subset O_1$, and $\mathrm{supp} f \subset \overline{O_0} \subset U$. Hence, the proof is complete if we prove continuity by showing that $f = g$. We first note that $f_r(x) > g_s(x)$ if $r > s$ and $x \in O_r, x \notin \overline{O_s}$. But $r > s$ implies $O_r \subset O_s$, which is a contradiction. Hence $f_r \leq g_s$ for all r, s and hence $f \leq g$. Finally, assume that there exists x such that $f(x) < g(x)$. There are $r, s \in \mathbb{Q}$ such that $f(x) < r < s < g(x)$. The first inequality implies that $x \notin O_r$ while the third inequality implies that $x \in \overline{O_s}$, and both together are in contradiction with the second inequality. Hence $f = g$. □

In a metric space, a Urysohn's function can be given explicitly as

$$f(x) = \frac{d(x, U^c)}{d(x, U^c) + d(x, K)}$$

where $d(x, E) = \inf\{d(x, y), y \in E\}$ is the distance of the point x to the set $E \subset S$ (in fact, this yields a slightly weaker result since $\mathrm{supp}(f) = \overline{U}$ in the case).

We conclude with two useful consequences of the lemma.

Proposition I.5.3. *Let $(S,\mathcal{T})$ be a LCH space, let K be compact and let $\{O_i : 1 \le i \le n\}$ be a finite open cover of K. There exists functions $\{f_i \in C_c(S) : 1 \le i \le n\}$ such that*

(i) $\sum_{i=1}^n f_i(x) = 1$ *for all* $x \in K$
(ii) $f_i \prec O_i$ *for all* $1 \le i \le n$

The family $\{f_i : 1 \le i \le n\}$ is called a *partition of unity* on K that is subordinate to $\{O_i : 1 \le i \le n\}$.

Proof. Let $x \in K$. By assumptions, there are i_x such that $x \in O_{i_x}$. Moreover, $\{x\}$ is a compact, hence there is a neighbourhood N_x with compact closure such that $x \in N_x \subset \overline{N}_x \subset O_{i_x}$ by Proposition I.5.1. By compactness, there are $x_1, \ldots, x_m \in K$ such that $K \subset \bigcup_{j=1}^m N_{x_j} \subset \bigcup_{j=1}^m \overline{N_{x_j}}$. For $1 \le i \le n$, let $K_i = \bigcup_j \overline{N_{x_{i_j}}}$ where $\overline{N_{x_{i_j}}} \subset O_i$. Then K_i is compact (as a finite union of compact sets) and $K_i \subset O_i$, so that there is a compactly supported continuous g_i such that $K_i \prec g_i \prec O_i$ by Urysohn's lemma. Since $K \subset \cup_{i=1}^n K_i$, we have that $\sum_{i=1}^n g_i \ge 1$ on K so it remains to properly normalize the g_i's. The set $W = \{x : \sum_{i=1}^n g_i(x) > 0\}$ is open (as the preimage of an open set by a continuous function) so that by Urysohn' lemma again, there is f such that $K \prec f \prec W$. Let $g_{n+1} = 1 - f$. Then by construction $\sum_{i=1}^{n+1} g_i > 0$, so that $f_i = g_i / \sum_{j=1}^{n+1} g_j$ is well-defined on S for $1 \le i \le n$. Clearly, $\mathrm{supp}(f_i) = \mathrm{supp}(g_i) \subset O_i$. Finally, $g_{n+1} = 0$ on K implies that $\sum_{i=1}^n f_i = 1$ on K. $\square$

Proposition I.5.4 (Tietze's extension). *Let $(S,\mathcal{T})$ be a LCH space, let K be compact and let $f \in C(K)$. There exists $F \in C_c(S)$ such that $F(x) = f(x)$ for all $x \in K$.*

Proof. Since f is continuous on a compact space, it is bounded and we assume without loss that $-1 \le f \le 1$ on K. Let V be as in the proof of Urysohn's lemma be open with compact closure and such that $K \subset V$. The sets $K^{\pm} = \{x \in K : f(x) \gtrless 1/3\}$ are disjoint closed subsets of K and hence compact. Applying Urysohn's lemma first to K^+ and $V \setminus K^-$, second to K^- and $V \setminus K^+$, taking the difference and rescaling, there is a function $f_1 \in C_c(S)$ such that $f_1 = 1/3$ on K^+, $f_1 = -1/3$ on K^-, and $-1/3 \le f_1 \le 1/3$ and $\mathrm{supp}(f_1) \subset V$. Hence $-2/3 \le f - f_1 \le 2/3$ on K. We repeat this with $f - f_1$ replacing f to obtain $f_2 \in C_c(S)$

with $\mathrm{supp}(f_2) \subset V$, such that $|f_2| \leq (1/3)(2/3)$ on S and $|f - f_1 - f_2| \leq (2/3)^2$ on K. This procedure provides a sequence $(f_n)_{n\in\mathbb{N}}$ in $C_c(S)$ such that $|f_n| \leq (1/3)(2/3)^{n-1}$ on S and $|f - \sum_{j=1}^n f_j| \leq (2/3)^n$ on K. This shows that the series $F = \sum_{j=1}^\infty f_j$ converges uniformly on S, hence F is continuous, and it converges to f on K. Moreover, $\mathrm{supp}(F) \subset \overline{V}$. $\square$

CHAPTER II

Normed vector spaces

1. Basic definitions and results

Definition II.1.1. A *normed linear space* $(V, \|\cdot\|)$ is a vector space V over $\mathbb{C}$ (or $\mathbb{R}$) equipped with a norm $\|\cdot\| : V \to [0, \infty)$ such that

(i) $\|v\| \geq 0$ for all $v \in V$ and $\|v\| = 0 \Leftrightarrow v = 0$,

(ii) $\|\lambda v\| = |\lambda|\|v\|$ for all $v \in V, \lambda \in \mathbb{C}$,

(iii) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$ (Minkowski's inequality).

Functional analysis is often interested in mappings between normed linear spaces. An important and simple class is that of bounded linear transformations.

Definition II.1.2. Let $(V_1, \|\cdot\|_1), (V_2, \|\cdot\|_2)$ be two normed linear spaces. A *bounded linear transformation* is a function $T : V_1 \to V_2$ such that

(i) $T(\lambda v + w) = \lambda T(v) + T(w)$ for all $v, w \in V_1, \lambda \in \mathbb{C}$

(ii) There exists $C \geq 0$ such that $\|Tv\|_2 \leq C\|v\|_1$ for all $v \in V_1$

The *norm* of T is the smallest such constant, namely

$$\|T\| = \sup\left\{\frac{\|Tv\|_2}{\|v\|_1} : v \in V_1, v \neq 0\right\}.$$

The set of all bounded linear transformations is a vector space denoted $\mathcal{L}(V_1, V_2)$, and the norm just defined is referred to as the *operator norm*. We briefly check that the triangle inequality holds:

$$\begin{aligned}
\|M + T\| &\leq \sup\left\{\frac{\|Mv\|_2 + \|Tv\|_2}{\|v\|_1} : v \in V_1, v \neq 0\right\} \\
&\leq \sup\left\{\frac{\|Mv\|_2}{\|v\|_1} : v \in V_1, v \neq 0\right\} + \sup\left\{\frac{\|Tv\|_2}{\|v\|_1} : v \in V_1, v \neq 0\right\} \\
&= \|M\| + \|T\|,
\end{aligned}$$

by the triangle inequality of the norm $\|\cdot\|_2$ and the property of the supremum.

Any normed linear space $(V, \|\cdot\|)$ is a metric space, with the metric being

$$d(v, w) = \|v - w\|.$$

If not otherwise stated, the topology on a normed linear space is always the one induced by the norm. In particular, a map $T : V_1 \to V_2$ between to normed linear spaces is continuous at v_0 if for any $\epsilon > 0$, there is $\delta > 0$ such that $\|v - v_0\|_1 < \delta$ implies $\|Tv - Tv_0\|_2 < \epsilon$ and T is continuous if it is continuous at all $v_0 \in V$.
Interestingly, linearity implies that boundedness and continuity are equivalent:

Proposition II.1.3. *Let $T : V_1 \to V_2$ be a linear transformation between two normed linear spaces $(V_1, \|\cdot\|_1), (V_2, \|\cdot\|_2)$. The following are equivalent:*

(i) *T is continuous at $v_0 \in V_1$*
(ii) *T is continuous everywhere*
(iii) *T is bounded*

Proof. (ii)$\Rightarrow$(i) is trivial. If (i) holds, there is $r > 0$ such that $\|v - v_0\|_1 < 2r^{-1}$ implies $\|Tv - Tv_0\|_2 < 1$. For any $w \in V_1$, the vector $v = \frac{w}{r\|w\|_1} + v_0$ is such that $\|v - v_0\|_1 = r^{-1}$ and so

$$\|Tw\|_2 = r\|w\|_1 \|T(v - v_0)\|_2 = r\|w\|_1 \|Tv - Tv_0\|_2 \leq r\|w\|_1,$$

which is (iii). Finally, assuming (iii), $\|Tv_1 - Tv_2\|_2 = \|T(v_1 - v_2)\|_2 \leq r\|v_1 - v_2\|_1$, so that (iii) implies (ii). $\square$

In $\mathbb{R}^n$, the closed unit ball is compact. Interestingly, this fact turns out to be characteristic of finite-dimensional normed linear spaces:

Theorem II.1.4. *Let V be a normed linear space. Then the set $\mathcal{B}_1 = \{v \in V : \|v\| \leq 1\}$ is compact if and only if V is finite dimensional.*

Proof. If V is finite dimensional, then it is isometrically isomorphic to the Euclidean space $\mathbb{C}^N$ for some $N \in \mathbb{N}$. The closed unit ball in $\mathbb{C}^N$ is compact, and hence so is the closed unit ball in V.
Reciprocally, let V be infinite-dimensional. We construct a sequence $(w_n)_{n\in\mathbb{N}}$ in $\mathcal{B}_1$ recursively as follows. Let $w_1 \in \mathcal{B}_1$ be arbitrary. Given $\{w_1, \ldots, w_n\}$, let W_n be their span, which

is finite-dimensional and hence closed. Since V is infinite-dimensional, $V \setminus W_n \neq \emptyset$. We claim that there exists $w_{n+1} \in V$ such that

$$\|w_{n+1}\| = 1, \qquad \|w_{n+1} - w\| > \frac{1}{2} \quad (w \in W_n).$$

It follows that $\|w_{j'} - w_j\| > 1/2$ for all $j, j' \in \mathbb{N}$ so that the sequence $(w_n)_{n\in\mathbb{N}}$ in $\mathcal{B}_1$ has no convergent subsequence and hence $\mathcal{B}_1$ is not compact (Recall that the norm induces a metric topology which is first countable, and compactness implies sequential compactness in first countable spaces). To prove the claim, let $x \in V \setminus W_n$. Since W_n is closed, $\delta_0 = \inf\{\|x - w\| : w \in W_n\} > 0$. In particular, there is $w_0 \in W_n$ such that $\|x - w_0\| < 2\delta_0$. We let $w_{n+1} = \frac{x-w_0}{\|x-w_0\|}$, and note that $\|w_{n+1}\| = 1$ and that

$$\inf_{w\in W_n} \|w_{n+1} - w\| = \inf_{w\in W_n} \frac{\|x - w_0 - w\|}{\|x - w_0\|} = \frac{\inf_{w\in W_n} \|x - w\|}{\|x - w_0\|} > \frac{1}{2},$$

where we simply renamed $w\|x - w_0\| \to w$ in the first equality and similarly $w - w_0 \to w$ in the second, since W_n is a linear space. $\square$

For completeness, we prove that $\delta_0 > 0$. By definition of the supremum, there is a sequence $(x_j)_{j\in\mathbb{N}}$ in W_n such that $\lim_{n\to\infty} \|v - x_j\| = \delta_0$. The sequence is bounded in the closed set W_n, hence by compactness there is a convergent subsequence: there is $w_0 \in W_n$ such that $x_{j_k} \to w_0$. We conclude by the continuity of the norm that $\|v - w_0\| = \lim_{k\to\infty} \|v - x_{j_k}\| = \delta_0$ and hence $\delta_0 > 0$ since $v \neq w_0$.

Here is one of the most important definitions of the course:

Definition II.1.5. A *Banach space* is a complete normed linear space.

Recall that a normed vector space is *complete* if every Cauchy sequence is convergent.

We start our study of Banach spaces with a equivalent characterization of completeness.

Theorem II.1.6. *A normed linear space $(V, \|\cdot\|)$ is complete if and only if every absolutely convergent series is convergent.*

Proof. Let V be complete, let $(\sum_{n=1}^N \|v_n\|)_{N\in\mathbb{N}}$ be convergent and denote $S_N = \sum_{n=1}^N v_n$ for all $N \in \mathbb{N}$. Then for any $M < N$, $\|S_N - S_M\| \leq \sum_{n=M+1}^N \|v_n\| \leq \sum_{n=M+1}^\infty \|v_n\|$, which converges to 0 as $M \to \infty$. Hence $(S_N)_{N\in\mathbb{N}}$ is Cauchy and therefore convergent. Reciprocally, let $(w_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. There are $n_1 < n_2 < \ldots$ such that $\|w_n - w_m\| < 2^{-j}$ for all

$n, m \geq n_j$. We define $z_1 = w_{n_1}$ and recursively $z_j = w_{n_j} - w_{n_{j-1}}$ for $j \geq 2$. The corresponding series is telescopic so that $\sum_{j=1}^{N} z_j = w_{n_N}$. On the other hand $\sum_{j=1}^{\infty} \|z_j\| \leq \|z_1\| + 1$, namely $\sum z_n$ is absolutely convergent. By assumption, $\sum_{j=1}^{\infty} z_j$ is convergent so that the subsequence $(w_{n_N})_{N\in\mathbb{N}}$ converges, say to w. It remains to prove that the full sequence is convergent. We have

$$\|w_n - w\| \leq \|w_n - w_{n_N}\| + \|w_{n_N} - w\|.$$

The first term vanishes because $(w_n)_{n\in\mathbb{N}}$ is Cauchy by assumption, and the second as well by the convergence of the subsequence just proved. Hence $(w_n)_{n\in\mathbb{N}}$ is convergent and $(V, \|\cdot\|)$ is complete. $\square$

2. L^p spaces

We now start a long example and discuss L^p spaces. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with a positive σ-finite measure μ, and let $1 \leq p < \infty$ (σ-finite means that Ω is a countable union of measurable sets with finite measure). Recall that

$$L^p(\Omega, \mu) = \{[f] : f : \Omega \to \mathbb{C} \text{ is measurable and } |f|^p \text{ is } \mu\text{-summable}\},$$

where $[f]$ denotes the equivalence class of functions that are equal to f μ-a.e. We shall from now on simply write $L^p(\Omega)$ since the measure is fixed. Since $x \mapsto |x|^p$ is convex (a fortiori midpoint convex) for all $p \geq 1$, we have that $|x+y|^p \leq 2^{p-1}(|x|^p + |y|^p)$ for any $x, y \in \mathbb{C}$, and hence $L^p(\Omega)$ is a vector space. It is a normed linear space when equipped with the norm

$$\|f\|_p = \Big(\int_\Omega |f(x)|^p d\mu(x)\Big)^{\frac{1}{p}}.$$

The first two properties of the norm follow immediately from the properties of the integral and the definition of the equivalence classes. We shall come back to the triangle inequality later.

The definition of $L^\infty(\Omega)$ is somewhat different:

$$L^\infty(\Omega, \mu) = \{[f] : f : \Omega \to \mathbb{C} \text{ is measurable and } \exists M \text{ s.t } |f(x)| \leq M, \mu\text{-a.e.}\}.$$

The corresponding norm, also called the *essential supremum* of f, is given by

$$\|f\|_\infty = \inf\{M : |f(x)| \leq M \text{ for } \mu\text{-almost every } x \in \Omega\}.$$

Of course, this can also be written as $\|f\|_\infty = \inf\{M : \mu(\{|f(x)| > M\}) = 0\}$. In particular, $|f(x)| \leq \|f\|_\infty$ for μ-almost every $x \in \Omega$.

The central inequality in the analysis of L^p spaces is *Jensen's inequality*. Recall that a function $J : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if $J(\lambda x + (1-\lambda)y) \leq \lambda J(x) + (1-\lambda)J(y)$. J is strictly convex at x if $J(x) < \lambda J(y) + (1-\lambda)J(z)$ whenever $x = \lambda y + (1-\lambda)z$.

THEOREM II.2.1. *Let $J : \mathbb{R} \to \mathbb{R}$ be convex and $f : \Omega \to \mathbb{R}$ be s.t. $f \in L^1(\Omega)$. Assume that $\mu(\Omega) < \infty$. Denote $\mu(f) = \mu(\Omega)^{-1} \int_\Omega f d\mu \in \mathbb{R}$. Then*

$$J(\mu(f)) \leq \mu(J \circ f).$$

If J is strictly convex at $\mu(f)$, then equality holds iff f is constant.

PROOF. By convexity, there is $a \in \mathbb{R}$ such that

$$J(t) \geq J(\mu(f)) + a(t - \mu(f)) \tag{2.1}$$

for all $t \in \mathbb{R}$. ($t \mapsto J(\mu(f)) + a(t - \mu(f))$ is called a support line of J at $\mu(f)$). Substituting $f(x)$ for t and integrating over Ω yields the first claim. If f is constant, the $f(x) = \mu(f)$ for all $x \in \Omega$ and equality holds. If J is strictly convex at $\mu(f)$, the inequality (2.1) is strict either for all $t > \mu(f)$ or for all $t < \mu(f)$. But $f(x) - \mu(f)$ takes on both positive and negative values if f is not constant. □

The following inequality due to Hölder, the importance of which in analysis cannot be overstated, is now a simple corollary of Jensen's.

THEOREM II.2.2. *Let $1 \leq p \leq q \leq \infty$ and q be such that $p^{-1} + q^{-1} = 1$. Let $f \in L^p(\Omega), g \in L^q(\Omega)$. Then $fg \in L^1(\Omega)$ and*

$$\left| \int_\Omega fg \, d\mu \right| \leq \int_\Omega |f||g| \, d\mu \leq \|f\|_p \|g\|_q.$$

The indices p, q are called *dual* when $p^{-1} + q^{-1} = 1$.

PROOF. Since $\left|\int_\Omega fg \, d\mu\right| \leq \int_\Omega |f||g| \, d\mu$, we assume w.l.o.g. that $f \geq 0$, $g \geq 0$. The cases $p = \infty$ or $q = \infty$ are immediate consequences of the properties of the integral. We now assume $1 < p, q < \infty$. Let $P = \{x \in \Omega : g(x) > 0\}$. Then $\int_\Omega g d\mu = \int_P g d\mu$ and similarly $\int_\Omega fgd\mu = \int_P fgd\mu$, while $\int_\Omega fd\mu = \int_{\Omega\setminus P} fd\mu + \int_P fd\mu \geq \int_P fd\mu$. The measure $d\nu(x) = g(x)^q d\mu(x)$ is well-defined on P and finite with $\nu(P) = \|g\|_q^q$. Let

$$F(x) = \frac{f(x)}{g(x)^{q/p}} \qquad (x \in P).$$

Now,

$$\nu(F) = \frac{1}{\|g\|_q^q}\int_P f(x)g(x)^{q-q/p}d\mu(x) = \frac{1}{\|g\|_q^q}\int_P f(x)g(x)d\mu(x)$$

since $p^{-1} = 1 - q^{-1}$. Since $J(t) = |t|^p$ is convex, we apply Jensen's inequality to get

$$\frac{\|f\|_p^p}{\|g\|_q^q} = \nu(J \circ F) \geq J(\nu(F)) = \frac{1}{\|g\|_q^{pq}}(\int_P f(x)g(x)d\mu(x))^p$$

which is the claim since p, q are dual indices. $\square$

A functional analytic point of view on this result is the following: Any function $f \in L^p(\Omega)$ defines a bounded linear map

$$T_f : L^q(\Omega) \to \mathbb{C}, \qquad T_f(g) = \int_\Omega fg\, d\mu,$$

since $|T_f(g)| \leq \|f\|_p\|g\|_q$ for all $g \in L^q(\Omega)$.

We are now equipped to prove a general version of Minkowski's inquality, which is the missing element in the proof that $\|\cdot\|_p$ are indeed norms.

Theorem II.2.3. *Let f be a nonnegative function on $\Omega \times \Upsilon$ that is $\mu \times \nu$-measurable, and let $1 \leq p < \infty$. Then*

$$\left(\int_\Omega \left(\int_\Upsilon f(x,y)d\nu(y)\right)^p d\mu(x)\right)^{\frac{1}{p}} \leq \int_\Upsilon \left(\int_\Omega f(x,y)^p d\mu(x)\right)^{\frac{1}{p}} d\nu(y) \tag{2.2}$$

In particular, the left hand side is finite whenever the right hand side is finite.

A particularly simple way of expressing the inequality is as follows: If $x \mapsto f(x,y)$ is in $L^p(\Omega,\mu)$ for ν-almost every y and if $y \mapsto \|f(\cdot,y)\|_p$ is in $L^1(\Upsilon,\nu)$, then $y \mapsto f(x,y)$ is in $L^1(\Upsilon,\nu)$ for μ-almost every x, the function $x \mapsto \int_\Upsilon f(x,y)d\nu(y)$ is in $L^p(\Omega,\mu)$ and

$$\left\|\int_\Upsilon f(\cdot,y)d\nu(y)\right\|_p \leq \int_\Upsilon \|f(\cdot,y)\|_p d\nu(y).$$

Corollary II.2.4. *For $g, h \in L^p(\Omega)$,*

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

Proof of the corollary. The identity $|g(x) + h(x)| \le |g(x)| + |h(x)|$ immediately yields the claim for $p = 1$ or $p = \infty$. The same inequality shows that it suffices to prove the claim for non-negative functions. Let $1 < p < \infty$, we apply the theorem to f defined by $f(x, 1) = |g(x)|$, $f(x, 2) = |h(x)|$ on $\Omega \times \{1, 2\}$, where $\{1, 2\}$ is equipped with the measure $\nu(\{1\}) = 1 = \nu(\{2\})$. $\square$

Proof of Theorem II.2.3. The function

$$F(x) = \int_\Upsilon f(x, y)\, d\nu(y)$$

is measurable by Fubini's theorem. We assume that $\int_\Omega F^p d\mu > 0$, since otherwise the inequality is trivially satisfied. Assuming that the left hand side of (2.2) is finite, it reads

$$\begin{aligned}
\int_\Omega F^p d\mu &= \int_\Omega \left(\int_\Upsilon f(x, y)\, d\nu(y) \right) F(x)^{p-1} d\mu(x) \\
&= \int_\Upsilon \left(\int_\Omega f(x, y) F(x)^{p-1}\, d\mu(x) \right) d\nu(y) \\
&\le \int_\Upsilon \left(\int_\Omega f(x, y)^p d\mu(x) \right)^{\frac{1}{p}} \left(\int_\Omega F^p\, d\mu \right)^{\frac{p-1}{p}} d\nu(y)
\end{aligned}$$

by Tonelli's theorem and by Hölder's inequality with $1/p + (p-1)/p = 1$. But this is the claim after dividing by $(\int_\Omega F^p d\mu)^{\frac{p-1}{p}}$. Note that if the left hand side were not finite, the argument would hold for a suitably truncated version of f, and hence the claim would follow by monotone convergence. $\square$

So far, we have proved that $L^p(\Omega)$ is a normed vector space, and that any element in $L^q(\Omega)$, where p, q are dual indices, defines a bounded linear functional on $L^p(\Omega)$. We now prove that $L^p(\Omega)$ are Banach spaces.

Theorem II.2.5. *Let $1 \le p \le \infty$. Then $L^p(\Omega)$ is complete.*

Proof. Case $1 \le p < \infty$. Let $(f_j)_{j\in\mathbb{N}}$ be so that $\sum_j f_j$ is absolutely convergent in $L^p(\Omega)$, and let $B = \sum_{j=1}^\infty \|f_j\|_p$. The sequence $G_n = \sum_{j=1}^n |f_j|$ is increasing pointwise, and let $G = \sum_{j=1}^\infty |f_j|$ (as usual in this sort of argument, $G(x)$ may be equal to $+\infty$). Moreover, $\|G_n\|_p \le \sum_{j=1}^n \|f_j\|_p \le B$ by Minkowski's inequality. Hence, monotone convergence applied to G_n^p implies that $G \in L^p(\Omega)$ and

$$\int_\Omega G^p d\mu = \lim_{n\to\infty} \int_\Omega G_n^p d\mu.$$

In particular, $G(x) < \infty$ for μ-almost every x. Furthermore, the numerical series $\sum_{j=1}^n f_j(x)$ is convergent for μ-almost every x. Let $F(x)$ be its limit. Since $|F(x)| \leq G(x)$ and $G \in L^p(\Omega)$, we have that $F \in L^p(\Omega)$. Moreover,

$$\Big|F(x) - \sum_{j=1}^n f_j(x)\Big|^p = \lim_{m\to\infty} \Big|\sum_{j=n+1}^m f_j(x)\Big|^p \leq \lim_{m\to\infty} \Big(\sum_{j=n+1}^m |f_j(x)|\Big)^p \leq G(x)^p,$$

and dominated convergence implies that

$$\lim_{n\to\infty} \|F - \sum_{j=1}^n f_j\|_p^p = \int_\Omega \lim_{n\to\infty} |F - \sum_{j=1}^n f_j|^p d\mu = 0,$$

namely $\sum_{j=1}^n f_j \to F$ in the L^p-norm, which concludes the proof with Theorem II.1.6.
The case $p = \infty$. Let $(f_j)_{j\in\mathbb{N}}$ be a Cauchy sequence in $L^\infty(\Omega)$. For each $j, k \in \mathbb{N}$, there is a set of measure zero $N_{j,k}$ such that $|f_j(x) - f_k(x)| \leq \|f_j - f_k\|_\infty$ for all $x \in \Omega \setminus N_{j,k}$. As a countable union of sets of measure zero, the set $N = \cup_{j,k\in\mathbb{N}} N_{j,k}$ has measure zero. For any $x \in \Omega \setminus N$, the sequence $(f_j(x))_{j\in\mathbb{N}}$ is Cauchy and therefore convergent, say to $f(x)$. It follows that $f_j \to f$ uniformly on $\Omega \setminus N$, and further $f_j \to f$ in the L^∞ norm. $\square$

The definition of L^p spaces is really made in order for them to be Banach. The fact that they are, strictly speaking, not sets of functions but of equivalence classes of functions is necessary for the norm to be well-defined. One may further wonder whether there could not be 'simpler' spaces of functions that would be complete. One way to see that this is not possible is the following, which shows that any $f \in L^p(\Omega)$ can be approximated in the L^p-norm by a C^∞ function. Recall that the *convolution* of two functions f, g is given by

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy$$

Proposition II.2.6. *Let $j \in C_c^\infty(\mathbb{R}^n)$ with $j \geq 0$ and $\int_{\mathbb{R}^n} j = 1$. For any $\epsilon > 0$, let*

$$j_\epsilon(x) = \epsilon^{-n} j\left(\frac{x}{\epsilon}\right).$$

Let $\Omega \subset \mathbb{R}^n$ be open and let $f \in L^p(\Omega)$ for $1 \leq p < \infty$. Let

$$f_\epsilon(x) = (j_\epsilon * \tilde{f})(x) \qquad (x \in \Omega),$$

where $\tilde{f}$ is the extension of f by 0 to $\mathbb{R}^n$. Then
(i) $f_\epsilon \in L^p(\Omega) \cap C^\infty(\mathbb{R}^n)$ and
(ii) $f_\epsilon \to f$ in $L^p(\Omega)$, as $\epsilon \to 0$.

This is a consequence of Young's inequality:

$$\|f * g\|_p \leq \|f\|_q \|g\|_r \qquad 1 + \frac{1}{p} = \frac{1}{q} + \frac{1}{r}.$$

Before coming back to the general theory of Banach spaces, we turn to another example.

Example 4. We equip $C^1([0,1])$ with the norm

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

and claim that it is a Banach space. Recall indeed that if a sequence $(f_n)_{n\in\mathbb{N}}$ of differentiable functions converges pointwise to f, and is such that $(f_n')_{n\in\mathbb{N}}$ converges uniformly to g, then $f \in C^1([0,1])$, with $f' = g$, and f_n converges uniformly to f. With this, we note that if $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(C^1([0,1]), \|\cdot\|_{W^{1,\infty}})$, then both $(f_n)_{n\in\mathbb{N}}$ and $(f_n')_{n\in\mathbb{N}}$ are Cauchy sequences with respect to $\|\cdot\|_\infty$ and hence they converge uniformly, to f, resp. g. The result above implies that $g = f'$. By induction the result would extend to $C^k([0,1])$ equipped with the norm $\|f\|_{W^{k,\infty}} = \sum_{j=0}^k \|f^{(j)}\|_\infty$.

3. Linear functionals and the Hahn-Banach theorem

Recall that $\mathcal{L}(V_1, V_2)$ is the normed vector space of bounded linear transformations between two vector spaces V_1, V_2.

Proposition II.3.1. *If V_2 is a Banach space, then so is $\mathcal{L}(V_1, V_2)$.*

Proof. Let $(T_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{L}(V_1, V_2)$. For each $v \in V_1$, the sequence $(T_n v)_{n\in\mathbb{N}}$ is Cauchy in V_2 since $\|(T_n - T_m)v\|_2 \leq \|T_n - T_m\| \|v\|_1$. Since V_2 is complete, there is $w \in V_2$ such that $\lim_{n\to\infty} T_n v = w$. This defines a map $T : V_1 \to V_2$ by $v \mapsto Tv = w$. We check that it is a bounded linear transformation and that $\|T_n - T\| \to 0$ as $n \to \infty$. Linearity follows from the linearity of the limit. Next, we note that $|\|T_n\| - \|T_m\|| \leq \|T_n - T_m\|$ so that $(\|T_n\|)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. Let C denote its limit. Then,

$$\|Tv\|_2 = \lim_{n\to\infty} \|T_n v\|_2 \leq \lim_{n\to\infty} \|T_n\| \|v\|_1 = C\|v\|_1,$$

proving the continuity of the limiting T. Finally,

$$\|T - T_n\| = \sup\left\{\frac{\|(T-T_n)v\|_2}{\|v\|_1} : 0 \neq v \in V_1\right\}$$

but $\|(T-T_n)v\|_2 = \lim_{m\to\infty}\|(T_m-T_n)v\|_2 \leq \|v\|_1 \lim_{m\to\infty}\|T_m-T_n\|$, which yields the claim since $(T_n)_{n\in\mathbb{N}}$ is Cauchy. $\square$

In the case $V_1 = V_2 = V$, the Banach space $\mathcal{L}(V,V)$ often denoted $\mathcal{L}(V)$ has an additional structure, namely an associative product given by composition. Then for any $S, T \in \mathcal{L}(V)$,

$$\|STv\|_V \leq \|S\|\|Tv\|_V \leq \|S\|\|T\|\|v\|_V$$

which shows that

$$\|ST\| \leq \|S\|\|T\|.$$

The algebra $\mathcal{L}(V)$ has a unit, namely the identity operator $v \mapsto v$. Altogether, $\mathcal{L}(V)$ is a *unital Banach algebra*.

Since $\mathbb{C}$ is a Banach space, the above shows that the space

$$V^* = \mathcal{L}(V, \mathbb{C})$$

is a Banach space for any normed linear space V, and it is called the *dual space* of V. An element of V^* is a *bounded linear functional* on V. V^* is naturally equipped with the operator norm

$$\|\ell\|_{V^*} = \sup\left\{\frac{|\ell(v)|}{\|v\|_V} : 0 \neq v \in V\right\}.$$

The topology induced by this norm is strong. Although it is useful as it makes V^* into a Banach space, it is often convenient to consider weaker topolgies on V^*. We will come back to this later.

Example 5. We have already discussed that Hölder's inequality implies $L^q(\Omega) \subset L^p(\Omega)^*$ whenever (p,q) are dual indices. Indeed: for any $f \in L^q(\Omega)$, the map $T_f(g) = \int_\Omega fg d\mu$ is a bounded linear functional $L^p(\Omega) \to \mathbb{C}$ with $\|T_f\| \leq \|f\|_q$. Since $\bar{f}|f|^{q-2} \in L^p(\Omega)$ with $\|\bar{f}|f|^{q-2}\|_p = \|f\|_q^{q/p} = \|f\|_q^{q-1}$, and $T_f(\bar{f}|f|^{q-2}) = \|f\|_q^q$, we conclude that $\|T_f\| = \|f\|_q$. In fact, they are all bounded linear functionals, provided $p < \infty$, which is the claim of the following theorem of Riesz. The case $p = \infty$ is different, in the sense that $L^1(\Omega)$ is a strict subset of $L^\infty(\Omega)^*$, while $L^1(\Omega)^* = L^\infty(\Omega)$.

First of all, we note that bounded linear functionals separate points.

Lemma II.3.2. *Let $1 \le p < \infty$. If $f \in L^p(\Omega)$ is such that $\ell(f) = 0$ for all $\ell \in L^p(\Omega)^*$, then $f = 0$.*

Of course, this implies that if $f \neq g$ in $L^p(\Omega)$, then there is $\ell \in L^p(\Omega)^*$ such that $\ell(f) \neq \ell(g)$.

Proof. Let

$$g(x) = \begin{cases} |f(x)|^{p-2}\overline{f(x)} & f(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

For $1 < p < \infty$, $f \in L^p(\Omega)$ implies $g \in L^q(\Omega)$ since $q(p-1) = p$. If $p = 1$, then $|g(x)| = 1$ whenever $f(x) \neq 0$ and 0 otherwise so that $g \in L^\infty(\Omega)$. Therefore in both cases T_g is a well-defined linear functional, so that by assumption, $0 = T_g(f) = \|f\|_p^p$. Hence $f = 0$ indeed. $\square$

Theorem II.3.3. *Let $1 < p < \infty$. Then $L^q(\Omega)$ is isometrically isomorphic to $L^p(\Omega)^*$. The same holds for $p = 1$ provided Ω is σ-finite.*

Note that an isomorphism of Banach spaces is an invertible linear map $T : V \to W$ such that both T, T^{-1} are bounded. It is isometric if $\|Tv\|_W = \|v\|_V$. Here, the isomorphism is given by $f \mapsto T_f$, which is isometric.

Proof of the theorem. *Case $p > 1$.* Let ℓ be a non-zero element of $L^p(\Omega)^*$. We explicitly construct a function $\lambda \in L^q(\Omega)$ such that

$$\ell(f) = \int_\Omega \lambda f \, d\mu. \tag{3.1}$$

Let $\mathcal{N}_\ell = \ell^{-1}(\{0\})$ be the kernel of ℓ. By continuity, $\mathcal{N}_\ell$ is closed. It is also convex: if $f, g \in \mathcal{N}_\ell$ then $\ell(\lambda f + (1-\lambda)g) = 0$ by linearity. Therefore, for any function $f \notin \mathcal{N}_\ell$, there is $h \in \mathcal{N}_\ell$ such that

$$\|f - h\|_p = \inf\{\|f - k\|_p : k \in \mathcal{N}_\ell\}.$$

(this is a fact for closed convex sets that would require a proof, but we will admit this rather intuitive fact). Let now $k \in \mathcal{N}_\ell$, and let $k(t) = (1-t)h + tk$ which is in $\mathcal{N}_\ell$ for all $t \in [0,1]$

by convexity. By definition of h, the function $[0,1] \ni t \mapsto F(t) = \|f - k(t)\|_p$ has a minimum at $t = 0$. Since it is differentiable, we must have that $F'(0) \geq 0$, namely

$$\int_\Omega |f-h|^{p-2}\left[(\overline{f}-\overline{h})(h-k)+(f-h)(\overline{h}-\overline{k})\right]d\mu \geq 0.$$

for all $k \in \mathcal{N}_\ell$ (recall that $(d/dt)\|f+tg\|_p|_{t=0} = (p/2)\int_\Omega |f|^{p-2}(\bar{f}g + f\bar{g})d\mu)$. Since $\mathcal{N}_\ell$ is a linear space and $h \in \mathcal{N}_\ell$, we conclude that

$$\operatorname{Re}\int_\Omega \varphi\tilde{k}\,d\mu \geq 0 \qquad \varphi = |f-h|^{p-2}(\overline{f}-\overline{h})$$

for all $\tilde{k} \in \mathcal{N}_\ell$. For any $k \in \mathcal{N}_\ell$, all of $\pm k, \pm ik$ are in $\mathcal{N}_\ell$, so that $\int_\Omega \varphi k\,d\mu = 0$ for all $k \in \mathcal{N}_\ell$. For any $g \in L^p(\Omega)$, let

$$g_1 = \frac{\ell(g)}{\ell(f-h)}(f-h) \quad \text{and} \quad g_2 = g - g_1,$$

which is well-defined since $\ell(f-h) = \ell(f) \neq 0$. The decomposition is so that $g_2 \in \mathcal{N}_\ell$, and hence, by the above,

$$\int_\Omega \varphi g\,d\mu = \int_\Omega \varphi g_1\,d\mu = \ell(g)\cdot I, \qquad I = \frac{1}{\ell(f-h)}\int_\Omega \varphi(f-h)\,d\mu,$$

and we note that $\int_\Omega \varphi(f-h)\,d\mu = \|f-h\|_p^p \neq 0$. Since $f,h \in L^p(\Omega)$ implies that $\varphi \in L^q(\Omega)$, the choice $\lambda = \varphi/I$ concludes the proof of the claim. To conclude, we show that λ is the unique function satisfying (3.1). Indeed, let $\lambda' \in L^q(\Omega)$ be another one. Then

$$\int_\Omega (\lambda-\lambda')g\,d\mu = 0$$

for all $g \in L^p(\Omega)$. But the choice $g = |\lambda-\lambda'|^{p-2}(\overline{\lambda}-\overline{\lambda'})$ yields $0 = \|\lambda-\lambda'\|_p^p$ and hence $\lambda = \lambda'$.

Case $p = 1$. We first assume that μ is a finite measure: $\mu(\Omega) < +\infty$. Let $\ell \in L^1(\Omega)^*$. Let $1 \leq p < +\infty$; For any $f \in L^p(\Omega)$, Hölder's inequality yields that

$$|\ell(f)| \leq C_\ell\|f\|_1 \leq C_\ell(\mu(\Omega))^{\frac{1}{q}}\|f\|_p,$$

namely the restriction of ℓ to $L^p(\Omega)$ is again bounded. By the above, for any $p > 1$, there is a unique $v_p \in L^q(\Omega)$ such that $\ell \restriction L^p(\Omega) = T_{v_p}$. By Hölder's inequality again, $L^r(\Omega) \subset L^p(\Omega)$ whenever $r \geq p$ since

$$\|f\|_p^p \leq (\mu(\Omega))^{1-\frac{p}{r}}\|f\|_r^p.$$

Therefore, the uniqueness of v_p implies that $v_r = v_p$ for all $r, p > 1$, and we denote it v. We claim that it is the function we seek. First of all, with $f = |v|^{q-2}\bar{v}$,

$$\|v\|_q^q = \ell(f) \leq C_\ell(\mu(\Omega))^{\frac{1}{q}}\|f\|_p = C_\ell(\mu(\Omega))^{\frac{1}{q}}\|v\|_q^{q-1}.$$

Hence $\|v\|_q \leq C_\ell(\mu(\Omega))^{\frac{1}{q}}$ for any $1 < q < +\infty$. Assume now that $\mu(\{x \in \Omega : |v(x)| > C_\ell + \epsilon\}) = M > 0$. Then

$$\|v\|_q \geq \Big(\int_{\{x\in\Omega:|v(x)|>C_\ell+\epsilon\}} |v(x)|^q\Big)^{\frac{1}{q}} > (C_\ell + \epsilon)M^{\frac{1}{q}}$$

and this is strictly larger than $C_\ell(\mu(\Omega))^{\frac{1}{q}}$ for q large enough since $\lim_{q\to\infty} \frac{M}{\mu(\Omega)} = 1$. This is a contradiction with the previous bound on $\|v\|_q$. Hence $M = 0$ for all $\epsilon > 0$ and we conclude that $v \in L^\infty(\Omega)$ with $\|v\|_\infty \leq C_\ell$. For any $f \in L^1(\Omega)$, then $|T_v(f)| < +\infty$. For any $f \in L^1(\Omega)$, then the functions f^n defined for any $n \in \mathbb{N}$ by $f^n(x) = f(x)$ whenever $|f(x)| \leq n$ and $f(x) = 0$ otherwise are all in $L^p(\Omega)$ for all $1 < p < +\infty$. Moreover $f^n \to f$ pointwise and $|f^n(x)| < |f(x)|$, so by dominated convergence $f^n \to f$ in $L^1(\Omega)$; Similarly, $vf^n \to vf$ in $L^1(\Omega)$. We conclude that

$$\ell(f) = \lim_{n\to\infty} \ell(f^n) = \lim_{n\to\infty} \int_\Omega vf^n d\mu = \int_\Omega vf d\mu$$

where the first bound is by continuity of ℓ (w.r.t. the $L^1(\Omega)$ norm), the second equality is because $f^n \in L^p(\Omega)$ and the third is by dominated convergence. Hence $\ell(f) = T_v(f)$ indeed. For the case of a σ-finite space Ω, the argument can be repeated on every Ω_j, where $\Omega = \cup_{j=1}^\infty \Omega_j$, with $f = \sum_{j=1}^\infty \chi_j f$, where χ_j is the characteristic function of Ω_j. The proof then proceeds with functions v_j and $v = \sum_{j=1}^\infty v_j$ and the observation that $\|v_j\|_\infty \leq C$ is independent of j. $\square$

We note that in the cases $1 < p < \infty$, the above implies that

$$L^p(\Omega)^{**} = (L^p(\Omega)^*)^* \simeq L^p(\Omega).$$

A space that is equal to its bidual is called reflexive.

More precisely, let V be a normed vector space. Then V is embedded in V^{**} through $\mathcal{I} : V \to V^{**}$ given by

$$\mathcal{I}(v)(\ell) = \ell(v) \tag{3.2}$$

for all $\ell \in V^*, v \in V$. As a consequence of the Hahn-Banach theorem to come, $\mathcal{I}$ is an isometry, namely $\|\mathcal{I}(v)\|_{V^{**}} = \|v\|_V$ for all $v \in V$. The space V is *reflexive* if $\mathcal{I}$ is surjective. Note further that since $V^{**} = (V^*)^*$ is complete, the space V is complete whenever it is reflexive. Finally, we point out that it is common to identify $\mathcal{I}(v)$ with v and in the case of a reflexive space V with V^{**} although this leads to abuse of notations.

We now turn to one of the pillars of functional analysis, the *Hahn-Banach theorem*. There are various versions of it, and many rather immediate corollaries that are very useful. Vaguely put, it allows for the extension of a linear functional defined on a subset of a Banach space to the whole of the space. It is however non-constructive and it requires the axiom of choice. We first recall Zorn's lemma.

A relation on a set S that is reflexive, transitive and antisymmetric is called a *partial order*. We denote it by $x \prec y$. 'Partial' refers here to the fact that a pair x, y of elements of S does not need to satisfy $x \prec y$ or $y \prec x$. A *linearly ordered* set is such that for any pair x, y, either $x \prec y$ or $y \prec x$. An element $m \in S$ is a *maximal element* if $m \prec x$ implies $m = x$. Finally, an element $p \in S$ is an *upper bound* for $X \subset S$ if $x \prec p$ for all $x \in X$.

Zorn's Lemma. *Let S be a nonempty partially ordered set such that every linearly ordered subset has an upper bound in S. Then each linearly ordered subset has an upper bound that is also a maximal element of S.*

We start with the real version of Hahn-Banach.

THEOREM II.3.4. *Let X be a real vector space, let $p : X \to \mathbb{R}$ be a convex function. Let $Y \subset X$ be a subspace, and let $\lambda : Y \to \mathbb{R}$ be a real linear functional such that $\lambda(x) \leq p(x)$ for all $x \in Y$. Then there exists a real linear functional $\ell : X \to \mathbb{R}$ such that $\ell(x) = \lambda(x)$ whenever $x \in Y$ and*

$$\ell(x) \leq p(x)$$

for all $x \in X$.

PROOF. We present the proof in the setting of p being sublinear, namely $p(x + y) \leq p(x) + p(y)$ and $p(ax) = ap(x)$ for all $x, y \in X$ and $a > 0$.

Step 1. Extending λ along one direction. Let $z \in X \setminus Y$, and let $\tilde{Y} = \{az + y : a \in \mathbb{R}, y \in Y\}$.

We shall construct $\tilde{\lambda}(z)$ and define

$$\tilde{\lambda}(az + y) = a\tilde{\lambda}(z) + \lambda(y).$$

For any $y_1, y_2 \in Y$, and any $a, b > 0$,

$$a\lambda(y_1) + b\lambda(y_2) = (a+b)\lambda\left(\frac{a}{a+b}y_1 + \frac{b}{a+b}y_2\right).$$

We now bound λ by p, and since the latter is defined everywhere, we can add and subtract $ab/(a+b)z$ in its argument. By convexity, we then obtain

$$a\lambda(y_1) + b\lambda(y_2) \le ap(y_1 - bz) + bp(y_2 + az),$$

or equivalently

$$b^{-1}(\lambda(y_1) - p(y_1 - bz)) \le a^{-1}(p(y_2 + az) - \lambda(y_2)).$$

Hence, there exists a (not necessarily unique) real number c such that

$$\sup\left\{b^{-1}(\lambda(y_1) - p(y_1 - bz)) : b > 0, y_1 \in Y\right\}$$
$$\le c \le \inf\left\{a^{-1}(p(y_2 + az) - \lambda(y_2)) : a > 0, y_2 \in Y\right\}$$

and we define $\tilde{\lambda}(z) = c$. The second inequality implies that $\tilde{\lambda}(x) \le p(x)$ for any $x = az + y \in \tilde{Y}$ with $a > 0$,

$$\tilde{\lambda}(az + y) = a(\tilde{\lambda}(z) + \lambda(y/a)) \le a\big(p(y/a + z) - \lambda(y/a) + \lambda(y/a)\big) = p(y + az),$$

where we used the positive homogeneity of g, while the first one does in the case $a < 0$.
Step 2. Extending λ to all of X. The set

$$\mathcal{S} = \{(V, \ell) : V \subset X \text{ is a subspace, and } \ell : V \to \mathbb{R} \text{ is linear and } \ell(x) \le p(x)\}$$

which is not empty since $(Y, \lambda) \in \mathcal{S}$, is equipped with the partial order

$$(V, \ell) \prec (V', \ell') \text{ iff } V \subset V' \text{ and } \ell(x) = \ell'(x) \text{ for all } x \in V,$$

namely ℓ' extends ℓ from V to V'. For any $\mathcal{S}_{\mathcal{I}} = \{(V_\alpha, \ell_\alpha) : \alpha \in \mathcal{I}\}$ linearly ordered set in $\mathcal{S}$, the element

$$\left(\bigcup_{\alpha \in \mathcal{A}} V_\alpha, \tilde{\ell}\right), \qquad \tilde{\ell}(x) = \ell_\alpha(x) \text{ whenever } x \in V_\alpha,$$

is a well-defined upper bound of $\mathcal{S}_{\mathcal{I}}$ in $\mathcal{S}$. Zorn's lemma now implies that there exists a maximal element (X', ℓ) of $\mathcal{S}$. In fact, $X' = X$, since otherwise ℓ could be extended by Step 1. $\square$

The complex version of Hahn-Banach is now a simple corollary.

THEOREM II.3.5. *Let X be a complex vector space, let $p : X \to \mathbb{R}$ be a function such that*

$$p(\alpha x + \beta y) \leq |\alpha| p(x) + |\beta| p(y) \qquad \alpha, \beta \in \mathbb{C},\ |\alpha| + |\beta| = 1.$$

Let $Y \subset X$ be a subspace, and let $\lambda : Y \to \mathbb{C}$ be a complex linear functional such that $|\lambda(x)| \leq p(x)$ for all $x \in Y$. Then there exists a complex linear functional $\ell : X \to \mathbb{C}$ such that $\ell(x) = \lambda(x)$ whenever $x \in Y$ and

$$|\ell(x)| \leq p(x)$$

for all $x \in X$.

PROOF. Let $\Lambda(x) = \mathrm{Re}\lambda(x)$, which is real linear. Since

$$\Lambda(\mathrm{i}x) = \mathrm{Re}(\mathrm{i}\lambda(x)) = -\mathrm{Im}\lambda(x),$$

we have that $\lambda(x) = \Lambda(x) - \mathrm{i}\Lambda(\mathrm{i}x)$. Now, Λ is bounded by p on Y and p is convex (for real α, β) so that it has a real linear extension $L \leq p$ on X (here X and Y are both seen as real vector spaces). The linear functional $\ell(x) = L(x) - \mathrm{i}L(\mathrm{i}x)$ extends λ and it is complex linear since $\ell(\mathrm{i}x) = \mathrm{i}\ell(x)$. Finally, let $x \in X$ and $\alpha = \ell(x)/|\ell(x)|$. Then $|\ell(x)| = \bar{\alpha}\ell(x) = \ell(\bar{\alpha}x)$, and since this is real, we conclude that

$$|\ell(x)| = L(\bar{\alpha}x) \leq p(\bar{\alpha}x) \leq p(x),$$

by the assumption on p since $|\bar{\alpha}| = 1$. $\square$

Note that the Hahn-Banach theorem does not require the full structure of a Banach space. However, if X is a normed vector space, then the norm itself and related functions are good p-functions. This yields a number of useful corollaries, valid both in the real and complex case.

COROLLARY II.3.6. *Let X be a normed vector space and let Y be a subspace. Let $\lambda \in Y^*$. There exists $\ell \in X^*$ such that $\lambda(x) = \ell(x)$ for $x \in Y$ and $\|\ell\|_{X^*} = \|\lambda\|_{Y^*}$.*

Proof. Apply H-B to $p(x) = \|\lambda\|_{Y^*}\|x\|_X$. Since ℓ is an extension of λ, we have $\|\ell\|_{X^*} \geq \|\lambda\|_{Y^*}$. On the other hand by H-B $|\ell(x)| \leq \|\lambda\|_{Y^*}\|x\|_X$, namely $\|\ell\|_{X^*} \leq \|\lambda\|_{Y^*}$. $\square$

Corollary II.3.7. *Let X be a normed vector space, let $x \in X$ and $\zeta \in \mathbb{C}$. There exists $\ell \in X^*$ such that $\ell(x) = \zeta\|x\|_X$ and $\|\ell\|_{X^*} = |\zeta|$.*

Proof. Follows from the previous corollary with $Y = \{ax : a \in \mathbb{C}\}$ and $\lambda(ax) = a\zeta\|x\|_X$, for which $\|\lambda\|_{Y^*} = |\zeta|$. $\square$

This implies that bounded linear functionals separate points in X:

Corollary II.3.8. *Let X be a normed vector space. For any $y_1 \neq y_2 \in X$, there exists $\ell \in X^*$ such that $\ell(y_1) \neq \ell(y_2)$*

Proof. This follows from the previous corollary with $\zeta = 1$, and $x = y_1 - y_2 \neq 0$. Then, which implies $\ell(y_1) - \ell(y_2) = \ell(x) = \|x\| \neq 0$. $\square$

Finally, the last result shows that the norm in a normed vector space can be computed using linear functionals, which is often a very useful tool.

Corollary II.3.9. *Let X be a normed vector space. For all $x \in X$,*

$$\|x\|_X = \sup\left\{|\ell(x)| : \ell \in X^*, \|\ell\|_{X^*} = 1\right\}.$$

Proof. By Corollary II.3.7 with $\zeta = 1$, there is $\ell \in X^*$ such that $\ell(x) = \|x\|_X$ and $\|\ell\|_{X^*} = 1$ proving $\leq$ above. The inequality $\geq$ is by definition of the norm in X^*, since $|\ell(x)| \leq \|\ell\|_{X^*}\|x\|_X$. $\square$

4. The Baire category theorem and its corollaries

We now turn to the second pillar of functional analysis and its corollaries, the principle of uniform boundedness, Corollary II.4.3 and the open mapping theorem, Corollary II.4.4.
A subset S of a metric space M is *nowhere dense* if $(\overline{S})^o = \emptyset$. Since, for any set $(X^o)^c = \overline{X^c}$, we conclude that $\overline{(\overline{S})^c} = ((\overline{S})^o)^c = M$, namely, $(\overline{S})^c$ is dense. For example, $\mathbb{Z}$ is nowhere dense in $\mathbb{R}$; so is the Cantor set.
Recall further that $D \subset M$ is dense if $\overline{D} = M$, and recall that $\overline{D}$ is the set of $x \in M$ such that $N_x \cap D \neq \emptyset$ for any open neighbourhood N_x of x.

Lemma II.4.1. *$D \subset M$ is dense if and only if $D \cap O \neq \emptyset$ for every non-empty open set O.*

Proof. If D is dense and O is open and not empty, then for any $x \in O$, we have that $x \in \overline{D}$. Hence, every open neighbourhood of x intersects D, in particular O itself. Reciprocally, assume that $D \cap O \neq \emptyset$ for every non-empty open set O. For any $x \in M$, let N_x be an open neighbourhood of x (in particular N_x is not empty) and hence $N_x \cap D \neq \emptyset$. It follows that $x \in \overline{D}$. □

Theorem II.4.2. *Let M be a complete metric space.*
(i) If $(U_n)_{n\in\mathbb{N}}$ is a sequence of open, dense sets in M, then $\cap_{n\in\mathbb{N}} U_n$ is dense in M.
(ii) M is not a countable union of nowhere dense sets.

Proof. Let $S \subset M$ be a nonempty open set. Since U_1 is dense, $U_1 \cap S$ is open and non-empty, so there is an open metric ball $B_{r_1}(x_1) \subset U_1 \cap S$ with $r_1 < 1/2$. Inductively, there are balls $B_{r_n}(x_n)$ with $r_n < 1/2^n$ such that $\overline{B_{r_n}(x_n)} \subset U_n \cap B_{r_{n-1}}(x_{n-1})$. By construction, the sequence $(x_n)_{n\in\mathbb{N}}$ is Cauchy since for any $n, m > N$, $x_n, x_m \in B_{r_N}(x_N)$. Hence it is convergent and let x be its limit. For any $N \in \mathbb{N}$,

$$x \in \overline{B_{r_N}(x_N)} \subset U_N \cap B_{r_1}(x_1) \subset U_N \cap S,$$

so that $S \cap (\cap_{n\in\mathbb{N}} U_n) \neq \emptyset$. Since S was arbitrary, (i) is proved by Lemma II.4.1.
(ii) Let now $(V_n)_{n\in\mathbb{N}}$ be a sequence of nowhere dense sets. Then $((\overline{V_n})^c)_{n\in\mathbb{N}}$ is a sequence of open, dense sets, and so their intersection is dense in M, in particular nonempty. Hence,

$$\cup_{n\in\mathbb{N}} V_n \subset \cup_{n\in\mathbb{N}} \overline{V_n} = (\cap_{n\in\mathbb{N}} \overline{V_n}^c)^c \neq M,$$

concluding the proof. □

In other words, if $M = \cup_{n\in\mathbb{N}} U_n$, then at least one of $\overline{U_n}$ must have a nonempty interior.
The name of the theorem comes from the following: A set is called meager or of the first category if it is a countable union of nowhere dense sets; otherwise it is of the second category. Baire's theorem shows that a complete metric space is of the second category. In the corollaries below, we could replace the assumption of spaces being Banach by the spaces being of the second category.
We now turn to the Principle of Uniform Boundedness.

Corollary II.4.3. *Let X be a Banach space, Y a normed linear space, and let $\mathcal{F}$ be a family of bounded linear transformations from X to Y. If, for each $x \in X$, the set $\{\|Tx\|_Y : T \in \mathcal{F}\}$ is bounded, then the set $\{\|T\| : T \in \mathcal{F}\}$ is bounded.*

In other words, if there is a bound on $\|Tx\|_Y$ that is uniform in x, pointwise in T (that is just the boundedness of T) and a bound on $\|Tx\|_Y$ that is uniform in T, pointwise in x, then there is a bound that is uniform in (x, T), hence the name of the theorem.

Proof. For $n \in \mathbb{N}$, let $E_n = \{x : \|Tx\| \leq n, \forall T \in \mathcal{F}\}$, which is a closed set. By assumption, for each $x \in X$, there is n_x such that $x \in E_n$ for all $n \geq n_x$, namely $X = \cup_{n \in \mathbb{N}} E_n$. The Baire category theorem implies that there is n_0 such that E_{n_0} has nonempty interior. Let $\overline{B_r(x_0)} \subset E^o_{n_0}$. If $x \in \overline{B_r(0)}$, then $x + x_0 \in \overline{B_r(x_0)}$ and hence

$$\|Tx\| \leq \|T(x + x_0)\| + \|Tx_0\| \leq 2n_0,$$

namely $\overline{B_r(0)} \subset E_{2n_0}$. In other words, $\|x\| \leq r$ implies $\|Tx\| \leq 2n_0$, hence $\|T\| \leq \frac{2n_0}{r}$. $\square$

This implies for example that if X, Y are both Banach spaces and $b : X \times Y \to \mathbb{C}$ is bilinear and separately continuous, then b is jointly continuous. It suffices to prove continuity at $(0, 0)$. Let $(x_n, y_n) \to (0, 0)$ as $n \to \infty$ and let $T_n(y) = b(x_n, y)$. Since $b(x_n, \cdot)$ is continuous, $\{T_n : n \in \mathbb{N}\}$ is a family of bounded linear functionals. Since $x_n \to 0$, and $b(\cdot, y)$ is continuous, $\{|T_n(y)| : n \in \mathbb{N}\}$ is bounded for each $y \in Y$. The principle of uniform boundedness implies that there exists C such that

$$|T_n(y)| \leq C\|y\|$$

uniformly in n and hence

$$|b(x_n, y_n)| = |T_n(y_n)| \leq C\|y_n\| \to 0,$$

as $n \to 0$.

This is of course a property that arises from linearity, as it is well-known not to hold for example for functions $f : \mathbb{R}^2 \to \mathbb{R}$.

Corollary II.4.4. *Let X, Y be Banach spaces, and let $T \in \mathcal{L}(X, Y)$ be surjective. For any open set $S \subset X$, $T(S)$ is open in Y.*

In other words, a surjective bounded linear map between Banach spaces is an open map.

PROOF. Let B_0^X be the open unit ball. We first claim that $T(B_0^X)$ contains an open ball around $0 \in Y$. Let B_1^X be the open ball of radius $1/2$ around $0 \in X$. Since $X = \cup_{n\in\mathbb{N}} nB_1^X$ (here $\lambda A = \{\lambda x : x \in A\}$) and T is surjective and linear

$$Y = T(X) = \bigcup_{n\in\mathbb{N}} nT(B_1^X).$$

By Baire's theorem, we conclude that there is n_0 such that $\overline{n_0 T(B_1^X)}^{\,o}$ is nonempty. In particular, it contains an open ball and so does $\overline{T(B_1^X)}$, namely there is $\epsilon > 0, y_0 \in Y$ such that

$$B_\epsilon^Y(y_0) \subset \overline{T(B_1^X)}, \tag{4.1}$$

or equivalently $B_\epsilon^Y \subset \overline{T(B_1^X)} - y_0$. Let now $y \in \overline{T(B_1^X)} - y_0$, namely $y + y_0 \in \overline{T(B_1^X)}$ as well as $y_0 \in \overline{T(B_1^X)}$. There are sequences $(x_j')_{j\in\mathbb{N}}$ and $(x_j'')_{j\in\mathbb{N}}$ in B_1^X such that

$$Tx_j' \to y_0, \qquad Tx_j'' \to y_0 + y \qquad (j \to \infty).$$

We have that $x_j = x_j'' - x_j' \in B_0^X$, and of course $Tx_j \to y$ as $j \to \infty$. It follows that $y \in \overline{T(B_0^X)}$. Since this holds for all such y, we conclude that $\overline{T(B_1^X)} - y_0 \subset \overline{T(B_0^X)}$, and furthermore $B_\epsilon^Y(0) \subset \overline{T(B_0^X)}$, see (4.1). If B_n^X denotes the ball of radius 2^{-n}, linearity implies that $\overline{T(B_n^X)} = 2^{-n}\overline{T(B_0^X)}$, and hence

$$B_{2^{-n}\epsilon}^Y \subset \overline{T(B_n^X)}. \tag{4.2}$$

We finally show that $B_{\epsilon/2}^Y \subset T(B_0^X)$ (no closure!). Let $y \in B_{\epsilon/2}^Y$. By the above, $B_{\epsilon/2}^Y \subset \overline{T(B_1^X)}$. In particular, there is $x_1 \in B_1^X$ such that

$$\|y - Tx_1\| < \epsilon/4.$$

We now assume inductively that there are $x_1, \ldots, x_{n-1}$ such that $x_j \in B_j^X$ and

$$\Big\|y - \sum_{j=1}^{n-1} Tx_j\Big\| < 2^{-n}\epsilon, \tag{4.3}$$

namely, the left hand side belongs to $B_{2^{-n}\epsilon}^Y$. By (4.2), there is $x_n \in B_n$ such that $\|(y - \sum_{j=1}^{n-1} Tx_j) - Tx_n\| < 2^{-(n+1)}\epsilon$. With this, the sequence $S_n = \sum_{j=1}^n x_j$ is Cauchy, hence convergent, say to x. In fact, $\|x\| \leq \sum_{j=1}^\infty \|x_j\| < \sum_{j=1}^\infty 2^{-j} = 1$, namely $x \in B_0$. Moreover, $TS_n \to Tx$ since T is continuous, and (4.3) shows that $y = Tx$ indeed.

Let now $O \subset X$ be open and let $y \in T(O)$, namely $y = Tx$ for $x \in O$. There is $r > 0$ such that $B_r^X(x) \subset O$, or equivalently $x + B_r^X(0) \subset O$. By linearity, this implies that $y + T(B_r^X(0)) \subset T(O)$. By the above, there is $\delta > 0$ such that $B_\delta^Y(0) \subset T(B_r^X(0))$, and hence $B_\delta^Y(y) = y + B_\delta^Y(0) \subset T(O)$. Since y is arbitrary, this proves that $T(O)$ is open. $\square$

As discussed earlier, a bijective map being open is equivalent to its inverse being continuous. We therefore immediately get the following inverse mapping theorem.

Corollary II.4.5. *Let V, W be Banach spaces and $T \in \mathcal{L}(V, W)$ be bijective. Then $T^{-1} \in \mathcal{L}(W, V)$.*

Finally, we discuss the closed graph theorem, which is the last important corollary of the Baire category theorem. For any two normed linear spaces V, W and any mapping $T : V \to W$, the *graph* of T is the set

$$\Gamma(T) = \{(v, w) \in V \times W : w = Tv\}.$$

We equip $V \times W$ with the norm $\|(v, w)\| = \|v\| + \|w\|$.

Corollary II.4.6. *Let V, W be Banach spaces and $T : V \to W$ be a linear map. Then T is bounded if and only if $\Gamma(T)$ is closed.*

Note that T is implicitly assumed to be defined on all of V.

Proof. Assume first that $\Gamma(T)$ is closed. Since T is linear, $\Gamma(T)$ is a subspace of $V \times W$, and since it is closed it is complete. The projections $\pi_1(v, Tv) = v$ and $\pi_2(v, Tv) = Tv$ are continuous since

$$\|\pi_1(v, Tv)\| = \|v\| \leq \|v\| + \|Tv\|, \qquad \|\pi_2(v, Tv)\| = \|Tv\| \leq \|v\| + \|Tv\|,$$

and π_1 is a bijection. Hence its inverse is bounded. But then $T = \pi_2 \circ \pi_1^{-1}$ is bounded. Reciprocally, let $T \in \mathcal{L}(V, W)$ and let $(v_n, w_n)_{n \in \mathbb{N}}$ be a sequence of elements of $\Gamma(T)$ that is convergent, with $(v, w) = \lim_{n\to\infty}(v_n, w_n)$. By continuity $w = \lim_{n\to\infty} w_n = \lim_{n\to\infty} Tv_n = Tv$, and hence $(v, w) \in \Gamma(T)$. Hence $\Gamma(T)$ is closed. $\square$

In principle, continuity of T requires that $v_n \to v$ implies $Tv_n \to w$ and $w = Tv$. With the closed graph theorem, it suffices to show that $v_n \to v$ and $Tv_n \to w$ imply $w = Tv$, which is a simpler task.

Let us briefly make an excursion into unbounded linear operators. A linear operator T between two normed vector spaces is *closed* if $\Gamma(T)$ is closed. The above theorem shows that if T is closed and unbounded, then it cannot be defined on all of V. Such operators are in fact very common. Let us consider $V = C^0([0,1];\mathbb{R})$ equipped with $\|\cdot\|_\infty$, and $T = d/dx$ defined on $D(T) = C^1([0,1];\mathbb{R})$. Let $(f_n)_{n\in\mathbb{N}}$ be the sequence $f_n(x) = x^n$. Then $\|f_n\|_\infty = 1$ for all $n \in \mathbb{N}$ but

$$\|Tf_n\|_\infty = n\|f_{n-1}\|_\infty = n,$$

proving that $\sup\{\|Tf\|_\infty/\|f\|_\infty : f \in D(T)\} = \infty$, namely T is unbounded. However, let (f_n, Tf_n) be a convergent sequence in $D(T) \times V$, and let (f, g) be its limit. Then $g = Tf$, namely $\Gamma(T)$ is closed indeed.

We can now extend Corollary II.4.5 to unbounded operators.

Theorem II.4.7. *Let $T : D(T) \subset V \to W$ be a linear, closed and bijective map. There exists $S \in \mathcal{L}(W, V)$ such that*

$$TS = 1_W, \qquad ST = 1\restriction_{D(T)}.$$

Proof. As in the proof of Corollary II.4.6 with the projections being defined on $\Gamma(T)$. In particular, $\pi_1 : \Gamma(T) \to D(T)$ and $\pi_2 : \Gamma(T) \to W$ and both are again bounded and bijective. It follows that π_2^{-1} is bounded and we let $S = \pi_1 \circ \pi_2^{-1}$. □

We conclude the example of $T = d/dx$. We consider a slightly limited domain

$$\tilde{D}(T) = \{f \in C^1([0,1];\mathbb{R}) : f(0) = 0\},$$

on which T remains closed, but $T : \tilde{D}(T) \to C^0([0,1];\mathbb{R})$ now injective. It is also surjective: if $g \in C^0([0,1];\mathbb{R})$, then $\int_0^x g(y)dy \in \tilde{D}(T)$ and $g(x) = T\int_0^x g(y)dy$. This also shows that its inverse is given by

$$(Sg)(x) = \int_0^x g(y)dy.$$

It is easy to check that S is bounded indeed since

$$\|Sg\|_\infty \leq \sup\{x \sup\{|g(y)| : y \in [0,x]\} : x \in [0,1]\} \leq \sup\{|g(x)| : x \in [0,1]\}. = \|g\|_\infty.$$

5. Weak topologies, the Banach-Alaoglu theorem and corollaries

We turn to one of the main reasons to discuss non-metric topologies in the first part, namely weak topologies on Banach spaces and the Banach-Alaoglu theorem.

DEFINITION II.5.1. Let V be a Banach space. The V^*-weak topology on V is usually referred to as *weak topology*, and it is the weakest topology on V such that every bounded linear functional $\ell : V \to \mathbb{C}$ is continuous.

Note that by definition of V^*, every element is continuous with respect to the metric topology induced by the norm. The weak topology is the weakest topology on V with respect to which this still holds. It is generated by sets of the form $\ell^{-1}(B_\epsilon(z))$, with $\ell \in V^*$ and $z \in \mathbb{C}, \epsilon > 0$. A neighbourhood base at v_0 is given by sets

$$N_{v_0}(\ell_1, \dots, \ell_n, \epsilon) = \{v \in V : |\ell_j(v) - \ell_j(v_0)| < \epsilon; 1 \leq j \leq n\}, \qquad \ell_1, \dots, \ell_n \in V^*, \epsilon > 0.$$

Importantly, a sequence $(v_n)_{n\in\mathbb{N}}$ *converges weakly* if and only if

$$\ell(v_n) \to \ell(v) \qquad (n \to \infty)$$

for any $\ell \in V^*$. Weak convergence is usually denoted $v_n \rightharpoonup v$. As per (iv) below, weak limits are unique.

PROPOSITION II.5.2. *(i) If V is infinite dimensional, the weak topology is not metrizable.*
(ii) The weak topology is weaker than the norm topology.
(iii) Weakly convergent sequences are norm bounded.
(iv) The weak topology is Hausdorff.

PROOF. We only prove (ii-iv). (i) follows from the fact that the weak topology is first countable if and only if V is finite dimensional. (ii) follows by definition, since any $\ell \in V^*$ is contiuous in the norm topology.
(iii) Let $(v_n)_{n\in\mathbb{N}}$ be a weakly convergent sequence. Let $V_n \in V^{**}$ be defined by

$$V_n(\ell) = \ell(v_n).$$

By assumption, the set $\{|V_n(\ell)| : n \in \mathbb{N}\}$ is bounded for any $\ell \in V^*$. By the principle of uniform boundedness, the set $\{\|V_n\|_{V^{**}} : n \in \mathbb{N}\}$ is bounded, which concludes the proof since

$\|V_n\|_{V^{**}} = \sup\{|\ell(v_n)| : \ell \in V^*, \|\ell\|_{V^*} = 1\} = \|v_n\|_V$ by Hahn-Banach.
(iv) Since linear functionals separate, for any $v \neq w$ in V, there is $\ell \in V^*$ such that $\ell(v) \neq \ell(w)$. Hence there is $\epsilon > 0$ such that $B_\epsilon(\ell(v)) \cap B_\epsilon(\ell(w)) = \emptyset$. The preimages under ℓ of these discs are open in V, disjoint and contain v, respectively w. $\square$

Remark II.5.3. (i) If V is infinite dimensional, then the weak topology is strictly weaker that the norm topology. For example, the weak closure of the unit sphere is in this case the whole unit ball.
(ii) Let $(v_n)_{n\in\mathbb{N}}$ be a weakly convergent sequence and let v be its limit. Then by Corollary II.3.7, there is $\ell \in V^*$ such that $\|\ell\| = 1$ and $\ell(v) = \|v\|$ so that

$$\|v\| = |\ell(v)| = \liminf_{n\to\infty} |\ell(v_n)| \leq \liminf_{n\to\infty} \|v_n\|. \tag{5.1}$$

(iii) It is sometimes easier to establish $\ell(v_n) \to \ell(v)$ only on a dense subset D of V^*. We claim that it is sufficient to prove weak convergence, provided $\{\|v_n\| : n \in \mathbb{N}\}$ is bounded. Indeed, let $\ell \in V^*$ and $(\ell_n)_{n\in\mathbb{N}}$ be a sequence in D converging to ℓ. Then

$$|\ell(v_n) - \ell(v)| \leq |\ell(v_n) - \ell_j(v_n)| + |\ell_j(v) - \ell(v)| + |\ell_j(v_n) - \ell_j(v)|.$$

The first two terms are bounded by $\sup\{\|v\|+\|v_n\| : n \in \mathbb{N}\}\|\ell-\ell_j\|$ and the last one vanishes as $n \to \infty$ by the above so that

$$\limsup_{n\to\infty} |\ell(v_n) - \ell(v)| \leq C\|\ell - \ell_j\|$$

which converges to zero as $j \to \infty$.
(iv) Some comments on weak convergence in L^p-spaces.
Let $g \in C_c^\infty(\mathbb{R})$ and let $f_n(x) = g(x+n)$. Then $\|f_n\|_p = \|g\|_p$ for all $n \in \mathbb{N}$ and in particular $(f_n)_{n\in\mathbb{N}}$ does not converge to zero in the norm topology. Moreover, for any $h \in C_c^\infty(\mathbb{R})$, we see that $\int_\mathbb{R} hf_n = 0$ for n large enough since the supports are eventually disjoint. Since $C_c^\infty(\mathbb{R})$ is dense in $L^q(\mathbb{R})$, we conclude that $f_n \rightharpoonup 0$ by the above remark. This 'escape to infinity' is the first type of possible mechanisms by which $(f_n)_{n\in\mathbb{N}}$ converges weakly but not strongly. We briefly discuss the other two. The second mechanism is related to 'oscillation to infinity', and we use a priori knowledge of Fourier analysis. Any function $f \in L^2((-\pi,\pi);\mathbb{R})$

has a Fourier representation as

$$\|f\|_2^2 = 2\pi \sum_{n=-\infty}^{+\infty} (s_n^2 + c_n^2), \qquad s_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx, \quad c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx.$$

In particular, $\lim_{n\to\infty} s_n \to 0$. Since L^2 is its own dual, this shows that the sequence $(\sin(nx))_{n\in\mathbb{N}}$ converges weakly to 0. However, $\int_{-\pi}^{\pi}\sin^2(nx)dx = \pi$, showing again that the sequence does not converge to zero in the norm topology. Note that the same holds in any L^p space, $1 < p < \infty$. The third general type of weak but not strong convergence is concentration. Let $g \in C_c^\infty(\mathbb{R})$ and let $f_n(x) = n^{1/p}g(nx)$. Then $\|f_n\|_p = \|g\|_p$ so that f_n does not converge strongly to zero. However, for any $h \in C_c^\infty(\mathbb{R})$,

$$\int_{\mathbb{R}} h(x)f_n(x)dx = n^{\frac{1}{p}}\int_{\mathbb{R}} h(x)g(nx)dx = n^{\frac{1}{p}-1}\int_{\mathbb{R}} h(y/n)g(y)dy \to 0.$$

Indeed, the integral converges to $h(0)\int_{\mathbb{R}} g$ by dominated convergence, and $1/p-1 = -1/q < 0$. Again, this shows that $f_n \rightharpoonup 0$ by density of $C_c^\infty(\mathbb{R})$ in $L^q(\mathbb{R})$.

A similar construction provides a topology on V^*. Indeed any $v \in V$ is a linear functional on V^* through $\ell \mapsto \ell(v)$. This family of functionals provides the weak-* topology.

Definition II.5.4. Let V be a Banach space. The *weak-* topology* is the weakest topology on V^* such that every map $\ell \mapsto \ell(v), v \in V$ is continuous.

Convergence in the weak-* topology is denoted $\ell_n \overset{*}{\rightharpoonup} \ell$.

Let us quickly comment on terminology. We shall see later that the dual of $C_0(X)$, the space of continuous functions vanishing at infinity on a LCH space X, is isomorphic to the space $M(X)$ of complex Radon measures. If we equip $M(X)$ with the weak-* topology, a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ converges if and only if $\int_X f d\mu_n \to \int_X f d\mu$ for any $f \in C_0(X)$. In probability, this topology is sometimes referred to as the vague topology. One further speaks of 'weak convergence' of measures, which is really the 'weak-* convergence' of measures.

One of the reasons of introducing the weak-* topology is the following Banach-Alaoglu theorem. It shows that while the unit ball is not compact in the norm topology, see Theorem II.1.4, it is in the weak-* topology. The proof relies on *Tychonoff's theorem*, which itself is about compactness. Let $\{S_\alpha : \alpha \in I\}$ be a family of sets, and let $S = \bigtimes_{\alpha\in I} S_\alpha$. Let

$\pi_\alpha : S \to S_\alpha$ be the canonical projection, and let $\Pi = \{\pi_\alpha : \alpha \in I\}$. The *product topology* on S is Π-weak topology, making all canonical projections continuous.

THEOREM II.5.5. *Let $\{X_\alpha : \alpha \in I\}$ be a collection of compact sets. Then $\bigtimes_{\alpha\in I} X_\alpha$ is compact in the product topology.*

With this,

THEOREM II.5.6. *Let V be a normed vector space. The unit ball in V^* is weakly-* compact.*

The product topology is the natural topology to study functionals in V^*. Indeed, if for $v \in V$, we set $B_v = \{\lambda \in \mathbb{C} : |\lambda| \le \|v\|\}$, then each B_v is compact and so is $\bigtimes_{v\in V} B_v$ in the product topology. But an element of $\bigtimes_{v\in V} B_v$ is nothing else than a bounded map $b : V \to \mathbb{C}$ such that $|b(v)| \le \|v\|$. To prove the theorem, one needs to show (i) that the relative topology on the unit ball in V^* (which is a subset of $\bigtimes_{v\in V} B_v$) is indeed the weak-* topology, and (ii) that the unit ball is closed.
Since this uses both Tychonoff's theorem (unproved in this course) and nets, we rather prove the sequential compactness version of Banach-Alaoglu.

THEOREM II.5.7. *Let V be a separable normed vector space, and let $(\ell_n)_{n\in\mathbb{N}}$ be a bounded sequence in V^*. There is $\ell \in V^*$ and a subsequence $(\ell_{n_k})_{k\in\mathbb{N}}$ such that $\ell_{n_k} \xrightarrow{*} \ell$ as $k \to \infty$.*

PROOF. Let $(v_j)_{j\in\mathbb{N}}$ be dense in V. The sequence $(\ell_n(v_1))_{n\in\mathbb{N}}$ is bounded in $\mathbb{C}$, hence there is a subsequence n_k^1 converging to z_1. Repeating this inductively with $v_2, v_3, \ldots$, we obtain subsequences n_k^j such that n^j is a subsequence of n^{j-1} for any $j \in \mathbb{N}$ and $\ell_{n_k^j}(v_j) \to z_j$ as $k \to \infty$. Define $\ell(v_j) = z_j$ and let m_j be the diagonal sequence, namely $m_j = n_j^j$. Then $\ell_{m_j}(v_j) \to z_j$ as $j \to \infty$ since by construction $\ell_{m_j}(v_j)$ is a subsequence of $\ell_{n_k^j}(v_j)$. Now, ℓ is linear on the span L of $\{v_n : n \in \mathbb{N}\}$ and it is bounded

$$|\ell(v)| = \lim_{j\to\infty} |\ell_{m_j}(v)| \le \limsup_{n\to\infty} \|\ell_n\|_{V^*} \|v\|_V$$

for any $v \in L$. Since L is dense, there is a bounded linear extension of ℓ to all of $V = \overline{L}$. It remains to check the weak-* convergence. Let $v \in V$; by density of $(v_j)_{j\in\mathbb{N}}$, there is a subsequence such that $\lim_{j\to\infty}, v_{n_j} = v$. For any $k, j \in \mathbb{N}$,

$$|\ell_{m_k}(v) - \ell(v)| \le |\ell_{m_k}(v) - \ell_{m_k}(v_{n_j})| + |\ell(v_{n_j}) - \ell(v)| + |\ell_{m_k}(v_{n_j}) - \ell(v_{n_j})|.$$

The first two terms are bounded by $\sup\{\|\ell_m\|_{X^*} + \|\ell\|_{X^*} : m \in \mathbb{N}\}\|v - v_{n_j}\|$ while the last one vanishes as $k \to \infty$ by definition of ℓ so that $\limsup_{k\to\infty} |\ell_{m_k}(v) - \ell(v)| \leq C\|v - v_{n_j}\|$ which converges to zero as $j \to \infty$. $\square$

We now turn to an application to the calculus of variations. First of all, the compactness just proved yields a weak Bolzano-Weierstrass theorem, namely that bounded sets are weakly sequentially compact. We shall use the fact that in a reflexive space, the weak and weak-* topologies are equivalent.

Proposition II.5.8. *Let V be a reflexive Banach space, and let $(v_n)_{n\in\mathbb{N}}$ be a bounded sequence in V. Then $(v_n)_{n\in\mathbb{N}}$ has a weakly convergent subsequence.*

Proof. The set $L = \overline{\mathrm{span}\{v_n : n \in \mathbb{N}\}}$ is separable and reflexive. Then L^* is separable. We consider the bounded sequence $(\mathcal{I}(v_n))_{n\in\mathbb{N}}$ in L^{**} (see (3.2)). By Banach-Alaoglu, there is a weakly-* convergent subsequence, namely a $v \in L$ (by reflexivity) such that for any $\ell \in L^*$,

$$\mathcal{I}(v_{n_k})(\ell) \to \mathcal{I}(v)(\ell), \qquad \text{namely} \qquad \ell(v_{n_k}) \to \ell(v)$$

as $k \to \infty$. Let now $\ell \in V^*$. Then $\ell \restriction_L \in L^*$. Since $(v_n)_{n\in\mathbb{N}}, v \in L$, we conclude that $\ell(v_{n_k}) \to \ell(v)$ as $k \to \infty$ for any $\ell \in V^*$, namely $v_{n_k} \rightharpoonup v$. $\square$

In the proof above, we used the following fact: If a normed vector space X is such that X^* is separable, then so is X. This implies that if L is reflexive and separable, then L^* is separable. We prove the claim. Let $(\ell_n)_{n\in\mathbb{N}}$ be a dense sequence in X^*, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X so that

$$\|x_n\| = 1, \qquad \ell_n(x_n) + 1/n \geq \|\ell_n\| \qquad (n \in \mathbb{N}). \tag{5.2}$$

Then $S = \overline{\mathrm{span}\{x_n : n \in \mathbb{N}\}} \subset X$ is separable, and we claim that $S = X$. Assume by contradiction that there is $x_0 \in X \setminus S$. By Hahn-Banach, there is $\ell \in X^*$ such that $\ell(x_0) = 1$ and $\ell \restriction S = 0$. By assumption, there is a convergent subsequence such that $\ell_{n_k} \to \ell$ as $k \to \infty$. Then,

$$0 \neq \|\ell\| = \lim_{k\to\infty} \|\ell_{n_k}\| \leq \limsup_{k\to\infty} \ell_{n_k}(x_{n_k})$$

by (5.2). However, $|\ell_{n_k}(x_{n_k})| = |(\ell_{n_k} - \ell)(x_{n_k})| \leq \|\ell_{n_k} - \ell\|$ converges to zero, which is a contradiction.

EXAMPLE 6. Let $(f_n)_{n\in\mathbb{N}}$ be a bounded sequence in $L^p(\Omega)$ for $1 < p < \infty$. Since L^p spaces are separable, there is a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ and a $f \in L^p(\Omega)$ such that

$$\int_\Omega f_n g d\mu \to \int_\Omega f g d\mu$$

for any $g \in L^q(\Omega)$.

We now prove the existence of a closest point to a closed convex set, a fact that was used in the proof of Theorem II.3.3. First of all, the need that the weak and norm closures of a convex set are equal.

LEMMA II.5.9. *Let V be a real normed vector space and $S \subset V$. Denote $\overline{S}$, respectively $\overline{S}^w$ the closure of S with respect to the norm, respectively weak topology. Then,*
(i) $\overline{S} \subset \overline{S}^w$,
(ii) if S is convex, then $\overline{S} = \overline{S}^w$.

PROOF. (i) Since the weak topology is weaker than the strong topology, it has fewer open sets and therefore also fewer closed sets. The inclusion follows from the definition of the closure as the smallest closet set containing S.
(ii) If $\overline{S} \neq \overline{S}^w$, there is $v_0 \in \overline{S}^w \setminus \overline{S}$. Applying the hyperplane separation theorem to the compact $A = \{v_0\}$ and the closed set $\overline{S}$, there is $\ell \in V^*$ such that

$$\ell(v_0) < \inf\{\ell(v) : v \in \overline{S}\} \leq \inf\{\ell(v) : v \in S\}.$$

Hence $v_0 \notin \overline{S}^w$ which is a contradiction. Hence $\overline{S} = \overline{S}^w$. $\square$

THEOREM II.5.10. *Let V be a reflexive real Banach space, and let $S \subset V$ be a nonempty convex and closed subset. Let $v_0 \in V \setminus S$. There exists $s_0 \in S$ such that*

$$\|v_0 - s_0\| = \inf\{\|v_0 - s\| : s \in S\}$$

PROOF. By definition of the infimum, there is a minimizing sequence $(s_n)_{n\in\mathbb{N}}$ in S such that $\|v_0 - s_n\| \to \inf\{\|v_0 - s\| : s \in S\}$. Since it is bounded, it has a weakly convergent subsequence $s_{n_k} \to s_0$ as $k \to \infty$, and $s_0 \in \overline{S}^w$. Since S is convex, its weak closure is equal to its norm closure, hence $s_0 \in \overline{S} = S$. But then

$$\|v_0 - s_0\| \leq \liminf_{n\to\infty} \|v_0 - s_n\| = \inf\{\|v_0 - s\| : s \in S\}$$

by (5.1), concluding the proof. $\square$

Let now $S \subset V$ and $F : S \to \mathbb{R}$. The function F is *weakly sequentially lower semicontinuous* at $v_0 \in S$ if

$$F(v_0) \leq \liminf_{n\to\infty} F(v_n)$$

for every sequence $(v_n)_{n\in\mathbb{N}}$ in S that converges weakly to v. It is moreover called *coercive* on S with respect to $\|\cdot\|$ whenever $F(v) \to +\infty$ as $\|v\| \to \infty$. By (5.1), norms are weakly sequentially lower semicontinuous.

A consequence of the weak Bolzano-Weierstrass theorem is the following principle of the calculus of variations.

THEOREM II.5.11. *Let V be a reflexive Banach space and let $S \subset V$ be non empty and weakly closed. Let $F : S \to \mathbb{R}$ be coercive and weakly sequentially lower semicontinuous. There exists $v_0 \in S$ such that $F(v_0) = \inf\{F(v) : v \in S\}$. If, moreover, S is convex and F is strictly convex, v_0 is the unique minimizer of F.*

PROOF. We consider a minizing sequence $(v_n)_{n\in\mathbb{N}}$ in S, namely $F(v_n) \to f = \inf\{F(v) : v \in S\}$. Since F is coercive, the sequence $(v_n)_{n\in\mathbb{N}}$ is bounded and has a weakly convergent subsequence $(v_{n_k})_{k\in\mathbb{N}}$ by Proposition II.5.8. Since S is weakly sequentially closed, the limit v_0 belongs to S, in particular $F(v_0) \geq f$. But

$$F(v_0) \leq \liminf_{k\to\infty} F(v_{n_k}) = f,$$

by weak sequential lower semicontinuity. This shows existence. For uniqueness, let v_0, v_1 be two minimizers, namely $F(v_0) = f = F(v_1)$. Let $v_t = (1-t)v_0 + tv_1$ for $t \in [0,1]$, which is in S by assumption. But then,

$$F(v_t) < (1-t)F(v_0) + tF(v_1) = f = \inf\{F(v) : v \in S\},$$

a contradiction. $\square$

6. Linear operators between Banach spaces

First of all, we recall that $\mathcal{L}(V,W)$ is a Banach space whenever W is complete. In this case, if $(T_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}(V,W)$ such that $\sum_{j=1}^{\infty} \|T_j\|$ is convergent in $\mathcal{L}(V,W)$, then so is the series $\sum_{j=1}^{\infty} T_j$. For example if $T \in \mathcal{L}(V)$ where V is Banach, then

$$\exp(T) = 1 + \sum_{j=1}^{\infty} \frac{T^n}{n!}$$

is a well-defined operator in $\mathcal{L}(V)$. This follows from

$$1 + \sum_{j=1}^{\infty} \frac{\|T^j\|}{n!} \leq \exp(\|T\|)$$

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

EXAMPLE 7. Let V be a Banach space and let $T \in \mathcal{L}(V)$ be such that $\|T\| < 1$. Then $1 - T$ is invertible and

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

The convergence of the series follows, similarly to above, from the convergence of the geometric series. Let $S_n = \sum_{j=0}^{n} T_j$. Then

$$(1-T)S_n = S_n(1-T) = S_n - (S_{n+1} - 1) = 1 - T^{n+1}.$$

Letting $n \to \infty$ yields the claim. The series for the inverse is called the *Neumann series.*

PROPOSITION II.6.1. *Let V be a Banach space and let $T \in \mathcal{L}(V)$. Then the following limit exists*

$$r_T = \lim_{n\to\infty} \|T^n\|^{1/n} = \inf\{\|T^n\|^{1/n} : n \in \mathbb{N}\} \leq \|T\|.$$

In the setting of the theorem, r_T is called the *spectral radius.*

PROOF. Let $\epsilon > 0$. There is $k \in \mathbb{N}$ such that

$$\|T^k\|^{1/k} \leq \inf\{\|T^n\|^{1/n} : n \in \mathbb{N}\} + \epsilon.$$

For any $n \in \mathbb{N}$, let $n = kl + m$ with $m < k$. Then, $\|T^n\|^{1/n} \leq \|T^{kl}\|^{1/n}\|T^m\|^{1/n} \leq \|T^k\|^{l/n}\|T\|^{m/n}$, which converges to $\|T^k\|^{1/k}$ as $n \to \infty$, since $l/n \to 1/k, m/n \to 0$. Together with the initial estimate, we conclude

$$\limsup_{n\to\infty} \|T^n\|^{1/n} \leq \inf\{\|T^n\|^{1/n} : n \in \mathbb{N}\} + \epsilon \leq \liminf_{n\to\infty} \|T^n\|^{1/n} + \epsilon.$$

Since this holds for all $\epsilon > 0$, we have that $\limsup_{n\to\infty} \|T^n\|^{1/n} \leq \liminf_{n\to\infty} \|T^n\|^{1/n}$ so that the limit exists. The last bound is immediate. $\square$

Finally, we denote the set of invertible operators in a Banach space V by

$$\mathrm{Gl}(V) = \{T \in \mathcal{L}(V) : T \text{ is invertible}\}.$$

Note that by the open mapping theorem, $T^{-1} \in \mathcal{L}(V)$. We then have:

THEOREM II.6.2. *Let V be a Banach space. Then $\mathrm{Gl}(V)$ is an open subspace of $\mathcal{L}(V)$.*

PROOF. Let $T_0 \in \mathrm{Gl}(V)$ and let $T \in \mathcal{L}(V)$ be such that

$$\|T - T_0\| < \|T_0^{-1}\|^{-1}.$$

Then $T = T_0(1 - T_0^{-1}(T_0 - T))$ and since $\|T_0^{-1}(T - T_0)\| < 1$, we conclude that $(1 - T_0^{-1}(T_0 - T)) \in \mathrm{Gl}(V)$, which yields the claim since T_0 in invertible. $\square$

DEFINITION II.6.3. Let V be a Banach space and $T \in \mathcal{L}(V)$. The *resolvent set* $\rho(T)$ of T is the set of $z \in \mathbb{C}$ such that $z \cdot 1 - T \in \mathrm{Gl}(V)$. The *spectrum* of T is the set $\sigma(T) = \mathbb{C} \setminus \rho(T)$.

In the following, we will prefer the notation $z - T$. The operator $R_z(T) = (z - T)^{-1}$ is called the *resolvent* of T at z. The name follows from the fact that the operator provides a solution of the linear system $Tv - zv = w$, namely $v = -R_z(T)w$.

PROPOSITION II.6.4. *Let V be a Banach space and $T \in \mathcal{L}(V)$. Then,*
(i) $\sigma(T)$ is closed,
(ii) the function $z \mapsto R_z(T)$ defined on $\rho(T)$ is analytic.

PROOF. (i) Let $z_0 \in \rho(T)$. Then $\|(z - T) - (z_0 - T)\| = |z - z_0|$. It follows from Theorem II.6.2 that $z - T$ is invertible for $|z - z_0|$ small enough. Hence $\rho(T)$ is open, namely $\sigma(T)$ is closed.
(ii) As in the proof of Theorem II.6.2, $(z - T) = (z_0 - T)(1 + (z - z_0)(z_0 - T)^{-1})$ is invertible whenever $|z - z_0| < \|(z_0 - T)^{-1}\|^{-1}$. Its inverse is given by the Neumann series, namely

$$(z - T)^{-1} = \left[1 + \sum_{j=1}^{\infty}(z - z_0)^j (T - z_0)^{-j}\right](z_0 - T)^{-1} = \sum_{j=0}^{\infty}(z_0 - T)^{-(j+1)}(z - z_0)^j \quad (6.1)$$

namely $R_z(T)$ can be expanded in a power series around each point z_0 of $\rho(T)$. $\square$

We note that the convergence of (6.1) implies that the distance $d(z_0)$ to the spectrum is at least $\|(z_0 - T)^{-1}\|^{-1}$, or equivalently

$$\|R_{z_0}(T)\| \geq \frac{1}{d(z_0)}.$$

THEOREM II.6.5. *Let V be a Banach space and $T \in \mathcal{L}(V)$. Then*
(i) $\sigma(T)$ is a nonempty, bounded and closed subset of $\mathbb{C}$,
(ii) $r_T = \max\{|z| : z \in \sigma(T)\}$.

Note that (ii) justifies the name spectral radius given to the limit in Proposition II.6.1.

PROOF. (i) Since $(z - T) = z(1 - z^{-1}T)$, the Neumann series

$$(z - T)^{-1} = \sum_{j=0}^{\infty} T^j z^{-j-1} \tag{6.2}$$

ensures that $R_z(T)$ exists if $\|z^{-1}T\| < 1$, namely if $|z| > \|T\|$. In particular, $\sigma(T) \subset B_{\|T\|}(0)$. It remains to prove that the spectrum is not empty. If $T = 0$, then $\{0\} \subset \sigma(T)$. Otherwise, $\|T\| > 0$. We assume by contradiction that $\sigma(T) = \emptyset$ and so $\rho(T) = \mathbb{C}$. Then the function $\mathbb{C} \ni z \mapsto R_z(T) \in \mathcal{L}(V)$ is entire. If $|z| \geq 2\|T\|$, then

$$\|R_z(T)\| \leq \frac{1}{\|T\|}$$

where we used the Neumann series to get

$$\|(z - T)^{-1}\| \leq \frac{1}{|z|} \sum_{j=0}^{\infty} \frac{1}{|z|}^{j} \|T\|^j = \frac{1}{|z| - \|T\|}.$$

Moreover, $z \mapsto R_z(T)$ is bounded on the compact disc $|z| \leq 2\|T\|$ since it is continuous. Hence it is an entire and bounded function, therefore it is a constant function by Liouville's theorem. Therefore $z \mapsto (z - T)$ is a constant function, which is a contradiction.
(ii) Let $n, k \in \mathbb{N}$. There are $0 \leq r < k$ and $q \in \mathbb{N}$ such that $n = kq + r$, and so $T^n = (T^k)^q T^r$, which yields

$$\left\| \sum_{n=0}^{\infty} T^n z^{-n+1} \right\| \leq \sum_{r=0}^{k-1} \|T\|^r |z|^{-r+1} \sum_{q=0}^{\infty} (\|T^k\| |z|^{-k})^q.$$

It follows that the series is convergent if $\|T^k\||z|^{-k} < 1$, namely $|z| > \|T^k\|^{\frac{1}{k}}$. Hence $z \in \sigma(T)$ implies $|z| < \|T^k\|^{\frac{1}{k}}$, and hence $|\sigma(T)| \leq r_T$, where we use the notation $|\sigma(T)| = \max\{|z| : z \in \sigma(T)\}$.

Let $\delta > 0$ and let Γ_δ be the contour in $\mathbb{C}$ winding once around $\sigma(T)$ given by $|z| = |\sigma(T)| + \delta$. For any $z \in \Gamma_\delta$, (6.2) yields for any $n \in \mathbb{N} \cup \{0\}$ that

$$\frac{1}{2\pi\mathrm{i}} \oint_{\Gamma_\delta} (z - T)^{-1} z^n dz = \sum_{j=0}^{\infty} T^j \frac{1}{2\pi\mathrm{i}} \oint_{\Gamma_\delta} z^{n-j-1} dz = T^n, \tag{6.3}$$

from which we obtain that

$$\|T^n\| \leq (|\sigma(T)| + \delta)^{n+1} \sup\{\|(z - T)^{-1}\| : z \in \Gamma_\delta\}.$$

Taking the nth root and the limit $n \to \infty$ now yields $r_T \leq |\sigma(T)| + \delta$ and therefore $r_T \leq |\sigma(T)|$ since $\delta > 0$ is arbitrary. □

In a finite dimensional setting, the spectrum is made of eigenvalues, namely $z \in \mathbb{C}$ for which there is $v \in V, v \neq 0$, such that $Tv = zv$. Indeed, the existence of a non-zero solution of the equation implies that $z - T$ has a non-trivial kernel and is therefore not invertible. The next example illustrates that in general the spectrum contains more than just eigenvalues.

Example 8. We consider a measure space (X, μ) with finite measure, a function $g \in L^\infty(X, \mu)$ and the linear operator M_g on $L^2(X, \mu)$ defined by

$$(M_g \psi)(x) = g(x) \psi(x)$$

It is well defined and bounded since $\|M_g \psi\|_2^2 \leq \|g\|_\infty\|^2 \psi\|_2^2$. In fact, if M_g maps $L^2(X, \mu)$ to itself, then $g \in L^\infty(X, \mu)$. Indeed, if $\psi_n \to \psi$ and $M_g \psi_n \to \phi$ in $L^2(X, \mu)$, then by the Riesz-Fischer theorem, there is a convergent subsequence $(\psi_{n_j})_{j \in \mathbb{N}}$ along which both convergences hold pointwise almost everywhere. But then the function $g\psi_{n_j}$ converge pointwise a.e. to both $g\psi$ and ϕ, hence $g\psi = \phi$ in $L^2(X, \mu)$. Therefore, M_g is closed and so it is bounded by the closed graph theorem, namely there is $C \geq 0$ such that $\|M_g \psi\|_2 \leq C \|\psi\|_2$ for all $\psi \in L^2(X, \mu)$. Let $\ell > 0$ and let χ_ℓ be the characteristic function of the superlevel set $X_\ell = \{x \in X : |g(x)| \geq \ell\}$. Then

$$\ell^2 \mu(X_\ell) \leq \int_X |\chi_\ell(x)|^2 |g(x)|^2 d\mu(x) \leq C^2 \int_X |\chi_\ell(x)|^2 d\mu(x) = C^2 \mu(X_\ell).$$

It follows that $\ell^2 > C^2$ implies $\mu(X_\ell) = 0$, namely $\|g\|_\infty \leq \ell$ indeed.

We can now characterize the resolvent set of M_g. $z \in \rho(M_g)$ if and only if $(z - M_g)$ is invertible, or equivalently if and only if $(z - g)^{-1}\psi$ is in $L^2(X, \mu)$ for all $\psi \in L^2(X, \mu)$. In other words, the operator $M_{(z-g)^{-1}} \in \mathcal{L}(L^2(X,\mu))$, which we have just shown to hold if and only if $(z-g)^{-1} \in L^\infty(X,\mu)$. Hence $z \in \rho(M_g)$ if and only if there is $c \geq 0$ such that $\mu\{x \in X : |(z-g(x))^{-1}| > c\} = 0$ or equivalently

$$\mu\{x \in X : |(z-g(x))| < c^{-1}\} = 0$$

We conclude that $z \in \sigma(M_g)$ if and only if $\mu\{x \in X : |(z-g(x))| \leq \epsilon\} > 0$ for any $\epsilon > 0$, namely z is in the *essential range* of the function g (which coincides with the closure of its range if g is continuous).

Concretely, if (X,μ) is $[0,1]$ with the Lebesgue measure and $g(x) = x$, we have that $\sigma(M_g) = [0,1]$. However, $z \in [0,1]$ is not an eigenvalue since the equation $x\psi(x) = z\psi(x)$ for all $x \in [0,1]$ implies $\psi = 0$ almost everywhere, namely, there is no eigenvector for the spectral value z.

We now turn to a cornerstone of operator theory, here in a limited framework. If $T \in \mathcal{L}(V)$, then

$$P(T) = \sum_{n=0}^{N} a_n T^n \tag{6.4}$$

is well-defined in $\mathcal{L}(V)$ for any polynomial $P(z) = \sum_{n=0}^{N} a_n z^n$. The goal is to extend the mapping $P \to P(T)$ to more general functions, thereby defining a *functional calculus*. Let f be a function that is analytic in a domain Ω containing $\sigma(T)$. Let Γ be a simple contour in $\Omega \cup \rho(T)$ winding once around $\sigma(T)$. Let

$$f(T) = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma (z-T)^{-1} f(z) dz. \tag{6.5}$$

Note that both the set of analytic functions and $\mathcal{L}(V)$ are unital algebras.

Theorem II.6.6. *(i) If f is a polynomial, then (6.5) and (6.4) are the same operator.*
(ii) The map $f \mapsto f(T)$ is a homomorphism of unital algebras.
(iii) $\sigma(f(T)) = f(\sigma(T))$.

Note that one could further prove that $(f \circ g)(T) = f(g(T))$.

Proof. (i) If f is a polynomial, then (6.3) used in (6.5) yields (6.4).
(ii) The mapping $f \mapsto f(T)$ is linear. The fact that units are mapped onto each other is (i)

in the case of the constant polynomial. In order to prove that $(fg)(T) = f(T)g(T)$, we first note the following resolvent identity

$$(z-T)^{-1} - (w-T)^{-1} = (w-z)(z-T)^{-1}(w-T)^{-1}$$

which follows from $(w-T)-(z-T) = (w-z)$ after multiplication by $(w-T)^{-1}$ from the right and by $(z-T)^{-1}$ from the left. Let Θ, Γ be two contours as above such that Θ is completely in the interior of Γ. Then

$$\begin{aligned} f(T)g(T) &= \oint_\Theta \oint_\Gamma (\theta-T)^{-1}(\gamma-T)^{-1} f(\theta)g(\gamma)\mathfrak{d}\gamma\mathfrak{d}\theta \\ &= \oint_\Theta \oint_\Gamma \left((\gamma-T)^{-1} - (\theta-T)^{-1}\right) \frac{f(\theta)g(\gamma)}{\theta-\gamma}\mathfrak{d}\gamma\mathfrak{d}\theta \end{aligned}$$

by the resolvent identity, where we denote $\mathfrak{d}\theta = (2\pi\mathrm{i})^{-1}d\theta$. The first term vanishes since $\oint_\Theta \frac{f(\theta)}{\theta-\gamma}d\theta = 0$ because Γ lies outside of the interior of Θ. The second term reduces to

$$f(T)g(T) = -\oint_\Theta \left(\oint_\Gamma \frac{g(\gamma)}{\theta-\gamma}\mathfrak{d}\gamma\right)(\theta-T)^{-1}f(\theta)\mathfrak{d}\theta = -\oint_\Theta (\theta-T)^{-1}f(\theta)g(\theta)\mathfrak{d}\theta = (fg)(T).$$

(iii) Let $\mu \in f(\sigma(T))$, namely $\mu = f(\lambda)$ for some $\lambda \in \sigma(T)$. Since f is analytic, the function $F(\zeta) = (\zeta-\lambda)^{-1}(f(\zeta)-f(\lambda))$ is analytic in Ω so that $F(T)$ is a well-defined element of $\mathcal{L}(V)$. By (ii) applied to $(\zeta-\lambda)F(\zeta)$, we conclude that $(T-\lambda)F(T) = f(T)-\mu$. But $\lambda \in \sigma(T)$ implies that the left hand side is not invertible, and hence $f(T)-\mu$ is not invertible, namely $\mu \in \sigma(f(T))$. Hence $f(\sigma(T)) \subset \sigma(f(T))$. Let now $\mu \notin f(\sigma(T))$, namely $f(\lambda)-\mu \neq 0$ for all $\lambda \in \sigma(T)$. Then $g(\lambda) = (f(\lambda)-\mu)^{-1}$ is analytic in an open neighbourhood of $\sigma(T)$, and hence $g(T)$ is well-defined in $\mathcal{L}(V)$. But then $(f(\lambda)-\mu)g(\lambda) = 1$ implies by (ii) that $(f(T)-\mu)g(T) = 1$, proving that $g(T)$ is the inverse of $f(T)-\mu$, and hence $\mu \notin \sigma(f(T))$. This shows that $\sigma(f(T)) \subset f(\sigma(T))$ and concludes the proof. □

Definition II.6.7. Let V, W be Banach spaces and $T \in \mathcal{L}(V,W)$. The *adjoint* of T is the operator $T' \in \mathcal{L}(W^*, V^*)$ defined by

$$(T'\ell)(v) = \ell(Tv)$$

for any $\ell \in W^*, v \in V$.

Proposition II.6.8. *The map $T \mapsto T'$ is an isometric isomorphism between $\mathcal{L}(V,W)$ and $\mathcal{L}(W^*, V^*)$.*

PROOF. Linearity is immediate. We note that

$$\|T\|_{\mathcal{L}(V,W)} = \sup\{\|Tv\|_W : \|v\| = 1\} = \sup\Big\{\sup\{|\ell(Tv)| : \|\ell\|_{W^*} = 1\} : \|v\|_V = 1\Big\}$$

by Corollary II.3.9. Replacing $\ell(Tv)$ by $(T'\ell)(v)$ and taking the supremum over v first, we obtain

$$\|T\|_{\mathcal{L}(V,W)} = \sup\{\|T'\ell\|_{V^*} : \|\ell\|_{W^*} = 1\} = \|T'\|_{\mathcal{L}(W^*,V^*)}$$

indeed. □

A set K in a Banach space V is called precompact if its closure is compact. Equivalently, K is precompact if every sequence in K has a subsequence that is convergent in V.

DEFINITION II.6.9. Let V, W be Banach spaces. An operator $T \in \mathcal{L}(V, W)$ is called *compact* if T maps bounded sets of V into precompact sets of W.

EXAMPLE 9. Let T be such that its range is finite dimensional. Then T is compact. Indeed, any $w = Tv$ can be written as $w = \sum_{j=1}^{N} \tau_j(v) e_j$ for a linearly independent set $\{e_j : j = 1, \ldots, N\}$ in W and $\tau_j(v) \in \mathbb{C}$. If $\{v_n\}$ is a bounded sequence in V, then $\{\tau_1(v_n) : n \in \mathbb{N}\}$ is a bounded set in $\mathbb{C}$ since T is bounded, and it has a convergent subsequence. Repeating this recursively for $j = 2, \ldots, N$, we obtain a subsequence $\{n_k\}$ such that $\{\tau_j(v_{n_k})\}$ is convergent for all j. Together, these define a convergent subsequence of $\{Tv_n\}$.

THEOREM II.6.10. *Let V, W be Banach spaces.*
(i) The set of compact operators is a vector subspace of $\mathcal{L}(V, W)$.
(ii) Let X be a Banach space and let $T \in \mathcal{L}(W, X)$. If $S \in \mathcal{L}(V, W)$ is compact, then TS is compact.
(iii) Let U be a Banach space and let $T \in \mathcal{L}(U, V)$. If $S \in \mathcal{L}(V, W)$ is compact, then ST is compact.
(iv) Let $\{T_n\}$ be a sequence of compact operators in $\mathcal{L}(V, W)$ such that $T_n \to T$ in the uniform topology. Then T is compact.

PROOF. (i) is a simple exercise.
(ii) follows from the fact that the image of a precompact set under a bounded linear map is precompact.

(iii) follows from the fact that a bounded set is mapped to a bounded set by a bounded linear map.
(iv) Let $\epsilon > 0$. There is $N \in \mathbb{N}$ such that $n \geq N$ implies $\|T_n - T\| < \frac{\epsilon}{2}$. Let B be a bounded subset of V. Since T_n is compact, the set $T_n B$ can be covered by a finite number of balls of radius $\frac{\epsilon}{2}$. Hence the set TB is covered by a finite number of balls of radius ϵ (with the same centers), and hence it is precompact. It follows that T is compact. $\square$

Here is an important mapping property of compact operators

THEOREM II.6.11. *A compact operator maps weakly convergent sequences into norm convergent sequences.*

PROOF. Let $v_n \stackrel{w}{\rightharpoonup} v$. By Proposition II.5.2, the set $\{\|v_n\|_V : n \in \mathbb{N}\}$ is bounded, compactness the set $\{w_n = Tv_n : n \in \mathbb{N}\}$ is precompact since T is a compact operator. Let also $w = Tv$. Then for any $\ell \in W^*$,

$$\ell(w_n) - \ell(w) = (T'\ell)(w_n) - (T'\ell)(w) \longrightarrow 0$$

as $n \to \infty$, namely w_n converges weakly to w. Since $\{w_n\}$ has a norm convergent subsequence, we conclude that it converges in norm to w. $\square$

We will continue the study of compact operators in the setting of Hilbert spaces.
To conclude, we get back to possibly unbounded operators. Let V, W be normed vector spaces and let $\Gamma \subset V \times W$ be a linear subspace. Γ is called a *linear graph* if

$$\big((v, w_1) \in \Gamma, (v, w_2) \in \Gamma\big) \Rightarrow w_1 = w_2,$$

or equivalently $(0, w) \in \Gamma$ implies $w = 0$. Clearly, the graph $\Gamma(T)$ of a linear operator $T : D(T) \subset V \to W$ is a linear graph. Reciprocally, a linear graph Γ defines a unique linear operator $T : D(T) \subset V \to W$ such that $\Gamma(T) = \Gamma$ through

$$D(T) = \pi_1(\Gamma), \qquad Tv = \pi_2((\{v\} \times W) \cap \Gamma) \text{ for } v \in D(T).$$

DEFINITION II.6.12. Let $T : D(T) \subset V \to W$ and $S : D(S) \subset V \to W$ be linear operators with graphs $\Gamma(T), \Gamma(S)$. S is an *extension of* T, denoted $T \subset S$, if $\Gamma(T) \subset \Gamma(S)$.

Equivalently,

$$D(T) \subset D(S) \qquad S\restriction_{D(T)} = T.$$

DEFINITION II.6.13. A linear operator $T : D(T) \subset V \to W$ is *closable* if $\overline{\Gamma(T)}$ is a linear graph. The corresponding operator $\overline{T} \supset T$ with $\Gamma(\overline{T}) = \overline{\Gamma(T)}$ is called the *closure* of T.

Note that

$$D(\overline{T}) = \{v \in V : \exists (v_n)_{n\in\mathbb{N}} \text{ in } D(T), w \in W : (v_n, Tv_n) \to (v, w) \text{ as } n \to \infty\}.$$

In particular, $D(T) \subset D(\overline{T})$ and the inclusion $D(\overline{T}) \subset \overline{D(T)}$ is in general strict.

PROPOSITION II.6.14. *An operator $T : D(T) \subset V \to W$ is* closable *if and only if for any sequence $(v_n, w_n) \in \Gamma(T)$, the convergence $v_n \to 0, w_n = Tv_n \to w$ implies $w = 0$.*

PROOF. T is closable if and only if $\overline{\Gamma(T)}$ is a linear graph, namely $(0, w) \in \overline{\Gamma(T)}$ implies $w = 0$. □

In particular, if $T : D(T) \subset V \to W$ is a bounded linear operator, then it is closable. Indeed, $(v_n, Tv_n) \to (0, w)$ implies

$$\|w\| = \lim_{n\to\infty} \|Tv_n\| \leq \lim_{n\to\infty} \|T\|\|v_n\| = 0.$$

EXAMPLE 10. (i) Let $V = L^2(\mathbb{R}; \mathbb{R}), W = \mathbb{R}$ and let

$$D(T) = \{f \in L^2(\mathbb{R}; \mathbb{R}) : \text{supp}(f) \text{ is compact}\},$$

and

$$Tf = \int_{-\infty}^{\infty} f(x)dx$$

Then the sequence $(f_n)_{n\in\mathbb{N}}$ in $D(T)$ given by $f_n(x) = n^{-1}\chi_{[0,n]}(x)$ converges to zero since $\|f_n\|_2^2 = n^{-1}$. However

$$Tf_n = \int_{-\infty}^{\infty} f_n(x)dx = 1$$

for all $n \in \mathbb{N}$. Hence T is not closable by the lemma.

(ii) Let $\Omega \subset \mathbb{R}^n$ be open and $V = L^2(\Omega) = W$. Let $D(\Delta) = C_c^\infty(\Omega)$. We claim that the

Laplacian $\Delta : D(\Delta) \to L^2(\Omega)$ is closable. Indeed, let $(f_n, g_n)_{n\in\mathbb{N}}$ be a sequence in $\Gamma(\Delta)$ such that $f_n \to 0, g_n \to g$ as $n \to \infty$. For any $\varphi \in C_c^\infty(\Omega)$,

$$\int_\Omega g_n \varphi dx = \int_\Omega \Delta f_n \varphi dx = \int_\Omega f_n \Delta\varphi dx$$

by Gauss-Green's theorem and the compact support of the functions. Letting $n \to \infty$, we obtain $\int_\Omega g\varphi dx = 0$ for any $\varphi \in C_c^\infty(\Omega)$. But $C_c^\infty(\Omega)$ is a dense subset of $L^2(\Omega)^* \simeq L^2(\Omega)$ so that $\int_\Omega ghdx = 0$ for all $h \in L^2(\Omega)$ and hence $g = 0$ since linear functionals separate. With Proposition II.6.14, it follows that Δ is closable. The determination of $\overline{\Delta}$ and $D(\overline{\Delta})$ is a separate issue. In the present case, it can be explicitly characterized, namely

$$D(\overline{\Delta}) = \{f \in L^2(\Omega) : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{N}_0^n : |\alpha| \leq 2\}.$$

This space is usually denoted $H^2(\Omega)$, or $W^{2,2}(\Omega)$ and is one of the *Sobolev spaces*.

CHAPTER III

Hilbert spaces

1. Definitions and elementary results

Hilbert spaces are complete linear spaces, that are equipped with a metric and have a geometric structure.

DEFINITION III.1.1. Let V be a complex vector space. An *inner product* on V is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$ such that
(i) $\langle v, v\rangle \geq 0$ with equality iff $v = 0$ (positivity)
(ii) $\langle v, w_1 + \alpha w_2\rangle = \langle v, w_1\rangle + \alpha\langle v, w_2\rangle$ (physicists' linearity)
(iii) $\langle w, v\rangle = \overline{\langle v, w\rangle}$ (symmetry)

Of course, this implies that $\langle v_1 + \beta v_2, w\rangle = \langle v_1, w\rangle + \overline{\beta}\langle v_2, w\rangle$. A complex vector space V equipped with an inner product is called an *inner product space*.

EXAMPLE 11. The space $V = C([0,1])$ of continuous complex-valued functions on $[0,1]$ is an inner product space with

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

We say that $v, w \in V$ are *orthogonal* if $\langle v, w\rangle = 0$, and a family $\{v_\alpha : \alpha \in \mathcal{I}\}$ is *orthonormal* if $\langle v_\alpha, v_\beta\rangle = \delta_{\alpha,\beta}$. We first prove what Pythagoras already knew.

THEOREM III.1.2. *Let V be an inner product space and let $\{v_n\}_{n=1}^N$ be an orthonormal set. Then for any $v \in V$,*

$$\|v\|^2 = \sum_{n=1}^N |\langle v_n, v\rangle|^2 + \|v - \sum_{n=1}^N \langle v_n, v\rangle v_n\|^2$$

where we denoted $\|v\|^2 = \langle v, v\rangle$.

PROOF. Trivially,

$$v = \sum_{n=1}^N \langle v_n, v\rangle v_n + (v - \sum_{n=1}^N \langle v_n, v\rangle v_n)$$

and the two terms are orthogonal to each other. Hence, the cross terms in $\langle v, v\rangle$ vanish, proving the claim since $\|v_n\|^2 = 1$. $\square$

A simple but useful consequence of this is

$$\|v\|^2 \geq \sum_{n=1}^{N} |\langle v_n, v\rangle|^2 \tag{1.1}$$

for any orthonormal set $\{v_n\}_{n=1}^{N}$, which is sometimes referred to as *Bessel's inequality.* The following inequality of Schwarz is crucial:

Corollary III.1.3. *Let V be an inner product space. For any $v, w \in V$,*

$$|\langle v, w\rangle| \leq \|v\|\|w\|$$

Proof. The case $w = 0$ trivially holds. And if $w \neq 0$, the claim is precisely Bessel's inequality applied to the orthonormal set $\{w/\|w\|\}$. $\square$

We now justify the notations used:

Proposition III.1.4. *Let V be an inner product space. Then V is a normed linear space with norm $\|v\| = \langle v, v\rangle^{1/2}$.*

Proof. By definition,

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\mathrm{Re}\langle v, w\rangle.$$

Since $\mathrm{Re}\langle v, w\rangle \leq |\langle v, w\rangle|$, the triangle inequality follows by Schwarz' inequality:

$$\|v + w\|^2 \leq \|v\|^2 + \|w\|^2 + 2\|v\|\|w\| = (\|v\| + \|w\|)^2.$$

This proves the triangle inequality. All other properties of the norm follow immediately from those of the inner product. $\square$

Hence, an inner product is naturally endowed with a metric

$$d(v, w) = \langle v - w, v - w\rangle^{\frac{1}{2}}.$$

We also note the following *parallelogram identity*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2,$$

as well as the *polarization identity*

$$4\langle v,w\rangle = \|v+w\|^2 - \|v-w\|^2 - \mathrm{i}\|v+\mathrm{i}w\|^2 + \mathrm{i}\|v-\mathrm{i}w\|^2. \tag{1.2}$$

Note that the parallelogram identity is specific of a norm arising from an inner product. In fact, if a norm satisfies the parallelogram identity, then it can be used to *define* an inner product through (1.2).

DEFINITION III.1.5. A *Hilbert space* is a complete inner product space.

We recall that a surjective linear map $U : V_1 \to V_2$ between to inner product spaces is called *unitary* if for all $v, w \in V_1$,

$$\langle Uv, Uw\rangle_{V_2} = \langle v, w\rangle_{V_1}.$$

In particular, $\|Uv\|_2 = \|v\|_1$, which also shows that U is necessarily injective. Two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ are called *isomorphic* if there is a unitary operator $U : \mathcal{H}_1 \to \mathcal{H}_2$.

EXAMPLE 12. Let (Ω, μ) be a finite measure space. While $C(\Omega)$ is an inner product space, it is not a Hilbert space. However, Hölder's inequality yields that $f, g \in L^2(\Omega, \mu)$ implies $\overline{f}g \in L^1(\Omega, \mu)$ so that

$$\langle f, g\rangle = \int_\Omega \overline{f(x)}g(x)d\mu(x)$$

is a well-defined inner product on $L^2(\Omega, \mu)$. Since $L^2(\Omega, \mu)$ is complete, it is a Hilbert space. In fact, it is the completion of $C(\Omega)$ in the L^2-norm.

2. Projections in Hilbert space; the Riesz lemma

In a Hilbert space, Theorem II.5.10 can be proved by elementary means without using the Banach-Alaoglu theorem.

PROPOSITION III.2.1. *Let $\mathcal{K}$ be a closed convex subset of $\mathcal{H}$. Then there is a unique $v_0 \in \mathcal{K}$ of minimal norm.*

PROOF. Let $v, w \in \mathcal{K}$. Applying the parallelogram identity to $v/2, w/2$, we obtain

$$\frac{1}{4}\|v-w\|^2 = \frac{1}{2}\|v\|^2 + \frac{1}{2}\|w\|^2 - \|(v+w)/2\|^2$$

If $\delta = \inf\{\|v\| : v \in \mathcal{K}\}$, the fact that $(v+w)/2 \in \mathcal{K}$ by convexity implies that

$$\|v - w\|^2 \leq 2\|v\|^2 + 2\|w\|^2 - 4\delta^2. \tag{2.1}$$

If $\|v\| = \|w\| = \delta$, the right hand side vanishes and hence $v = w$, proving uniqueness of the minimizer. To prove existence, we consider a minimizing sequence $(v_n)_{n\in\mathbb{N}}$ in $\mathcal{K}$, namely such that $\|v_n\| \to \delta$ as $n \to \infty$. Then (2.1) for $v = v_n$ and $w = v_m$ implies that $(v_n)_{n\in\mathbb{N}}$ is Cauchy and hence converges in $\mathcal{H}$. Since $\mathcal{K}$ is closed, $\bar{v} = \lim_{n\to\infty} v_n \in \mathcal{K}$, and by continuity of the norm, $\|\bar{v}\| = \lim_{n\to\infty} \|v_n\| = \delta$. □

Here is a natural application: Let $\mathcal{K}$ be a closed subspace and let $v \in \mathcal{H}$. The proposition applied to the closed convex set $\mathcal{K} - v$ yields a $k_0 \in \mathcal{K}$ such that

$$\|k_0 - v\| = \inf\{\|k - v\| : k \in \mathcal{K}\}$$

namely, k_0 is the unique element in $\mathcal{K}$ closest to v.
With this, many 'intuitive' properties known from planar geometry hold in a general Hilbert space. If $\mathcal{K}$ is a subspace of $\mathcal{H}$,

$$\mathcal{K}^\perp = \{v \in \mathcal{H} : \langle w, v\rangle = 0 \text{ for all } w \in \mathcal{K}\}$$

THEOREM III.2.2. *Let $\mathcal{K}$ be a closed subspace of $\mathcal{H}$. Any $v \in \mathcal{H}$ has a unique decomposition*

$$v = k + k^\perp \qquad k \in \mathcal{K}, k^\perp \in \mathcal{K}^\perp$$

Moreover, k is the point in $\mathcal{K}$, and $k^\perp$ the point in $\mathcal{K}^\perp$, closest to v.

In particular, if $\mathcal{K} \neq \mathcal{H}$, then $\mathcal{K}^\perp$ is not just $\{0\}$.

PROOF. Since $\mathcal{K}$ is convex, so is the translated set $v + \mathcal{K}$. Hence there is a element $k^\perp \in v + \mathcal{K}$ of smallest norm, and let $k = v - k^\perp$. Clearly $k \in \mathcal{K}$. Since the norm of $k^\perp$ is minimal and since $k^\perp - \lambda w \in v + \mathcal{K}$ for all $\lambda \in \mathbb{C}$ and all $w \in \mathcal{K}$ with $\|w\| = 1$,

$$\|k^\perp\|^2 \leq \|k^\perp - \lambda w\|^2.$$

Hence,

$$0 \leq -\lambda\langle k^\perp, w\rangle - \bar{\lambda}\langle w, k^\perp\rangle + |\lambda|^2.$$

The choice $\lambda = \langle w, k^\perp \rangle$ yields $0 \leq -|\langle w, k^\perp \rangle|^2$, namely $\langle w, k^\perp \rangle = 0$ and therefore $k^\perp \in \mathcal{K}^\perp$. It remains to prove that $\|v - k\|$ is minimal. For any $w \in \mathcal{K}$,

$$\|v - w\|^2 = \|k^\perp + k - w\|^2 = \|k^\perp\|^2 + \|k - w\|^2$$

which is minimal if $w = k$. □

With this, we define the bounded linear maps $P, P^\perp \in \mathcal{L}(\mathcal{H})$ by

$$v \mapsto Pv = k, \qquad v \mapsto P^\perp v = k^\perp.$$

Since $\langle Pv, P^\perp v \rangle = 0$ by construction, we have that

$$\|v\|^2 = \|Pv\|^2 + \|P^\perp v\|^2,$$

and so $\|P\| = 1 = \|P^\perp\|$ provided $\mathcal{K} \neq \emptyset$ and $\mathcal{K} \neq \mathcal{H}$. Moreover, $P^2 = P$ and similarly for $P^\perp$. P, resp. $P^\perp$ are the *orthogonal projections* of $\mathcal{H}$ onto $\mathcal{K}$, resp. $\mathcal{K}^\perp$.
We have already shown that the dual space of $L^2(\Omega)$ is itself, namely any bounded linear functional on $L^2(\Omega)$ is of the form

$$f \mapsto \int_\Omega \bar{g} f d\mu$$

for some $g \in L^2(\Omega)$. In a general Hilbert space, the Cauchy-Schwarz inequality implies that any element $v \in \mathcal{H}$ defines a bounded linear functional through $w \mapsto \langle v, w \rangle$. That all bounded linear functionals are of this form is usually referred to as Riesz' lemma:

Proposition III.2.3. *For any $\ell \in \mathcal{H}^*$, there is $v \in \mathcal{H}$ such that*

$$\ell(w) = \langle v, w \rangle,$$

and $\|\ell\|_{\mathcal{H}^*} = \|v\|_{\mathcal{H}}$.

Proof. If $\ell = 0$, choose $v = 0$. Otherwise the subspace $\mathcal{K} = \mathrm{Ker}(\ell)$ is closed by continuity of ℓ, and $\mathcal{K}^\perp \neq \{0\}$. Let $\tilde{v} \in \mathcal{K}^\perp$ with $\|\tilde{v}\| = 1$. Then by linearity

$$\ell(\ell(w)\tilde{v} - \ell(\tilde{v})w) = 0,$$

namely $\ell(w)\tilde{v} - \ell(\tilde{v})w \in \mathcal{K}$ and so $\langle \tilde{v}, (\ell(w)\tilde{v} - \ell(\tilde{v})w) \rangle = 0$, or equivalently

$$\ell(w) = \ell(\tilde{v})\langle \tilde{v}, w \rangle.$$

The choice $v = \overline{\ell(\tilde{v})}\, \tilde{v}$ concludes the proof. □

An important consequence of the representation theorem is the following result of Lax-Milgram.

THEOREM III.2.4. *Let $t : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ be sesquilinear (antilinear in the first variable, linear in the second one), bounded namely there is $K > 0$ such that*

$$|t(x, y)| \leq K\|x\|\|y\|$$

for all $x, y \in \mathcal{H}$, and there is $k > 0$ such that

$$|t(x, x)| \geq k\|x\|^2$$

for all $x \in \mathcal{H}$. Then there is a unique $T \in \mathrm{Gl}(\mathcal{H})$ such that

$$t(x, y) = \langle Tx, y\rangle$$

for all $x, y \in \mathcal{H}$. Moreover, $\|T\| \leq K$ and $\|T^{-1}\| \leq k^{-1}$.

PROOF. For any $x \in \mathcal{H}$, the map $\ell_x : \mathcal{H} \to \mathbb{R}$ given by $\ell_x(y) = t(x, y)$ is linear and bounded with $\|\ell_x\|_{\mathcal{H}^*} = \sup\{|t(x, y)|/\|y\| : 0 \neq y \in \mathcal{H}\} \leq K\|x\|$. Hence there is $v_x \in \mathcal{H}$ such that $t(x, y) = \ell_x(y) = \langle v_x, y\rangle$ for all $y \in \mathcal{H}$. Define $Tx = v_x$, which is linear by antilinearity of t in the first variable. It is moreover bounded since

$$\|Tx\| = \|v_x\| = \|\ell_x\|_{\mathcal{H}^*} \leq K\|x\|.$$

Now, for any $x \in \mathcal{H}$,

$$\|Tx\|\|x\| \geq |\langle Tx, x\rangle| = |t(x, x)| \geq k\|x\|^2$$

showing that $x \neq 0$ implies $Tx \neq 0$, namely T is injective. We denote $R = \mathrm{Ran}(T)$, which is a closed subspace. Indeed, let $(x_n)_{n\in\mathbb{N}}$ be a sequence such that $(Tx_n)_{n\in\mathbb{N}}$ converges. Then by the above

$$k\|x_n - x_m\|^2 \leq \|Tx_n - Tx_m\|\|x_n - x_m\|,$$

namely $k\|x_n - x_m\| \leq \|Tx_n - Tx_m\|$, which converges to 0 when $n, m \to \infty$. If x is the limit of the Cauchy sequence $(x_n)_{n\in\mathbb{N}}$, then by continuity $\lim_{n\to\infty} Tx_n = Tx \in R$, proving that R is closed. If T is not surjective, let $v_0 \in \mathcal{H} \setminus R$. Then by Theorem III.2.2, $v_0 = r + r^\perp$ with $r^\perp \neq 0$ and

$$0 < k\|r^\perp\|^2 \leq t(r^\perp, r^\perp) = \langle Tr^\perp, r^\perp\rangle = 0$$

since $r^\perp \in R^\perp$. Since this is a contradiction, we conclude that T is surjective and hence has a bounded inverse by the open mapping theorem. To estimate its norm, let $z = T^{-1}x$ for which we have that

$$k\|z\|^2 \leq |\langle Tz, z\rangle| \leq \|x\|\|z\|,$$

namely $\|T^{-1}x\| \leq k^{-1}\|x\|$ upon division by $\|z\|$. □

Corollary III.2.5. *Let t be as above and let $\ell \in \mathcal{H}^*$. There is a unique $x_\ell \in \mathcal{H}$ such that $t(x_\ell, y) = \ell(y)$ for all $y \in \mathcal{H}$. Moreover, $\ell \mapsto x_\ell$ is continuous.*

Proof. With $T \in \mathrm{Gl}(\mathcal{H})$ be given by Theorem III.2.4, we have $\langle Tx, y\rangle = t(x, y)$ for all $x, y \in \mathcal{H}$. On the other hand, the Riesz representation yields that there is $x_0 \in \mathcal{H}$ such that $\ell(y) = \langle x_0, y\rangle$ for all $y \in \mathcal{H}$. The choice $x_\ell = T^{-1}x_0$ yields the first claim. Moreover, the second part of Riesz' lemma yields that $\|x_\ell\| \leq \|T^{-1}\|\|x_0\| \leq k^{-1}\|\ell\|_{\mathcal{H}^*}$. □

We note that the Riesz lemma and its corollaries do not use the fact that $\mathcal{H}$ is a complex Hilbert space, and would also hold in a real Hilbert space. In that case, the sesquilinear form of Lax-Milgram is a bilinear form. We shall use this in the following example.

Example 13. Lax-Milgram is a very useful theorem to obtain existence of (weak) solutions to PDEs such as

$$\begin{cases} -u''(x) + (2 + \sin(x))u(x) = f(x) & x \in (-1, 1) \\ u(-1) = 0 = u(1) \end{cases} \tag{2.2}$$

where $f \in L^2((-1, 1))$. Note that if f is not continuous, then the equation cannot have a solution $u \in C^2((-1, 1))$. Formally, (2.2) is equivalent to u solving

$$\int_{-1}^{1} \big(u'(x)\phi'(x) + (2 + \sin(x))u(x)\phi(x)\big)dx = \int_{-1}^{1} f(x)\phi(x)dx$$

for all $\phi \in C_c^\infty((-1, 1))$. This form of the equation makes sense as soon as $u, u' \in L^2((-1, 1))$. This formulation does not encode the boundary condition. In order to do so, we pick the inner product space $C_c^\infty((-1, 1))$ equipped with the inner product

$$\langle \psi, \phi\rangle_{H^1} := \langle \psi, \phi\rangle_{L^2} + \langle \psi', \phi'\rangle_{L^2},$$

and complete it to obtain a Hilbert space, denoted $H_0^1((-1,1))$; the lower index is a reminder of the boundary conditions and the upper index tells of the number of derivatives that are required to be square integrable. Hölder's inequality then insures that

$$t(u,v) = \int_{-1}^{1} \big(u'(x)v'(x) + (2+\sin(x))u(x)v(x)\big)dx$$

is a well-defined bounded bilinear form on $H_0^1((-1,1))$ such that

$$|t(u,v)| \leq \|u'\|_{L^2}\|v'\|_{L^2} + 3\|u\|_{L^2}\|v\|_{L^2} \leq 3\|u\|_{H^1}\|v\|_{H^1}.$$

It is moreover coercive with $|t(u,u)| \geq \|u\|_{H^1}^2$. Finally, let $\ell \in \mathcal{H}^*$ be given by

$$\ell(v) = \int_{-1}^{1} f(x)v(x)dx$$

for which $|\ell(v)| \leq \|f\|_{L^2}\|v\|_{L^2} \leq \|f\|_{L^2}\|v\|_{H^1}$. Corollary III.2.5 now yields a unique $u \in H_0^1$ such that $t(u,v) = \ell(v)$, namely

$$\int_{-1}^{1} \big(u'(x)v'(x) + (2+\sin(x))u(x)v(x)\big)dx = \int_{-1}^{1} f(x)v(x)dx$$

for all $v \in H_0^1$. Such a solution is called a *weak solution* of (2.2) and it is given 'for free' by Lax-Milgram.

3. Orthonormal bases

Definition III.3.1. An *orthonormal basis* of $\mathcal{H}$ is a maximal orthonormal set S, namely an orthonormal set such that no other orthonormal set contains S as a proper subset.

Since orthonormal sets can be partially ordered by inclusion and the union of ordered orthonormal sets is an upper bound, the following theorem is a consequence of Zorn's lemma.

Theorem III.3.2. *Every Hilbert space has an orthonormal basis.*

Let now A be an arbitrary nonempty set, and let μ_c be the counting measure on $(A, \mathcal{P}(A))$, namely

$$\mu_c(B) = \begin{cases} |B| & \text{if } B \text{ is finite} \\ +\infty & \text{otherwise} \end{cases}$$

for any $B \in \mathcal{P}(A)$. The space $L^2(A, \mu_c)$ is usually denoted $l^2(A)$ and for any function $f \in L^1(A, \mu_c)$, the integral $\int_A f(\alpha) d\mu_c(\alpha)$ is denoted $\sum_{\alpha \in A} f(\alpha)$. With these definitions, $l^2(A)$ is a Hilbert space with inner product

$$\langle f, g \rangle_{l^2(A)} = \sum_{\alpha \in A} \overline{f(\alpha)} g(\alpha)$$

which is well-defined since $\overline{f} g \in L^1(A, \mu_c)$ provided $f, g \in L^2(A, \mu_c)$. With these definitions, we can state the following result, which allows for a use of orthonormal bases in the spirit of finite-dimensional inner product spaces.

Theorem III.3.3. *Let $\mathcal{H}$ be a Hilbert space and $S = \{v_\alpha : \alpha \in A\}$ be an orthonormal basis. For any $w \in \mathcal{H}$,*

$$w = \sum_{\alpha \in A} \langle v_\alpha, w \rangle v_\alpha$$

and

$$\|w\|^2 = \sum_{\alpha \in A} |\langle v_\alpha, w \rangle|^2. \tag{3.1}$$

Note that the proof shows that for any given $w \in \mathcal{H}$, there are only countably many non-zero terms in the sums, and that the first series converges with respect to the topology of the Hilbert space norm.

Proof. We first prove that the set $A_w = \{\alpha \in A : \langle v_\alpha, w \rangle \neq 0\}$ is at most countable. Indeed,

$$A_w = \bigcup_{n=1}^{\infty} A_w(n) \qquad A_w(n) = \{\alpha \in A : |\langle v_\alpha, w \rangle| \geq \frac{1}{n}\}$$

By Bessel's inequality, $|A_w(n)| \leq n^2 \|w\|^2$, so that A_w is a countable union of finite, possibly empty, sets. Leaving the case of a finite A_w as an exercise, we label $A_w = (\alpha_n)_{n \in \mathbb{N}}$. We further note that the sequence $(\Sigma_N)_{N \in \mathbb{N}}$ given by

$$\Sigma_N = \sum_{n=1}^{N} |\langle v_{\alpha_n}, w \rangle|^2$$

is monotone increasing and bounded above by $\|w\|^2$, hence it is convergent. Let $(w_N)_{N \in \mathbb{N}}$ be the sequence in $\mathcal{H}$ defined by

$$w_N = \sum_{n=1}^{N} \langle v_{\alpha_n}, w \rangle v_{\alpha_n}.$$

It is Cauchy by the above since

$$\|w_M - w_N\|^2 = \sum_{n=N+1}^{M} |\langle v_{\alpha_n}, w\rangle|^2.$$

Let $\tilde{w}$ be its limit. On the one hand,

$$\langle w - \tilde{w}, v_{\alpha_n}\rangle = \lim_{N\to\infty} \langle w - \sum_{j=1}^{N} \langle v_{\alpha_j}, w\rangle v_{\alpha_j}, v_{\alpha_n}\rangle = 0 \tag{3.2}$$

by orthonormality, while for any $\alpha \notin A_w$, both $\langle w, v_\alpha\rangle = 0$ by definition of A_w and $\langle v_{\alpha_n}, v_\alpha\rangle = 0$ for all $n \in \mathbb{N}$ by orthogonality. Hence $w - \tilde{w}$ is orthogonal to all $v \in S$ and hence $w - \tilde{w} = 0$ by maximality of S, proving the first claim of the proposition. The second follows easily from this:

$$\begin{aligned} \|w\|^2 - \sum_{\alpha\in A} |\langle v_\alpha, w\rangle|^2 &= \lim_{N\to\infty} \Big(\|w\|^2 - \sum_{j=1}^{N} |\langle v_{\alpha_j}, w\rangle|^2\Big) \\ &= \lim_{N\to\infty} \Big\|w - \sum_{j=1}^{N} \langle v_{\alpha_j}, w\rangle v_{\alpha_j}\|^2 = 0, \end{aligned}$$

where the second equality follows from (3.2). □

The coefficients

$$\hat{w}(\alpha) = \langle v_\alpha, w\rangle$$

are called *Fourier coefficients* of w with respect to the set $\{v_\alpha : \alpha \in A\}$, and (3.1) is referred to as *Parseval's identity.* In a (complex) Hilbert space, it is equivalently (by the polarization identity) formulated as

$$\langle v, w\rangle = \sum_{\alpha\in A} \overline{\hat{v}(\alpha)}\hat{w}(\alpha).$$

In case that there are countably many elements in S and the index set A can be taken as $\mathbb{N}$, we write $l^2(\mathbb{N}) = l^2$. We conclude with

Corollary III.3.4. *A (complex) Hilbert space $\mathcal{H}$ is separable if and only if it has a countable basis S. If $|S| = N < \infty$, then $\mathcal{H}$ is isomorphic to $\mathbb{C}^N$. Otherwise, $\mathcal{H}$ is isomorphic to l^2.*

Proof. Let $\{v_n : n \in \mathbb{N}\}$ be a countable dense set. We construct recursively a set $\tilde{v}_n$ as follows. Let $\tilde{v}_1 = v_1$. Let $n_0 = \min\{n \in \mathbb{N} : v_n \notin \operatorname{span}\{\tilde{v}_1, \ldots, \tilde{v}_{n-1}\}\}$, and let $\tilde{v}_n = v_{n_0}$. By construction, $\{\tilde{v}_n : n \in \mathbb{N}\}$ is a linearly independent set and its span is the same as the span of $\{v_n : n \in \mathbb{N}\}$, namely dense in $\mathcal{H}$. The Gram-Schmidt orthogonalization applied to $\{\tilde{v}_n : n \in \mathbb{N}\}$ yields a countable orthonormal basis of $\mathcal{H}$. Conversely, given a countable orthonormal basis $\{u_n : n \in \mathbb{N}\}$ of $\mathcal{H}$, Theorem (III.3.3) and its proof show that the countable set of finite linear combinations, with coefficients in a countable dense set of $\mathbb{C}$, of u_n's is dense in $\mathcal{H}$. Hence $\mathcal{H}$ is separable.

Let now $\mathcal{H}$ be separable, and let $\{v_n : n \in \mathbb{N}\}$ be a orthonormal basis. The map

$$U : \mathcal{H} \to l^2, \qquad v \mapsto (\langle v_n, v\rangle)_{n\in\mathbb{N}}$$

is a well-defined isometry by (3.1) and it is onto: Indeed, if $f \in l^2$, then $v = \sum_{n\in\mathbb{N}} f_n v_n$ is a well-defined vector in $\mathcal{H}$ by Theorem (III.3.3) such that $Uv = f$. The finite-dimensional case is similar and elementary. $\square$

We conclude this chapter with some remarks about Fourier series. First of all, we note that the set of functions

$$u_n(t) = \mathrm{e}^{int} \qquad (n \in \mathbb{Z})$$

is an orthonormal set in $L^2(\mathbb{T})$, namely the set of 2π-periodic functions such that

$$\|f\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(t)|^2 dt$$

is finite. It further holds (see exercises) that the set of finite linear combinations of u_n's is dense in $L^2(\mathbb{T})$. Hence the set $\{u_n : n \in \mathbb{N}\}$ is an orthonormal basis of the Hilbert space $L^2(\mathbb{T})$. Then, Theorem III.3.3 shows that any $f \in L^2(\mathbb{T})$ has a representation as

$$f(t) = \sum_{n\in\mathbb{Z}} \hat{f}_n \mathrm{e}^{int}, \quad \text{where} \quad \hat{f}_n = \langle u_n, f\rangle_2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-int} f(t) dt,$$

and reciprocally that given a square integrable double sequence $a \in l^2(\mathbb{Z})$, the series $\sum_{n\in\mathbb{Z}} a_n \mathrm{e}^{int}$ is finite and defines a function in $L^2(\mathbb{T})$. Parseval's identity reads

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{f(t)} g(t) dt = \sum_{n\in\mathbb{Z}} \overline{\hat{f}_n} \hat{g}_n.$$

This shows the power of the abstract approach. However, it also exhibits its restriction in that the only notion of convergence available here is that of the Hilbert space topology, so that questions such as pointwise convergence of Fourier series are out of reach.

4. Spectral theory of self-adjoint compact operators

We first consider the notion of the adjoint of an operator, defined for general Banach spaces in Definition II.6.7, in the case $T \in \mathcal{L}(\mathcal{H})$, namely $T' \in \mathcal{L}(\mathcal{H}^*)$. By Riesz' lemma, the map $C : \mathcal{H} \to \mathcal{H}^*$ given by $\psi \mapsto \langle \psi, \cdot \rangle$ is a surjective isometry. We define

$$T^* = C^{-1} T' C.$$

With this, T^* satisfies the following identity: For any $\psi, \phi \in \mathcal{H}$,

$$\langle \psi, T\phi \rangle = (C\psi)(T\phi) = (T'C\psi)(\phi) = \langle C^{-1}T'C\psi, \phi \rangle = \langle T^*\psi, \phi \rangle. \tag{4.1}$$

The operator T^* is the *Hilbert space adjoint* of T. Its basic propoerties are summarized in the following proposition.

PROPOSITION III.4.1. *(i)* $T \mapsto T^*$ *is an antilinear automorsphism of* $\mathcal{L}(\mathcal{H})$
(ii) $(TS)^* = S^*T^*$
(iii) $(T^*)^* = T$
(iv) $\|T^*T\| = \|T\|^2$

PROOF. (i) follows from Proposition II.6.8 and the fact that C is antilinear. (ii,iii) are immediate calculations. For (iv), we first note that $\|T^*T\| \leq \|T^*\|\|T\| = \|T\|^2$ by (i). Reciprocally,

$$\|T^*T\| \geq \sup\{|\langle \psi, T^*T\psi \rangle| : \|\psi\| = 1\} = \sup\{\|T\psi\|^2 : \|\psi\| = 1\} = \|T\|^2$$

concluding the proof. □

DEFINITION III.4.2. An operator $T \in \mathcal{L}(\mathcal{H})$ is called *self-adjoint* if $T = T^*$.

EXAMPLE 14. The operator $(T\psi)(x) = f(x)\psi(x)$ defined on $L^2(\mathbb{R})$, where f is a real-valued function in $L^\infty(\mathbb{R})$ is bounded and self-adjoint. It is bounded since $\|T\psi\| \leq \|f\|_\infty \|\psi\|$. It is self-adjoint since

$$\langle \psi, T\phi \rangle = \int_{\mathbb{R}} \overline{\psi(x)} f(x) \phi(x) dx = \int_{\mathbb{R}} \overline{f(x)\psi(x)} \phi(x) dx = \langle T\psi, \phi \rangle.$$

for any $\psi, \phi \in L^2(\mathbb{R})$.

We immediately note that eigenvalues of bounded self-adjoint operators are real since $T\psi = z\psi$ implies that $z\langle \psi, \psi\rangle = \langle \psi, T\psi\rangle = \langle T\psi, \psi\rangle = \overline{z}\langle \psi, \psi\rangle$. Moreover, if ψ, ϕ are eigenvectors of a self-adjoint T for different (real) eigenvalues z, w, then

$$z\langle \psi, \phi\rangle = \langle T\psi, \phi\rangle = \langle \psi, T\phi\rangle = w\langle \psi, \phi\rangle$$

namely $\langle \psi, \phi\rangle = 0$: eigenspaces for different eigenvalues are orthogonal.
We now come back to compact operators.

Proposition III.4.3. *The identity is compact on $\mathcal{H}$ if and only if $\mathcal{H}$ is finite dimensional.*

Proof. If $\mathcal{H}$ is finite dimensional, then the identity is a finite rank operator, hence compact. Reciprocally, assume that $\mathcal{H}$ is not finite dimensional. Then the closed unit ball in $\mathcal{H}$, which is a bounded set, is mapped to itself, which is not compact by Theorem II.1.4. $\square$

The first result is a very useful approximation statement.

Theorem III.4.4. *Let $\mathcal{H}$ be a separable Hilbert space and let $T \in \mathcal{L}(\mathcal{H})$. Then T is compact if and only if T is the uniform limit of a sequence of finite rank operators.*

Proof. Let $\{\phi_n : n \in \mathbb{N}\}$ be an orthonormal basis of $\mathcal{H}$, and let

$$z_n = \sup\{\|T\psi\| : \|\psi\| = 1, \langle \phi_j, \psi\rangle = 0\ \forall j = 1, \ldots, n\}$$

The non-negative sequence $\{z_n\}$ is decreasing, hence it has a limit $z \geq 0$. By definition of z_n, there is ψ_n, orthogonal to $\{\phi_1, \ldots, \phi_n\}$, such that $\|\psi_n\| = 1$ and $\|T\psi_n\| \geq z/2$ for n large enough. By Theorem III.3.3 and Cauchy-Schwarz, $|\langle \phi, \psi_n\rangle| = \left|\sum_{j=n+1}^{\infty} \langle \phi_j, \psi_n\rangle \langle \phi, \phi_j\rangle\right| \leq (\sum_{j=n+1}^{\infty} |\langle \phi, \phi_j\rangle|^2)^{\frac{1}{2}}$ which converges to zero. Hence ψ_n converges weakly to 0, therefore $T\psi_n \to 0$ in norm by Theorem II.6.11 and hence $z = 0$.
Now for any $\psi \in \mathcal{H}$,

$$T\psi = T\sum_{j=1}^{\infty} \langle \phi_j, \psi\rangle \phi_j = \sum_{j=1}^{\infty} \langle \phi_j, \psi\rangle T\phi_j$$

because T is bounded. Hence $z_n = \|T - \sum_{j=1}^{n} \langle \phi_j, \cdot\rangle T\phi_j\|$. We conclude that

$$\sum_{j=1}^{n} \langle \phi_j, \cdot\rangle T\phi_j \longrightarrow T$$

in norm. This concludes the proof since the right hand side has finite rank. The other implication holds in a general Banach space, see Theorem II.6.10. $\square$

We now turn to the problem of solving equations.

THEOREM III.4.5. *Let Ω be an open connected subset of $\mathbb{C}$ and let $f : \Omega \to \mathcal{L}(\mathcal{H})$ be an analytic function such that $f(z)$ is a compact operator for all $z \in \Omega$. Then either*
(i) $(1 - f(z))^{-1}$ exists for no $z \in \Omega$ or
(ii) $(1 - f(z))^{-1}$ exists for all $z \in \Omega \setminus S$ where S is a discrete subset of Ω. For $z \in S$, the equation $f(z)\psi = \psi$ has a nonzero solution in $\mathcal{H}$.

PROOF. Let $z_0 \in \Omega$. By continuity, there is $r > 0$ such that $\|f(z_0) - f(z)\| < \frac{1}{2}$ for all $|z - z_0| < r$. By Theorem III.4.4, there is a finite rank operator F such that $\|F - f(z_0)\| < \frac{1}{2}$. Hence, for all $z \in B_r(z_0)$, $\|f(z) - F\| < 1$. The Neumann series then ensures that $1 - f(z) + F$ is invertible and that $z \mapsto (1 - f(z) + F)^{-1}$ is analytic.
Since F is finite rank, there are $\phi_1, \dots, \phi_N$ and linearly independent vectors $\psi_1, \dots, \psi_N$ such that $F = \sum_{j=1}^N \langle \phi_j, \cdot \rangle \psi_j$. Let

$$g(z) = F(1 - f(z) + F)^{-1}.$$

We see that $F = g(z)(1 - f(z) + F)$ and so

$$(1 - f(z)) = (1 - g(z))(1 - f(z) + F).$$

It follows that, for any $z \in B_r(z_0)$, the operator $1 - f(z)$ is invertible if and only if $1 - g(z)$ is invertible and that $f(z)\psi = \psi$ has a nonzero solution if and only if $g(z)\phi = \phi$ has a nonzero solution (with $\phi = (1 - f(z) + F)\psi$).
Now, $g(z)$ is a finite rank operator explicitly given by

$$g(z)\xi = \sum_{j=1}^N \langle \phi_j, (1 - f(z) + F)^{-1}\xi \rangle \psi_j = \sum_{j=1}^N \langle \zeta_j(z), \xi \rangle \psi_j$$

where $\zeta_j(z) = ((1 - f(z) + F)^{-1})^* \phi_j$. If $g(z)\phi = \phi$ has a solution (in particular, $1 - g(z)$ is not invertible), then ϕ is in the range of $g(z)$ and therefore $\phi = \sum_{j=1}^N \beta_j \psi_j$ and the β_j's solve

$$\beta_j = \sum_{k=1}^N \langle \zeta_j(z), \psi_k \rangle \beta_k \tag{4.2}$$

Reciprocally, if $\beta_1, \ldots, \beta_N$ solve this equation, then $\phi = \sum_{j=1}^N \beta_j \psi_j$ is a solution of $g(z)\phi = \phi$. But (4.2) is a simple matrix equation, which has a nonzero solution if and only if the determinant $d(z)$ of $(\delta_{jk} - \langle \zeta_j(z), \psi_k \rangle)$ equals zero. Since $z \mapsto \zeta_j(z)$ is analytic in $B_r(z_0)$, so is $d(z)$, and so either $S_r(z_0) = \{z \in B_r(z_0) : d(z) = 0\}$ is a discrete set in $B_r(z_0)$ with no limit point, or $S_r(z_0) = B_r(z_0)$.

If $z \notin S_r(z_0)$, namely $d(z) \neq 0$, then for any ψ, the equation $(1 - g(z))\phi = \psi$ has the solution $\phi = \psi + \sum_{j=1}^N \alpha_j \psi_j$, provided the α_j's solve

$$\alpha_j - \langle \zeta_j(z), \psi \rangle - \sum_{k=1}^N \alpha_k \langle \zeta_j(z), \psi_k \rangle = 0.$$

This matrix equation has a unique solution since $d(z) \neq 0$, which proves that $z \notin S_r(z_0)$ implies that $1 - g(z)$ is invertible.

This proves the theorem locally. It global version is obtained using the connectedness of Ω and we skip the details here. $\square$

The Fredholm alternative, Theorem III.4.5 has important consequences:

Corollary III.4.6. *If T is a compact operator on $\mathcal{H}$, then either $(1-T)^{-1}$ exists or $T\psi = \psi$ has a non-zero solution.*

Proof. Consider $f(z) = zT$ and use the Fredholm alternative at $z = 1$. $\square$

We note that while the result above is true for finite-dimensional matrices, it is not true in general. Indeed, the operator T on $\mathcal{H} = L^2([0,1])$ defined by $(T\psi)(x) = 2x\psi(x)$ is so that $T\psi = \psi$ has no nonzero solution in $\mathcal{H}$ and $(1 - T)^{-1}$ does not exist as a bounded operator on $\mathcal{H}$, see also Example 8.

The second corollary is about the spectrum of a compact operator

Theorem III.4.7. *Let T be a compact operator on $\mathcal{H}$. Then $\sigma(T)$ is a set having no limit point except possibly at $z = 0$. Moreover, any nonzero $z \in \sigma(T)$ is an eigenvalue of finite multiplicity.*

Proof. Consider again $f(z) = zT$. By the Fredholm alternative with $\Omega = \mathbb{C}$, the set $S = \{z : zT\psi = \psi \text{ has a nonzero solution}\}$ is discrete: Indeed, (i) of the alternative is excluded since $(1 - zT)^{-1}$ exists at $z = 0$. Now, the identity

$$(z - T)^{-1} = z^{-1}(1 - z^{-1}T)^{-1}$$

valid for all $z \in \mathbb{C} \setminus \{0\}$ shows that $z \in \rho(T)$ if and only if $1 - z^{-1}T$ is invertible, namely $z^{-1} \notin S$. Equivalently $z \in \sigma(T)$ if and only if $z^{-1} \in S$, showing that $\sigma(T)$ is a discrete set, except possibly at $\{0\}$. Moreover, if $z \in \sigma(T) \setminus \{0\}$, then $T\psi = z\psi$ has a nonzero solution, and hence z is an eigenvalue. If the corresponding eigenspace $\mathcal{H}_z$ was infinite dimensional, then $z^{-1}T \restriction \mathcal{H}_z = 1_{\mathcal{H}_z}$ which is not compact, a contradiction. Hence all eigenspaces are finite dimensional. $\square$

Theorem III.4.8. *Let T be a self-adjoint compact operator on an infinite dimensional Hilbert space $\mathcal{H}$. Then there is an orthonormal basis $\{\psi_n : n \in \mathbb{N}\}$ such that $T\psi_n = z_n\psi_n$ and $\lim_{n\to\infty} z_n = 0$.*

Proof. For each eigenvalue of T, we pick an orthonormal basis of the corresponding finite-dimensional eigenspace. The collection $\{\psi_n\}$ of these vectors is an orthonormal set since eigenvectors for different eigenvalues of a self-adjoint operator are orthogonal. We prove that it is a complete set. Let $\mathcal{K}$ be the closure of their span. Then $T\mathcal{K} \subset \mathcal{K}$. Moreover, if $\phi \in \mathcal{K}^\perp$, then $\langle \psi, T\phi\rangle = \langle T\psi, \phi\rangle = 0$ for all $\psi \in \mathcal{K}$, namely $T\mathcal{K}^\perp \subset \mathcal{K}^\perp$. Let $\tilde{T} = T \restriction (\mathcal{K}^\perp)$. Then $\tilde{T}$ is self-adjoint and compact. If $z \neq 0$ is in $\sigma(\tilde{T})$, then it is an eigenvalue of $\tilde{T}$, and therefore an eigenvalue of T. Hence the spectral radius of $\tilde{T}$ is zero since any corresponding eigenvector belongs to $\mathcal{K}$. But $\|\tilde{T}\|^2 = \|\tilde{T}^2\|$ since $\tilde{T}$ is self-adjoint and so $\|\tilde{T}\|^{2^n} = \|\tilde{T}^{2^n}\|$, which implies that

$$r_{\tilde{T}} = \|\tilde{T}\|.$$

Hence $\tilde{T} = 0$. If $\phi \in \mathcal{K}^\perp$ is non-zero, then $\tilde{T}\phi = 0$, namely $\phi \in \mathcal{K}$ a contradiction. Hence $\mathcal{K}^\perp = \{0\}$, and therefore $\mathcal{K} = \mathcal{H}$.

Since $\mathcal{H}$ is infinite dimensional and each eigenspace is finite dimensional, the set $\{z_n\}$ must be infinite. It is bounded since it is a subset of the spectrum of T, hence $\{z_n : n \in \mathbb{N}\}$ has a limit point. Theorem III.4.7 implies that it must be 0. $\square$

By Parseval's identity, $T\phi = \sum_{n=1}^{\infty} \langle \psi_n, T\phi \rangle \psi_n = \sum_{n=1}^{\infty} z_n \langle \psi_n, \phi \rangle \psi_n$, and so

$$T = \sum_{n=1}^{\infty} z_n \langle \psi_n, \cdot \rangle \psi_n$$

for any compact self-adjoint operator $T \in \mathcal{L}(\mathcal{H})$ on a separable Hilbert space. This is the generalization of the *spectral theorem* for self-adjoint matrices on $\mathbb{C}^N$. A similar representation holds for a not necessarily self-adjoint compact operator.

THEOREM III.4.9. *Let T be a compact operator. Then there are orthonormal bases $\{\psi_n : n \in \mathbb{N}\}$ and $\{\phi_n : n \in \mathbb{N}\}$ and non-negative real numbers t_n such that*

$$T = \sum_{n=1}^{\infty} t_n \langle \phi_n, \cdot \rangle \psi_n$$

where the sum is convergent in norm.

The t_n's are called the *singular values* of the operator T.

PROOF. The operator TT^* is compact and self-adjoint because T is compact and T^* is bounded. Hence Theorem III.4.8 imples the existence of an ONB $\{\psi_n : n \in \mathbb{N}\}$ and real numbers z_n such that $z_n \to 0$ and $TT^*\psi_n = z_n\psi_n$. Note that $0 \leq \|T^*\psi_n\|^2 = z_n$, namely z_n are non-negative, so we let $t_n = \sqrt{z}_n$. Assuming first that they are all positive (and so $T^*\psi_n \neq 0$), we let $t_n\phi_n = T^*\psi_n$ and check that $\langle \phi_m, \phi_n \rangle = \frac{1}{t_n t_m} \langle \psi_m, TT^*\psi_n \rangle = \frac{z_n}{t_n t_m} \delta_{n,m} = \delta_{n,m}$, namely $\{\psi_n : n \in \mathbb{N}\}$ is an orthonormal set. For any $\xi \in \mathcal{H}$,

$$\sum_{n=1}^{\infty} t_n \langle \phi_n, \xi \rangle \psi_n = \sum_{n=1}^{\infty} \langle T^*\psi_n, \xi \rangle \psi_n = \sum_{n=1}^{\infty} \langle \psi_n, T\xi \rangle \psi_n = T\xi$$

by Parseval's identity. If $\mathrm{Ker}(T)$ is non-trivial, namely there are n such that $z_n = 0$, the argument above applies to $\mathrm{Ker}(T)^\perp$, and it is complemented by picking an arbitrary ONB of $\mathrm{Ker}(T)$. □

Theorem III.4.8 further allows one to define a *functional calculus* for self-adjoint compact operators. Let $\{\psi_n : n \in \mathbb{N}\}$ be the orthonormal basis exhibited in the theorem with eigenvalues $\{z_n : n \in \mathbb{N}\}$, and let f be a complex-valued bounded function defined on $\sigma(T)$. We define an operator $f(T)$ by letting, for any $\phi \in \mathcal{H}$,

$$\phi = \sum_{j=1}^{\infty} \alpha_j \psi_j, \qquad f(T)\phi = \sum_{j=1}^{\infty} f(z_j) \alpha_j \psi_j. \tag{4.3}$$

We have:

THEOREM III.4.10. *Let T be a self-adjoint compact operator. For any complex-valued bounded function f defined on $\sigma(T)$, let $f(T)$ be defined in (4.3). Then*
(i) The operator associated with the constant function 1 is the identity.
(ii) The operator associated with the function $f(z) = z$ is T itself.
*(iii) The map $f \mapsto f(T)$ is an isometric *-isomorphism of the algebra of bounded functions on $\sigma(T)$ into the algebra $\mathcal{L}(\mathcal{H})$; in particular,*

$$\|f(T)\| = \sup\{|f(z)| : z \in \sigma(T)\}.$$

and $\bar{f}(T) = (f(T))^*$.
(iv) If f is real-valued, then $f(T)$ is self-adjoint.

PROOF. (i,ii) are immediate. For (iii), we see that

$$\begin{aligned} f(T)g(T)\phi &= \sum_{n=1}^{\infty} f(z_n)\langle\psi_n, g(T)\phi\rangle = \sum_{n,m=1}^{\infty} f(z_n)g(z_m)\langle\psi_n, \psi_m,\rangle\langle\psi_m, \phi\rangle \\ &= \sum_{n=1}^{\infty}(fg)(z_n)\langle\psi_n, \phi\rangle = (fg)(T). \end{aligned}$$

The isometric property follows from

$$\|f(T)\|^2 = \sup\left\{\sum_{j=1}^{\infty} |f(z_j)|^2|\alpha_j|^2 : \sum_{j=1}^{\infty} |\alpha_j|^2 = 1\right\} = \sup\left\{|f(z_j)|^2 : j \in \mathbb{N}\right\}.$$

Here, we used Theorem III.3.3 in the first equality, and the fact that expression obtained is a convex combination of $\{|f(z_j)|^2\}$. Moreover,

$$\langle\overline{f}(T)\phi, \psi\rangle = \sum_{n=1}^{\infty} f(z_n)\overline{\langle\psi_n, \phi\rangle}\langle\psi_n, \psi\rangle = \langle\phi, \sum_{n=1}^{\infty} f(z_n)\langle\psi_n, \psi\rangle\psi_n\rangle = \langle\phi, f(T)\psi\rangle.$$

Finally, (iv) follows from the last part of (iii). □

5. The Hellinger-Toeplitz theorem

In operator theory, one encounters many unbounded linear operators T that are a priori defined only on dense subsets of a Hilbert space $\mathcal{H}$. For example $-\mathrm{i}\frac{d}{dx}$ on the dense subset $C_c^\infty(\mathbb{R})$ of $L^2(\mathbb{R})$. In this case, the domain is part of the definition of an operator, and the definitions and results above, such as (4.1) must be handled with care. Let T be a linear map with dense domain $D(T)$. Its adjoint is defined as follows. A vector $\phi \in \mathcal{H}$ is in $D(T^*)$ is the linear functional

$$D(T) \ni \psi \mapsto \langle \phi, T\psi \rangle \in \mathbb{C}$$

is bounded. In that case, the Riesz lemma yields a vector $\zeta_\phi \in \mathcal{H}$ such that

$$\langle \phi, T\psi \rangle = \langle \zeta_\phi, \psi \rangle$$

for all $\psi \in D(T)$. One then defines $T^*\phi = \zeta_\phi$ for any $\phi \in D(T^*)$. In other words,

$$\langle T^*\phi, \psi \rangle = \langle \phi, T\psi \rangle \qquad (\phi \in D(T^*), \psi \in D(T)).$$

An operator T with domain $D(T)$ is *self-adjoint* if $T = T^*$, in particular $D(T) = D(T^*)$. It is called *symmetric* if $\langle T\phi, \psi \rangle = \langle \phi, T\psi \rangle$ holds for all $\phi, \psi \in D(T)$. In the latter case, we write $T \subset T^*$, which means that $D(T) \subset D(T^*)$ and $T\psi = T^*\psi$ for all $\psi \in D(T)$, and T^* is an *extension* of T.

Here is a brief example, without the technical details. We consider $-\mathrm{i}\frac{d}{dx}$ on various domains in $L^2([0,1])$. The maximal domain is $D(\tilde{T}) = \{\psi \in L^2([0,1]) : \psi' \in L^2([0,1])\}$ where the derivative must be understood in the sense of distributions. We check that $\tilde{T}$ is unbounded but not self-adjoint. Indeed, for any $z \in \mathbb{C}$, the equation $-\mathrm{i}\psi' = z\psi$ has a non-zero solution, namely $\psi(x) = \mathrm{e}^{\mathrm{i}zx}$, and both ψ, ψ' are square integrable so that $\psi \in D(\tilde{T})$. Hence $\sigma(\tilde{T}) = \mathbb{C}$ and every $z \in \sigma(\tilde{T})$ is an eigenvalue. Let us now consider $D(T_0) = \{\psi \in L^2([0,1]) : \psi(0) = \psi(1) = 0\}$, so that $T_0 \subset \tilde{T}$. We note that the existence of a distributional derivative in $L^2([0,1])$ implies, in one dimension, that ψ is in fact continuous, so that the pointwise boundary condition makes sense. T_0 is a symmetric operator since

$$\int_0^1 \overline{\phi(x)}\mathrm{i}\psi'(x)dx = \int_0^1 \overline{\mathrm{i}\phi'(x)}\psi(x)dx$$

for all $\phi, \psi \in D(T_0)$ by integration by parts. It is however not self-adjoint, and one can verify that $(T_0)^* = \tilde{T}$. Moreover, T_0 has no eigenvalues. Finally, we consider the one-parameter family of operator T_θ labelled by $\theta \in [0, 2\pi)$ given by $D(T_\theta) = \{\psi \in L^2([0,1]) : \psi(1) = \mathrm{e}^{\mathrm{i}\theta}\psi(0)\}$. One checks that they are all self-adjoint, namely

$$T_0 \subset T_\theta = (T_\theta)^* \subset \tilde{T} = (T_0)^*.$$

Here, the set of eigenvalues is discrete $\{\theta + 2\pi n : n \in \mathbb{N}\}$ for the eigenfunctions $\psi_n(x) = \mathrm{e}^{\mathrm{i}(\theta+2\pi n)x}$.

That the subtlety between symmetry and self-adjointness arises only in the case of unbounded operators follows from the following Hellinger-Toeplitz theorem.

THEOREM III.5.1. *Let T be an everywhere defined linear operator on a Hilbert space $\mathcal{H}$ such that $\langle \phi, T\psi \rangle = \langle T\phi, \psi \rangle$ for all $\phi, \psi \in \mathcal{H}$. Then T is bounded.*

PROOF. By the closed graph theorem, it suffices to prove that $\Gamma(T)$ is closed. Let $(\psi_n, T\psi_n) \to (\psi, \phi)$ in $\mathcal{H} \times \mathcal{H}$. For any $\xi \in \mathcal{H}$,

$$\langle \xi, \phi \rangle = \lim_{n\to\infty} \langle \xi, T\psi_n \rangle = \lim_{n\to\infty} \langle T\xi, \psi_n \rangle = \langle T\xi, \psi \rangle = \langle \xi, T\psi \rangle.$$

Since linear functionals separate, $\phi = T\psi$. $\square$

CHAPTER IV

The Riesz-Markov theorem

1. Facts from measure theory

- A *σ-algebra* on a set S is a collection $\mathcal{A}$ of subsets of S that is closed under countable unions and complements. It follows that $\emptyset, S \in \mathcal{A}$. In a topological space, the Borel σ-algebra $\mathcal{B}_S$ is generated by the open sets, namely it is the smallest σ-algebra containing all open sets.
- If S is a set and $\mathcal{A}$ is a σ-algebra on S, a *measure* on $(S, \mathcal{A})$ is a countably additive set function $\mu : \mathcal{A} \to [0, \infty]$ such that $\mu(\emptyset) = 0$, namely $\mu(\cup_{i=1}^{\infty} M_i) = \sum_{i=1}^{\infty} \mu(M_i)$ for any family of disjoint sets $M_i \in \mathcal{A}$. In particular, a measure is monotonous and countably subadditive for any family of sets $M_i \in \mathcal{A}$. A measure is *complete* is all subsets of null sets are measurable, and any measure can be completed to a complete measure.
- An *outer measure* on a set S is a monotonous, countably subadditive set function $\mu^* : \mathcal{P}(S) \to [0, \infty]$ such that $\mu^*(\emptyset) = 0$, namely $X \subset Y \Rightarrow \mu^*(X) \leq \mu^*(Y)$ and $\mu^*(\cup_{i=1}^{\infty} M_i) \leq \sum_{i=1}^{\infty} \mu^*(M_i)$ for any family of sets $M_i \in \mathcal{P}(S)$. The *μ^*-measurable sets* are the $X \in \mathcal{P}(S)$ such that $\mu^*(Y) = \mu^*(Y \cap X) + \mu^*(Y \cap X^c)$ for all $Y \in \mathcal{P}(S)$.
- *Construction Lemma*: Let $\mathcal{E} \subset \mathcal{P}(S)$, and let $\rho : \mathcal{E} \to [0, \infty]$ be such that $\emptyset \in \mathcal{E}, S \in \mathcal{E}$ and $\rho(\emptyset) = 0$. For any $M \in \mathcal{P}(S)$, let
$$\mu^*(M) = \inf \Big\{ \sum_{i=1}^{\infty} \rho(E_i) : E_i \in \mathcal{E} \text{ and } M \subset \bigcup_{i=1}^{\infty} E_i \Big\}.$$
Then μ^* is an outermeasure.
- *Carathéodory's theorem*: Let μ^* be an outer measure and let $\mathcal{A}$ be the set of μ^*-measurable sets. Then $\mathcal{A}$ is a σ-algebra and $\mu^* \restriction \mathcal{A}$ is a complete measure.

From here on, X is a locally compact Hausdorff space, and it is always understood to be equipped with its Borel σ-algebra $\mathcal{B}_X$.

A measure μ is *outer regular* on $A \in \mathcal{B}_X$ if

$$\mu(A) = \inf\{\mu(O) : O \text{ open and } A \subset O\},$$

inner regular on $A \in \mathcal{B}_X$ if

$$\mu(A) = \sup\{\mu(K) : K \text{ compact and } K \subset A\},$$

and *regular* if it is both inner and outer regular on all Borel sets. Finally, a *Radon measure* on X is a Borel measure that is

(i) inner regular on all open sets

(ii) outer regular on all Borel sets

(iii) finite on all compact sets

We immediately point out that by (ii), a Radon measure is completely determined by its value on open sets.

2. The representation theorem

We have already seen the Riesz representation theorem, Theorem II.3.3, in the context of L^p-spaces: There is a one-to-one correspondence between bounded linear functional on L^p and functions in L^q, where p, q are dual indices. In a similar fashion, we now turn to linear functional over $C_c(X)$, the set of continuous functions over S with compact support. A linear functional I on $C_c(X)$ is called *positive* if

$$f \geq 0 \quad \Rightarrow \quad I(f) \geq 0.$$

Let now μ be a Radon measure. Since $\mu(K) < \infty$ for any compact K, we have that $C_c(X) \subset L^1(X, \mu)$, which can be rephrased as: The map $I_\mu : C_c(X) \to \mathbb{C}$ defined by

$$I_\mu(f) = \int_X f d\mu$$

is a positive linear functional. Just as in the L^p case, it turns out that every positive linear functional on $C_c(X)$ is of the above form for a unique Radon measure.

First of all, we note that a positive linear functional is locally bounded.

Proposition IV.2.1. *Let I be a positive linear functional on $C_c(X)$. For any compact $K \subset X$, there is a constant C_K such that $|I(f)| \leq C_K \|f\|_\infty$ for all $f \in C_c(X)$ supported in K.*

Proof. Without loss of generality, we assume that f is real-valued. By Urysohn's Lemma, there exists $\phi \in C_c(X)$ such that $K \prec \phi$, and $\text{supp}(f) \subset K$ implies that $|f| \leq \|f\|_\infty \phi$. It follows that both $\|f\|_\infty \phi \pm f$ are positive functions and hence $|I(f)| \leq I(\phi)\|f\|_\infty$, by the linearity and positivity of I. □

We are now ready to state and prove a first version of the Riesz-Markov representation theorem.

Theorem IV.2.2. *Let I be a positive linear functional on $C_c(X)$. There is a unique Radon measure μ such that*

$$I(f) = \int_X f d\mu \tag{2.1}$$

for all $f \in C_c(X)$.

Note that the proof will establish the following properties:

$$\mu(O) = \sup\{I(f) : f \in C_c(X), f \prec O\}$$

for any open set O (this is in fact how the measure will be defined), and

$$\mu(K) = \inf\{I(f) : f \in C_c(X), K \prec f\} \tag{2.2}$$

for any compact set K.

Proof. For any open set O, let

$$\mu(O) = \sup\{I(f) : f \in C_c(X), f \prec O\} \tag{2.3}$$

and let

$$\mu^*(M) = \inf\{\mu(O) : O \text{ open and } M \subset O\} \tag{2.4}$$

for any set $M \in \mathcal{P}(X)$. Since, by the definition of μ, $O_1 \subset O_2$ implies $\mu(O_1) \leq \mu(O_2)$, we conclude that the infimum in the definition of $\mu^*(O)$ is reached at O, namely $\mu^*(O) = \mu(O)$ for any open set O. We will use this repeatedly below.

We first establish that μ^* is an outer measure, using the construction lemma. Let $O = \cup_{i=1}^{\infty} O_i$ be a countable union of open sets. Let $f \in C_c(X)$ be such that $f \prec O$, with $\mathrm{supp} f = K$. By compactness, $K \subset \cup_{i=1}^{n} O_i$ so that Proposition I.5.3 yields a partition of unity $\{g_i \in C_c(X) : 1 \leq i \leq n\}$ on K such that $g_i \prec O_i$. In particular, $f = \sum_{i=1}^{n} f g_i$ and $f g_i \prec O_i$, so that by definition (2.3) of μ,

$$I(f) = \sum_{i=1}^{n} I(f g_i) \leq \sum_{i=1}^{n} \mu(O_i) \leq \sum_{i=1}^{\infty} \mu(O_i).$$

Taking the supremum of all such f, we conclude that $\mu(O) \leq \sum_{i=1}^{\infty} \mu(O_i)$. For any $M \in \mathcal{P}(X)$, we therefore have by (2.4) that

$$\mu^*(M) = \inf \Big\{ \sum_{i=1}^{\infty} \mu(O_i) : O_i \text{ open and } M \subset \bigcup_{i=1}^{\infty} O_i \Big\}.$$

This proves the claim since the construction lemma ensures that the expression on the right hand side defined an outer measure.

The next step is to show that every open set in μ^*-measurable. Let O be open and let M be such that $\mu^*(M) < \infty$. Then by subadditivity, $\mu^*(M) \leq \mu^*(M \cap O) + \mu^*(M \setminus O)$ so it suffices to prove the opposite bound. If M is open, so is $O \cap M$, so for any $\epsilon > 0$, there is by (2.3) an $f \in C_c(X)$ such that $f \prec O \cap M$ and $I(f) > \mu(O \cap M) - \epsilon$. By the same argument applied to the open set $M \setminus \mathrm{supp} f$, there is a $g \in C_c(X)$ such that $g \prec M \setminus \mathrm{supp} f$ and $I(g) > \mu(M \setminus \mathrm{supp} f) - \epsilon$. Since $f + g \prec M$ and $M \setminus \mathrm{supp} f \supset M \setminus O$,

$$\mu^*(M) = \mu(M) \geq I(f) + I(g) > \mu^*(O \cap M) + \mu^*(M \setminus O) - 2\epsilon,$$

which yields the desired inequality since ϵ is arbitrary. If M is arbitrary, the definition of μ^* implies that for any $\epsilon > 0$, there is an open set $U \supset M$ such that $\mu(U) < \mu^*(M) + \epsilon$ and we conclude by the above that

$$\mu^*(M) + \epsilon > \mu(U) \geq \mu^*(U \cap O) + \mu^*(U \setminus O) \geq \mu^*(M \cap O) + \mu^*(M \setminus O),$$

which yields the claim for a general set M.

By Carathéodory's theorem and since $\mathcal{B}_X$ is the smallest σ-algebra containing all open sets, every Borel set is μ^*-measurable and $\mu = \mu^* \restriction \mathcal{B}_X$ is a Borel measure (note that the μ here is an extension of μ defined in (2.3)). By construction, it is furthermore outer regular. Let K be compact and let $f \in C_c(X)$ be such that $K \prec f$. For any $\epsilon > 0$, the set

$O_\epsilon = \{x : f(x) > 1 - \epsilon\} \supset K$ is open by continuity. For any $g \prec O_\epsilon$, we have that $(1-\epsilon)^{-1}f - g$ is a positive function, so that $I(g) \leq (1-\epsilon)^{-1}I(f)$. Taking the supremum over such g yields by (2.3)

$$\mu(K) \leq \mu(O_\epsilon) \leq (1-\epsilon)^{-1}I(f)$$

for any compact set K and hence $\mu(K) \leq I(f)$ since ϵ is arbitrary. For any open $O \supset K$, Urysohn's lemma provides a $f \in C_c(X)$ such that $K \prec f \prec O$. Hence $I(f) \leq \mu(O)$ and we conclude that (2.2) holds since μ is outer regular on K.

With (2.2), we conclude by Proposition IV.2.1 that $\mu(K) < \infty$ for any compact set.

Equation (2.2) further implies inner regularity on open sets. Let O be open and let $\epsilon > 0$. By (2.3), there is $f \in C_c(X)$ such that $f \prec O$ and $I(f) > \mu(O) - \epsilon$. Let $K = \text{supp} f$ and let $K \prec g$. Then $g - f$ is positive and hence $I(g) \geq I(f) \leq \mu(O) - \epsilon$. Since this holds for any such g, we conclude that $\mu(K) \geq \mu(O) - \epsilon$ and hence μ is inner regular on O.

We have now constructed a Borel measure μ and established that it is indeed a Radon measure. It remains to prove the identity (2.8). First of all, it suffices by linearity to prove it for functions $0 \leq f \leq 1$. Let $K_0 = \text{supp} f$, let $N \in \mathbb{N}$ and for any $1 \leq j \leq N$, let

$$K_j = \{x \in X : f(x) \geq jN^{-1}\}$$

for which $K_j \subset K_{j-1}$, $(1 \leq j \leq N)$, and let

$$f_j(x) = \begin{cases} 0 & \text{if } x \notin K_{j-1} \\ f(x) - (j-1)N^{-1} & \text{if } x \in K_{j-1} \setminus K_j \\ N^{-1} & \text{if } x \in K_j \end{cases}$$

Clearly, $f = \sum_{j=1}^N f_j$ and $N^{-1}\chi_{K_j} \leq f_j \leq N^{-1}\chi_{K_{j-1}}$ and hence

$$N^{-1}\mu(K_j) \leq \int_X f_j d\mu \leq N^{-1}\mu(K_{j-1}). \tag{2.5}$$

Summing over j yields

$$N^{-1}\sum_{j=1}^{N}\mu(K_j) \leq \int_X f d\mu \leq N^{-1}\sum_{j=0}^{N-1}\mu(K_j). \tag{2.6}$$

Let now O be an open set containing K_{j-1}. Then $K_j \prec Nf_j \prec O$. By (2.2) $\mu(K_j) \leq NI(f_j)$, while by definition (2.3) of μ, $NI(f_j) \leq \mu(O)$. By outer regularity, we conclude that

$$N^{-1}\mu(K_j) \leq I(f_j) \leq N^{-1}\mu(K_{j-1}),$$

which yields upon summation over j the same bound as (2.6) but for $I(f)$. Together, these inequalities imply that

$$\left| I(f) - \int_X f d\mu \right| \leq \frac{\mu(K_0) - \mu(K_N)}{N} \leq \frac{\mu(\mathrm{supp} f)}{N}.$$

This concludes the proof of (2.8) since $\mathrm{supp} f$ is finite and N arbitrary.
To conclude the proof of the theorem, it remains to show uniqueness. Let ν be a Radon measure such that $I(f) = \int_X f d\nu$. Let O be open and $K \subset O$ compact. By Urysohn's lemma, there is $K \prec f \prec O$, which implies that $\nu(K) \leq \int_X f d\nu = I(f) \leq \nu(O)$ by the properties of the integral and the fact that $0 \leq f \leq 1$. By inner regularity on open sets, we conclude that ν satisfies (2.3). Hence, it is equal to μ on open sets, and further on all sets by outer regularity. $\square$

Our last goal is to discuss an extension of the above theorem to the dual of $C_0(X)$, the space of continuous functions vanishing at infinity on a LCH space X. The first step is the following result, showing that functions vanishing at infinity are exactly the uniform limits of compactly supported functions.

Lemma IV.2.3. *Let X be a LCH space. Then $\overline{C_c(X)} = C_0(X)$, where the closure is in the uniform topology.*

Proof. Since the uniform topology is a metric topology, it suffices to consider sequences. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $C_c(X)$ that converges to f uniformly. Let $\epsilon > 0$ and $n_0 \in \mathbb{N}$ such that $\sup\{|f_{n_0}(x) - f(x)| : x \in X\} < \epsilon$. It follows that $|f(x)| < \epsilon$ for x outside of the compact support of f_{n_0}. This shows that $f \in C_0(X)$, namely $\overline{C_c(X)} \subset C_0(X)$. Reciprocally, let $f \in C_0(X)$ and let $n \in \mathbb{N}$. There is a compact K_n such that $|f(x)| < 1/n$ for all $x \in X \setminus K_n$. By Urysohn's lemma, there is $g_n \in C_c(X)$ such that $K_n \prec g_n$. Then $(g_n f)_{n\in\mathbb{N}}$ is a sequence in $C_c(X)$ such that

$$\sup\{|g_n(x)f(x) - f(x)| : x \in X\} = \sup\{|1 - g_n(x)||f(x)| : x \in X \setminus K_n\} < 1/n$$

namely $g_n f \to f$ uniformly. Hence $\overline{C_c(X)} \supset C_0(X)$, concluding the proof. $\square$

It immediately follows that any positive linear functional I on $C_c(X)$ extends uniquely to a bounded positive linear functional on $C_0(X)$, if and only if it is globally bounded with respect to the uniform topology. But the Riesz representation, in particular (2.3), implies

$$\mu(X) = \sup\{I(f) : f \in C_c(X), 0 \leq f \leq 1\} \tag{2.7}$$

showing that I is bounded if and only if $\mu(X) < \infty$, in which case $\|I\| = \mu(X)$. We have just proved:

PROPOSITION IV.2.4. *Let X be a LCH space. Let I be a bounded positive linear functional on $C_0(X)$. There is a unique finite Radon measure μ such that*

$$I(f) = \int_X f d\mu \tag{2.8}$$

for all $f \in C_0(X)$.

It remains to remove the positivity condition. Analogously to the Jordan decomposition of measures, general real-valued bounded linear functionals decompose into a positive and negative part, and complex-valued functionals decompose real and imaginary parts, which in turn decompose into positive and negative parts, to which the proposition above can be applied.

LEMMA IV.2.5. *Let $I : C_0(X; \mathbb{R}) \to \mathbb{R}$ be a real-valued bounded linear functional on $C_0(X; \mathbb{R})$. There exist positive bounded linear functionals $I^{\pm} \in C_0(X; \mathbb{R})^*$ such that $I = I^+ - I^-$.*

PROOF. Let f be a non-negative continuous function vanishing at infinity and let

$$I^+(f) = \sup\{I(g) : g \in C_0(X; \mathbb{R}), 0 \leq g \leq f\}.$$

In particular, $I(f) \leq I^+(f)$. Since $I(0) = 0$, we conclude that $I^+(f) \geq 0$. Taking the supremum of $|I(g)| \leq \|I\|\|g\|_\infty$ over $0 \leq g \leq f$ yields

$$0 \leq I^+(f) \leq \|I\|\|f\|_\infty.$$

By linearity of I, $I^+(rf) = rI^+(f)$ for $r \geq 0$. Moreover,

$$\begin{aligned}
I^+(f_1+f_2) &= \sup\{I(g) : g \in C_0(X;\mathbb{R}), 0 \leq g \leq f_1+f_2\} \\
&\geq \sup\{I(g_1+g_2) : g_{1,2} \in C_0(X;\mathbb{R}), 0 \leq g_1 \leq f_1, 0 \leq g_2 \leq f_2\} \\
&= I^+(f_1) + I^+(f_2).
\end{aligned}$$

On the other hand, let $0 \leq g \leq f_1+f_2$. If $g_1 = \min\{f_1, g\}$ then $0 \leq g_1 \leq f_1$, while $g_2 = g - g_1$ satisfies $0 \leq g_2 \leq f_2$. Since I is linear,

$$I(g) = I(g_1) + I(g_2) \leq I^+(f_1) + I^+(f_2),$$

which implies $I^+(f_1+f_2) \leq I^+(f_1) + I^+(f_2)$ by taking the supremum. Altogether, I^+ is a positive, bounded, linear, functional on the set of non-negative functions. For an arbitrary $f \in C_0(X;\mathbb{R})$, let $f = f_+ - f_-$ be its decomposition into positive and negative parts, and let $I^+(f) = I^+(f_+) - I^+(f_-)$. On $C_0(X;\mathbb{R})$, this is linear and bounded since $|I^+(f)| \leq \|I\| \max\{\|f_+\|_\infty, \|f_-\|_\infty\} = \|I\|\|f\|_\infty$, namely $\|I^+\| \leq \|I\|$. It remains to define $I^- = I^+ - I \in C_0(X;\mathbb{R})^*$, which is a positive functional. $\square$

Let now $I \in C_0(X)^*$ be a bounded complex-linear functional over the complex-valued continuous functions vanishing at infinity. For any $f \in C_0(X)$, we write $f = u + \mathrm{i}v$, where $u, v \in C_0(X;\mathbb{R})$, so that $I(f) = J(u) + \mathrm{i}J(v)$ is completely determined by its real-linear restriction $J = I \restriction C_0(X;\mathbb{R})$. By Lemma IV.2.5, $J = J^+ + J^-$ and hence there are two finite Radon measures μ^+, μ_- such that

$$I(f) = \int_X (u^+ + \mathrm{i}v^+) d\mu^+ - \int_X (u^- + \mathrm{i}v^-) d\mu^- \qquad (f = u + \mathrm{i}v)$$

by Proposition 2.8.

We have reached the final form of the Riesz-Markov theorem. We denote $M(X)$ the space of all complex Radon measures, namely of set of complex Borel measure such that the real and imaginary parts are finite, signed Radon measures. Given $\mu \in M(X)$, we let

$$I_\mu(f) = \int_X f d\mu.$$

which is a bounded linear functional on $C_0(X)$. Note that $M(x)$ is a Banach space when equipped with the norm of total variation $\|\mu\| = |\mu|(X)$.

THEOREM IV.2.6. *Let X be a LCH space. Then $C_0(X)^*$ is isometrically isomorphic to $M(X)$.*

PROOF. For any $\mu \in M(X)$,

$$\left|\int_X f d\mu\right| \leq \int_X |f| d|\mu \leq \|f\|_\infty \|\mu\|$$

showing that $I_\mu \in C_0(X)^*$. We have just proved that the map $\mu \to I_\mu$ is surjective with $\|I_\mu\| \leq \|\mu\|$. By the open mapping theorem, it is invertible with bounded inverse. We skip the argument showing that $\|\mu\| \leq \|I_\mu\|$. $\square$

If X is compact, then $C_0(X) = C(X)$ so that

COROLLARY IV.2.7. *Let X be a compact Hausdorff space. Then $C(X)^*$ is isometrically isomorphic to $M(X)$.*