KU LEUVEN

DEPARTMENT OF MATHEMATICS

Functional Analysis

Course G0B03A

Stefaan Vaes and Mateusz Wasilewski

Celestijnenlaan 200B – box 2400
B-3001 Heverlee, Belgium
stefaan.vaes@kuleuven.be and mateusz.wasilewski@kuleuven.be

Version : September 2019

Contents

General information and references

These lecture notes accompany the course *Functional Analysis*. The notes are based on the books [Con] and [Ped]. The notes do not provide a complete and finished text. There are a lot of exercises and parts of arguments that you have to fill in yourself. The main places where you have to work yourself are indicated by  . Often we add a reference to one of the text books [Con] and [Ped] where detailed proofs and a lot of extra material can be found.

Functional analysis is the analysis of infinite dimensional vector spaces, typically spaces of functions of all possible kinds, and of linear operators on them, for instance differential operators. Sometimes the results and topics of this functional analysis course might have a flavor of “abstract nonsense”. But these abstract results are used all the time in all possible mathematical theories. In this course itself we focus on one particular application: analysis on infinite discrete groups and the notion of amenability.

The course consists of 11 lectures of two hours. The lecture notes are structured accordingly. During the lectures you are often asked to fill in proofs yourself. It is therefore very useful to prepare the lecture by browsing the corresponding chapter and checking whether you master all prerequisites.

The lecture notes start with a 0’th lecture. This 0’th lecture contains material that you should know from earlier courses. This material will only be very quickly reminded during the first real lecture 1.

These lecture notes were seriously revised in September 2011 and in September 2019. There could be misprints or mistakes in these notes. We would be very glad if you could indicate them, during the course or by e-mail at stefaan.vaes@kuleuven.be and mateusz.wasilewski@kuleuven.be.

Bibliography

[Con] John Conway, A course in functional analysis. Second Edition. *Springer-Verlag,* Graduate Texts in Mathematics **96**, New York, 1990.

[Ped] Gert K. Pedersen, Analysis Now. *Springer-Verlag,* Graduate Texts in Mathematics **118**, New York, 1989.

Lecture 0

Banach spaces

This zero'th lecture only serves to remind everybody of known concepts concerning Banach spaces. We will very quickly go through it and immediately go on with the real first lecture.

0.1 Normed spaces, Banach spaces

I guess everybody is acquainted with the notion of a *norm on a vector space.*

Definition 0.1. Let X be a vector space over $\mathbb{C}$. A map

$$X \to [0, +\infty) : x \mapsto \|x\|$$

is called a *norm* on X if the following conditions hold.

- $\|x\| = 0$ if and only if $x = 0$,
- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$ (the *triangle inequality*),
- $\|\lambda x\| = |\lambda| \, \|x\|$ for all $x \in X$ and $\lambda \in \mathbb{C}$.

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

The concept of a normed space is extremely general, as shown by the following variety of examples.

Example 0.2. (i) The vector space $\mathbb{C}^n$ can be equipped with several different norms.

- $\|(z_1, \ldots, z_n)\|_\infty = \max\{|z_1|, \ldots, |z_n|\}$.
- For all $1 \leq p < \infty$, we have the norm

$$\|(z_1, \ldots, z_n)\|_p = (|z_1|^p + \cdots + |z_n|^p)^{1/p} .$$

In the case $p = 2$, we retrieve the usual Euclidean norm on $\mathbb{C}^n$. If you never did so, it is probably not so easy to prove the triangle inequality when $p > 1$. This can be done in a way analogous to Theorem 0.9 below.

(ii) Consider the vector space $C([0,1],\mathbb{C})$ of continuous functions from $[0,1]$ to $\mathbb{C}$. Because a continuous function on a compact set is automatically bounded, we define the following norm:
$$\|f\|_\infty = \sup\{|f(x)| \mid x \in [0,1]\} .$$
(iii) Set $X = \{z : \mathbb{N} \to \mathbb{C} \mid z(n) = 0 \text{ for } n \text{ large enough } \}$. This means that X consists of the sequences in $\mathbb{C}$ that are 0 from a certain point onwards. Define the norm
$$\|z\| = \sum_{n=0}^{\infty} |z(n)| .$$

Normed spaces fall apart in two different families. This is already clear in the previous examples. In the first two examples we have the impression to have defined a very natural normed space. In the last example, that is not the case: it would be much more natural to consider all functions $\mathbb{N} \to \mathbb{C}$ that are absolutely summable, i.e. for which $\sum_{n=0}^{\infty} |z(n)| < \infty$.

The precise difference between examples 1 and 2 on the one hand and example 3 on the other hand, lies in the notion of *completeness*. Before introducing this concept, we recall a few definitions.

Definition 0.3. Let $(X, \|\cdot\|)$ be a normed space.

- We say that a sequence (x_n) in X is *convergent*, if there exists $x \in X$ such that for all $\varepsilon > 0$, there exists n_0 satisfying $\|x - x_n\| < \varepsilon$ for all $n \geq n_0$.
- We say that a sequence (x_n) in X is a *Cauchy sequence*, if for all $\varepsilon > 0$, there exists n_0 satisfying $\|x_n - x_m\| < \varepsilon$ for all $n, m \geq n_0$.
- Look up, in earlier courses or in the literature, the concepts of *open* subsets, *closed* subsets and *dense* subsets of a normed space. In Lecture 7 we will recall basic notions of general topology. In the first lectures only the norm topology on a normed space will be used.

Exercise 1. Take $X = \mathbb{C}^n$ with any of the norms defined in Example 0.2.(i). Let (z_k) be a sequence in $\mathbb{C}^n$ and write $z_k = (z_{k1}, \ldots, z_{kn})$. Prove that (z_n) converges to $y = (y_1, \ldots, y_n) \in \mathbb{C}^n$ if and only if $(z_{ki})_k$ converges to y_i in $\mathbb{C}$ for all $i = 1, \ldots, n$.

Exercise 2. Consider $X = C([0,1],\mathbb{C})$ as in Example 0.2.(ii). Prove that the sequence f_n converges to f if and only if $f_n \to f$ uniformly on $[0,1]$.

Definition 0.4. A normed space X is called *complete* if every Cauchy sequence in X is convergent. A complete normed space is called a *Banach space.*

Proposition 0.5. *The normed space $X = C([0,1],\mathbb{C})$ with norm $\|\cdot\|$ defined in Example 0.2.(ii), is a Banach space.*

Proof. Prove yourself this proposition according to the following steps. See [Con, Example III.1.6] for details.

1. Realize that the only nontrivial part consists in proving the completeness of X.
2. Take a Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ in X. Prove that for every $x \in [0,1]$ the sequence $(f_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{C}$. By the completeness of $\mathbb{C}$, denote its limit by $f(x)$.

3. Prove that $f_n \to f$ uniformly on $[0, 1]$. Deduce that $f \in X$ and that $f_n \to f$ in the normed space X.

□

The following principle makes it easy to prove the noncompleteness of certain normed spaces. Roughly the criterion says that if a sequence (x_n) in a normed space X converges in some larger normed space Y to a limit x that lies outside X, then X is not complete.

Proposition 0.6. *Let Y be a normed space and $X \subset Y$ a vector subspace. Assume that (x_n) is a sequence in X such that in the normed space Y the sequence (x_n) converges to x with $x \notin X$. Then X is not complete.*

Proof. Assume that X is complete. We prove that $x \in X$. Since (x_n) is a convergent sequence in Y, it also is a Cauchy sequence. So it is a Cauchy sequence in X as well. Since X is complete, $x_n \to z$ for some $z \in X$. It follows that in Y the sequence (x_n) also converges to z. Since a convergent sequence can only have one limit, we conclude that $z = x$. Hence $x \in X$. □

Example 0.7. Consider the vector space $X = \{z : \mathbb{N} \to \mathbb{C} \mid z(n) = 0 \text{ for } n \text{ large enough } \}$ equipped with the norm

$$\|z\|_1 = \sum_{n=0}^{\infty} |z(n)| \; .$$

Then X is not complete. Indeed, we can view $X \subset \ell^1(\mathbb{N})$, where

$$\ell^1(\mathbb{N}) = \{z : \mathbb{N} \to \mathbb{C} \mid \sum_{n\in\mathbb{N}} |z(n)| < \infty\}$$

with the obvious norm $\|\cdot\|_1$. Define the element $x \in \ell^1(\mathbb{N})$ given by $x(n) = (1 + n^2)^{-1}$. Note that $x \notin X$. Define the sequence $(x_k)_{k\in\mathbb{N}}$ in X given by

$$x_k(n) = \begin{cases} x(n) & \text{if } \; 0 \leq n \leq k \; , \\ 0 & \text{if } \; n > k \; . \end{cases}$$

Prove that $x_k \to x$. By Proposition 0.6 it follows that X is not complete.

0.2 Further examples of Banach spaces

Whenever $p \geq 1$ and $x : \mathbb{N} \to \mathbb{C}$, we define

$$\|x\|_p := \Big(\sum_{n\in\mathbb{N}} |x(n)|^p\Big)^{1/p} \; .$$

We also define

$$\|x\|_\infty = \sup_{n\in\mathbb{N}} |x(n)| \; .$$

For general sequences $x : \mathbb{N} \to \mathbb{C}$, it is of course possible that $\|x\|_p = \infty$ or $\|x\|_\infty = \infty$.

Definition 0.8. Define for $p \geq 1$,

$$\ell^p(\mathbb{N}) = \{x : \mathbb{N} \to \mathbb{C} \mid \|x\|_p < +\infty\} ,$$
$$\ell^\infty(\mathbb{N}) = \{x : \mathbb{N} \to \mathbb{C} \mid \|x\|_\infty < \infty\} ,$$
$$c_0(\mathbb{N}) = \{x : \mathbb{N} \to \mathbb{C} \mid \lim_{n\to\infty} |x(n)| = 0\} .$$

Prove yourself that $\ell^1(\mathbb{N})$ equipped with $\| \cdot \|_1$ and $\ell^\infty(\mathbb{N})$ equipped with $\| \cdot \|_\infty$ are Banach spaces. Prove also that $c_0(\mathbb{N})$ is a closed vector subspace of $\ell^\infty(\mathbb{N})$ equipped with $\| \cdot \|_\infty$. It is less obvious to prove for $p > 1$ that $\ell^p(\mathbb{N})$ is a vector space and that $\| \cdot \|_p$ is a norm on it. In order to do so, you need the Hölder and Minkowski inequalities in the following theorem.

Theorem 0.9. *Let $x, y : \mathbb{N} \to \mathbb{C}$ and take $1 < p, q < \infty$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Then the following inequalities hold.*

$$\|xy\|_1 \leq \|x\|_p \, \|y\|_q \qquad \text{(Hölder inequality)}$$
$$\|x + y\|_p \leq \|x\|_p + \|y\|_p \qquad \text{(Minkowski inequality)}$$

Proof. Look this up in an earlier course or in the literature. □

Proposition 0.10. *For all $1 \leq p \leq \infty$ the vector space $\ell^p(\mathbb{N})$ equipped with $\| \cdot \|_p$ is a Banach space.*

Proof. It follows from the Minkowski inequality that $\ell^p(\mathbb{N})$ is a vector space and that $\| \cdot \|_p$ is a norm on this vector space. Choose a Cauchy sequence (x_k) in $\ell^p(\mathbb{N})$. As in the proof of Proposition 0.5, we find a function $x : \mathbb{N} \to \mathbb{C}$ such that $x_k \to x$ pointwise. We have to prove that $x \in \ell^p(\mathbb{N})$ and that $\|x - x_k\|_p \to 0$. Choose $\varepsilon > 0$. Take k_0 such that $\|x_r - x_k\|_p \leq \varepsilon$ for all $k, r \geq k_0$. We claim that $\|x - x_k\|_p \leq \varepsilon$ for all $k \geq k_0$. To prove this claim, fix a $k \geq k_0$. Fix $N \in \mathbb{N}$. For every $r \geq k_0$ we have

$$\sum_{n=0}^{N} |x_r(n) - x_k(n)|^p \leq \|x_r - x_k\|_p^p \leq \varepsilon^p .$$

Taking the limit for $r \to \infty$, it follows that

$$\sum_{n=0}^{N} |x(n) - x_k(n)|^p \leq \varepsilon^p .$$

Since this holds for all $N \in \mathbb{N}$, we also have $\|x - x_k\|_p \leq \varepsilon$, as in the claim. In particular $x \in \ell^p(\mathbb{N})$ and $\|x - x_k\|_p \leq \varepsilon$ for all $k \geq k_0$. This completes the proof of the proposition. □

0.3 Dual Banach space

Let X be a Banach space. We study linear maps $\omega : X \to \mathbb{C}$. Linear maps from X to $\mathbb{C}$ will be called *functionals* on X. As to be expected in a functional analysis course, we are not particularly interested in arbitrary functionals, but impose a continuity condition.

Proposition 0.11. *Let X be a Banach space and $\omega : X \to \mathbb{C}$ a linear map. The following conditions are equivalent.*

- *ω is continuous in* 0.
- *ω is continuous.*
- *There exists an $M > 0$ such that $|\omega(x)| \leq M\|x\|$ for all $x \in X$.*

Proof. Prove Proposition 0.11 yourself or see [Con, Proposition III.2.1]. □

Definition 0.12. Define

$$X^* = \{\omega \mid \omega : X \to \mathbb{C} \;\text{ is a continuous linear map }\} \; .$$

When $\omega \in X^*$, we set

$$\|\omega\| = \sup\{|\omega(x)| \mid x \in X, \|x\| \leq 1\} \; .$$

Note that Proposition 0.11 implies that $\|\omega\|$ is finite for all $\omega \in X^*$.

Exercise 3. Prove that $\|\omega\| \leq M$ if and only if $|\omega(x)| \leq M\|x\|$ for all $x \in X$.

It is clear that X^* is a vector space for the obvious operations

$$(\omega_1 + \omega_2)(x) = \omega_1(x) + \omega_2(x) \quad \text{and} \quad (\lambda\omega)(x) = \lambda\omega(x) \; .$$

Proposition 0.13. *Equipped with the norm $\omega \mapsto \|\omega\|$, the vector space X^* is a Banach space.*

Proof. Prove Proposition 0.13 yourself or see [Con, Proposition III.5.4]. □

Theorem 0.14. *Let $1 < p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. In the following precise sense, the dual of $\ell^p(\mathbb{N})$ is $\ell^q(\mathbb{N})$.*

- *For every $y \in \ell^q(\mathbb{N})$ the formula*

  $$\omega_y : \ell^p(\mathbb{N}) \to \mathbb{C} : \omega_y(x) = \sum_{n=0}^{\infty} y(n)\, x(n)$$

  yields a well defined element $\omega_y \in \ell^p(\mathbb{N})^*$.
- *We have $\|\omega_y\| = \|y\|_q$.*
- *Every $\omega \in \ell^p(\mathbb{N})^*$ is of the form $\omega = \omega_y$ for a unique $y \in \ell^q(\mathbb{N})$.*

In more advanced terminology, the map

$$\ell^q(\mathbb{N}) \to \ell^p(\mathbb{N})^* : y \mapsto \omega_y$$

is an isometric isomorphism.

Proof. Prove the theorem yourself according to the following steps.

1. Use the Hölder inequality to prove that for $y \in \ell^q(\mathbb{N})$ we have that ω_y is a well defined linear functional on $\ell^p(\mathbb{N})$ and that $\|\omega_y\| \leq \|y\|_q$.

2. Fix $y \in \ell^q(\mathbb{N})$ with $\|y\|_q = 1$ and prove that $\|\omega_y\| \geq 1$. To do so, write in a smart way an element $x \in \ell^p(\mathbb{N})$ (in terms of y) satisfying $\|x\|_p = 1$ and $\omega_y(x) = 1$. Together with the previous point we have $\|\omega_y\| = 1$ whenever $\|y\|_q = 1$. Deduce that $\|\omega_y\| = \|y\|_q$ for every $y \in \ell^q(\mathbb{N})$.

3. Finally choose an arbitrary $\omega \in \ell^p(\mathbb{N})^*$. Define the elements $\delta_n \in \ell^p(\mathbb{N})$ given by $\delta_n(m) = 1$ if $m = n$ and $\delta_n(m) = 0$ if $m \neq n$. Define $y : \mathbb{N} \to \mathbb{C}$ given by $y(n) = \omega(\delta_n)$. Use the following steps to prove that $y \in \ell^q(\mathbb{N})$ and that $\omega_y = \omega$.

   (a) Whenever $z : \mathbb{N} \to \mathbb{C}$ is a function and $N \in \mathbb{N}$, denote by $(z)_N$ the truncated function $(z)_N(k) = z(k)$ if $0 \leq k \leq N$ and $(z)_N(k) = 0$ if $k > N$. Observe that $(y)_N \in \ell^q(\mathbb{N})$. By point 2 we have $\|(y)_N\|_q = \|\omega_{(y)_N}\|$. Check that $\omega_{(y)_N}(x) = \omega((x)_N)$ for all $x \in \ell^p(\mathbb{N})$ and deduce that $\|\omega_{(y)_N}\| \leq \|\omega\|$ for all $N \in \mathbb{N}$. Finally deduce that $\|y\|_q \leq \|\omega\|$.

   (b) By construction ω and ω_y are continuous linear maps from $\ell^p(\mathbb{N})$ to $\mathbb{C}$ that coincide on the functions $x \in \ell^p(\mathbb{N})$ that are finitely supported. Prove that the finitely supported elements are dense in $\ell^p(\mathbb{N})$ and deduce that $\omega = \omega_y$.

□

Exercise 4. Write isometric isomorphisms $\ell^1(\mathbb{N}) \cong c_0(\mathbb{N})^*$ and $\ell^\infty(\mathbb{N}) \cong \ell^1(\mathbb{N})^*$.

0.4 The Banach spaces $\mathrm{L}^p(A, \lambda)$

We consider the Lebesgue measure λ on $\mathbb{R}^n$. For every measurable subset $A \subset \mathbb{R}^n$, we construct a Banach space $\mathrm{L}^p(A, \lambda)$. We do not provide proofs in this section and you are not supposed to be able to give them yourselves. Either you have seen proofs in a course on measure theory, either you just accept the statements at face value. Those who followed a course on abstract measure theory, can of course replace everywhere (A, λ) by an abstract measure space.

Definition 0.15. Let $A \subset \mathbb{R}^n$ be a Borel set and $f : A \to \mathbb{C}$ a Borel measurable function. For $p \geq 1$, we define

- $\|f\|_p = \left(\int_A |f|^p \, d\lambda \right)^{\frac{1}{p}}$.

- $\mathcal{L}^p(A, \lambda) = \{ f : A \to \mathbb{C} \mid f \text{ is Borel measurable and } \|f\|_p < \infty \}$.

Note that $\mathcal{L}^1(A, \lambda)$ consists exactly of the integrable functions from A to $\mathbb{C}$.

If $1 < p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, the following *Hölder inequality* holds for all Borel measurable functions $f, g : A \to \mathbb{C}$:

$$\|fg\|_1 \leq \|f\|_p \, \|g\|_q \, .$$

For all $p \geq 1$ and all Borel measurable functions $f, g : A \to \mathbb{C}$, the following *Minkowski inequality* holds

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

The Minkowski inequality implies that $\mathcal{L}^p(A, \lambda)$ is a vector space. We have the following vector subspace of $\mathcal{L}^p(A, \lambda)$:

$$\mathcal{L}_0(A, \lambda) := \{f : A \to \mathbb{C} \mid f \text{ is Borel measurable and } f(x) = 0 \text{ for almost all } x \in A\} .$$

Note that $\|f\|_p = 0$ if and only if $f \in \mathcal{L}_0(A, \lambda)$.

We define $\mathrm{L}^p(A, \lambda)$ as the quotient of $\mathcal{L}^p(A, \lambda)$ by the subspace $\mathcal{L}_0(A, \lambda)$. Intuitively, this means that we identify two functions in $\mathcal{L}^p(A, \lambda)$ once they are equal almost everywhere.

Because the integral of a function remains the same when we change the function on a set of measure zero, one can check that the map $f \mapsto \|f\|_p$ yields a well defined norm on $\mathrm{L}^p(A, \lambda)$. Here we already started to make a big abuse of notation: in principle, the elements of $\mathrm{L}^p(A, \lambda)$ are *equivalence classes of measurable functions,* but we nevertheless write elements of $\mathrm{L}^p(A, \lambda)$ as functions, always keeping in mind that we identify functions that are equal almost everywhere.

One can prove that $\mathrm{L}^p(A, \lambda)$ is a *Banach space.*

0.5 Completion of a normed space (optional)

We have seen above that a normed space is called *complete* if every Cauchy sequence has a limit. There is a procedure to *complete* an arbitrary normed space by adding in a certain sense all limits of Cauchy sequences. Look again at the normed space X in Example 0.2.(iii). Nobody will be surprised that $\ell^1(\mathbb{N})$ is the completion of X. The Banach space $\ell^1(\mathbb{N})$ satisfies the following abstract properties with respect to X.

- $\ell^1(\mathbb{N})$ is a Banach space.
- $X \hookrightarrow \ell^1(\mathbb{N})$ in a way preserving the norm.
- If we consider X as a vector subspace of $\ell^1(\mathbb{N})$, we have that X is dense in $\ell^1(\mathbb{N})$.

In Theorem 0.17 below, we prove that such a completion exists for any normed space. Moreover, the completion is essentially unique, but we have to give a careful formulation of this uniqueness. And if we want to understand uniqueness of the completion, we first have to say what it means to be 'the same normed space'.

Definition 0.16. Let X and Y be normed spaces and $\pi : X \to Y$ a linear map.

- We call π an *isometry* if $\|\pi(x)\| = \|x\|$ for all $x \in X$.
- We call π an *isometric isomorphism* if π is a bijective isometry.

In words, an isometric isomorphism is a map between normed spaces preserving all the available structure: a bijective linear map preserving the norm.

The formulation of the following theorem might scare some of you. It is one of the most abstract results in these notes. The proof of the theorem might remind some of you of one of the constructions of the field of real numbers out of the field of rational numbers. Indeed, one can view $\mathbb{R}$ as a completion of $\mathbb{Q}$.

Theorem 0.17. *Let X be a normed space. Then, there exist a Banach space Y and a linear map $\pi : X \to Y$ satisfying*

- *π is isometric;*
- *$\pi(X)$ is a dense subspace of Y.*

The pair (Y, π) is unique in the following sense: if (Y', π') satisfy the same properties, there exists a unique isometric isomorphism $\theta : Y \to Y'$ satisfying $\theta \circ \pi = \pi'$.

We view X as a subspace of Y through π and call Y the *completion* of X.

Proof. We leave the proof of the uniqueness statement as an exercise: one first defines θ_0 from $\pi(X)$ to Y' by the formula $\theta_0(\pi(x)) = \pi'(x)$ for all $x \in X$. One next takes the unique isometric map from Y to Y' extending θ_0.

We now prove the existence of Y and π. Define $\mathcal{Y}$ as the vector space of all Cauchy sequences in X. Addition in $\mathcal{Y}$ is defined as $(x_n)+(y_n) = (x_n+y_n)$ and we observe that the sum of two Cauchy sequences is again a Cauchy sequence. Scalar multiplication is defined analogously. Define the vector subspace $\mathcal{Y}_0$ of $\mathcal{Y}$ consisting of sequences (x_n) converging to 0. Define Y as the quotient vector space $Y := \frac{\mathcal{Y}}{\mathcal{Y}_0}$ and denote the quotient map by $q : \mathcal{Y} \to Y$.

We claim that the formula

$$\|y\| = \lim_{n\to\infty} \|x_n\| \quad \text{whenever } y = q((x_n)_{n\in\mathbb{N}}) \text{ and } (x_n) \in \mathcal{Y}$$

yields a well defined norm on Y. In order to prove this claim, you have to check the following facts.

1. The right hand side makes sense (i.e. is convergent) for all Cauchy sequences (x_n) in X.
2. The right hand side is independent of the representative $(x_n)_{n\in\mathbb{N}}$ that we have chosen for $y \in Y$. This means that you have to prove that

   $$\lim_{n\to\infty} \|x_n\| = \lim_{n\to\infty} \|y_n\|$$

   whenever (x_n) and (y_n) are Cauchy sequences and $q((x_n)_{n\in\mathbb{N}}) = q((y_n)_{n\in\mathbb{N}})$.
3. The map $y \mapsto \|y\|$ defines a norm on Y.

Define $\pi : X \to Y$ as follows: given $x \in X$, consider the constant sequence $x_n = x$ for all n and set $\pi(x) = q((x_n)_{n\in\mathbb{N}})$. Check that $\|\pi(x)\| = \|x\|$. Prove that $\pi(X)$ is dense in Y by proving the following statement: whenever $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X, the sequence $(\pi(x_n))_{n\in\mathbb{N}}$ converges to $q((x_n)_{n\in\mathbb{N}})$. Note that the previous statement is slightly subtle: first of all (x_n) is a Cauchy sequence in X and hence, $q((x_n)_{n\in\mathbb{N}})$ is an element of Y; on the other hand, $(\pi(x_n))_{n\in\mathbb{N}}$ is a sequence in Y.

It remains to prove that Y is complete. Let $(y_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in Y. Take for every n, an element $x_n \in X$ such that $\|y_n - \pi(x_n)\| < 1/n$. It follows that

$$\|x_n - x_m\| = \|\pi(x_n) - \pi(x_m)\| \le \|\pi(x_n) - y_n\| + \|y_n - y_m\| + \|y_m - \pi(x_m)\| \le \frac{1}{n} + \frac{1}{m} + \|y_n - y_m\| \, .$$

Deduce from this that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X. Define $y = q((x_n)_{n\in\mathbb{N}})$. We already proved that $\pi(x_n) \to y$ in Y. Since $\|y_n - \pi(x_n)\| \to 0$, it follows that $y_n \to y$. So, we have proved that (y_n) is convergent. □

0.6 Exercises

Exercise 5. Prove that a closed vector subspace of a Banach space is again a Banach space.

Exercise 6. Let X and Y be Banach spaces. Consider the vector space

$$X \oplus Y := \{(x, y) \mid x \in X, y \in Y\}$$

with the obvious componentwise vector space operations. Define the norms

$$\|(x, y)\|_{\max} := \max\{\|x\|, \|y\|\} \quad \text{and} \quad \|(x, y)\|_{\text{sum}} := \|x\| + \|y\| \, .$$

Prove that $\|\cdot\|_{\max}$ and $\|\cdot\|_{\text{sum}}$ are norms on $X \oplus Y$. We denote the corresponding normed spaces as $X \oplus_{\max} Y$ and $X \oplus_{\text{sum}} Y$.

Prove that $X \oplus_{\max} Y$ and $X \oplus_{\text{sum}} Y$ are Banach spaces. Write an isometric isomorphism[1]

$$(X \oplus_{\max} Y)^* \to X^* \oplus_{\text{sum}} Y^* \, .$$

[1] An isometric isomorphism is a linear bijection that preserves the norm.

Lecture 1

Hilbert spaces

1.1 Definition

We introduce the special class of Banach spaces called *Hilbert spaces.*

Definition 1.1. Let H, K be vector spaces over the field $\mathbb{C}$ of complex numbers. A map $H \times K \to \mathbb{C} : (x, y) \mapsto \langle x, y\rangle$ is called a *sesquilinear form* if $\langle \cdot, \cdot \rangle$ is linear in the first variable and anti-linear in the second variable:

- $\langle \lambda x + \mu y, z\rangle = \lambda\langle x, z\rangle + \mu\langle y, z\rangle$ for all $x, y \in H$, $z \in K$, $\lambda, \mu \in \mathbb{C}$;
- $\langle x, \lambda y + \mu z\rangle = \overline{\lambda}\langle x, y\rangle + \overline{\mu}\langle x, z\rangle$ for all $x \in H$, $y, z \in K$, $\lambda, \mu \in \mathbb{C}$.

Suppose now that $H = K$. A *Hermitian form* is a sesquilinear form that is *symmetric:* $\overline{\langle y, x\rangle} = \langle x, y\rangle$.

A Hermitian form is said to be *positive* if $\langle x, x\rangle \geq 0$ for all $x \in H$ and *positive-definite* if $\langle x, x\rangle > 0$ for all $x \in H$ with $x \neq 0$.

The following are standard examples of positive-definite Hermitian forms.

Example 1.2. (i) Take $H = \mathbb{C}^n$ and $\langle (x_1, \dots, x_n), (y_1, \dots, y_n)\rangle = x_1\overline{y_1} + \cdots + x_n\overline{y_n}$.

(ii) Consider the Banach space $\ell^2(\mathbb{N})$ as in Definition 0.8. The Hölder inequality allows to define

$$\langle x, y\rangle = \sum_{n=0}^{\infty} x(n)\overline{y(n)} \quad \text{for all} \quad x, y \in \ell^2(\mathbb{N}) .$$

(iii) On $C([0, 1])$, the vector space of continuous functions from $[0, 1]$ to $\mathbb{C}$, we can define

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

(iv) The previous example is not very natural. Its more natural version is given as follows. Let $A \subset \mathbb{R}^n$ be a Borel set and define on the vector space $\mathrm{L}^2(A, \lambda)$ the inner product

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

Check that $\langle \cdot, \cdot \rangle$ is well defined, keeping in mind that $\mathrm{L}^2(A, \lambda)$ is defined as a quotient of $\mathcal{L}^2(A, \lambda)$ by identifying functions equal almost everywhere.

Proposition 1.3. *Let $\langle \cdot, \cdot \rangle$ be a positive Hermitian form on H. Define $\|x\| = \sqrt{\langle x, x\rangle}$ for all $x \in H$. For all $x, y \in H$, the following holds.*

1. *Cauchy-Schwartz inequality:* $|\langle x, y\rangle| \leq \|x\| \, \|y\|$.
2. *Minkowski inequality:* $\|x + y\| \leq \|x\| + \|y\|$.
3. *Parallelogram law:* $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$.
4. *Polarization formula:* $\langle x, y\rangle = \dfrac{1}{4}\displaystyle\sum_{k=0}^{3} i^k \|x + i^k y\|^2$.

If $\langle \cdot, \cdot \rangle$ is a positive-definite Hermitian form, $\| \cdot \|$ defines a norm on H.

Proof. Prove Proposition 1.3 yourself or see [Con, I.1.4]. □

Definition 1.4. 123 We call H a *Hilbert space* if H is a complex vector space equipped with a positive-definite Hermitian form $\langle \cdot, \cdot \rangle$ in such a way that H is complete with respect to the norm defined by the Hermitian form.

Example 1.5. The positive-definite Hermitian forms in Example 1.2 turn $\mathbb{C}^n$, $\ell^2(\mathbb{N})$ and $\mathrm{L}^2(A, \lambda)$ into Hilbert spaces. But the positive-definite Hermitian form on $C([0, 1])$ in 1.2.(iii) does not define a complete norm on $C([0, 1])$ (check this by viewing $C([0, 1]) \subset \mathrm{L}^2([0, 1])$ and using Proposition 0.6!).

1.2 Orthogonal projections and Riesz theorem

In Definition 0.12 we introduced the dual Banach space X^* of a Banach space X. In Theorem 0.14, we have seen that the dual Banach space of $\ell^p(\mathbb{N})$ is isomorphic with $\ell^q(\mathbb{N})$ when $\frac{1}{p} + \frac{1}{q} = 1$. It follows in particular that the dual Banach space of $\ell^2(\mathbb{N})$ is isomorphic with $\ell^2(\mathbb{N})$ itself. The same phenomenon appears for arbitrary Hilbert spaces[1]. That is the content of the Riesz representation theorem that we prove below. We first need to study the notion of *orthogonality.*

We have seen that $\mathbb{C}^n$ is a Hilbert space. From geometry we know that every vector subspace $K \subset \mathbb{C}^n$ has an orthogonal complement $K^\perp$ and that every vector in $x \in \mathbb{C}^n$ has a unique decomposition of the form $x = y + z$ with $y \in K$ and $z \in K^\perp$. The same result holds for Hilbert spaces.

Definition 1.6. Let H be a Hilbert space and A a subset of H. We set

$$A^\perp = \{x \in H \mid \langle x, y\rangle = 0 \quad \text{for all} \quad y \in A\}\,.$$

We call $A^\perp$ the *orthogonal complement* of A. We say that the two vectors $x, y \in H$ are *orthogonal* if $\langle x, y\rangle = 0$. We denote this by $x \perp y$.

Check that orthogonal vectors x, y satisfy $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ (known as Pythagoras' theorem).

[1]Further on, it will become clear that an arbitrary Hilbert space is in fact not that arbitrary and very often isometrically isomorphic with $\ell^2(\mathbb{N})$.

Theorem 1.7. *Let H be a Hilbert space and $K \subset H$ a closed vector subspace of H. Every vector $x \in H$ has a unique decomposition of the form $x = y + z$ with $y \in K$ and $z \in K^\perp$.*

The proof of Theorem 1.7 is based on the following lemma.

Lemma 1.8. *Let H be a Hilbert space and $S \subset H$ a nonempty convex closed subset. Let $x \in H$. There exists a unique $y \in S$ satisfying*

$$\|x - y\| = \inf\{\|x - a\| \mid a \in S\} .$$

Proof. Prove the lemma yourself according to the following steps. Details can be found in [Con, Theorem I.2.5].

1. Fix $x \in H$ and put $d = \inf\{\|x - a\| \mid a \in S\}$. Take a sequence (a_n) in S such that $\|x - a_n\| \to d$. The main idea is to prove that the sequence (a_n) converges and that its limit is the required $y \in S$. Apply the parallelogram law to the vectors $(x - a_n)/2$ and $(x - a_m)/2$ and deduce that $\|(a_n - a_m)/2\|$ is small for n, m large. Conclude that (a_n) is a Cauchy sequence in H. Denote its limit by y. Observe that $y \in S$.

2. Prove that $\|x - y\| = d$. It remains to prove the uniqueness of such a $y \in S$. Assume that $y' \in S$ and $\|x - y'\| = d$. Apply the parallelogram law to the vectors $(x - y)/2$ and $(x - y')/2$. Conclude that $(y - y')/2 = 0$, i.e. $y = y'$.

□

Proof of Theorem 1.7. Prove Theorem 1.7 yourself using the following scheme. Details can be found in [Con, Theorem I.2.6].

1. Prove the uniqueness of the decomposition $x = y + z$.

2. Fix $x \in H$. Observe that K is a closed and convex subset of H. Apply Lemma 1.8 and denote by $y \in K$ the element that is closest to x. Put $z = x - y$. It remains to prove that $z \in K^\perp$. Fix $a \in K$. You have to prove that $z \perp a$. To prove this write explicitly that for all $t \in \mathbb{R}$ the distance from $y + ta$ to x is longer than the distance from y to x.

□

Corollary 1.9. *Let H be a Hilbert space and $K \subset H$ a vector subspace. Then $(K^\perp)^\perp$ equals the closure of K.*

Proof. See exercise 5. □

An immediate consequence of Theorem 1.7 is the Riesz representation theorem, for which a proof is given in [Con, I.3.4].

Theorem 1.10 (Riesz Representation theorem)**.** *Let H be a Hilbert space and $\omega : H \to \mathbb{C}$ a continuous linear map. There exists a unique vector $y \in H$ satisfying*

$$\omega(x) = \langle x, y\rangle \quad \text{for all} \quad x \in H .$$

The vector y satisfies $\|\omega\| = \|y\|$.

Definition 1.11. Let $K \subset H$ be a closed vector subspace of a Hilbert space. Theorem 1.7 implies that every vector $x \in H$ has a unique decomposition $x = y + z$ with $y \in K$ and $z \in K^{\perp}$. We call y the *orthogonal projection of* x *on* K and we write $y = P_K(x)$. The map

$$P_K : H \to K$$

is called the *orthogonal projection on* K.

Exercise 1. Use the uniqueness of the orthogonal decomposition $x = y + z$ to prove that P_K is a linear map.

Exercise 2. Let $x \in H$. Prove that $P_K(x)$ is the element of K that lies closest to x, i.e.

$$\|x - P_K(x)\| = \inf\{\|x - y\| \mid y \in K\} .$$

In words, we can say that $P_K(x)$ is the *best approximation* of x by an element from K.

1.3 Orthonormal families in Hilbert spaces

If K is a closed vector subspace of a Hilbert space H, we have seen that the orthogonal projection $P_K(x)$ of $x \in H$ onto K corresponds to the *best approximation* of x by an element from K. This has several quite concrete applications, but before being useful, one needs an *explicit formula to compute* $P_K(x)$. This will be given by using *orthonormal bases.* As a result, concrete orthonormal bases for a Hilbert space, will give concrete approximation procedures. Typical examples include approximations by *Fourier series* or the usage of *wavelet bases* used in JPEG compressions of images.

Definition 1.12. A family of vectors $(e_i)_{i \in I}$ in a Hilbert space H is called an *orthonormal family* if

- $\|e_i\| = 1$ for all $i \in I$,
- $\langle e_i, e_j \rangle = 0$ if $i \neq j$.

In words: an orthonormal family is a family of two by two orthogonal vectors having norm one.

Example 1.13. (i) The standard basis vectors $e_1, \dots, e_n$ in $\mathbb{C}^n$ form an orthonormal family for the standard inner product.

(ii) Similarly, we define unit vectors $e_n \in \ell^2(\mathbb{N})$ by $e_n(k) = 1$ if $k = n$ and $e_n(k) = 0$ if $k \neq n$. Check that $\{e_n \mid n \in \mathbb{N}\}$ is an orthonormal family in $\ell^2(\mathbb{N})$.

(iii) Consider the Hilbert space $\mathrm{L}^2([0, 2\pi], \lambda)$. Check that the vectors $(e_n)_{n \in \mathbb{Z}}$ defined by

$$e_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$$

form an orthonormal family.

Definition 1.14. An *orthonormal basis* of a Hilbert space H is a *maximal orthonormal family,* i.e. an orthonormal family $(e_i)_{i \in I}$ in H that cannot be enlarged: if $x \in H$ and $x \perp e_i$ for all $i \in I$, then $x = 0$.

Proposition 1.15. *An orthonormal family* $(e_i)_{i\in I}$ *in a Hilbert space* H *is maximal if and only if the linear span* $\text{span}\{e_i \mid i \in I\}$ *is a* dense *vector subspace of* H.

Note that by definition $\text{span}\{e_i \mid i \in I\}$ consists of all *finite* linear combinations of vectors e_i, $i \in I$. A priori infinite linear combinations do not make sense because they should be regarded as limits of series. Recall that a basis of a vector space X is a family of linearly independent vectors whose linear span *is the whole of* X. The linear span of an orthonormal basis of a Hilbert space *is not necessarily the whole of* X. So in general, an orthonormal basis of a Hilbert space H is not a basis of H as a vector space. We will see that the only case where it actually is a vector space basis, is the case where H is finite dimensional.

Proof of Proposition 1.15. Let $(e_i)_{i\in I}$ be an orthonormal family in a Hilbert space H. Denote $K_0 := \text{span}\{e_i \mid i \in I\}$ and denote by K the closure of K_0. Check yourself that $x \in K^\perp$ if and only if $x \perp e_i$ for all $i \in I$.

So if $(e_i)_{i\in I}$ is a maximal orthonormal family, we conclude that $K^\perp = \{0\}$ and hence, using Corollary 1.9, that $H = (\{0\})^\perp = (K^\perp)^\perp = K$.

Conversely, if $K = H$, it follows that $K^\perp = \{0\}$ and hence that $(e_i)_{i\in I}$ is a maximal orthonormal family. □

Example 1.16.
(i) In Example 1.13.(ii), we defined the orthonormal family $(e_n)_{n\in\mathbb{N}}$ in $\ell^2(\mathbb{N})$. Prove that this is an orthonormal basis.
(ii) The orthonormal family $(e_n)_{n\in\mathbb{Z}}$ in $\mathrm{L}^2([0, 2\pi], \lambda)$ that we defined in Example 1.13.(iii) is also an orthonormal basis. This is not entirely trivial to prove and we refer to a course in Fourier analysis. Indeed, if $\xi \in \mathrm{L}^2([0, 2\pi], \lambda)$, we have

$$\langle \xi, e_n \rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} \xi(x) e^{-inx} \, dx \, .$$

Up to a conventional normalization, this expression coincides with the n-th Fourier coefficient of ξ. The fact that $(e_n)_{n\in\mathbb{Z}}$ is an orthonormal basis follows from the fact that a function vanishes almost everywhere if all its Fourier coefficients are zero.

Our *main questions* now are the following.

- Does every Hilbert space admit an orthonormal basis?
- Given an orthonormal basis $(e_i)_{i\in I}$ for H, we know from Proposition 1.15 that every vector $x \in H$ lies in the closure of the linear span of the vectors $(e_i)_{i\in I}$. But, is there an *explicit way* to write x as a limit of finite linear combinations of the $(e_i)_{i\in I}$?

Both questions are answered in the following sections. We first need a set theoretic intermezzo.

1.4 Zorn's lemma

We all know how to find an orthonormal basis in a finite dimensional Hilbert space: use the Gram-Schmidt orthogonalization procedure. This is still possible in infinite dimensional Hilbert spaces, up to a certain extent (for *separable* Hilbert spaces, see below). Every Hilbert space that you encounter in real life is separable. So there is no real need to prove in general the existence of an orthonormal basis. It is however a good excuse to discuss Zorn's lemma that we will use often in the final lectures of this course.

The need for Zorn's lemma can be felt as follows. Assume that H is a Hilbert space and that you want to prove the existence of an orthonormal basis. Start with a norm one vector e_0. If the span of e_0 is the whole of H, you are done. Otherwise you can add a norm one vector e_1 that is orthogonal to e_0. Either H is the span of $\{e_0, e_1\}$ or you continue. If the process never stops, you end up with an orthonormal family $(e_n)_{n\in\mathbb{N}}$. Maybe this family is maximal and you are done. Otherwise you can add one more vector, and again a vector, and maybe again an infinite number of vectors. You might be done at that point, or still not... How long do you have to continue? The answer is provided by Zorn's lemma.

Terminology 1.17. Let $(I, \leq)$ be a partially ordered set.

- The subset $J \subset I$ is called a *totally ordered subset* (or *chain*) if for all $i, j \in J$ we have $i \leq j$ or $j \leq i$.
- If $J \subset I$ and $i \in I$, we say that i *is an upper bound of* J if $j \leq i$ for all $j \in J$. *Note that we do not require that* $i \in J$.
- We call $i \in I$ a *maximal element of* I if the following condition holds: if $i \leq j$ and $j \in I$, then $i = j$. *Note that if* i *is a maximal element* I*, this does not mean that* $j \leq i$ *for all* $j \in I$.

Theorem 1.18 (Zorn's Lemma)**.** *If* $(I, \leq)$ *is a partially ordered set such that every totally ordered subset of* I *admits an upper bound, then* I *admits a maximal element.*

It is probably not a good idea to give Zorn's Lemma the status of a 'theorem' or a 'lemma'. Indeed, it is rather an axiom of set theory. One can show that Zorn's Lemma is equivalent with the Axiom of Choice which says the following: if $(X_i)_{i\in I}$ is a family of nonempty subsets of X, there exists a function $f : I \to X$ such that $f(i) \in X_i$ for all $i \in I$. In words: when we have a family of nonempty sets, we can choose one element in every set.

The Axiom of Choice has some intuitive evidence. This is already less the case for Zorn's Lemma, the idea being the following: take an element $i_0 \in I$. If i_0 is a maximal element, we are done. If i_0 is not a maximal element, we find $i_1 \geq i_0$ with $i_1 \neq i_0$. We continue like this and find a strictly increasing sequence $i_0 < i_1 < \cdots$. This strictly increasing sequence in I is a totally ordered subset. So, it admits an upper bound j_0. We continue the same game with j_0, and so on, and so on. The trouble is that this seemingly inductive type argument, never comes to an end. And indeed, the Axiom of Choice is also equivalent with the Principle of Transfinite Induction, a kind of induction beyond countability.

So, is the Axiom of Choice *true*? We know that the Axiom of Choice cannot be proven, neither disproved from the other Zermelo-Fraenkel axioms of set theory. On the other hand, the Axiom of Choice does have some consequences that are hard to believe, the most well known being the Banach-Tarski paradox: it is possible to partition the unit ball of $\mathbb{R}^3$ into a finite number of subsets and then to reassemble this finite number of pieces by only using rotations and translations, in such a way that we end up with ... *two* balls of radius 1.

1.5 Existence and construction of orthonormal bases

Theorem 1.19. *Every Hilbert space H admits an orthonormal basis.*

Proof. Define I to be the set of orthonormal families $\mathcal{F} \subset H$. Define the partial order $\leq$ on I given by $\mathcal{F} \leq \mathcal{F}'$ if $\mathcal{F} \subset \mathcal{F}'$. Prove yourself that $(I, \leq)$ satisfies the conditions of Zorn's lemma. So $(I, \leq)$ admits a maximal element. Prove that this maximal element is an orthonormal basis of H. □

Hilbert spaces that one encounters in real life are *separable,* meaning that they are not too big in the following precise sense.

Definition 1.20. A metric space (X, d) is called separable if it admits a countable dense subset.

Proposition 1.21. *A Hilbert space H is separable if and only if it admits an orthonormal basis with at most countably many vectors, i.e. of the form $\{e_1, \dots, e_n\}$ or of the form $(e_n)_{n\in\mathbb{N}}$.*

Proof. Prove the proposition yourself according to the following steps. If H is finite dimensional or admits an orthonormal basis $(e_n)_{n\in\mathbb{N}}$, use finite linear combinations of the basis vectors with coefficients in $\mathbb{Q} + i\mathbb{Q}$ to conclude that H is separable.

Conversely assume that H is separable and infinite dimensional. Choose a countable dense subset of H. Remove redundancies and find a sequence of linearly independent vectors $(x_n)_{n\in\mathbb{N}}$ in H such that $\operatorname{span}\{x_n \mid n \in \mathbb{N}\}$ is dense in H. Apply the Gram Schmidt orthonormalization procedure to the vectors $(x_n)_{n\in\mathbb{N}}$ and prove that the resulting orthonormal family $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis. □

1.6 Parseval and Plancherel formulae

Exercise 3. Let $e_1, \dots, e_n$ be an orthonormal family in H and denote $K = \operatorname{span}\{e_1, \dots, e_n\}$. Prove the following statements.

1. $\displaystyle \Big\| \sum_{k=1}^{n} \lambda_k e_k \Big\|^2 = \sum_{k=1}^{n} |\lambda_k|^2 \quad \text{for all} \quad \lambda_1, \dots, \lambda_n \in \mathbb{C}\ .$

2. The vectors $e_1, \dots, e_n$ are linearly independent.

3. For all $x \in H$, we have $\displaystyle P_K(x) = \sum_{k=1}^{n} \langle x, e_k \rangle e_k\ .$

Proposition 1.22. *Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in H. Define K as the closure of $\operatorname{span}\{e_n \mid n \in \mathbb{N}\}$. For all $x \in H$, the sequence $\displaystyle\sum_{k=0}^{n} \langle x, e_k \rangle e_k$ is convergent and we have*

$$P_K(x) = \sum_{k=0}^{\infty} \langle x, e_k \rangle e_k \quad \textit{and} \quad \| P_K(x) \|^2 = \sum_{k=0}^{\infty} |\langle x, e_k \rangle|^2\ .$$

In particular, we have Bessel's inequality

$$\sum_{k=0}^{\infty} |\langle x, e_k \rangle|^2 \leq \|x\|^2 \,.$$

Proof. Define $K_n = \operatorname{span}\{e_0, \ldots, e_n\}$. Because of Exercise 3, we only have to prove that the sequence $P_{K_n}(x)$ is convergent, with limit $P_K(x)$. We first prove that $(P_{K_n}(x))_{n\in\mathbb{N}}$ is a Cauchy sequence. Since

$$\sum_{k=0}^{n} |\langle x, e_k \rangle|^2 = \|P_{K_n}(x)\|^2 \leq \|x\|^2 \,,$$

it follows that $\sum_{k=0}^{\infty} |\langle x, e_k \rangle|^2 < \infty$. Moreover, for $m > n$, we have

$$\|P_{K_m}(x) - P_{K_n}(x)\|^2 = \sum_{k=n+1}^{m} |\langle x, e_k \rangle|^2 \,,$$

implying that $(P_{K_n}(x))_{n\in\mathbb{N}}$ is a Cauchy sequence.

Set $y = \lim_n P_{K_n}(x)$. By definition $y \in K$. It remains to prove that $x - y \in K^{\perp}$ or, equivalently, $\langle x - y, e_k \rangle = 0$ for all $k \in \mathbb{N}$. But for all $n \geq k$, we have $\langle x - P_{K_n}(x), e_k \rangle = 0$. Letting $n \to \infty$, we are done. □

Proposition 1.22 has the following immediate consequence.

Proposition 1.23. *Let H be a Hilbert space with orthonormal basis $(e_n)_{n\in\mathbb{N}}$. For all $x \in H$, we have*

$$x = \sum_{n=0}^{\infty} \langle x, e_n \rangle e_n \qquad \text{(Plancherel formula)}$$

$$\|x\|^2 = \sum_{n=0}^{\infty} |\langle x, e_n \rangle|^2 \qquad \text{(Parseval equality)}$$

1.7 Exercises

Exercise 4. Suppose that $(X, \|\cdot\|)$ is a normed space satisfying the parallelogram law. Follow the steps below and prove that the polarization formula defines a positive-definite Hermitian form on X satisfying $\|x\| = \sqrt{\langle x, x \rangle}$.

1. Prove that $\langle y, x \rangle = \overline{\langle x, y \rangle}$ and that $\langle ix, y \rangle = i\langle x, y \rangle$.
2. Prove that $\langle x, 2y \rangle = 2\langle x, y \rangle$.
3. Prove that $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$.
4. Prove that $\langle qx, y \rangle = q\langle x, y \rangle$ first when $q \in \mathbb{N}$ and next when $q \in \mathbb{Q}$.

5. Take limits and prove that $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for all $\lambda \in \mathbb{R}$. Deduce that the same holds for $\lambda \in \mathbb{C}$.

Exercise 5. Prove Corollary 1.9 yourself using the following scheme.

1. First prove Corollary 1.9 in the case where $K \subset H$ is a *closed* vector subspace. Realize that $K \subset (K^{\perp})^{\perp}$ is trivial. Conversely, take $x \in (K^{\perp})^{\perp}$. Use Theorem 1.7 to decompose $x = y + z$ with $y \in K$ and $z \in K^{\perp}$. Use that $x \in (K^{\perp})^{\perp}$ to conclude that $z = 0$.

2. Denote by $\bar{K}$ the closure of K. Prove that $K^{\perp} = (\bar{K})^{\perp}$.

3. Use the previous two steps to conclude.

Exercise 6. Use Zorn's lemma to prove that every vector space admits a basis (a maximal linearly independent subset).

Exercise 7.

Definition 1.24. Let $(x_i)_{i\in I}$ be a family of elements of a Banach space X. We call $(x_i)_{i\in I}$ *unconditionally summable* if there exists an $x \in X$ satisfying the following condition: for all $\varepsilon > 0$, there exists a finite subset $I_0 \subset I$ such that

$$\Big\| x - \sum_{i\in I_1} x_i \Big\| < \varepsilon \quad \text{for all finite subset } I_1 \subset I \text{ with } I_0 \subset I_1 .$$

We write $x = \sum_{i\in I} x_i$.

Those who are familiar with the notion of a *net* (generalized sequence, see Section 7.4 in Lecture 7), will recognize that unconditional summability of $(x_i)_{i\in I}$ is the same as convergence of the net $\big(\sum_{i\in I_0} x_i\big)_{I_0 \subset I}$ indexed by the finite subsets $I_0 \subset I$.

Let $(e_i)_{i\in I}$ be an orthonormal family in H. Define K as the closure of $\operatorname{span}\{e_i \mid i \in I\}$. Prove that for all $x \in H$, the family $(\langle x, e_i\rangle e_i)_{i\in I}$ is unconditionally summable and

$$P_K(x) = \sum_{i\in I} \langle x, e_i\rangle e_i \ .$$

Lecture 2

Bounded operators on a Hilbert space

Hilbert spaces as such are not really interesting mathematical objects, but linear maps between Hilbert spaces definitely are. A linear map from H to H is called an *operator on H* and such operators appear in different contexts.

- In a mathematical approach to quantum mechanics, operators play the role of *observables.*
- In the study of differential equations, one encounters *differential operators.* We will not meet them in this functional analysis course, because they are unbounded (this notion is introduced below). Nevertheless, the theory of differential operators is a major application of functional analysis.
- When studying groups (and more specifically Lie groups), one is interested in *group representations* by means of Hilbert space operators.
- The bounded operators on a Hilbert space form an *algebra.* This algebra has many interesting subalgebras, leading to the study of operator algebras: *C^*-algebras and von Neumann algebras.* This branch of mathematics has many relations with group theory, but also group actions, mathematical physics, etc. We refer to the course *Spectral Theory and Operator Algebras* for details.

2.1 Definition and first examples

Definition 2.1. Let H and K be Hilbert spaces. An *operator* T from H to K is a linear map $T : H \to K$.

Example 2.2. (i) Let $H = \ell^2(\mathbb{N})$. Define the operator

$$T : H \to H : (Tx)(n) = x(n+1) \quad \text{for all } \; x \in \ell^2(\mathbb{N}), n \in \mathbb{N} \; .$$

Check that $T(x)$ belongs to $\ell^2(\mathbb{N})$ whenever $x \in \ell^2(\mathbb{N})$.

(ii) Let $H = \mathrm{L}^2([0,1],\lambda)$. Because a square integrable function on $[0,1]$ is automatically integrable, we can define the following *integral operator,* which is sometimes called the *Volterra operator.*

$$T : H \to H : (Tf)(x) = \int_0^x f(y)\, dy \quad \text{for all} \quad f \in \mathrm{L}^2([0,1],\lambda), x \in [0,1] \; .$$

Prove that Tf belongs to $\mathrm{L}^2([0,1])$ by proving that for all $f \in \mathrm{L}^2([0,1],\lambda)$, Tf is in fact a continuous function on $[0,1]$.

(iii) Fix $\lambda \in \ell^\infty(\mathbb{N})$ and set $H = \ell^2(\mathbb{N})$. Define the *multiplication operator*

$$M_\lambda : H \to H : (M_\lambda x)(n) = \lambda(n)x(n) \quad \text{for all} \quad x \in \ell^2(\mathbb{N}), n \in \mathbb{N} \; .$$

Exactly as in the study of functionals on a Banach space, we are not interested[1] in arbitrary operators on a Hilbert space, but we impose a continuity condition.

Proposition 2.3. *Let H and K be Hilbert spaces and $T : H \to K$ an operator. Then the following conditions are equivalent.*

1. *T is continuous.*
2. *T is continuous in* 0.
3. *T is bounded, i.e. there exists $M \geq 0$ such that $\|Tx\| \leq M\|x\|$ for all $x \in H$.*

An operator satisfying one of these equivalent conditions is called a bounded operator.

Exercise 1. Prove Proposition 2.3. Prove that all operators in Example 2.2 are bounded. Prove that every linear map $\mathbb{C}^n \to \mathbb{C}^m$ is bounded.

Definition 2.4. The set of bounded operators from H to K is denoted as $\mathrm{B}(H,K)$. We also use the short-hand notation $\mathrm{B}(H) = \mathrm{B}(H,H)$.

The norm of $T \in \mathrm{B}(H,K)$ is defined as the smallest positive real number $M \geq 0$ satisfying $\|Tx\| \leq M\|x\|$ for all $x \in H$, i.e.

$$\|T\| = \sup\{\|Tx\| \mid x \in H, \|x\| \leq 1\} \; .$$

Exercise 2. Prove that $\mathrm{B}(H,K)$ equipped with the norm of Definition 2.4 is a Banach space. Also prove that the composition ST of $T \in \mathrm{B}(H_1,H_2)$ and $S \in \mathrm{B}(H_2,H_3)$ belongs to $\mathrm{B}(H_1,H_3)$ and satisfies

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

Remark 2.5. It follows that $\mathrm{B}(H)$ is a vector space and that composition of operators defines a product on $\mathrm{B}(H)$, turning $\mathrm{B}(H)$ into an *algebra.* The precise definition of an algebra A over $\mathbb{C}$ goes as follows: A is at the same time a ring and a vector space over $\mathbb{C}$ such that the multiplication map $A \times A \to A : (a,b) \mapsto ab$ is bilinear.

What *we have to keep in mind,* is that we can make little computations with the elements of $\mathrm{B}(H)$. If for instance $S,T \in \mathrm{B}(H)$, the expression $S^3T + \pi S - iT^2$ defines again an operator in $\mathrm{B}(H)$. Moreover, the algebra $\mathrm{B}(H)$ has a *unit element*, given by the operator

$$1 : H \to H : 1x = x \quad \text{for all} \quad x \in H \; .$$

[1] It should be stressed that specific types of noncontinuous operators (so called unbounded operators) are of crucial importance, but are not treated in this functional analysis course. Examples include the differential operators.

2.2 Sesquilinear forms, bounded operators, adjoints

In 1.1, we defined the notion of a *sesquilinear form.* A sesquilinear form $s : H \times K \to \mathbb{C}$ is called *bounded* if there exists $M \geq 0$ such that

$$|s(x, y)| \leq M \, \|x\| \, \|y\| \; .$$

The smallest M satisfying the above inequality is called the *norm of* s and denoted by $\|s\|$. So,

$$\|s\| = \sup\{|s(x, y)| \mid x \in H, y \in K, \|x\| \leq 1, \|y\| \leq 1\} \; .$$

The importance of bounded sesquilinear forms lies in the following lemma.

Lemma 2.6. *Let H, K be Hilbert spaces. Define for every bounded operator $T : H \to K$, the sesquilinear form*

$$s_T : H \times K \to \mathbb{C} : s_T(x, y) = \langle Tx, y\rangle \; .$$

The map $T \mapsto s_T$ is a bijection between $\mathrm{B}(H, K)$ *and the bounded sesquilinear forms $H \times K \to \mathbb{C}$, satisfying $\|s_T\| = \|T\|$.*

Proof. It is easy to check that for every $T \in \mathrm{B}(H, K)$, the sesquilinear form s_T is bounded with $\|s_T\| \leq \|T\|$. Observing that for all $z \in K$ we have

$$\|z\| = \sup\{|\langle z, y\rangle| \mid y \in K, \|y\| \leq 1\} \; ,$$

it follows that

$$\|s_T\| = \sup\{\|T(x)\| \mid x \in H, \|x\| \leq 1\} = \|T\| \; .$$

Conversely assume that $s : H \times K \to \mathbb{C}$ is a bounded sesquilinear form. Fix $x \in H$. One checks that the formula $K \to \mathbb{C} : y \mapsto \overline{s(x, y)}$ is a bounded linear functional on K. By the Riesz representation theorem 1.10 there is a unique vector in K, that we denote by $T(x)$, such that

$$\overline{s(x, y)} = \langle y, T(x)\rangle$$

for all $y \in K$. The uniqueness of the vector $T(x)$ implies that $x \mapsto T(x)$ is a linear map from H to K. The boundedness of s implies that T is bounded. By construction $s = s_T$. □

The relation between sesquilinear forms and operators allows to define the *adjoint* of a bounded operator. It is a generalization of the Hermitian adjoint of a matrix.

Theorem 2.7. *Let H, K be Hilbert spaces and $T : H \to K$ a bounded operator. There exists a unique bounded operator $T^* : K \to H$ satisfying*

$$\langle T^* y, x\rangle = \langle y, Tx\rangle \quad \text{for all } \; x \in H, y \in K \; .$$

We call T^* the adjoint *of T. We have the following.*

1. $(\lambda T + \mu S)^* = \overline{\lambda} T^* + \overline{\mu} S^*$ *for all $S, T \in \mathrm{B}(H, K)$ and $\lambda, \mu \in \mathbb{C}$.*
2. $(T^*)^* = T$ *for all $T \in \mathrm{B}(H, K)$.*

3. $(ST)^* = T^*S^*$ *for all* $T \in \mathrm{B}(H_1, H_2)$ *and* $S \in \mathrm{B}(H_2, H_3)$.
4. $\|T^*\| = \|T\|$ *for all* $T \in \mathrm{B}(H, K)$.

Proof. Let $T \in \mathrm{B}(H, K)$. Check that

$$s : K \times H \to \mathbb{C} : s(y, x) = \langle y, Tx \rangle$$

defines a bounded sesquilinear form. Lemma 2.6 yields a unique operator from K to H that we denote by T^* and that satisfies

$$s(y, x) = \langle T^* y, x \rangle \quad \text{for all } \; x \in H, y \in K \; .$$

This proves the existence of T^*. Uniqueness of T^* is obvious and the proof of the remaining statements is left as an exercise. □

Exercise 3. Let $A \in \mathrm{M}_{n,m}(\mathbb{C})$ be an $n \times m$ matrix defining the linear map

$$T : \mathbb{C}^m \to \mathbb{C}^n : (Tx)_i = \sum_{j=1}^{m} A_{ij} x_j \; .$$

Prove that the matrix of the adjoint T^* corresponds to the Hermitian adjoint of A.

Proposition 2.8. *For all* $T \in \mathrm{B}(H, K)$, *we have*

- $\|T^*T\| = \|T\|^2$.
- $\operatorname{Ker} T = (\operatorname{Im} T^*)^{\perp}$.

Proof. First observe that $\|T^*T\| \leq \|T^*\| \, \|T\| = \|T\|^2$. On the other hand, for every $x \in H$ we have

$$\|T(x)\|^2 = \langle T(x), T(x) \rangle = \langle T^*Tx, x \rangle \leq \|T^*Tx\| \, \|x\| \leq \|T^*T\| \, \|x\|^2 \; .$$

It follows that $\|T\|^2 \leq \|T^*T\|$.

Finally we have $x \in \operatorname{Ker} T$ iff $T(x) = 0$ iff $\langle T(x), y \rangle = 0$ for all $y \in K$ iff $\langle x, T^*(y) \rangle = 0$ for all $y \in K$ iff $x \perp T^*(K)$. □

2.3 Integral operators

The following proposition provides us with a large family of bounded operators, the so called *integral operators*.

Proposition 2.9. *Let* $A \subset \mathbb{R}^n$ *be a Borel set and* $K : A \times A \to \mathbb{C}$ *a measurable function. Suppose that one of the following conditions hold.*

(a) K *is square integrable.*

(b) There exist $c_1, c_2 \geq 0$ such that

$$\int_A |K(x,y)|\, dy \leq c_1 \quad \textit{for almost all} \;\; x \in A \;,$$
$$\int_A |K(x,y)|\, dx \leq c_2 \quad \textit{for almost all} \;\; y \in A \;.$$

Then, the formula

$$(Tf)(x) = \int_A K(x,y) f(y)\, dy \quad \textit{for all} \;\; f \in L^2(A,\lambda)$$

defines a bounded operator T on $L^2(A,\lambda)$. Under condition (a), we have $\|T\| \leq \|K\|_2$, while under condition (b), we have $\|T\| \leq \sqrt{c_1 c_2}$.

We call T the integral operator with kernel K. *The adjoint T^* is again an integral operator with kernel $K^*(x,y) = \overline{K(y,x)}$.*

Proof. We first study case (a). So, let $K \in \mathrm{L}^2(A \times A, \lambda)$. The Hölder inequality implies that

$$|(Tf)(x)|^2 \leq \left(\int_A |K(x,y)|^2\, dy\right)\left(\int_A |f(y)|^2\, dy\right) .$$

It follows that $\|Tf\|_2 \leq \|K\|_2 \, \|f\|_2$. So, T is a bounded operator and $\|T\| \leq \|K\|_2$. The formula for T^* is left as an exercise.

For case (b), see [Con, II.1.6]. The estimates to be made are more subtle. □

2.4 Orthogonal projections

In Definition 1.11, we have already seen that every closed vector subspace K of a Hilbert space H admits an *orthogonal projection* $P_K : H \to K$. For every $x \in H$, the vector $P_K(x) \in K$ is the best approximation of x by a vector in K. We already observed that $\|P_K(x)\| \leq \|x\|$ and it follows that P_K is a bounded operator.

The following proposition gives an abstract characterization of orthogonal projections.

Proposition 2.10. *A bounded operator P on a Hilbert space H is an orthogonal projection onto a closed vector subspace of H if and only if $P = P^*$ and $P^2 = P$.*

Proof. Suppose first that $P = P_K$ for some closed vector subspace $K \subset H$. It is clear that $P^2 = P$. Take $x, y \in H$ and write $x = x_1 + x_2$, $y = y_1 + y_2$ with $x_1, y_1 \in K$ and $x_2, y_2 \in K^\perp$. It follows that $Px = x_1$ and $Py = y_1$. So,

$$\langle Px, y \rangle = \langle x_1, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle = \langle x_1 + x_2, y_1 \rangle = \langle x, Py \rangle \; .$$

Hence, $P = P^*$.

Suppose conversely that $P = P^*$ and $P^2 = P$. Define $K = \operatorname{Im} P$ and observe that $K = \{x \in H \mid Px = x\}$, proving that K is closed. We claim that $P = P_K$. Since $x = Px + (x - Px)$, it is sufficient to prove that $x - Px \in K^\perp$ for all $x \in H$. But this follows because

$$\langle Py, x - Px \rangle = \langle y, P(x - Px) \rangle = 0 \quad \text{for all} \;\; x, y \in H \; .$$

□

2.5 Unitary operators

Definition 2.11. A bounded operator $T : H \to K$ is called *unitary* if T is a bijection satisfying $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in H$.

In other words, a unitary operator is "an isomorphism in the category of Hilbert spaces", which is just an expensive way of saying that a unitary operator between two Hilbert spaces is nothing else than a bijection that preserves all structure (vector space, sesquilinear form, norm).

Example 2.12. (i) The operator

$$U : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z}) : (Ux)(n) = x(n-1) \quad \text{for all } \ x \in \ell^2(\mathbb{Z}), n \in \mathbb{Z}$$

is unitary.

(ii) The operator

$$V : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}) : (Vx)(n) = \begin{cases} x(n-1) & \text{if } \ n \geq 1\,, \\ 0 & \text{if } \ n = 0\,. \end{cases}$$

is isometric, but not unitary.

(iii) The Fourier transform

$$\mathcal{F} : \mathrm{L}^2([0, 2\pi], \lambda) \to \ell^2(\mathbb{Z}) : (\mathcal{F} f)(n) = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-inx}\, dx$$

is unitary.

Exercise 4. Prove the statements in Examples 2.12.(i) and 2.12.(ii). The remaining Example 2.12.(iii) is usually proved in a course on Fourier analysis. If you admit Example 1.13.(iii), you can prove it yourself though.

2.6 Self-adjoint operators

Definition 2.13. 230 A bounded operator $T \in \mathrm{B}(H)$ is called *self-adjoint* if $T = T^*$.

If $H = \mathbb{C}^n$ with the standard scalar product, bounded operators on H correspond to matrices. Self-adjoint operators correspond to so called *Hermitian matrices,* i.e. matrices A that are equal to the complex conjugate of their transpose.

2.7 Exercises

Exercise 5. Let P, Q be orthogonal projections on closed vector subspaces of H. Prove that the following statements are equivalent.

1. The operator $P + Q$ is an orthogonal projection.

2. $PQ = 0$.

3. The subspaces Im P and Im Q are orthogonal.

Exercise 6. Let $T : H \to K$ be a bijective linear map that preserves the norm: $\|T(x)\| = \|x\|$ for all $x \in H$. Use the polarization formula to prove that T is unitary.

Exercise 7. Let $T : H \to K$ be a bounded operator and $(e_n)_{n\in\mathbb{N}}$ an orthonormal basis for H.

1. Prove that T is an isometry if and only if $T^*T = 1$,

2. Prove that T is unitary if and only if $T^*T = 1$ and $TT^* = 1$.

3. Prove that T is unitary if and only if $(Te_n)_{n\in\mathbb{N}}$ is an orthonormal basis of K.

4. Let $(f_n)_{n\in\mathbb{N}}$ be an orthonormal basis for K. Prove that there exists a unique unitary operator $U : H \to K$ satisfying $Ue_n = f_n$ for all $n \in \mathbb{N}$.

Exercise 8. Prove that the operator T in Example 2.2.(i) has adjoint

$$T^* : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}) : (T^*y)(n) = \begin{cases} y(n-1) & \text{if } n \geq 1 \,, \\ 0 & \text{if } n = 0 \,. \end{cases}$$

Exercise 9. Prove that the Volterra operator T defined in 2.2.(ii) has adjoint given by

$$(T^*f)(x) = \int_x^1 f(y)\, dy \,.$$

Exercise 10. Prove that the multiplication operator M_λ defined in 2.2.(iii) has adjoint $M_{\overline{\lambda}}$ where $\overline{\lambda}(n) = \overline{\lambda(n)}$.

Lecture 3

Compact operators

Before starting the actual contents of this lecture we remind a number of properties about compact metric spaces.

3.1 Compact metric spaces

By definition (see 7.9 below) a metric space (X, d) is called *compact* if every covering of X by a family of open subsets admits a finite subcovering, i.e. a finite subfamily that still covers the whole of X. This definition is not very intuitive, but makes sense for arbitrary topological spaces. For metric spaces, several more natural conditions are equivalent with compactness.

Theorem 3.1. *Let (X, d) be a metric space and $K \subset X$. Then, the following statements are equivalent.*

(i) Every sequence in K has a subsequence converging to an element of K.

(ii) Every infinite subset of K has an accumulation point in K.

(iii) The metric space (K, d) is complete and for every $\varepsilon > 0$, there exists a finite subset $I \subset X$ such that

$$K \subset \bigcup_{x \in I} B(x, \varepsilon) .$$

(iv) K is compact.

Proof. Prove as an exercise that 1 and 2 are equivalent.

(i) $\Longrightarrow$ (iii). Suppose that (i) holds. We first prove that (K, d) is complete. Let (x_n) be a Cauchy sequence in K. Let (x_{n_k}) be a convergent subsequence with limit $x \in K$. Choose $\varepsilon > 0$. Take n_0 such that $d(x_n, x_m) < \varepsilon/2$ for all $n, m \geq n_0$. Take k such that $n_k \geq n_0$ and $d(x, x_{n_k}) < \varepsilon/2$. It follows that for all $n \geq n_0$ we have

$$d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon .$$

So $x_n \to x$ and (K, d) is complete. Next choose $\varepsilon > 0$. Suppose that there is no finite subset $I \subset X$ such that $K \subset \bigcup_{x \in I} B(x, \varepsilon)$. We can then choose inductively a sequence (x_n) in K satisfying

$$x_{n+1} \in K \setminus \Big(\bigcup_{i=0}^{n} B(x_i, \varepsilon) \Big) .$$

Let (x_{n_k}) be a subsequence of (x_n) converging to $x \in K$. Take k_0 such that $d(x, x_{n_k}) < \varepsilon/2$ for all $k \geq k_0$. It follows in particular that $d(x_{n_{k_0+1}}, x_{n_{k_0}}) < \varepsilon$. This is in contradiction with the construction of $x_{n_{k_0+1}}$.

(iii) $\Longrightarrow$ (ii). Suppose that (iii) holds and that $A \subset K$ is an infinite subset. We claim that we can inductively construct a decreasing sequence of infinite subsets $A_n \subset A$ as well as a sequence (x_n) in X such that $A_n \subset B(x_n, 1/n)$. Suppose that we have constructed $A \supset A_1 \supset \cdots \supset A_n$ and $x_1, \ldots, x_n$. Take a finite subset $I \subset X$ such that

$$K \subset \bigcup_{x \in I} B(x, \frac{1}{n+1}) .$$

Because A_n is infinite, the pigeon hole principle provides us with an infinite subset $A_{n+1} \subset A_n$ and an element $x_{n+1} \in I$ such that $A_{n+1} \subset B(x_{n+1}, \frac{1}{n+1})$. This proves the claim. Take mutually different elements $a_n \in A_n$. If $n, m \geq n_0$, it follows that $d(a_n, a_m) \leq 2/n_0$. So, (a_n) is a Cauchy sequence in K. Because of 3, the sequence (a_n) converges to $a \in K$. By construction, a is an accumulation point of A.

(i) $\Longrightarrow$ (iv). Suppose that (i) holds. Let $(\mathcal{U}_i)_{i \in I}$ be an open covering of K. Because we already know that (iii) holds, it is sufficient to prove the existence of $\varepsilon > 0$ with the following property: for all $x \in K$, there exists $i \in I$ such that $B(x, \varepsilon) \subset \mathcal{U}_i$. Suppose that the contrary holds. Take for every $n \geq 1$ an element $x_n \in K$ such that

$$B(x_n, \frac{1}{n}) \not\subset \mathcal{U}_i \quad \text{for all } \; i \in I .$$

Let (x_{n_k}) be a subsequence of (x_n) converging to $x \in K$. Take $i \in I$ such that $x \in \mathcal{U}_i$. Take $\delta > 0$ such that $B(x, \delta) \subset \mathcal{U}_i$. Finally take k such that $\frac{1}{n_k} < \frac{\delta}{2}$ and $d(x, x_{n_k}) < \frac{\delta}{2}$. It follows that

$$B(x_{n_k}, \frac{1}{n_k}) \subset B(x, \delta) \subset \mathcal{U}_i$$

yielding a contradiction.

(iv) $\Longrightarrow$ (ii). Finally suppose that K is compact and let $A \subset K$ be an infinite subset. Assume that A has no accumulation point in K. Whenever $x \in K$, we thus know that x is not an accumulation point of A, meaning that there exists $\varepsilon_x > 0$ such that

$$B(x, \varepsilon_x) \cap A \subset \{x\} .$$

Then, $\{B(x, \varepsilon_x) \mid x \in K\}$ is an open covering of K. Take a finite subcovering $\{B(x, \varepsilon_x) \mid x \in I\}$, with $I \subset K$ finite. It follows that

$$A = K \cap A = \Big(\bigcup_{x \in I} B(x, \varepsilon_x) \Big) \cap A = \bigcup_{x \in I} (B(x, \varepsilon_x) \cap A) \subset I .$$

This is a contradiction with the assumption that A is infinite. □

Corollary 3.2. *Let (X, d) be a complete metric space and $K_0 \subset X$. Then, the following two conditions are equivalent.*

- *The closure of K_0 is compact.*
- *For every $\varepsilon > 0$, there exists a finite subset $I \subset X$ such that*

$$K_0 \subset \bigcup_{x \in I} B(x, \varepsilon) .$$

3.2 Compact operators

Definition 3.3. Let H, K be Hilbert spaces. A bounded operator $T : H \to K$ is called *compact* if the set

$$\{Tx \mid x \in H, \|x\| \leq 1\}$$

has compact closure in K.

The set of compact operators from H to K is denoted by $\mathcal{K}(H, K)$ and the set of compact operators on H is denoted by $\mathcal{K}(H)$.

We define the *rank* of an operator $T : H \to K$ as the vector space dimension of $\operatorname{Im} T$. So, T *has finite rank* if the image of T is a finite-dimensional vector subspace of K. Since the unit ball of $\mathbb{C}^n$ is compact, it follows that all finite rank operators are compact. The converse does not hold, but every compact operator can approximated in the norm topology by finite rank operators, as we prove now.

Theorem 3.4. *Let H be a Hilbert space. Then $\mathcal{K}(H)$ is a closed two-sided ideal in* $\mathrm{B}(H)$*. This means that*

- *$\mathcal{K}(H)$ is a closed vector subspace of* $\mathrm{B}(H)$*;*
- *$ST \in \mathcal{K}(H)$ and $TS \in \mathcal{K}(H)$ for all $T \in \mathcal{K}(H)$ and $S \in$* $\mathrm{B}(H)$*.*

If $T \in$ $\mathrm{B}(H)$*, the following conditions are equivalent.*

- *T is compact.*
- *T^* is compact.*
- *There exists a sequence (T_n) of finite rank operators such that $\|T - T_n\| \to 0$.*

Proof. The proof can be found in [Con, II.4.2 and II.4.4]. You can however write a proof yourself according to the following steps. For simplicity we assume that H is separable.

Use the following scheme to prove that $\mathcal{K}(H)$ is a closed vector subspace of $\mathrm{B}(H)$. Take a sequence $T_n \in \mathcal{K}(H)$ and a $T \in \mathrm{B}(H)$ such that $\|T - T_n\| \to 0$. You have to prove that $T \in \mathcal{K}(H)$. Choose $\varepsilon > 0$. You have to prove that $\{Tx \mid \|x\| \leq 1\}$ can be covered by finitely many balls of radius ε.

1. Take n such that $\|T - T_n\| < \varepsilon/2$.
2. Since T_n is compact, choose finitely many vectors $y_1, \dots, y_k \in H$ such that the balls $B(y_i, \varepsilon/2)$, $i = 1, \dots, k$, cover $\{T_n x \mid \|x\| \le 1\}$.
3. Prove that the balls $B(y_i, \varepsilon)$, $i = 1, \dots, k$, cover $\{Tx \mid \|x\| \le 1\}$.

Next take $T \in \mathcal{K}(H)$ and $S \in \mathrm{B}(H)$.

1. To prove that $ST \in \mathcal{K}(H)$ use that the continuous map $S : H \to H$ maps compact sets to compact sets.
2. To prove that $TS \in \mathcal{K}(H)$ use that the bounded operator $S : H \to H$ maps the unit ball of H into the ball around 0 with radius $\|S\|$.

We already observed that finite rank operators are compact. Since $\mathcal{K}(H)$ is closed inside $\mathrm{B}(H)$ all limits of finite rank operators are compact as well. Since the adjoint of a finite rank operator is still of finite rank, the only remaining property to prove is that every compact operator can be approximated by finite rank operators. Take $T \in \mathcal{K}(H)$. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of H. Denote by P_n the orthogonal projection of H onto $\operatorname{span}\{e_0, \dots, e_n\}$. We know from Proposition 1.22 that for all $y \in H$ we have $P_n y \to y$.

Prove as follows that $P_n y \to y$ *uniformly* on compact subsets of H. So let $K \subset H$ be a compact subset. Choose $\varepsilon > 0$.

1. Cover K with finitely many balls $B(y_i, \varepsilon/3)$, $i = 1, \dots, k$.
2. Take n_0 such that $\|P_n(y_i) - y_i\| < \varepsilon/3$ for all $n \ge n_0$ and all $i = 1, \dots, k$.
3. Prove that $\|P_n(y) - y\| < \varepsilon$ for all $n \ge n_0$ and all $y \in K$.

Deduce that $P_n T x \to Tx$ uniformly on the unit ball of H. Conclude that this means that $\|P_n T - T\| \to 0$. Since $P_n T$ is a finite rank operator, we are done. □

The next criterion proves that all integral operators with square integrable kernel are compact. In fact, operators satisfying (3.1) are called *Hilbert-Schmidt operators* and are studied in more detail in Section 3.6.

Proposition 3.5. *Let H be a Hilbert space with orthonormal basis $(e_n)_{n\in\mathbb{N}}$. If $T \in \mathrm{B}(H)$ and*

$$\sum_{n=0}^{\infty} \|Te_n\|^2 < \infty , \tag{3.1}$$

then T is a compact operator.

Proof. We claim that for all $S \in \mathrm{B}(H)$

$$\|S\|^2 \le \sum_{n=0}^{\infty} \|Se_n\|^2 .$$

Indeed, since $\|S\| = \|S^*\|$, it suffices to observe that

$$\|S^*x\|^2 = \sum_{n=0}^{\infty} |\langle e_n, S^*x\rangle|^2 = \sum_{n=0}^{\infty} |\langle Se_n, x\rangle|^2 \leq \sum_{n=0}^{\infty} \|Se_n\|^2 \, \|x\|^2 \,.$$

Suppose that $T \in \mathrm{B}(H)$ satisfies (3.1). Set

$$\varepsilon_n = \sum_{k=n+1}^{\infty} \|Te_k\|^2 \,.$$

Observe that $\varepsilon_n \to 0$. Moreover, if P_n denotes the orthogonal projection onto $\operatorname{span}\{e_0, \dots, e_n\}$, it follows that

$$\|T - TP_n\|^2 = \|T(1 - P_n)\|^2 \leq \sum_{k=0}^{\infty} \|T(1 - P_n)e_k\|^2 = \varepsilon_n \to 0 \,.$$

So, T is compact. □

Corollary 3.6. *An integral operator T defined by a square integrable kernel $K \in L^2(A \times A, \lambda)$ as in Proposition 2.9, is compact.*

Proof. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis for $\mathrm{L}^2(A, \lambda)$. Define $f_{m,n}(x, y) = \overline{e_m(x)}e_n(y)$ and check that $(f_{m,n})_{m,n\in\mathbb{N}}$ is an orthonormal family in $\mathrm{L}^2(A \times A, \lambda)$. (One can actually prove that $(f_{m,n})_{m,n\in\mathbb{N}}$ is an orthonormal basis but this is not needed to prove this corollary.) It follows that

$$\sum_{n=0}^{\infty} \|Te_n\|^2 = \sum_{n,m=0}^{\infty} |\langle Te_n, e_m\rangle|^2 = \sum_{n,m=0}^{\infty} |\langle K, f_{m,n}\rangle|^2 \leq \|K\|_2^2 < \infty \,.$$

By Proposition 3.5, the operator T is compact. □

3.3 Diagonalizable operators

In linear algebra, one proves that every Hermitian matrix can be diagonalized: there exists a basis of eigenvectors. We prove that this holds in general for *compact self-adjoint operators*. We introduce the following obvious definition.

Definition 3.7. Let $T \in \mathrm{B}(H)$ be a bounded operator.

- A vector $x \in H$ is called an *eigenvector* of T if $x \neq 0$ and $Tx = \lambda x$ for some $\lambda \in \mathbb{C}$. We call λ the *eigenvalue of* x.
- A complex number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of T if there exists an eigenvector with eigenvalue λ.
- We call T *diagonalizable* if H admits an orthonormal basis consisting of eigenvectors of T.

Example 3.8. The multiplication operator M_λ in Example 2.2.(iii) is diagonalizable for all $\lambda \in \ell^\infty(\mathbb{N})$. Indeed, the standard orthonormal basis $(e_n)_{n\in\mathbb{N}}$ for $\ell^2(\mathbb{N})$ consists of eigenvectors.

Exercise 1. Let H be a separable infinite dimensional Hilbert space and $T \in \mathrm{B}(H)$. Prove that T is diagonalizable if and only if there exists $\lambda \in \ell^\infty(\mathbb{N})$ and a unitary operator $U : \ell^2(\mathbb{N}) \to H$ such that $T = UM_\lambda U^*$, where again M_λ is the multiplication operator defined in Example 2.2.(iii).

3.4 Diagonalization of compact self-adjoint operators

The main aim of this section is to prove the following theorem.

Theorem 3.9. *Every compact self-adjoint operator is diagonalizable.*

To prove Theorem 3.9, we start with the modest lemma 3.10 below telling us that a compact self-adjoint operator has at least one eigenvector. But first of all, we need the following.

Lemma 3.10. *Let $T \in \mathrm{B}(H)$ be self-adjoint. Then,*

$$\|T\| = \sup\{|\langle Tx, x\rangle| \mid x \in H, \|x\| \leq 1\} .$$

Proof. Denote $M = \sup\{|\langle Tx, x\rangle| \mid x \in H, \|x\| \leq 1\}$. Check that $M \leq \|T\|$. Also observe that $\pm\langle Tx, x\rangle \leq M\|x\|^2$ for all $x \in H$ (this inequality makes sense: because T is self-adjoint, $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in H$).

Take $x, y \in H$. Then,

$$\langle T(x+y), x+y\rangle \leq M\|x+y\|^2 ,$$
$$-\langle T(x-y), x-y\rangle \leq M\|x-y\|^2 .$$

Adding these two inequalities and making a little computation, we arrive at

$$4\,\mathrm{Re}\langle Tx, y\rangle \leq 2M(\|x\|^2 + \|y\|^2) .$$

It follows that $\mathrm{Re}\langle Tx, y\rangle \leq M$ for all x, y with $\|x\|, \|y\| \leq 1$. Hence, $\|T\| \leq M$. □

Lemma 3.11. *Let T be a compact self-adjoint operator. Then T has an eigenvector with eigenvalue $\|T\|$ or $-\|T\|$.*

Proof. If $T = 0$, nothing has to be proven. Suppose $T \neq 0$.

Write $B = B(0,1) \subset H$. By Lemma 3.10, take a sequence (x_n) in B such that $|\langle Tx_n, x_n\rangle| \to \|T\|$. Theorem 3.1 ensures that we can pass to a subsequence and suppose that $Tx_n \to y$ and $\langle Tx_n, x_n\rangle \to \lambda$ with $\lambda = \pm\|T\|$. In particular, $\lambda \neq 0$. Check that $\|y\| = |\lambda|$ and hence, also $y \neq 0$.

Observe now that

$$\|Tx_n - \lambda x_n\|^2 = \|Tx_n\|^2 - 2\lambda\,\mathrm{Re}\langle Tx_n, x_n\rangle + \lambda^2\|x_n\|^2 \leq 2(\lambda^2 - \lambda\,\mathrm{Re}\langle Tx_n, x_n\rangle) \to 0 .$$

Because $Tx_n \to y$, it follows that $\lambda x_n \to y$ and hence $x_n \to \frac{1}{\lambda}y$. By continuity of T, we get $Tx_n \to \frac{1}{\lambda}Ty$ and so, $Ty = \lambda y$. □

Proof of Theorem 3.9. By Zorn's lemma take a maximal orthonormal family $(e_i)_{i\in I}$ of eigenvectors for T. Denote by $K \subset H$ the closure of the linear span of the vectors $e_i, i \in I$. We shall prove that $K = H$. Check that $TK \subset K$. Use the self-adjointness of T to deduce that $TK^\perp \subset K^\perp$. Denote by T_0 the restriction of T to $K^\perp$. View T_0 as a bounded operator on the Hilbert space $K^\perp$. Prove that T_0 is compact and self-adjoint. So if $K^\perp \neq \{0\}$, Lemma 3.11 provides a norm one eigenvector f for T_0. Then we can add f to the family $(e_i)_{i\in I}$ contradicting its maximality. So $K^\perp = \{0\}$ meaning that $K = H$. □

3.5 Positive operators (optional)

Definition 3.12. A bounded operator $T \in B(H)$ is called *positive* if

$$\langle Tx, x\rangle \geq 0 \quad \text{for all} \quad x \in H \; .$$

We denote the set of positive operators by $B(H)^+$. We also write $T \geq 0$ to denote that T is a positive operator.

Observe that a positive operator is automatically self-adjoint. Indeed, for all $x \in H$, we have

$$\langle Tx, x\rangle = \overline{\langle Tx, x\rangle} = \langle x, Tx\rangle$$

and polarization implies that $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$.

Exercise 2. Prove that $T = 0$ whenever $T \in B(H)^+$ and $-T \in B(H)^+$.

Definition 3.13. Because of the previous exercise, we can define a partial order on the set of self-adjoint operators, by putting

$$T \leq S \quad \text{if and only if} \quad S - T \in B(H)^+ \; .$$

Exercise 3. Prove the following statements.

1. If S is self-adjoint, we have $-\|S\|1 \leq S \leq \|S\|1$.
2. If $S \leq T$ and $R \in B(H)$, then $RSR^* \leq RTR^*$.

Observe that the first item of the previous exercise implies that every self-adjoint operator can be written as the difference of two positive operators. This provides the first indication that self-adjoint operators behave much like real numbers with the positive operators corresponding to the positive real numbers. This becomes more clear in the course *Spectral Theory and Operator Algebras*.

Another illustration is provided by the following theorem.

Theorem 3.14. *Let $T \in B(H)^+$ be a positive operator. There exists a unique positive operator $S \in B(H)^+$ satisfying $S^2 = T$. In words: every positive operator has a unique positive square root. We write $S = T^{1/2}$.*

An elegant proof of Theorem 3.14 can only be given by invoking the spectral theorem for arbitrary self-adjoint operators, which we will do in Lecture 4 (see 4.14. For compact operators we did prove a spectral theorem in 3.9 and this will allow us to prove easily Theorem 3.14 when T is compact.

Proof of Theorem 3.14 when T is compact. We first prove the existence of a positive square root S. By Theorem 3.9, take an orthonormal basis $(e_i)_{i\in I}$ of H satisfying $Te_i = \lambda_i e_i$ for all $i \in I$. By positivity of T, it follows that $\lambda_i \geq 0$ for all $i \in I$. Define $S \in B(H)$ such that $Se_i = \sqrt{\lambda_i} e_i$ for all $i \in I$. Check that $S \in B(H)^+$ and $S^2 = T$.

In order to prove uniqueness of S, it suffices to prove the following statement: whenever $R \in B(H)^+$, $R^2 = T$ and $x \in H$ is an eigenvector for T with eigenvalue $\lambda \geq 0$, then $Rx = \sqrt{\lambda}x$. First note that if $\lambda = 0$, we have $Tx = 0$ and hence

$$\|Rx\|^2 = \langle Rx, Rx\rangle = \langle Tx, x\rangle = 0$$

implying that also $Rx = 0$. If $\rho > 0$, first observe that $\mathrm{Ker}(R+\rho) = \{0\}$. Indeed, if $(R+\rho)x = 0$, it follows that

$$0 = \langle (R+\rho)x, x\rangle = \langle Rx, x\rangle + \rho\|x\|^2 \geq \rho\|x\|^2$$

implying that $x = 0$. But, when $Tx = \lambda x$ and $\lambda \neq 0$, the vector $(R - \sqrt{\lambda})x$ belongs to the kernel of $R + \sqrt{\lambda}$, implying that $Rx = \sqrt{\lambda}x$. □

3.6 Trace-class and Hilbert-Schmidt operators (optional)

The trace of an $n \times n$ matrix $A \in \mathrm{M}_n(\mathbb{C})$ is defined as

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii} \ .$$

If we consider A as an operator on $\mathbb{C}^n$, the formula for the trace can be rewritten as

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \langle Ae_i, e_i\rangle \ ,$$

where $e_1, \ldots, e_n$ denotes the standard orthonormal basis of $\mathbb{C}^n$.

All this makes the following definition not so surprising, but we have to take care because the Hilbert spaces involved are infinite dimensional and so, finite sums become series.

Throughout this section, all Hilbert spaces are supposed to be separable.

Definition 3.15. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis for a Hilbert space H. We define the map

$$\mathrm{Tr} : \mathrm{B}(H)^+ \to [0, +\infty] : \mathrm{Tr}(A) = \sum_{n=0}^{\infty} \langle Ae_n, e_n\rangle \ .$$

We call Tr the *trace* on $\mathrm{B}(H)$.

There is no problem in summing the series in Definition 3.15. Indeed, for the moment we are only dealing with positive operators and a series with positive terms can always be summed provided that we allow $+\infty$ as a value for the sum.

Proposition 3.16. *The trace* Tr *is independent of the choice of orthonormal basis. Moreover, the trace satisfies the following properties.*

1. $\mathrm{Tr}(T^*T) = \mathrm{Tr}(TT^*)$ *for all* $T \in \mathrm{B}(H)$.
2. *For all* $\lambda \geq 0$*, the set* $\{T \in \mathrm{B}(H)^+ \mid \mathrm{Tr}(T) \leq \lambda\}$ *is closed. In words: the trace is lower semicontinuous.*
3. *If* $0 \leq S \leq T$*, we have* $0 \leq \mathrm{Tr}(S) \leq \mathrm{Tr}(T)$.

Proof. Suppose that (e_n) and (f_n) are two orthonormal bases for H. Let $T \in \mathrm{B}(H)^+$ with positive square root S as in Theorem 3.14. Then the Parseval equality yields

$$\begin{aligned}\sum_{n=0}^{\infty} \langle T e_n, e_n \rangle &= \sum_{n=0}^{\infty} \|S e_n\|^2 = \sum_{n,m=0}^{\infty} |\langle S e_n, f_m \rangle|^2 \\ &= \sum_{n,m=0}^{\infty} |\langle S f_m, e_n \rangle|^2 = \sum_{m=0}^{\infty} \|S f_m\|^2 \\ &= \sum_{m=0}^{\infty} \langle T f_m, f_m \rangle \; ,\end{aligned}$$

proving that the definition of Tr is independent of the choice of orthonormal basis. The formula $\mathrm{Tr}(T^*T) = \mathrm{Tr}(TT^*)$ follows by an analogous application of the Parseval equality. The remaining statements are left as an exercise. □

Definition 3.17. We define

$$\begin{aligned}\mathcal{TC}(H)^+ &= \{T \in \mathrm{B}(H)^+ \mid \mathrm{Tr}(T) < \infty\} \; , \\ \mathcal{TC}(H) &= \operatorname{span} \mathcal{TC}(H)^+ \; , \\ \mathcal{HS}(H) &= \{T \in \mathrm{B}(H) \mid \mathrm{Tr}(T^*T) < \infty\} \; .\end{aligned}$$

The elements of $\mathcal{TC}(H)$ are called *trace-class operators.* We define for $T \in \mathcal{TC}(H)$,

$$\mathrm{Tr}(T) = \sum_{n=0}^{\infty} \langle T e_n, e_n \rangle$$

and observe that the series on the right hand side is absolutely summable.

The elements of $\mathcal{HS}(H)$ are called *Hilbert-Schmidt operators.*

Example 3.18. The integral operators T with square integrable kernel K (see Proposition 2.9) are Hilbert-Schmidt operators. Indeed, in the proof of Corollary 3.6 we checked that $\mathrm{Tr}(T^*T) \leq \|K\|_2^2$. In fact, equality holds if you accept that the orthogonal family $(f_{n,m})_{n,m \in \mathbb{N}}$ defined in the proof of Corollary 3.6 is in fact an orthonormal basis of $\mathrm{L}^2(A \times A, \lambda)$.

The next proposition is conceptually nontrivial: the set of Hilbert-Schmidt operators on H forms itself a Hilbert space.

Proposition 3.19. *(i) The Hilbert-Schmidt operators form a vector subspace of* $\mathrm{B}(H)$.

(ii) If $T, S \in \mathcal{HS}(H)$, *we have* $TS, ST \in \mathcal{TC}(H)$ *and* $\mathrm{Tr}(TS) = \mathrm{Tr}(ST)$.

(iii) Equipped with the inner product

$$\langle T, S \rangle = \mathrm{Tr}(TS^*) \; ,$$

the vector space $\mathcal{HS}(H)$ *becomes a Hilbert space. We denote the corresponding norm as*

$$\|T\|_2 = \sqrt{\mathrm{Tr}(T^*T)} \; .$$

(iv) *Further, $\mathcal{HS}(H)$ is a two-sided ideal in $\mathrm{B}(H)$, meaning that $ST, TS \in \mathcal{HS}(H)$ whenever $T \in \mathcal{HS}(H)$ and $S \in \mathrm{B}(H)$. Moreover, the following inequalities hold.*

$$\|ST\|_2 \leq \|S\|\,\|T\|_2 \quad \textit{and} \quad \|TS\|_2 \leq \|S\|\,\|T\|_2\ .$$

(v) *Finally, $\mathcal{HS}(H) \subset \mathcal{K}(H)$.*

Proof. (i) Take $S, T \in \mathrm{B}(H)$. Observe that

$$(S+T)^*(S+T) + (S-T)^*(S-T) = 2(S^*S + T^*T)\ .$$

Hence, $(S+T)^*(S+T) \leq 2(S^*S + T^*T)$ implying that $\mathcal{HS}(H)$ is a vector subspace of $\mathrm{B}(H)$.

(ii) If $T \in \mathcal{HS}(H)$, we have by definition and Proposition 3.16 that $T^*T, TT^* \in \mathcal{TC}(H)$ with $\mathrm{Tr}(T^*T) = \mathrm{Tr}(TT^*)$. By polarization, we get that $TS^*, S^*T \in \mathcal{TC}(H)$ for all $T, S \in \mathcal{HS}(H)$ and that $\mathrm{Tr}(TS^*) = \mathrm{Tr}(S^*T)$.

(iii) From 2 we know that $\langle T, S\rangle = \mathrm{Tr}(TS^*)$ yields a well defined positive Hermitian form on $\mathcal{HS}(H)$. In the proof of Proposition 3.5, we have already seen that $\mathrm{Tr}(T^*T) \geq \|T\|^2$, implying that the above Hermitian form is positive-definite. So, it remains to prove that $\mathcal{HS}(H)$ equipped with the norm $\|\cdot\|_2$ is complete. Suppose that (T_n) is a Cauchy sequence in $\mathcal{HS}(H)$. The inequality $\|T\| \leq \|T\|_2$ for all $T \in \mathcal{HS}(H)$, implies that (T_n) is also a Cauchy sequence in $\mathrm{B}(H)$. So, $T_n \to T$ in $\mathrm{B}(H)$. It remains to show that $\|T - T_n\|_2 \to 0$. Choose $\varepsilon > 0$. Take n_0 such that $\|T_m - T_n\|_2 \leq \varepsilon$ for all $n, m \geq n_0$. By lower semicontinuity of Tr, we can take the limit $m \to \infty$ and get $\|T - T_n\|_2 \leq \varepsilon$ for all $n \geq n_0$.

(iv) Let $T \in \mathcal{HS}(H)$ and $S \in \mathrm{B}(H)$. The inequality $(ST)^*ST = T^*S^*ST \leq \|S\|^2 T^*T$ implies that $ST \in \mathcal{HS}(H)$ and $\|ST\|_2 \leq \|S\|\,\|T\|_2$. The fact that $TS \in \mathcal{HS}(H)$ with $\|TS\|_2 \leq \|S\|\,\|T\|_2$ is proven similarly.

(v) This follows from Proposition 3.5. □

Exercise 4. Let M_λ be the multiplication operator defined in Example 2.2.(iii). Prove that

1. M_λ is compact if and only if $\lambda \in c_0(\mathbb{N})$;
2. M_λ is trace-class if and only if $\lambda \in \ell^1(\mathbb{N})$; compute $\mathrm{Tr}(M_\lambda)$ in that case;
3. M_λ is Hilbert-Schmidt if and only if $\lambda \in \ell^2(\mathbb{N})$.

Conclude that the inclusion $\mathcal{HS}(H) \subset \mathcal{K}(H)$ is strict when H is infinite dimensional.

Before making a more detailed study of trace-class operators, we introduce a new tool: the *polar decomposition* of a bounded operator. Recall that every complex number can be written as the product of a number of modulus 1 and a positive real number. In a certain sense, a similar thing can be done with bounded operators on a Hilbert space.

Definition 3.20. Let $T \in \mathrm{B}(H)$. Define the *absolute value* of T as

$$|T| = (T^*T)^{\frac{1}{2}}\ .$$

Theorem 3.21. *Let $T \in \mathrm{B}(H)$. Denote by P the orthogonal projection onto $(\operatorname{Ker} T)^\perp$. There exists a unique operator $U \in \mathrm{B}(H)$ satisfying*

$$T = U|T| \quad \textit{and} \quad U^*U = P \; .$$

We call the expression $T = U|T|$, the polar decomposition *of T.*

Proof. Check that the formula $U^*U = P$ is equivalent with the following statement

$$\|Ux\| = \|x\| \quad \text{for all} \;\; x \in (\operatorname{Ker} T)^\perp \quad \text{and} \quad Ux = 0 \quad \text{for all} \;\; x \in \operatorname{Ker} T \; .$$

Also observe that $\|Tx\| = \| \, |T|x \, \|$ for all $x \in H$. It follows in particular that $\operatorname{Ker} T = \operatorname{Ker} |T|$ and so $(\operatorname{Ker} T)^\perp$ coincides with the closure of $\operatorname{Im} |T|$. So, we uniquely define $U \in \mathrm{B}(H)$ satisfying $U^*U = P$ and $U|T|x = Tx$ for all $x \in H$. □

Exercise 5.

1. Let $\lambda \in \ell^\infty(\mathbb{Z})$. Define the operator

$$T : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z}) : (Tx)(n) = \lambda(n)x(n+1) \; .$$

   Prove that T is a bounded operator and compute its polar decomposition.

2. Let $U \in \mathrm{B}(H)$. Prove that the following statements are equivalent.

   (a) U^*U is an orthogonal projection.

   (b) UU^* is an orthogonal projection.

   (c) There exists a closed subspace $K \subset H$ such that the restriction of U to K is isometric and $Ux = 0$ for all $x \in K^\perp$.

   An operator satisfying one of these equivalent conditions is called a *partial isometry.* Note that the polar part U in the polar decomposition given by Theorem 3.21 is a partial isometry.

3. Let $T = U|T|$ be the polar decomposition of a bounded operator T. Prove that U is unitary if and only if $\operatorname{Ker} T = \operatorname{Ker}(T^*) = \{0\}$.

We gathered enough material to prove the following duality theorem. First recall Exercise 4, saying that the multiplication operator M_λ is compact if and only if $\lambda \in c_0(\mathbb{N})$ and is trace-class if and only if $\lambda \in \ell^1(\mathbb{N})$. Finally recall the Banach space duality $\ell^1(\mathbb{N}) \cong c_0(\mathbb{N})^*$ and $\ell^\infty(\mathbb{N}) \cong \ell^1(\mathbb{N})^*$. The same results holds true for operators: $\mathcal{TC}(H) \cong \mathcal{K}(H)^*$ and $\mathrm{B}(H) \cong \mathcal{TC}(H)^*$. A more precise statement is given now.

Theorem 3.22. *(i) The vector space $\mathcal{TC}(H)$ is a two-sided ideal in $\mathrm{B}(H)$ and $\operatorname{Tr}(ST) = \operatorname{Tr}(TS)$ for all $T \in \mathcal{TC}(H)$ and $S \in \mathrm{B}(H)$.*

(ii) A bounded operator $T \in \mathrm{B}(H)$ belongs to $\mathcal{TC}(H)$ if and only if $\operatorname{Tr}(|T|) < \infty$. Moreover, the formula

$$\|T\|_1 = \operatorname{Tr}(|T|)$$

defines a norm on $\mathcal{TC}(H)$ and equipped with this norm, $\mathcal{TC}(H)$ is a Banach space.

(iii) The maps

$$\mathcal{TC}(H) \to \mathcal{K}(H)^* : T \mapsto \omega_T \quad \textit{with} \quad \omega_T : \mathcal{K}(H) \to \mathbb{C} : \omega_T(S) = \mathrm{Tr}(ST) ,$$
$$\mathrm{B}(H) \to \mathcal{TC}(H)^* : S \mapsto \mu_S \quad \textit{with} \quad \mu_S : \mathcal{TC}(H) \to \mathbb{C} : \mu_S(T) = \mathrm{Tr}(ST) ,$$

are isometric isomorphisms.

(iv) We have the inclusions $\mathcal{TC}(H) \subset \mathcal{HS}(H) \subset \mathcal{K}(H)$ with corresponding inequalities

$$\|T\|_2 \le \sqrt{\|T\|\,\|T\|_1} \quad \textit{and} \quad \|S\| \le \|S\|_2$$

for all $T \in \mathcal{TC}(H)$ and $S \in \mathrm{B}(H)$.

Proof. (i) By linearity, we may assume that $T \in \mathcal{TC}(H)^+$. Set $R = T^{\frac{1}{2}}$. Then, $R \in \mathcal{HS}(H)$ and Proposition 3.19 implies firstly that $SR, RS \in \mathcal{HS}(H)$, secondly that $ST = (SR)R$ and $TS = R(RS)$ are trace-class operators and finally that

$$\mathrm{Tr}(ST) = \mathrm{Tr}((SR)R) = \mathrm{Tr}(R(SR)) = \mathrm{Tr}((RS)R) = \mathrm{Tr}(R(RS)) = \mathrm{Tr}(TS) .$$

(ii) Let $T = U|T|$ be the polar decomposition of T. Since $T = U|T|$ and $|T| = U^*T$, it follows from (i) that $T \in \mathcal{TC}(H)$ if and only if $|T| \in \mathcal{TC}(H)$. And the latter is by definition equivalent with $\mathrm{Tr}(|T|) < \infty$. The rest of (ii) will be proven below.

(iii) We first prove that for all $T \in \mathcal{TC}(H)$,

$$\mathrm{Tr}(|T|) = \sup\{|\mathrm{Tr}(ST)| \mid S \in \mathcal{K}(H), \|S\| \le 1\} = \sup\{|\mathrm{Tr}(ST)| \mid S \in \mathrm{B}(H), \|S\| \le 1\} . \tag{3.2}$$

Choose $S \in \mathrm{B}(H)$ with $\|S\| \le 1$. Take the polar decomposition $T = U|T|$. Using the inner product on $\mathcal{HS}(H)$ and the results of Proposition 3.19, we get

$$|\mathrm{Tr}(ST)| = |\langle SU|T|^{\frac{1}{2}}, |T|^{\frac{1}{2}}\rangle| \le \|SU|T|^{\frac{1}{2}}\|_2 \, \||T|^{\frac{1}{2}}\|_2 \le \|SU\| \, \mathrm{Tr}(|T|) \le \mathrm{Tr}(|T|) .$$

We have shown that

$$\sup\{|\mathrm{Tr}(ST)| \mid S \in \mathrm{B}(H), \|S\| \le 1\} \le \mathrm{Tr}(|T|) .$$

Next, let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of H and denote by P_n the orthogonal projection onto $\mathrm{span}\{e_0, \ldots, e_n\}$. It follows that

$$\mathrm{Tr}((P_n U^*)T) = \mathrm{Tr}(P_n|T|) = \sum_{k=0}^{n} \langle |T|e_k, e_k\rangle .$$

Because $P_n U^*$ is a compact operator with $\|P_n U^*\| \le 1$, we have shown that

$$\mathrm{Tr}(|T|) \le \sup\{|\mathrm{Tr}(ST)| \mid S \in \mathcal{K}(H), \|S\| \le 1\} .$$

Altogether, we have proven (3.2). It follows that the formula $\|T\|_1 = \mathrm{Tr}(|T|)$ defines a norm on $\mathcal{TC}(H)$ in such a way that the map $\Theta : \mathcal{TC}(H) \to \mathcal{K}(H)^* : T \mapsto \omega_T$ is an isometry. If we now prove that Θ is surjective, it follows that Θ is an isometric isomorphism and that $\mathcal{TC}(H)$ is a Banach space (concluding in particular the proof of (ii)).

Suppose that $\omega \in \mathcal{K}(H)^*$. Denote for all $x, y \in H$ by $\theta_{x,y}$ the rank one operator defined as

$$\theta_{x,y}(z) = \langle z, y \rangle x \; .$$

It follows that $H \times H \to \mathbb{C} : (x, y) \mapsto \omega(\theta_{x,y})$ is a bounded sesquilinear form on H. So, we find a bounded operator $T \in \mathrm{B}(H)$ such that

$$\omega(\theta_{x,y}) = \langle Tx, y \rangle = \mathrm{Tr}(T\theta_{x,y}) \quad \text{for all} \;\; x, y \in H \; .$$

We prove now that $T \in \mathcal{TC}(H)$. Since the finite rank operators are dense in $\mathcal{K}(H)$, it then follows that $\omega = \omega_T$, proving the surjectivity of Θ. Let $T = U|T|$ be the polar decomposition of T and let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of H. Denote by P_n the orthogonal projection onto $\mathrm{span}\{e_0, \ldots, e_n\}$. For every n, we have

$$\langle |T|e_n, e_n \rangle = \langle Te_n, Ue_n \rangle = \omega(\theta_{e_n,Ue_n}) = \omega(\theta_{e_n,e_n} U^*) \; .$$

It follows that

$$\sum_{k=0}^{n} \langle |T|e_k, e_k \rangle = \omega(P_n U^*) \leq \|\omega\|$$

for all n. Hence, $\mathrm{Tr}(|T|) < \infty$ and $T \in \mathcal{TC}(H)$.

To conclude the proof of (iii), it remains to study $S \mapsto \mu_S$. By the above results, $\|\mu_S\| \leq \|S\|$ for all $S \in \mathrm{B}(H)$. Moreover, $\langle Sx, y \rangle = \mathrm{Tr}(S\theta_{x,y}) = \mu_S(\theta_{x,y})$, implying that $\|S\| \leq \|\mu_S\|$. So, we obtain 3 once we have shown that every $\mu \in \mathcal{TC}(H)^*$ is of the form μ_S for some $S \in \mathrm{B}(H)$. But given μ, we find in exactly the same way as above, a bounded operator $S \in \mathrm{B}(H)$ such that

$$\langle Sx, y \rangle = \mu(\theta_{x,y}) \quad \text{for all} \;\; x, y \in H \; .$$

We claim that the finite rank operators are dense in $\mathcal{TC}(H)$, implying that $\mu = \mu_S$ and concluding the proof of (iii).

It is sufficient to prove that every $T \in \mathcal{TC}(H)^+$ can be approximated in $\|\cdot\|_1$-norm by finite rank operators. Let again $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of H and P_n the orthogonal projection onto $\mathrm{span}\{e_0, \ldots, e_n\}$. Then,

$$\|(1 - P_n)T(1 - P_n)\|_1 = \sum_{k=n+1}^{\infty} \langle Te_k, e_k \rangle \to 0 \; .$$

But $(1-P_n)T(1-P_n) = T-(P_nT+TP_n-P_nTP_n)$. Since $P_nT+TP_n-P_nTP_n$ is an operator of finite rank, we are done with 3.

(iv) Only the first inequality still has to be shown. But,

$$\|T\|_2^2 = \mathrm{Tr}(T^*T) = \mathrm{Tr}(|T|^{\frac{1}{2}}\,|T|\,|T|^{\frac{1}{2}}) \leq \|\,|T|\,\|\; \mathrm{Tr}(|T|) = \|T\|\,\|T\|_1 \; .$$

□

3.7 Exercises

Exercise 6. Let H be a Hilbert space. Prove that one of the following holds.

1. Either there exists an $n \in \mathbb{N}$ and a unitary operator $U : \mathbb{C}^n \to H$,
2. or there exists an isometry $V : \ell^2(\mathbb{N}) \to H$.

Deduce that the unit ball of the Hilbert space H is compact if and only if H is finite-dimensional. In fact, the same statement holds for arbitrary normed spaces, but is slightly harder to prove (cf. [Ped, 2.1.9]).

Exercise 7. Consider the following variant of the Volterra operator. Define

$$H = \{f \in \mathrm{L}^2([0, 2\pi]) \mid \int_0^{2\pi} f(x)\, dx = 0\} \; .$$

Define $T \in \mathrm{B}(H)$ by the formula

$$(Tf)(x) = i \int_0^x f(y)\, dy + \frac{i}{2\pi} \int_0^{2\pi} y f(y)\, dy \; .$$

Prove the following statements

1. T is a compact self-adjoint operator on H.
2. For all $f \in H$, Tf is a continuous function.
3. If $Tf = \lambda f$ with $\lambda \neq 0$, the function f is infinitely differentiable and satisfies $\lambda f' = if$.
4. Deduce that $\lambda \neq 0$ is an eigenvalue of T if and only if $\lambda = \frac{1}{n}$ with $n \in \mathbb{Z} \setminus \{0\}$. Compute the corresponding eigenvectors.
5. Define $e_n \in \mathrm{L}^2([0, 2\pi])$ by $e_n(x) = (2\pi)^{-1/2} e^{inx}$. Accept for the moment that $\operatorname{Ker} T = \{0\}$. Conclude that $(e_n)_{n \in \mathbb{Z} \setminus \{0\}}$ is an orthonormal basis for H and hence, $(e_n)_{n \in \mathbb{Z}}$ is an orthonormal basis for $\mathrm{L}^2([0, 2\pi])$.
6. Suppose that $f \in \operatorname{Ker} T$. We want to prove that $f = 0$. First prove that

$$\int_a^b f(x)\, dx = 0 \quad \text{for all } \; 0 \le a \le b \le 2\pi \; .$$

Use your knowledge of measure theory to deduce that $f(x) = 0$ for almost all x.

Lecture 4

Spectral theorem

The aim of this lecture is to prove that any self-adjoint operator can be put in a special form, just like any compact operator can be diagonalized. The motivating example is the following multiplication operator on the Hilbert space $L^2[0,1]$

$$(M_x f)(t) := t f(t).$$

The subscript x denotes here the identity function $f(x) \equiv x$ and should not be confused with the variable y. In a certain sense this operator is diagonal, if we view the interval $[0,1]$ as some sort of 'continuous' orthonormal system. But this operator does not have any eigenvalues, so we cannot study it using the same tools as the ones applied in the case of compact operators; we need something more general.

4.1 The spectrum

Definition 4.1. Let $T \in B(H)$. Define the spectrum $\sigma(T)$ and spectral radius $\rho(T)$ of T as

$$\begin{aligned} \sigma(T) &= \{\lambda \in \mathbb{C} \mid T - \lambda 1 \text{ is not invertible } \}\ , \\ \rho(T) &= \sup\{|\lambda| \mid \lambda \in \sigma(T)\}\ . \end{aligned}$$

The spectrum can be divided into 3 disjoint parts:

- *point spectrum* – the eigenvalues;
- *continuous spectrum* – $\lambda \in \sigma(T)$ such that λ is not an eigenvalue but the range of $T - \lambda 1$ is dense in H;
- *residual spectrum* – none of the above.

Lemma 4.2. *The residual spectrum of a self-adjoint operator T is empty.*

Proof. Suppose that λ belongs to the residual spectrum, so $(\mathrm{Im}(T - \lambda 1))^\perp \neq \{0\}$. But $(\mathrm{Im}(T - \lambda 1))^\perp = \ker(T - \overline{\lambda} 1)$ by Proposition 2.8, so $\overline{\lambda}$ is an eigenvalue of T. As eigenvalues of self-adjoint operators are real, $\lambda = \overline{\lambda}$ is an eigenvalue, so it does not belong to the residual spectrum. □

Lemma 4.3. *Let $T \in B(H)$ be self-adjoint. Then $\lambda \in \sigma(T)$ if and only if there exists a sequence of unit vectors $\xi_n \in H$ such that*

$$\|T\xi_n - \lambda\xi_n\| \to 0 \,. \tag{4.1}$$

So, contrary to the finite-dimensional and the compact case, elements of $\sigma(T)$ need not be eigenvalues, but they do arise from approximate eigenvectors.

Proof. The statement is clear for eigenvalues, so we may assume that λ belongs to the continuous spectrum.

If there is no sequence of unit vectors $\xi_n \in H$ satisfying (4.1), then we find $\varepsilon > 0$ such that $\|(T-\lambda 1)\xi\| \geq \varepsilon\,\|\xi\|$ for all $\xi \in H$. So, $T-\lambda 1$ has closed range and is an invertible operator from H onto this closed range. But the range is also, so it is all of H. Therefore $T-\lambda 1$ is an invertible operator, thus $\lambda \notin \sigma(T)$. □

Corollary 4.4. *If $T \in B(H)$ is self-adjoint and $\lambda \in \sigma(T)$ then there exists a sequence of unit vectors $\xi_n \in H$ such that $\langle T\xi_n, \xi_n\rangle \to \lambda$.*

Proof. By the previous corollary, there is a sequence of unit vectors $\xi_n \in H$ such that $\|T\xi_n - \lambda\xi_n\| \to 0$. It follows from Cauchy-Schwarz inequality that $\langle T\xi_n, \xi_n\rangle - \lambda \underbrace{\langle \xi_n, \xi_n\rangle}_{=1} \to 0$. □

Corollary 4.5. *Spectrum of a self-adjoint operator T is real.*

Proof. We $\langle T\xi, \xi\rangle \in \mathbb{R}$ for any $\xi \in H$. □

Lemma 4.6. *If $T \in B(H)$ is self-adjoint, then $\sigma(T)$ contains $\|T\|$ or $-\|T\|$.*

Proof. It follows from the proof of Lemma 3.11. □

Lemma 4.7. *If $S \in B(H)$ and $\|S\| < 1$, then $1 - S$ is invertible.*

Proof. The sequence $\sum_{k=0}^{n} S^k$ converges in operator norm and its limit is the inverse of $1-S$. □

Lemma 4.8. *For all $T \in B(H)$, we have $\rho(T) \leq \|T\|$.*

Proof. Let $\lambda \in \mathbb{C}$ with $|\lambda| > \|T\|$. Write $S = -\lambda^{-1}T$. By Lemma 4.7, $T - \lambda 1 = -\lambda(1-S)$ is invertible. So, $\lambda \notin \sigma(T)$. □

Proposition 4.9. *If $T \in B(H)$ is self-adjoint, then $\sigma(T)$ is a closed subset of $[-\|T\|, \|T\|]$ and $\rho(T) = \|T\|$.*

Proof. By Lemmas 4.6 and 4.8, it only remains to prove that $\sigma(T)$ is closed. Take $\lambda \notin \sigma(T)$. Write $S = T - \lambda 1$. Whenever $|\mu| < \|S^{-1}\|^{-1}$, we get that

$$T - (\lambda + \mu)1 = S - \mu 1 = S(1 - \mu S^{-1})$$

is invertible by Lemma 4.7. So, $\lambda + \mu \notin \sigma(T)$ for all μ small enough. □

Lemma 4.10. *Let $S, T \in B(H)$. Then ST and TS are both invertible if and only if S and T are both invertible.*

Proof. Assume that ST and TS are both invertible. Then, $(ST)^{-1}S$ and $S(TS)^{-1}$ are a left, respectively right inverse of T, so that T is invertible. Similarly, S is invertible. The converse is trivial. □

Lemma 4.11. *Let $p \in \mathbb{C}[X]$ be a polynomial and $T \in B(H)$. Then, $\sigma(p(T)) = p(\sigma(T))$.*

Proof. Since $p(T) - \lambda 1 = (p - \lambda)(T)$, it suffices to prove that $p(T)$ is invertible if and only if all zeros of p lie outside $\sigma(T)$. Writing $p(X) = (X - \lambda_1) \cdots (X - \lambda_n)$ with $\lambda_i \in \mathbb{C}$, this follows from Lemma 4.10. □

Lemma 4.12. *Let $T \in B(H)$ be self-adjoint and $p \in \mathbb{C}[X]$. Then,*

$$\|p(T)\| = \sup\{|p(\lambda)| \mid \lambda \in \sigma(T)\} .$$

Proof. Define the polynomial p^* by taking the complex conjugate of each coefficient of p. Write $q = p^* p$. Since $T = T^*$, we get that $q(T) = p(T)^* p(T)$. It follows that $q(T)$ is self-adjoint and that $\|p(T)\|^2 = \|q(T)\|$. By Proposition 4.9 and Lemma 4.11, we get that

$$\|p(T)\|^2 = \|q(T)\| = \rho(q(T)) = \sup\{|q(\lambda)| \mid \lambda \in \sigma(T)\} = \big(\sup\{|p(\lambda)| \mid \lambda \in \sigma(T)\}\big)^2 .$$

□

We can apply polynomial functions to operators and our next goal will be to show that in the case of self-adjoint operators we can also apply continuous functions defined on the spectrum; Lemma 4.12 will be a crucial tool.

4.2 Continuous functional calculus

Given a compact subset $K \subset \mathbb{R}$, we denote by $C(K)$ the space of bounded continuous functions $K \to \mathbb{C}$ with the supremum norm $\| \cdot \|_\infty$.

Proposition 4.13. *Let $T \in B(H)$ be self-adjoint. There is a unique unital $*$-homomorphism $\Phi\colon C(\sigma(T)) \to B(H)$ (it means that Φ is a linear map that satisfies $\Phi(fg) = \Phi(f)\Phi(g)$, $\Phi(\bar{f}) = (\Phi(f))^*$ and $\Phi(1) = 1$) such that $\|\Phi(f)\| = \|f\|_\infty$ and $\Phi(p) = p(T)$ for all $f \in C(\sigma(T))$ and $p \in \mathbb{C}[X]$.*

We write $f(T)$ instead of $\Phi(f)$. Then the following properties hold for all $f \in C(\sigma(T))$.

1. *If $S \in B(H)$ and $ST = TS$, then $S\, f(T) = f(T)\, S$.*

2. *If $\xi \in H \setminus \{0\}$ is an eigenvector of T with eigenvalue λ, then ξ is also an eigenvector of $f(T)$, with eigenvalue $f(\lambda)$.*

3. *$\sigma(f(T)) = f(\sigma(T))$.*

4. *When $f : \sigma(T) \to \mathbb{R}$ and $g : f(\sigma(T)) \to \mathbb{C}$ are continuous functions, we have that $g(f(T)) = (g \circ f)(T)$.*

Proof. The existence and uniqueness of Φ, as well as properties 1 and 2, follow immediately from Lemma 4.12 and the density of polynomial functions in $C(\sigma(T)$.

It remains to prove properties 3 and 4. Note that $f(\sigma(T))$ is a compact subset of $\mathbb{C}$. If $\lambda \notin f(\sigma(T))$, we define $g \in C(\sigma(T))$ by $g(x) = (f(x)-\lambda)^{-1}$. Then, $g(T)$ is the inverse of $f(T)-\lambda 1$, so that $\lambda \notin \sigma(f(T))$. Conversely, assume that $\lambda = f(\mu)$ with $\mu \in \sigma(T)$. By Lemma 4.3, we can take a sequence of unit vectors $\xi_n \in H$ such that $\|T\xi_n - \mu\xi_n\| \to 0$. It follows that $\|p(T)\xi_n - p(\mu)\xi_n\| \to 0$ for every $p \in \mathbb{C}[X]$. By density, it also follows that $\|f(T)\xi_n - f(\mu)\xi_n\| \to 0$. This implies that $f(T) - f(\mu)1$ is not invertible. So, $\lambda = f(\mu)$ belongs to $\sigma(f(T))$.

Property 4 is immediate when g is a polynomial function and thus holds in general by density. □

Recall that an operator $T \in B(H)$ is said to be positive if $\langle T\xi, \xi\rangle \geq 0$ for all $\xi \in H$. Note that positive operators are always self-adjoint. We include the following characterization of positive operators as a corollary of the continuous functional calculus (cf. 3.14).

Corollary 4.14. *Let $T \in B(H)$ be self-adjoint. Then the following statements are equivalent.*

(i) $T = S^2$ *for some self-adjoint* $S \in B(H)$.

(ii) $T = S^*S$ *for some* $S \in B(H)$.

(iii) T *is positive.*

(iv) $\sigma(T) \subset [0, +\infty)$.

Also, for every positive operator $T \in B(H)$, there exists a unique positive operator $S \in B(H)$ such that $T = S^2$.

Proof. The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) 3 are easy.

(iv) $\Rightarrow$ (i). Define $f \in C(\sigma(T))$ given by $f(x) = \sqrt{x}$ for all $x \in \sigma(T) \subset [0, +\infty)$. Write $S = f(T)$. Then $S = S^*$ and $S^2 = T$.

(iii) $\Rightarrow$ (iv). It follows from Lemma 4.4.

For the remaining statement, let $T \in B(H)$ be positive. Define the continuous function $f : [0, +\infty) \to \mathbb{R} : f(x) = \sqrt{x}$. Then, $f(T)$ is positive and $f(T)^2 = T$. When also $S \in B(H)$ is positive and $T = S^2$, we get that $f(T) = f(S^2) = f(g(S))$, where $g : \mathbb{R} \to \mathbb{R} : g(x) = x^2$. Since $\sigma(S) \subset [0, +\infty)$, we get that $f(g(x)) = x$ for all $x \in \sigma(S)$. Using property 4 in Proposition 4.13, it follows that $f(T) = (f \circ g)(S) = S$. □

4.3 Spectral theorem vol. 1 – mutliplication operator form

We will now show that any self-adjoint operator is unitarily equivalent to a multiplication operator on some measure space. To this end we are going to need the following result.

Theorem 4.15 (Riesz representation theorem)**.** *Let K be a compact space and let $\varphi\colon C(K) \to \mathbb{C}$ be a bounded functional. Then there exists a Borel measure μ on K such that $\varphi(f) = \int_X f d\mu$. Moreover, if φ is a positive functional, i.e. $\varphi(f) \geqslant 0$ for any $f \geqslant 0$, then μ is a positive measure.*

We start with a lemma.

Lemma 4.16. *Let $T \in B(H)$ be a self-adjoint operator. Assume that it admits a* cyclic vector v, *i.e. a unit vector such that* $\operatorname{span}\{T^n v : n \in \mathbb{N}\}$ *is dense in H. Then there is a measure μ on $\sigma(T)$ and unitary operator $U : H \to L^2(\sigma(T), \mu)$ such that $UTU^* = M_x$, where $M_x : L^2(\sigma(T), \mu) \to L^2(\sigma(T), \mu)$ is the operator given by $(M_x f)(t) := tf(t)$.*

Proof. By Proposition 4.13, we may consider a functional $\varphi_v : C(\sigma(T)) \to \mathbb{C}$ given by $\varphi_v(f) := \langle f(T)v, v\rangle$. By the Riesz representation theorem 4.15 we get a positive measure μ on $\sigma(T)$ such that $\langle f(T)v, v\rangle = \int_{\sigma(T)} f d\mu$. We may now consider the map $H \ni f(T)v \mapsto f \in L^2(\sigma(T), \mu)$, which is defined on a dense subspace of H and its range is also dense, because continuous functions are dense in the L^2-space. This map is isometric

$$\begin{aligned}\|f\|_2^2 &= \int_{\sigma(T)} |f|^2 d\mu = \langle |f|^2(T)v, v\rangle \\ &= \langle \overline{f}(T) f(T)v, v\rangle = \langle f(T)v, f(T)v\rangle \\ &= \|f(T)v\|^2.\end{aligned}$$

It follows that it extends to a unitary operator $U : H \to L^2(\sigma(T), \mu)$. We will check that the equality $UTU^* = M_x$ holds on the dense subspace formed by continuous functions. Indeed, we have $UTU^* f = UTf(T)v = U(x \cdot f)(T)v = x \cdot f = M_x f$. □

Exercise 1. Let $T = T^* \in M_n$ be an Hermitian matrix. Prove that it admits a cyclic vector iff it has no repeated eigenvalues.

Theorem 4.17. *Let $T \in B(H)$ be a self-adjoint operator. Then there exists a measure space (X, μ), a bounded measurable function $g : X \to \sigma(T)$ and a unitary operator $U : H \to L^2(X, \mu)$ such that $UTU^* = M_g$.*

Proof. Let $v \in H$ be a unit vector. Consider the subspace $H_1 := \overline{\operatorname{span}}\{T^n v : n \in \mathbb{N}\}$. By Lemma 4.16 there is a measure μ_1 on $\sigma(T)$ and a unitary operator $U_1 : H_1 \to L^2(\sigma(T), \mu_1)$. If $H_1 = H$ then we are done. If $H_1 \neq H$, we pick a unit vector in $H_1^\perp$ and repeat the construction; as $T(H_1) \subset H_1$ and T is self-adjoint, we also have $T(H_1^\perp) \subset H_1^\perp$. Since we can always continue the procedure, we will exhaust the whole space H at some point – we get an orthogonal decomposition $H = \bigoplus_{i=1}^n H_i$, where we allow $n = \infty$. We also have the corresponding unitary operators $U_k : H_k \to L^2(\sigma(T), \mu_k)$, which allow us to build a unitary $U : H \to \bigoplus_{i=1}^n L^2(\sigma(T), \mu_i)$, given by $U := \bigoplus_{i=1}^n U_i$. To finish the proof, we have to view the space $\bigoplus_{i=1}^n L^2(\sigma(T), \mu_i)$ as $L^2(X, \mu)$ for some measure space (X, μ). As X we take a disjoint union of n copies of $\sigma(T)$ and the measure μ will be given as disjoint union of the measures μ_i, i.e. the restriction of μ to the i-th copy of $\sigma(T)$ will be given by μ_i. Finally, the function g restricts to the identity function on each of the copies of $\sigma(T)$. □

We will now be able to extend the continuous functional calculus to Borel functional calculus.

4.4 Spectral theorem vol. 2 – Borel functional calculus

Let $K \subset \mathbb{R}$ be a compact subset. We denote by $B_b(K)$ the space of bounded Borel functions $f : K \to \mathbb{C}$ and we equip $B_b(K)$ with the supremum norm $\|\cdot\|_\infty$.

We say that a sequence $f_n \in B_b(K)$ converges *boundedly pointwise* to $f \in B_b(K)$ if $\sup_n \|f_n\|_\infty < \infty$ and $f_n(x) \to f(x)$ for every $x \in K$.

Theorem 4.18. *Let $T \in B(H)$ be self-adjoint. There is a unique unital $*$-homomorphism $\Phi : B_b(\sigma(T)) \to B(H)$ with the following properties.*

(i) $\Phi(f) = f(T)$ *for all* $f \in C(\sigma(T))$.

(ii) $\|\Phi(f)\| \leq \|f\|_\infty$ *for all* $f \in B_b(\sigma(T))$.

(iii) *If* $f_n \in B_b(\sigma(T))$ *and* $f_n \to f$ *boundedly pointwise, then* $\Phi(f_n) \to \Phi(f)$ *strongly, i.e.* $\lim_{n\to\infty} \Phi(f_n)v = \Phi(f)v$ *for any* $v \in H$.

We write $f(T)$ instead of $\Phi(f)$. Then the following properties hold for all $f \in B_b(\sigma(T))$.

1. *If $S \in B(H)$ and $ST = TS$, then $S\,f(T) = f(T)\,S$.*
2. *If $\xi \in H \setminus \{0\}$ is an eigenvector of T with eigenvalue λ, then ξ is also an eigenvector of $f(T)$, with eigenvalue $f(\lambda)$.*
3. *$\sigma(f(T)) \subset \overline{f(\sigma(T))}$.*
4. *When $f : \sigma(T) \to \mathbb{R}$ and $g : \overline{f(\sigma(T))} \to \mathbb{C}$ are bounded Borel functions, we have that $g(f(T)) = (g \circ f)(T)$.*

Proof. The existence of extension follows from Theorem 4.17. Indeed, if $UTU^* = M_g$ and $f : \sigma(T) \to \mathbb{C}$ is a Borel function, we define $f(T) := U^* M_{f\circ g} U$. It is clearly a unital $*$-homomorphism and(i) and (ii) hold. Let us prove (iii). Assume that the sequence $(f_n)_{n\in\mathbb{N}}$ converges boundedly pointwise to f. For any $v \in H$ we have

$$\begin{aligned} \|f(T)v - f_n(T)v\|^2 &= \|U^*(M_{f\circ g} - M_{f_n\circ g})Uv\|^2 \\ &= \int_X |f\circ g - f_n \circ g|^2 \cdot |Uv|^2 d\mu. \end{aligned}$$

Since $|f\circ g - f_n \circ g| \leqslant M$ for some constant $M > 0$, the integrand is dominated by the integrable function $M|Uv|^2$, so Lebesgue's dominated convergence theorem shows that $\lim_{n\to\infty} \|f(T)v - f_n(T)v\| = 0$. The Lemma 4.20 shows that the extension is unique.

To check properties 1. and 2. consider the class of Borel functions for which they hold. By Proposition 4.13 this class contains the continuous functions. It suffices to check that it is closed under bounded pointwise limits and then appeal to Lemma 4.20. But bounded pointwise convergence translates to strong convergence of the corresponding operators, so it is clear. Property 3. follows from the easy observation that if M_g is a mutliplication operator then its spectrum is contained in the closure of the image of g; otherwise you can easily write a formula for the inverse. Since g takes values in $\sigma(T)$, it follows that $\sigma(f(T)) = \sigma(M_{f\circ g}) \subset \overline{f(g(X)} \subset \overline{f(\sigma(T))}$. Now you can verify property 4. exactly the same as properties 1. and 2. □

We can now provide yet another formulation of the spectral theorem. In what follows we denote by $\mathcal{P}(H)$ the set of all orthogonal projections onto closed subspaces of H.

4.5 Spectral theorem vol. 3 – the spectral measure (optional)

Theorem 4.19. *Let $T \in B(H)$ be self-adjoint. Denote by $\mathcal{B}$ the Borel σ-algebra on $\sigma(T)$. There is a unique map $E : \mathcal{B} \to \mathcal{P}(H)$ with the following properties.*

1. $E(\emptyset) = 0$ *and* $E(\sigma(T)) = 1$.
2. $E(\mathcal{U} \cap \mathcal{V}) = E(\mathcal{U})\, E(\mathcal{V})$ *for all* $\mathcal{U}, \mathcal{V} \in \mathcal{B}$.
3. $E(\bigcup_n \mathcal{U}_n) = \sum_n E(\mathcal{U}_n)$ *if the* $\mathcal{U}_n \in \mathcal{B}$ *are disjoint.*
4. $E(\mathcal{U})\, T = T\, E(\mathcal{U})$ *for all* $\mathcal{U} \in \mathcal{B}$.
5. *For all* $\mathcal{U} \in \mathcal{B}$*, the spectrum of* $T|_{E(\mathcal{U})H}$ *is contained in the closure* $\overline{\mathcal{U}}$.

When $\mathcal{U}_n \in \mathcal{B}$ are disjoint, property 2 implies that the projections $E(\mathcal{U}_n)$ have orthogonal ranges, so that $\sum_n E(\mathcal{U}_n)$ is a well defined orthogonal projection, with the series converging in the strong topology.

The map $E : \mathcal{B} \to \mathcal{P}(H)$ in Theorem 4.19 is called the (projection valued) spectral measure of T. When H is finite dimensional, we have

$$E(\mathcal{U}) = \sum_{\lambda \in \mathcal{U} \cap \sigma(T)} E_\lambda \,,$$

where E_λ is the orthogonal projection onto the eigenspace with eigenvalue λ.

Proof. By Theorem 4.18, the map $E(\mathcal{U}) = 1_{\mathcal{U}}(T)$ satisfies properties 1–5.

When $F : \mathcal{B} \to \mathcal{P}(H)$ is another map satisfying properties 1–5, it follows from Theorem 4.18 that $E(\mathcal{U})\, F(\mathcal{V}) = F(\mathcal{V})\, E(\mathcal{U})$ for all $\mathcal{U}, \mathcal{V} \in \mathcal{B}$. To prove the uniqueness of E, it thus suffices to prove that $E = F$ whenever E and F satisfy properties 1–5 and E, F have commuting ranges.

Let $K \subset \sigma(T)$ be closed. We first prove that $E(K) = F(K)$. Let $L \subset \sigma(T) \setminus K$ be closed. We start by proving that $E(L)\, F(K) = 0$. The projections $E(L)$ and $F(K)$ commute, and they commute with T. By property 5, the spectrum of the restriction of T to $E(L)\, F(K)\, H$ is contained in $L \cap K = \emptyset$. So, $E(L)\, F(K) = 0$. Writing $\sigma(T) \setminus K$ as the increasing union of a sequence of closed subsets, it follows that also $E(\sigma(T) \setminus K)\, F(K) = 0$. This means that $(1 - E(K))\, F(K) = 0$ and thus, $F(K) = E(K)\, F(K)$. Since our assumptions on E and F are symmetric, we also have $E(K) = F(K)\, E(K) = E(K)\, F(K)$. So, $E(K) = F(K)$ for all closed subsets $K \subset \sigma(T)$.

Properties 1–3 imply that $\{\mathcal{U} \in \mathcal{B} \mid E(\mathcal{U}) = F(\mathcal{U})\}$ is a σ-algebra. It thus follows that $E(\mathcal{U}) = F(\mathcal{U})$ for all $\mathcal{U} \in \mathcal{B}$. □

4.6 Borel functional calculus strikes again

We will now give an alternative proof of Theorem 4.18, which does not rely on the Riesz representation theorem. We present the necessary prerequisites in Section 4.6. We will be following [Ped, Section 6.1].

Alternative proof of Theorem 4.18. By Lemma 4.20 below, the uniqueness of Φ is immediate.

By Lemma 4.23 below, for all $\xi, \eta \in H$, there is a unique linear map $\omega_{\xi,\eta} : B_b(\sigma(T)) \to \mathbb{C}$ that is continuous under bounded pointwise limits and that satisfies $\omega_{\xi,\eta}(f) = \langle f(T)\xi, \eta\rangle$ for all $f \in C(\sigma(T))$. Given $f \in B_b(\sigma(T))$, it also follows from Lemma 4.23 that $(\xi, \eta) \mapsto \omega_{\xi,\eta}(f)$ is a bounded sesquilinear form. We can therefore uniquely define $\Phi(f) \in B(H)$ satisfying

$$\langle \Phi(f)\xi, \eta\rangle = \omega_{\xi,\eta}(f) \quad \text{for all } \xi, \eta \in H.$$

The uniqueness of the extension of $\omega_{\xi,\eta}$ from $C(\sigma(T))$ to $B_b(\sigma(T))$ implies that $\omega_{\xi,\eta}(\overline{f}) = \overline{\omega_{\eta,\xi}(f)}$ for all $f \in B_b(\sigma(T))$. It follows that $\Phi(\overline{f}) = \Phi(f)^*$.

Let $g \in C(\sigma(T))$. Again, the uniqueness of the extension implies that $\omega_{g(T)\xi,\eta}(f) = \omega_{\xi,\eta}(fg)$ for all $f \in B_b(\sigma(T))$. It follows that $\Phi(fg) = \Phi(f)\Phi(g)$ for all $f \in B_b(\sigma(T))$ and $g \in C(\sigma(T))$. This means that

$$\omega_{\xi,\Phi(f)^*\eta}(g) = \omega_{\xi,\eta}(fg) \quad \text{for all } g \in C(\sigma(T)).$$

It follows that the same equality holds for all $g \in B_b(\sigma(T))$. We conclude that Φ is a $*$-homomorphism.

Since $\|\Phi(f)\xi\|^2 = \langle \Phi(f)^*\Phi(f)\xi, \xi\rangle = \langle \Phi(\overline{f}\, f)\xi, \xi\rangle = \omega_\xi(|f|^2) \leq \|\xi\|^2 \, \|f\|_\infty^2$, we conclude that $\|\Phi(f)\| \leq \|f\|_\infty$ for all $f \in B_b(\sigma(T))$.

The continuity of $\omega_{\xi,\eta}$ implies that $\Phi(f_n) \to \Phi(f)$ weakly whenever $f_n \to f$ boundedly pointwise. Then also,

$$\|\Phi(f_n)\xi - \Phi(f)\xi\|^2 = \langle \Phi(|f_n - f|^2)\xi, \xi\rangle = \omega_\xi(|f_n - f|^2) \to 0 \,.$$

This means that $\Phi(f_n) \to \Phi(f)$ strongly.

Let $S \in B(H)$ such that $ST = TS$. By Proposition 4.13, we have $S\, f(T) = f(T)\, S$ for all $f \in C(\sigma(T))$. It follows that $\omega_{\xi,S^*\eta}(f) = \omega_{S\xi,\eta}(f)$ for all $f \in C(\sigma(T))$ and hence, for all $f \in B_b(\sigma(T))$. We conclude that $S\, f(T) = f(T)\, S$ for all $f \in B_b(\sigma(T))$. We similarly prove that $f(T)\xi = f(\lambda)\,\xi$ whenever $T\xi = \lambda\,\xi$ and $\xi \neq 0$.

When $f \in B_b(\sigma(T))$ and $\lambda \notin \overline{f(\sigma(T))}$, the function $g : x \mapsto (f(x) - \lambda)^{-1}$ is bounded and Borel on $\sigma(T)$. Then, $g(T)$ is the inverse of $f(T) - \lambda 1$, proving that $\lambda \notin \sigma(f(T))$.

Finally, let $f : \sigma(T) \to \mathbb{R}$ be a bounded Borel function and define $K = \overline{f(\sigma(T))}$. When $p \in \mathbb{C}[X]$, we immediately get that $p(f(T)) = (p \circ f)(T)$. Approximating $g \in C(K)$ uniformly by polynomial functions, we conclude that $g(f(T)) = (g \circ f)(T)$ for all $g \in C(K)$. This means that for all ξ, η, the linear maps $C(K) \to \mathbb{C}$ given by

$$g \mapsto \langle g(f(T))\xi, \eta\rangle \quad \text{and} \quad g \mapsto \langle (g \circ f)(T)\xi, \eta\rangle$$

coincide on $C(K)$. By definition, the first one has a unique extension to a linear map on $B_b(K)$ that is continuous under bounded pointwise convergence on K and that is given by

$$B_b(K) \to \mathbb{C} : g \mapsto \langle g(f(T))\xi, \eta\rangle \,.$$

For the second one, denote by $\omega_{\xi,\eta}$ the unique extension of $C(\sigma(T)) \to \mathbb{C} : h \mapsto \langle h(T)\xi, \eta\rangle$ to a linear map $B_b(\sigma(T)) \to \mathbb{C}$ that is continuous under bounded pointwise convergence on $\sigma(T)$. Then also

$$B_b(K) \to \mathbb{C} : g \mapsto \omega_{\xi,\eta}(g \circ f)$$

is continuous under bounded pointwise convergence on K. We conclude that $\omega_{\xi,\eta}(g \circ f) = \langle g(f(T))\xi, \eta\rangle$, so that $(g \circ f)(T) = g(f(T))$. □

Some measure theoretic background

Using the approach of [Ped, Section 6.1], one can avoid using measure theory and the Riesz representation and give a reasonably short self-contained presentation. We reproduce part of [Ped, Section 6.1], adapted to our notations. The following results hold for arbitrary compact second countable spaces K, with the same proof. Even second countability can be removed, but one then has to replace all sequences by nets.

Lemma 4.20. *Let $K \subset \mathbb{R}$ be compact. Then $B_b(K)$ is the smallest vector space of functions from K to $\mathbb{C}$ containing $C(K,\mathbb{R})$ and being closed under bounded pointwise limits.*

Proof. Denote by V the smallest vector space of functions from K to $\mathbb{R}$ containing $C(K,\mathbb{R})$ and being closed under bounded pointwise limits. Then, $C(K) \subset V \subset B_b(K)$.

For every $g \in C(K)$, the set $\{f \in V \mid fg \in V\}$ is a vector space of functions containing $C(K)$ and being closed under bounded pointwise limits. So this set is equal to V. This means that $fg \in V$ for all $f \in V$ and $g \in C(K)$. We then similarly prove that V is an algebra. It follows that $\mathcal{B}_0 = \{\mathcal{U} \subset \mathbb{R} \mid 1_{\mathcal{U}} \in V\}$ is a σ-algebra. When $\mathcal{U} \subset K$ is open, we can write $1_{\mathcal{U}}$ as the limit of an increasing sequence in $C(K)$. It follows that $\mathcal{B}_0$ is the Borel σ-algebra. Since the linear span of $\{1_{\mathcal{U}} \mid \mathcal{U} \subset K \text{ Borel}\}$ is uniformly dense in $B_b(K)$, we get that $V = B_b(K)$. □

Theorem 4.21. *Let $K \subset \mathbb{R}$ be compact. Every $\mathbb{R}$-linear map $\omega : C(K,\mathbb{R}) \to \mathbb{R}$ that is positive, meaning that $\omega(f) \geq 0$ whenever $f \in C(K,[0,+\infty))$, has a unique extension to a $\mathbb{C}$-linear map $\widetilde{\omega} : B_b(K) \to \mathbb{C}$ that is continuous under bounded pointwise convergence. Moreover, $\widetilde{\omega}$ is positive, $\widetilde{\omega}(\overline{f}) = \overline{\widetilde{\omega}(f)}$ for all $f \in B_b(K)$ and $\|\widetilde{\omega}\| = \omega(1)$.*

Proof (taken from [Ped, Section 6.1]). If $f \in C(K,\mathbb{R})$, we have $-\|f\|_\infty 1 \leq f \leq \|f\|_\infty 1$ and conclude that $|\omega(f)| \leq \omega(1)\,\|f\|_\infty$.

Denote by $U(K,\mathbb{R})$ the space of bounded functions $f : K \to \mathbb{R}$ that can be written as the pointwise limit of a bounded increasing sequence in $C(K,\mathbb{R})$. Define the map

$$\omega_u : U(K,\mathbb{R}) \to \mathbb{R} : \omega_u(f) = \sup\{\omega(g) \mid g \in C(K,\mathbb{R}) \, , \; g \leq f\} \; .$$

Note that the following two properties hold by definition. If $g \in C(K,\mathbb{R})$, then $g \in U(K,\mathbb{R})$ and $\omega_u(g) = \omega(g)$. When $f, g \in U(K,\mathbb{R})$ and $f \leq g$, then $\omega_u(f) \leq \omega_u(g)$.

1. Whenever $f_n \in C(K,\mathbb{R})$ is a bounded increasing sequence and $f = \sup_n f_n$, we have $\omega_u(f) = \lim_n \omega(f_n)$.

To prove 1, fix $g \in C(K,\mathbb{R})$ with $g \leq f$. Then $g \wedge f_n$ is an increasing sequence of continuous functions converging pointwise to the continuous function g. Therefore, the convergence is uniform and $\omega(g) = \lim_n \omega(g \wedge f_n) \leq \lim_n \omega(f_n) \leq \omega_u(f)$. Since this holds for all $g \in C(K,\mathbb{R})$ with $g \leq f$, statement 1 follows.

2. If $f, g \in U(K,\mathbb{R})$ and $a, b \geq 0$, we have that $af + bg \in U(K,\mathbb{R})$ and $\omega_u(af + bg) = a\omega_u(f) + b\omega_u(g)$.

This follows immediately by writing f and g as the limit of a bounded increasing sequence of continuous functions and then applying 1.

3. If $f_n \in U(K,\mathbb{R})$ is a bounded increasing sequence and $f = \sup_n f_n$, then also $f \in U(K,\mathbb{R})$ and $\omega_u(f) = \lim_n \omega_u(f_n)$.

Note that $\omega_u(f_n)$ is a bounded increasing sequence, so that $\lim_n \omega_u(f_n)$ is well defined. Choose bounded increasing sequences $(g_{n,k})_k$ in $C(K,\mathbb{R})$ such that $f_n = \sup_k g_{n,k}$. Define $g_n \in C(K,\mathbb{R})$ given by

$$g_n = \bigvee_{m=1}^{n} g_{m,n} \; .$$

Then $(g_n)_n$ is an increasing sequence with $f = \sup_n g_n$. So, $f \in U(K,\mathbb{R})$. Also, $g_n \leq f_n \leq f$. By 1, we get that

$$\omega_u(f) = \lim_n \omega_u(g_n) \leq \lim_n \omega_u(f_n) \leq \omega_u(f) \; ,$$

ending the proof of 3.

We now similarly define $L(K,\mathbb{R}) = -U(K,\mathbb{R})$ and $\omega_l : L(K,\mathbb{R}) \to \mathbb{R} : \omega_l(f) = -\omega_u(-f)$. Then, $L(K,\mathbb{R})$ and ω_l satisfy 1, 2 and 3, replacing increasing sequences by decreasing sequences.

We extend ω_u and ω_l to arbitrary bounded functions $f : K \to \mathbb{R}$ by

$$\omega_u(f) = \inf\{\omega_u(g) \mid g \in U(K,\mathbb{R}), f \leq g\} \; ,$$
$$\omega_l(f) = \sup\{\omega_l(g) \mid g \in L(K,\mathbb{R}), g \leq f\} \; .$$

4. For every bounded function $f : K \to \mathbb{R}$, we have that $\omega_l(f) \leq \omega_u(f)$.

Let $g \in L(K,\mathbb{R})$, $h \in U(K,\mathbb{R})$ and $g \leq f \leq h$. We have to prove that $\omega_l(g) \leq \omega_u(h)$. Since $0 \leq h - g = h + (-g)$, it follows from 2 that $h - g \in U(K,\mathbb{R})$ and

$$0 \leq \omega_u(h-g) = \omega_u(h) + \omega_u(-g) = \omega_u(h) - \omega_l(g) \; ,$$

concluding the proof of 4.

Denote by $I(K,\mathbb{R})$ the set of bounded functions $f : K \to \mathbb{R}$ with the property that $\omega_u(f) = \omega_l(f)$. We denote this common value as $\widetilde{\omega}(f)$. By 4, we get that $I(K,\mathbb{R})$ consists of all bounded functions $f : K \to \mathbb{R}$ with the following property: for every $\varepsilon > 0$, there exist $g \in L(K,\mathbb{R})$ and $h \in U(K,\mathbb{R})$ such that $g \leq f \leq h$ and $\omega_u(h) - \omega_l(g) < \varepsilon$. Combining this characterization with 2, it follows that $I(K,\mathbb{R})$ is a real vector space and that $\widetilde{\omega} : I(K,\mathbb{R}) \to \mathbb{R}$ is an $\mathbb{R}$-linear map. Similarly, $I(K,\mathbb{R})$ is closed under $\vee$ and $\wedge$. When $f : K \to \mathbb{R}$ is a bounded positive function, we have $\omega_u(f) \geq 0$. Therefore, $\widetilde{\omega} : I(K,\mathbb{R}) \to \mathbb{R}$ is positive as well.

5. For every $f \in U(K,\mathbb{R})$, we have that $f \in I(K,\mathbb{R})$ and $\widetilde{\omega}(f) = \omega_u(f)$. Similarly, for every $f \in L(K,\mathbb{R})$, we have that $f \in I(K,\mathbb{R})$ and $\widetilde{\omega}(f) = \omega_l(f)$.

Given $f \in U(K,\mathbb{R})$ and $\varepsilon > 0$, we can find $g \in C(K,\mathbb{R})$ with $g \leq f$ and $\omega_u(f) - \omega(g) < \varepsilon$. Then $g \in L(K,\mathbb{R})$ and $\omega_l(g) = \omega(g)$. So, $f \in I(K,\mathbb{R})$ and $\widetilde{\omega}(f) = \omega_u(f)$. The second statement holds by symmetry, so that 5 is proven.

6. If $f_n \in I(K,\mathbb{R})$ is a bounded increasing sequence and $f = \sup_n f_n$, then also $f \in I(K,\mathbb{R})$ and $\widetilde{\omega}(f) = \lim_n \widetilde{\omega}(f_n)$.

By the positivity of $\widetilde{\omega}$, the sequence $\widetilde{\omega}(f_n)$ is bounded and increasing, so that its limit is well defined. Choose $\varepsilon > 0$ arbitrarily. Denote $g_0 = f_0$ and $g_n = f_n - f_{n-1}$ for all $n \geq 1$. Since $g_n \in I(K,\mathbb{R})$, we can choose $k_n \in U(K,\mathbb{R})$ such that $g_n \leq k_n$ and $\omega_u(k_n) \leq \widetilde{\omega}(g_n) + 2^{-n}\varepsilon$. Write $M = \|f\|_\infty$ and define

$$k := M \wedge \big(\sum_{n=0}^{\infty} k_n\big) = \sup_n \big(M \wedge \sum_{m=0}^{n} k_m\big) \; .$$

By 3, we get that $k \in U(K,\mathbb{R})$. By construction, $f \leq k$. Also by 3 and by the linearity of $\widetilde{\omega}$, we have that

$$\omega_u(k) \leq \sum_{n=0}^{\infty} \omega_u(k_n) \leq 2\varepsilon + \sum_{n=0}^{\infty} \widetilde{\omega}(g_n) = 2\varepsilon + \sup_n \widetilde{\omega}\Big(\sum_{m=0}^{n} g_m\Big) = 2\varepsilon + \sup_n \widetilde{\omega}(f_n) \ .$$

For every n, we have

$$\widetilde{\omega}(f_n) = \omega_l(f_n) \leq \omega_l(f) \leq \omega_u(f) \leq \omega_u(k) \leq 2\varepsilon + \sup_n \widetilde{\omega}(f_n) \ .$$

Taking the supremum over n, we conclude that

$$\sup_n \widetilde{\omega}(f_n) \leq \omega_l(f) \leq \omega_u(f) \leq 2\varepsilon + \sup_n \widetilde{\omega}(f_n) \ .$$

Since this holds for all $\varepsilon > 0$, we get that $f \in I(K,\mathbb{R})$ and that $\widetilde{\omega}(f) = \sup_n \widetilde{\omega}(f_n)$. This concludes the proof of 6.

By symmetry, 6 also holds for bounded decreasing sequences and their infimum.

7. If $f_n \in I(K,\mathbb{R})$ is a bounded sequence and $f = \liminf_n f_n$, then also $f \in I(K,\mathbb{R})$ and $\widetilde{\omega}(f) \leq \liminf_n \widetilde{\omega}(f_n)$.

Define $h_n = \inf_{m \geq n} f_m = \lim_m (f_n \wedge f_{n+1} \wedge \cdots \wedge f_m)$. The decreasing variant of 6 implies that $h_n \in I(K,\mathbb{R})$. Since $f = \sup_n h_n$, it follows from 6 that $f \in I(K,\mathbb{R})$ and that $\widetilde{\omega}(f) = \sup_n \widetilde{\omega}(h_n)$. We have $\widetilde{\omega}(h_n) \leq \widetilde{\omega}(f_m)$ for all $m \geq n$ and conclude that 7 holds.

8. If $f_n \in I(K,\mathbb{R})$ is a bounded sequence converging pointwise to f, then also $f \in I(K,\mathbb{R})$ and the sequence $\widetilde{\omega}(f_n)$ converges to $\widetilde{\omega}(f)$.

Applying 7 to the sequences f_n and $-f_n$, we find that $f \in I(K,\mathbb{R})$ and

$$\limsup_n \widetilde{\omega}(f_n) \leq \widetilde{\omega}(f) \leq \liminf_n \widetilde{\omega}(f_n) \ .$$

So, 8 holds.

We have proven that $I(K,\mathbb{R})$ is a real vector space of bounded functions on K that is closed under bounded pointwise convergence and that contains $C(K,\mathbb{R})$. By Lemma 4.20, we get that $B_b(K,\mathbb{R}) \subset I(K,\mathbb{R})$. The restriction of $\widetilde{\omega}$ to $B_b(K,\mathbb{R})$ is the (necessarily unique) extension of ω to an $\mathbb{R}$-linear map $B_b(K,\mathbb{R}) \to \mathbb{R}$ that is continuous under bounded pointwise convergence.

For $f \in B_b(K)$, we finally define $\widetilde{\omega}(f) = \widetilde{\omega}(\operatorname{Re} f) + i\widetilde{\omega}(\operatorname{Im} f)$. We get that $\widetilde{\omega} : B_b(K) \to \mathbb{C}$ is linear, positive, continuous under bounded pointwise convergence and that $\widetilde{\omega}(\overline{f}) = \overline{\widetilde{\omega}(f)}$ for all $f \in B_b(K)$.

Finally, whenever $f \in B_b(K)$, take $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ such that $|\widetilde{\omega}(f)| = \alpha\,\widetilde{\omega}(f)$. It follows that

$$|\widetilde{\omega}(f)| = \operatorname{Re}(\alpha\,\widetilde{\omega}(f)) = \operatorname{Re}\widetilde{\omega}(\alpha f) = \widetilde{\omega}(\operatorname{Re}(\alpha f)) \leq \omega(1)\,\|\operatorname{Re}(\alpha f)\| \leq \omega(1)\,\|f\|_\infty \ .$$

□

Remark 4.22. Theorem 4.21 provides the following construction for the Lebesgue measure on the interval [0, 1]. Define $\omega : C([0,1],\mathbb{R}) \to \mathbb{R}$ using the Riemann integral

$$\omega(f) = \int_0^1 f(x)\,dx \;.$$

Then extend to $\widetilde{\omega} : B_b([0,1]) \to \mathbb{C}$. The Lebesgue measure $\lambda(\mathcal{U})$ of a Borel set $\mathcal{U} \subset [0,1]$ is now given by

$$\lambda(\mathcal{U}) = \widetilde{\omega}(1_{\mathcal{U}}) \;.$$

Lemma 4.23. *Let $T \in B(H)$ be self-adjoint. For every $\xi, \eta \in H$, the functional $\omega_{\xi,\eta} : C(K) \to \mathbb{C} : \omega_{\xi,\eta}(f) = \langle f(T)\xi, \eta\rangle$ has a unique extension to a functional $\omega_{\xi,\eta} : B_b(K) \to \mathbb{C}$ that is continuous under bounded pointwise convergence. Moreover, $\|\omega_{\xi,\eta}\| \leq 4\,\|\xi\|\,\|\eta\|$. Also, for every $f \in B_b(K)$, the map $(\xi,\eta) \mapsto \omega_{\xi,\eta}(f)$ is sesquilinear.*

One can actually prove that the norm of $\|\omega_{\xi,\eta}\|$ is bounded by $\|\xi\|\,\|\eta\|$, but we do not need this in our approach to spectral theory.

Proof. For every $\xi \in H$, define $\omega_\xi : C(K) \to \mathbb{C} : \omega_\xi(f) = \langle f(T)\xi, \xi\rangle$. Note that ω_ξ is a positive linear map. By Theorem 4.21, ω_ξ has a unique extension to a positive linear map $\omega_\xi : B_b(K) \to \mathbb{C}$ that is continuous under bounded pointwise convergence and that satisfies $\|\omega_\xi\| = \|\xi\|^2$.

Given $\xi, \eta \in H$, define by polarization

$$\omega_{\xi,\eta} : B_b(K) \to \mathbb{C} : \omega_{\xi,\eta} = \frac{1}{4}\sum_{k=0}^{3} i^k\,\omega_{\xi+i^k\eta} \;. \tag{4.2}$$

Then $\omega_{\xi,\eta}$ is continuous under bounded pointwise convergence and $\omega_{\xi,\eta}(f) = \langle f(T)\xi,\eta\rangle$ for all $f \in C(K)$. This proves the existence of the extension. Its uniqueness follows from Lemma 4.20.

The uniqueness of the extension implies that for every $f \in B_b(K)$, the map $(\xi,\eta) \mapsto \omega_{\xi,\eta}(f)$ is sesquilinear. To finally prove that $\|\omega_{\xi,\eta}\| \leq 4\,\|\xi\|\,\|\eta\|$, we may thus replace ξ by $t\,\xi$ and η by $t^{-1}\,\eta$, so that we can assume that $\|\xi\| = \|\eta\|$. Then,

$$\|\omega_{\xi+i^k\eta}\| = \|\xi + i^k\eta\|^2 \leq (\|\xi\| + \|\eta\|)^2 = 4\,\|\xi\|^2 = 4\,\|\xi\|\,\|\eta\| \;.$$

It then follows from (4.2) that $\|\omega_{\xi,\eta}\| \leq 4\,\|\xi\|\,\|\eta\|$. □

Lecture 5

The Hahn-Banach extension theorem

We discuss in this and the following lectures Stefan Banach's classical theorems on, what we nowadays call, Banach spaces.

The three main theorems are the following.

- The Hahn-Banach theorem. It allows to extend a functional defined on a vector subspace of a Banach space to the whole space in a norm-preserving way. Convince yourself that such a thing is nontrivial. Indeed, try (and probably fail) to prove yourself that every Banach space different from $\{0\}$ admits a nonzero continuous functional.
- The open mapping theorem and the closed graph theorem. These theorems deduce continuity of linear maps out of seemingly weaker assumptions.
- The uniform boundedness principle. This theorem allows in certain cases to prove that a family of linear maps is uniformly bounded once it is pointwise bounded.

The Hahn-Banach theorem deals with arbitrary seminormed spaces and the proof is a quite easy application of Zorn's lemma. The two other theorems are very specific for Banach spaces and use the remarkable Baire category theorem. That theorem says that in a complete metric space, the intersection of countably many open dense subsets is still dense.

Important convention. As long as the opposite is not explicitly stated, all vector spaces are over the field $\mathbb{C}$ of complex numbers. Nevertheless, all vector spaces over $\mathbb{C}$ can be regarded as vector spaces over $\mathbb{R}$ and we sometimes exploit this.

5.1 Hahn-Banach extension theorem – a first version

Suppose that X is a vector space over the field $\mathbb{R}$ and that $Y \subset X$ a vector subspace. Assume that $\omega : Y \to \mathbb{R}$ is a linear functional. Can you extend ω to a linear functional from X to $\mathbb{R}$? The answer is yes and the proof is easy, but nonconstructive: choose a vector space basis for Y and extend this basis for Y to a basis for the whole of X; extend ω to X by imposing that ω equals 0 on the new basis vectors.

Now assume that X is a normed vector space over $\mathbb{R}$ and that $Y \subset X$ is a vector subspace. Assume that $\omega : Y \to \mathbb{R}$ is a *bounded* linear functional. Can you extend ω to a bounded linear functional from X to $\mathbb{R}$? Without increasing the norm of ω ? The above proof no longer works because a vector space basis for X has no relation at all with the norm on X. But the same idea still works: we add one by one new basis vectors, extend ω at every step and are careful enough to ensure that the norm never increases. The precise proof goes as follows.

Theorem 5.1. *Let X be a normed vector space over the field $\mathbb{R}$. Let $Y \subset X$ be a vector subspace and assume that $\omega : Y \to \mathbb{R}$ is a bounded linear map. Then ω can be extended to a bounded linear map $\widetilde{\omega} : X \to \mathbb{R}$ satisfying $\|\widetilde{\omega}\| = \|\omega\|$.*

We provide the following sketch of proof where you have to fill in the details yourself. A more general result will be stated below as Theorem 5.6, for which you can find a detailed proof in [Con, III.6.9 and paragraph after 6.9].

Proof. Let $x_0 \in X \setminus Y$. Use the following steps to prove that ω can be extended to a bounded linear map $\widetilde{\omega} : Y + \mathbb{R}x_0 \to \mathbb{R}$ without increasing the norm.

1. Replacing ω by a multiple, we may assume that $\|\omega\| = 1$.
2. Realize that we need to prove the existence of a real number $\alpha \in \mathbb{R}$ such that the linear map
$$\widetilde{\omega} : Y + \mathbb{R}x_0 \to \mathbb{R} : \widetilde{\omega}(y + tx_0) = \omega(y) + t\alpha \quad \text{for all} \quad y \in Y, t \in \mathbb{R}$$
satisfies $\|\widetilde{\omega}\| = 1$.
3. Note that to prove that $\|\widetilde{\omega}\| = 1$, it suffices to prove that $\widetilde{\omega}(y + tx_0) \leq \|y + tx_0\|$ for all $y \in Y, t \in \mathbb{R}$. So we need to find $\alpha \in \mathbb{R}$ such that
$$\omega(y) + t\alpha \leq \|y + tx_0\| \quad \text{for all} \quad y \in Y, t \in \mathbb{R} \ .$$
4. Consider separately the cases $t = 0, t > 0, t < 0$ and observe that it suffices to find an $\alpha \in \mathbb{R}$ such that
$$\begin{aligned} \omega(y) + \alpha \leq \|y + x_0\| \quad &\text{for all} \quad y \in Y \quad \text{and} \\ \omega(y) - \alpha \leq \|y - x_0\| \quad &\text{for all} \quad y \in Y \ . \end{aligned} \tag{5.1}$$
5. Prove that one can find an $\alpha \in \mathbb{R}$ satisfying (5.1) provided that
$$\omega(y) - \|y - x_0\| \leq -\omega(z) + \|z + x_0\| \quad \text{for all} \quad y, z \in Y \ .$$
6. The above inequality can be rewritten as
$$\omega(y + z) \leq \|y - x_0\| + \|z + x_0\| \quad \text{for all} \quad y, z \in Y \ .$$
Deduce this new inequality from the fact that $\|\omega\| = 1$.

So far we have proven that we can always extend a bounded functional to a vector space of one dimension higher without increasing the norm. We now return to our initial problem. Intuitively we add one by one more and more vectors to Y and extend ω at every step without increasing the norm. But how can we make sure that the procedure stops? For this we again need Zorn's lemma.

Denote by $\mathcal{I}$ the "set of extensions" of ω. More precisely $\mathcal{I}$ is defined as the set of pairs (Y_1, ω_1) where $Y_1 \subset X$ is a vector subspace containing Y and $\omega_1 : Y_1 \to \mathbb{R}$ is an $\mathbb{R}$-linear map that extends ω and satisfies $\|\omega_1\| = \|\omega\|$. Define the partial order $\leq$ on $\mathcal{I}$ by declaring $(Y_1, \omega_1) \leq (Y_2, \omega_2)$ if $Y_1 \subset Y_2$ and if ω_2 is an extension of ω_1. Check yourself that $(\mathcal{I}, \leq)$ satisfies the assumptions of Zorn's lemma.

By Zorn's lemma, take a maximal element $(\widetilde{Y}, \widetilde{\omega})$ in $\mathcal{I}$. It suffices to prove that $\widetilde{Y} = X$ because we then have found the required extension of ω. If $\widetilde{Y} \neq X$, we can find $x_0 \in X \setminus \widetilde{Y}$. The first half of the proof then allows to extend $\widetilde{\omega}$ to $\widetilde{Y} + \mathbb{R}x_0$ without increasing its norm. This would contradict the maximality of $(\widetilde{Y}, \widetilde{\omega})$. So, $\widetilde{Y} = X$. □

Theorem 5.2. *The same theorem as 5.1 holds over the field $\mathbb{C}$.*

Proof. Let X be a normed vector space over the field $\mathbb{C}$. Let $Y \subset X$ be a vector subspace and assume that $\omega : Y \to \mathbb{C}$ is a bounded linear map. View X as a normed space over $\mathbb{R}$. Then $\operatorname{Re}\omega : Y \to \mathbb{R}$ is a bounded $\mathbb{R}$-linear map with $\|\operatorname{Re}\omega\| \leq \|\omega\|$ (actually equality holds). By Theorem 5.1 we can choose an $\mathbb{R}$-linear extension $\mu : X \to \mathbb{R}$ of $\operatorname{Re}\omega$ to the whole of X satisfying $\|\mu\| = \|\operatorname{Re}\omega\|$. Define $\widetilde{\omega} : X \to \mathbb{C} : \widetilde{\omega}(x) = \mu(x) - i\mu(ix)$. Check yourself that $\widetilde{\omega}$ is $\mathbb{C}$-linear and that $\widetilde{\omega}(y) = \omega(y)$ for all $y \in Y$.

It remains to prove that $\|\widetilde{\omega}\| \leq \|\omega\|$. Choose $x \in X$ with $\|x\| \leq 1$. It suffices to prove that $|\widetilde{\omega}(x)| \leq \|\mu\|$. Take $\lambda \in \mathbb{C}$ with $|\lambda| = 1$ such that $|\widetilde{\omega}(x)| = \lambda\widetilde{\omega}(x)$. Since $|\widetilde{\omega}(x)|$ is a real number, it equals its real part and we find that

$$|\widetilde{\omega}(x)| = \operatorname{Re}|\widetilde{\omega}(x)| = \operatorname{Re}(\lambda\widetilde{\omega}(x)) = \operatorname{Re}\widetilde{\omega}(\lambda x) = \mu(\lambda x) \leq \|\mu\|\,\|\lambda x\| \leq \|\mu\| \,.$$

□

Corollary 5.3. *Let X be a normed space. If $x \in X \setminus \{0\}$, there exists $\omega \in X^*$ with $\|\omega\| = 1$ and $\omega(x) = \|x\|$. In more abstract language: denoting the dual of X^* by $X^{**} = (X^*)^*$, the map*

$$i : X \to X^{**} : i(x)(\omega) = \omega(x)$$

is an isometry.

Proof. Put $Y = \mathbb{C}x$ and define $\omega : Y \to \mathbb{C}$ given by $\omega(\lambda x) = \lambda\|x\|$ for all $\lambda \in \mathbb{C}$. Check that $\|\omega\| = 1$. Extend ω to the whole of X using Theorem 5.2. □

If X is a normed space and $Y \subset X$ a vector subspace, we define for $x \in X$,

$$d(x, Y) = \inf\{\|x - y\| \mid y \in Y\} \,.$$

Observe that $d(x, Y) = 0$ if and only if x belongs to the closure of Y.

We prove the following last corollary to the Hahn-Banach extension theorem.

Corollary 5.4. *Let X be a normed space and $Y \subset X$ a vector subspace. If $x \in X$ and $d(x, Y) = 1$, there exists a continuous functional $\omega \in X^*$ satisfying $\omega(x) = 1$, $\omega(y) = 0$ for all $y \in Y$ and $\|\omega\| = 1$.*

Proof. Define $\omega : Y + \mathbb{C}x \to \mathbb{C} : \omega(y + \lambda x) = \lambda$ for all $y \in Y, \lambda \in \mathbb{C}$. We only need to prove that $\|\omega\| = 1$, because then Theorem 5.2 provides an extension of ω to the whole of X without increasing its norm. So we have to prove that

$$|\lambda| \leq \|y + \lambda x\| \quad \text{for all} \quad y \in Y, \lambda \in \mathbb{C} \; .$$

Divide both sides of the inequality by $|\lambda|$ and deduce the inequality from the fact that $d(x, Y) \geq 1$. □

5.2 Hahn-Banach extension theorem – a second version

In Theorem 5.1 we extended linear functionals from subspaces of a normed space to the whole normed space without increasing the norm. In later lectures it will be useful to work with more general "pseudo-quasi-almost-norms" that we call sublinear maps.

Definition 5.5. Let X be a vector space over the field $\mathbb{R}$. A map $f : X \to \mathbb{R}$ is called *sublinear* if

- $f(x + y) \leq f(x) + f(y)$ for all $x, y \in X$,
- $f(tx) = tf(x)$ for all $x \in X$ and $t \geq 0$.

Observe that we do not assume that f takes its values in the *positive* real numbers and that we do not say anything about $f(tx)$ when $t < 0$. So, a sublinear map is more general than a seminorm and certainly much more general than a norm.

Theorem 5.6 (The Hahn-Banach extension theorem)**.** *Let X be a vector space over $\mathbb{R}$ and $f : X \to \mathbb{R}$ a sublinear map. If $Y \subset X$ is a vector subspace and $\omega : Y \to \mathbb{R}$ is an $\mathbb{R}$-linear map satisfying $\omega(x) \leq f(x)$ for all $x \in Y$, then ω can be extended to an $\mathbb{R}$-linear map $\widetilde{\omega} : X \to \mathbb{R}$ that satisfies $\widetilde{\omega}(x) \leq f(x)$ for all $x \in X$.*

Exercise 1. Prove yourself Theorem 5.6 following the same method as in the proof of Theorem 5.1. Details can be found in [Con, III.6.9 and paragraph after 6.9].

In the same way as Theorem 5.2 is a complex version of the real Theorem 5.1, we have the following result.

Corollary 5.7. *Let X be a vector space and $p : X \to [0, +\infty)$ a seminorm. If $Y \subset X$ is a vector subspace and $\omega : Y \to \mathbb{C}$ a linear map satisfying $|\omega(y)| \leq p(y)$ for all $y \in Y$, there exists a linear map $\widetilde{\omega} : X \to \mathbb{C}$ extending ω and satisfying $|\widetilde{\omega}(x)| \leq p(x)$ for all $x \in X$.*

Exercise 2. Prove yourself Corollary 5.7 in exactly the same way as we have proven Theorem 5.2.

5.3 Illustration: Banach limits

As an illustration of the Hahn-Banach theorem, we discuss the following: although a bounded sequence in $\mathbb{C}$ is not necessarily convergent, there exist (noncanonical) ways to attach a kind of limit to an arbitrary bounded sequence in $\mathbb{C}$. In order to interpret the following theorem, you have to realize that a bounded sequence in $\mathbb{C}$ is the same thing as an element of $\ell^\infty(\mathbb{N})$.

Theorem 5.8. *There exists a linear map $L : \ell^\infty(\mathbb{N}) \to \mathbb{C}$, called a* Banach limit*, satisfying the following properties.*

(i) $L(x) = \lim_{n\to\infty} x(n)$ *if this limit exists.*

(ii) $L(y) = L(x)$ *whenever* $y(n) = x(n+1)$ *for all* $n \geq 0$.

(iii) *If* $x(n) \geq 0$ *for all* n, *then* $L(x) \geq 0$.

(iv) $\|L\| = 1$.

It is crucial to keep in mind that a Banach limit L is by no means unique or canonical. To get some feeling for its properties, first make the following exercise.

Exercise 3. Let L be a Banach limit on $\ell^\infty(\mathbb{N})$.

1. Prove that $L(0, 1, 0, 1, 0, 1, \ldots) = \frac{1}{2}$.
2. Prove that there exist $x, y \in \ell^\infty(\mathbb{N})$ such that $L(xy) \neq L(x)L(y)$. So we can never arrange for a Banach limit to have the property that the limit of the product of two sequences is the product of the limits.
3. Prove that $L(xy) = L(x)L(y)$ if at least one of the sequences $(x(n))$ or $(y(n))$ is convergent.

Proof of Theorem 5.8. Complete the following sketch to give yourself a proof of Theorem 5.8. Details can be found in [Con, III.7].

Put $X = \ell^\infty(\mathbb{N}, \mathbb{R})$ (real-valued bounded sequences) equipped with the supremum norm $\|\cdot\|_\infty$. For a sequence $x \in \ell^\infty(\mathbb{N})$ we define its sequence of *Cesàro means* by $C_n x := \frac{x_1 + \cdots + x_n}{n}$. Consider the following subspace

$$Y := \{x \in \ell^\infty(\mathbb{N}, \mathbb{R}) : C_n x \text{ converges}\}.$$

On Y we can define the following functional $\widetilde{L}\colon Y \to \mathbb{R}$ given by $\widetilde{L}(x) := \lim_{n\to\infty} C_n x$. We clearly have $\|\widetilde{L}\| = 1$, so by the Hahn-Banach theorem (Theorem 5.1) we get an extension $\overline{L}\colon \ell^\infty(\mathbb{N}, \mathbb{R}) \to \mathbb{R}$. We can extend it uniquely to a complex linear functional $L\colon \ell^\infty(\mathbb{N}) \to \mathbb{C}$. We just need to verify that L is a Banach limit.

We get (i) because if the sequence (x_n) converges then also its sequence of Cesàro averages converges and the limits agree. If we take $x \in \ell^\infty(\mathbb{N})$ and $y \in \ell^\infty(\mathbb{N})$ is defined by $y_n := x_{n+1}$ then $C_n(x - y) = \frac{x_1 - x_{n+1}}{n} \to 0$, so $L(y) = L(x)$, which establishes (ii). If $x \geqslant 0$ and $\|x\| = c$ then $c1 - x \geqslant 0$, hence $\|c1 - x\| \leqslant c$. We therefore get[1] $c - L(x) = L(c1 - x) \leqslant c$, i.e. $L(x) \geqslant 0$, thus we get (iii). The last part (iv) is immediate. □

5.4 Exercises

Exercise 4. A Banach space is called *reflexive* if the isometric map $i : X \to X^{**}$ in Corollary 5.3 is an isomorphism. So, X is reflexive if and only if $X^{**} = i(X)$.

[1] We can write this inequality because x is real, hence $L(x) = \overline{L}(x)$, so we are dealing with real numbers.

1. Prove that the Banach space $\ell^p(\mathbb{N})$ with $1 < p < \infty$ is reflexive. A way of doing so, goes as follows. Take $1 < q < +\infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Theorem 0.14 yields an isomorphism $\pi_1 : \ell^p(\mathbb{N}) \to \ell^q(\mathbb{N})^*$ and an isomorphism $\pi_2 : \ell^q(\mathbb{N}) \to \ell^p(\mathbb{N})^*$. Dualizing π_2, you get an isomorphism $\pi_2^* : \ell^p(\mathbb{N})^{**} \to \ell^q(\mathbb{N})^*$. Prove that $(\pi_2^*)^{-1} \circ \pi_1 : \ell^p(\mathbb{N}) \to \ell^p(\mathbb{N})^{**}$ coincides with the embedding i.

2. Use the same technique as in 1 to prove that every Hilbert space is a reflexive Banach space.

3. Prove that $\ell^1(\mathbb{N})$ is not reflexive.

Lecture 6

Baire category, open mapping, closed graph, uniform boundedness

6.1 The Baire category theorem

The Baire category theorem says that in a complete metric space, the intersection of *countably many* open dense subsets, is still dense. Before proving this result and in order to get some intuition, you should think about the following facts.

- It is crucial to look at open sets: give an example of two dense subsets of $\mathbb{R}$ that have an empty intersection.
- Prove that a subset K of a metric space (X, d) is dense if and only if K intersects nontrivially every nonempty open subset.
- The intersection of two (and hence of finitely many) open dense subsets, is dense (and, of course, open). Prove this.
- It is crucial to look at only *countably many* open dense subsets: give an uncountable family of open dense subsets of $\mathbb{R}$ that have an empty intersection.

Theorem 6.1 (Baire category theorem)**.** *Let (X, d) be a complete metric space. If $(\mathcal{U}_n)$ is a sequence of open dense subsets of X, the intersection $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ is still dense.*

The strange name *category theorem* has the following origin: a subset of a topological space is said to be of the *first category* if it can be written as the union of countably many closed subsets with empty interior. It is said to be of the *second category* if it is not of the first category. An equivalent formulation of the Baire category theorem then says that in a complete metric space every subset of the first category has empty interior.

Proof of Theorem 6.1. Use the following hints to give yourself a proof of Theorem 6.1. Details can be found in [Ped, 2.2.2].

Put $K := \bigcap_{n\in\mathbb{N}} \mathcal{U}_n$. We have to prove that K is dense in X. So choose a nonempty open subset $\mathcal{V} \subset X$. We have to prove that $\mathcal{V} \cap K$ is nonempty. We use the notations

$$B(x, \varepsilon) := \{y \in X \mid d(y, x) < \varepsilon\} \quad \text{and} \quad \overline{B(x, \varepsilon)} := \{y \in X \mid d(y, x) \leq \varepsilon\} .$$

1. Note that $\mathcal{V} \cap \mathcal{U}_0$ is open and nonempty. So, choose $x_0 \in X$ and $\varepsilon_0 > 0$ such that
$$\overline{B(x_0, \varepsilon_0)} \subset \mathcal{V} \cap \mathcal{U}_0 \; .$$
Make sure that $0 < \varepsilon_0 < 1$.

2. Since $B(x_0, \varepsilon_0) \cap \mathcal{U}_1$ is open and nonempty, similarly choose $x_1 \in X$ and $0 < \varepsilon_1 < 1/2$ such that
$$\overline{B(x_1, \varepsilon_1)} \subset B(x_0, \varepsilon_0) \cap \mathcal{U}_1 \; .$$

3. Continue by choosing inductively $x_n \in X$ and $0 < \varepsilon_n < 1/(n+1)$ satisfying
$$\overline{B(x_n, \varepsilon_n)} \subset B(x_{n-1}, \varepsilon_{n-1}) \cap \mathcal{U}_n \; .$$

4. Prove that $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in X. Since (X, d) is complete, $x_n \to x$ for some $x \in X$. Prove that $x \in \mathcal{V} \cap K$.

□

Corollary 6.2. *Let (X, d) be a complete metric space and (K_n) a sequence of closed subsets of X with empty interior. Then the union $\bigcup_{n \in \mathbb{N}} K_n$ still has empty interior.*

Proof. A subset of a metric space has empty interior if and only if its complement is dense. So Corollary 6.2 is tautologically equivalent with Theorem 6.1 by taking complements. □

6.2 Illustration: there are many continuous functions that are nowhere differentiable

Maybe you already had the occasion to construct by hand a continuous function $f : [0,1] \to \mathbb{R}$ that is nowhere differentiable. Then you certainly remember that such a construction is nontrivial. We will see now that an application of the Baire category theorem yields rather easily that the continuous nowhere differentiable functions are *uniformly dense* in the space of all continuous functions.

Define the complete metric space
$$X = \{ f : [0,1] \to \mathbb{R} \mid f \text{ is continuous} \} \quad \text{with} \quad d(f,g) = \|f - g\|_\infty \; .$$

Define the subsets
$$K_n = \{ f \in X \mid \text{There exists } x \in [0,1] \text{ such that } |f(x) - f(y)| \leq n|x - y| \text{ for all } y \in [0,1] \} \; .$$

We prove two statements.

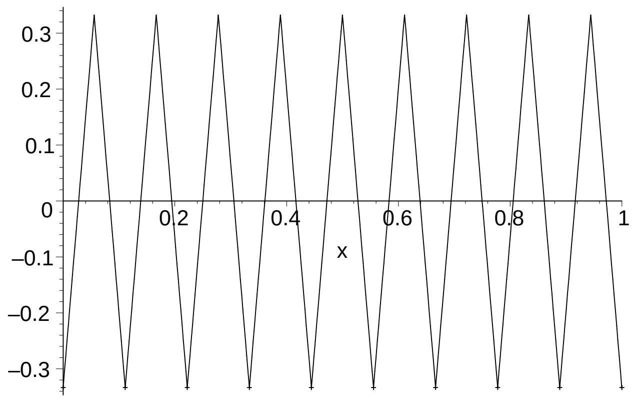
1. If $f : [0,1] \to \mathbb{R}$ is continuous on $[0,1]$ and differentiable in x, there exists $n \in \mathbb{N}$ such that $f \in K_n$. In other words, if f belongs to the complement of $\bigcup_{n \in \mathbb{N}} K_n$, then f is nowhere differentiable.

2. The subsets $K_n \subset X$ are closed and have empty interior.

Both statements together and Corollary 6.2 yield that the nowhere differentiable continuous functions are dense in X.

Exercise 1. Prove the first statement and prove that K_n is closed. *Hint:* use the compactness of $[0,1]$.

We now prove that K_n has empty interior. We define for every $n \geq 1$ a continuous function $h_n : [0,1] \to \mathbb{R}$ with the following properties: $\|h_n\|_\infty = \frac{1}{n}$ and $h_n \notin K_k$ if $k < 4n$. The graph of h_3 looks like



A formal definition of h_n can be given as

$$h_n : [0,1] \to \mathbb{R} : h_n(x) = \begin{cases} 4nx - \frac{4k-3}{n} & \text{if } x \in \left[\frac{k-1}{n^2}, \frac{2k-1}{2n^2}\right] \text{ and } k = 1, \dots, n^2 , \\ -4nx + \frac{4k-1}{n} & \text{if } x \in \left[\frac{2k-1}{2n^2}, \frac{k}{n^2}\right] \text{ and } k = 1, \dots, n^2 . \end{cases}$$

In order to prove that K_m has empty interior, we have to prove the following: given $m \in \mathbb{N}_0$, $f \in X$ and $\varepsilon > 0$, we have $B(f,\varepsilon) \not\subset K_m$. Because polynomial functions are dense in X by Weierstrass' theorem, we may assume that f is continuously differentiable. Take k such that $k \geq |f'(x)|$ for all $x \in [0,1]$. Take $n \in \mathbb{N}_0$ such that $4n > m+k$ and $1/n < \varepsilon$. Then, the function $f + h_n$ belongs to $B(f,\varepsilon)$. We claim that $f + h_n \notin K_m$. Indeed, if $f + h_n \in K_m$, it follows that $h_n = (f + h_n) - f$ belongs to K_{m+k}. Because $4n > m + k$, this is a contradiction.

6.3 Open mapping and closed graph theorem

Definition 6.3. A map $f : X \to Y$ between topological spaces is called *open* if $f(\mathcal{U})$ is open for every open subset $\mathcal{U} \subset X$.

From now on we will often use the following notation when X is a vector space, $\lambda \in \mathbb{C}$, $x \in X$ and $\mathcal{U}, \mathcal{V} \subset X$:

$$\begin{aligned} x + \mathcal{U} &:= \{x + u \mid u \in \mathcal{U}\} , \\ \mathcal{U} + \mathcal{V} &:= \{u + v \mid u \in \mathcal{U}, v \in \mathcal{V}\} , \\ \mathcal{U} - \mathcal{V} &:= \{u - v \mid u \in \mathcal{U}, v \in \mathcal{V}\} , \\ \lambda\mathcal{U} &:= \{\lambda u \mid u \in \mathcal{U}\} . \end{aligned}$$

Lemma 6.4. *Let X and Y be normed spaces and $T : X \to Y$ a linear map. Assume that 0 lies in the interior of $T(B(0,r))$ for some $r > 0$. Then T is open.*

Proof. Use the following scheme to prove yourself the lemma. So assume that $T : X \to Y$ is a linear map and that 0 lies in the interior of $T(B(0,r))$.

1. Observe that $B(0,\varepsilon) = (\varepsilon/r)B(0,r)$. Use the linearity of T to show that 0 lies in the interior of $T(B(0,\varepsilon))$ for every $\varepsilon > 0$.

2. Note that $B(x,\varepsilon) = x + B(0,\varepsilon)$ and hence that $T(x)$ lies in the interior of $T(B(x,\varepsilon))$ for every $x \in X$, $\varepsilon > 0$.

3. Deduce that for every open $\mathcal{V} \subset X$ and every $x \in \mathcal{V}$ we have that $T(x)$ lies in the interior of $T(\mathcal{V})$. Note that this precisely means that T is an open map.

□

Theorem 6.5 (Open mapping theorem)**.** *Let X and Y be Banach spaces and $T : X \to Y$ a* surjective *bounded operator. Then, T is open.*

Proof. The following scheme should allow you to prove the theorem yourself. Details can be found in [Con, III.12.1]. Denote by $\operatorname{int} K$ the interior of a subset $K \subset X$. Denote by $\operatorname{cl} L$ the closure of a subset $L \subset X$.

Part 1 of the proof. Use the following steps to prove that $0 \in \operatorname{int}\operatorname{cl} T(B(0,r))$ for every $r > 0$.

1. For $n \geq 1$ write $K_n = \operatorname{cl} T(B(0,n))$. Prove that $\bigcup_{n=1}^{\infty} K_n = Y$. Use Corollary 6.2 to deduce the existence of an $n \in \mathbb{N}$ such that $\operatorname{int} K_n$ is nonempty.

2. Deduce that $\operatorname{cl} T(B(0,1))$ has nonempty interior and take $y \in Y$, $\varepsilon > 0$ such that $B(y,\varepsilon) \subset \operatorname{cl} T(B(0,1))$.

3. Write $B(0,\varepsilon) = y - B(y,\varepsilon)$ and deduce that $B(0,\varepsilon) \subset \operatorname{cl} T(B(0,2))$.

4. Deduce part 1 of the proof.

Part 2 of the proof. Use the following steps to prove that $\operatorname{cl} T(B(0,1)) \subset T(B(0,2))$. Choose $y_0 \in \operatorname{cl} T(B(0,1))$.

1. Since 0 lies in the interior of $\operatorname{cl} T(B(0,1/2))$ and y_0 lies in the closure of $T(B(0,1))$, it follows that $y_0 - \operatorname{cl} T(B(0,1/2))$ intersects $T(B(0,1))$. So, take $y_1 \in \operatorname{cl} T(B(0,1/2))$ and $x_0 \in B(0,1)$ such that $y_0 - y_1 = T(x_0)$.

2. Continue inductively and prove that you can choose $y_{n+1} \in \operatorname{cl} T(B(0,2^{-n-1}))$ and $x_n \in B(0,2^{-n})$ such that $y_n - y_{n+1} = T(x_n)$.

3. Prove that the sequence $\sum_{k=0}^{n} x_k$ is a Cauchy sequence in X. Denote its limit by x. Prove that $x \in B(0,2)$.

4. Prove that $T(\sum_{k=0}^{n} x_k) = y_0 - y_{n+1}$.

5. Prove that $\|y_n\| \le 2^{-n}\|T\|$ and conclude that $y_n \to 0$. Deduce that $T(x) = y_0$. So, $y_0 \in T(B(0,2))$.

Combining parts 1 and 2, we conclude that 0 lies in the interior of $T(B(0,2))$. By Lemma 6.4 we are done. □

In the proof of the open mapping theorem, both the completeness of X and the completeness of Y are used. In Exercises 4 and 5 below we see that both completeness assumptions are crucial.

Corollary 6.6. *Let X and Y be Banach spaces. If $T : X \to Y$ is a bounded and bijective linear map, then the inverse T^{-1} is bounded as well.*

In Exercise 6 of Lecture 0 we put several norms on a direct sum $X \oplus Y$ of normed spaces X and Y. Since the norms $\|\cdot\|_{\text{sum}}$ and $\|\cdot\|_{\max}$ are equivalent (see Exercise 10 in Lecture 7), we do not specify a choice and simply speak about the Banach space $X \oplus Y$.

Definition 6.7. Let X and Y be Banach spaces. The *graph* of the map $f : X \to Y$ is the subset of $X \oplus Y$ defined as

$$\operatorname{graph} f = \{(x, f(x)) \mid x \in X\} .$$

We say that f has *closed graph* if graph f is a closed subset of $X \oplus Y$.

Theorem 6.8 (Closed graph theorem)**.** *Let X and Y be Banach spaces and $T : X \to Y$ a linear map. Then, T is bounded if and only if T has closed graph.*

Proof. We leave it as an exercise to prove that bounded operators have closed graph. Suppose conversely that graph T is closed. Then, graph $T \subset X \oplus Y$ is a Banach space and the restriction of $X \oplus Y \to X : (x, y) \mapsto x$ to graph T yields a bounded bijective linear map $p : \operatorname{graph} T \to X$. By Corollary 6.6 the inverse of p is bounded and so, T is bounded. □

6.4 The uniform boundedness principle

Theorem 6.9 (Uniform boundedness principle)**.** *Let X be a Banach space and Y a normed space. Let $\mathcal{A}$ be a family of bounded operators from X to Y satisfying*

$$\sup\{\|Ax\| \mid A \in \mathcal{A}\} < \infty \quad \textit{for all} \;\; x \in X \; .$$

Then, $\sup\{\|A\| \mid A \in \mathcal{A}\} < \infty$.

So, the uniform boundedness principle provides a conclusion that is hard to believe: if a family of bounded operators on a Banach space is pointwise bounded, then it is uniformly bounded. Nevertheless, the proof of the uniform boundedness principle is an easy application of the Baire category theorem.

Proof. Define the closed subsets $K_n \subset X$ as

$$K_n = \{x \in X \mid \|Tx\| \le n \;\text{ for all }\; T \in \mathcal{A}\} .$$

The assumption of the theorem says that $X = \bigcup_{n\in\mathbb{N}} K_n$. Since K_n is closed, Corollary 6.2 implies the existence of at least one n such that K_n has a nonempty interior. Since $K_n - K_n \subset K_{2n}$, it follows that 0 lies in the interior of K_{2n}. Take $\rho > 0$ such that $B(\rho) \subset K_{2n}$. It follows that $\|T\| \le 2n/\rho$ for all $T \in \mathcal{A}$. □

In Exercise 7 you will see that the completeness of X is a crucial assumption.

As a consequence of the uniform boundedness principle, we get the following result: if a *sequence* of bounded operators on a Banach space is pointwise convergent, the limiting operator is necessarily bounded.

Theorem 6.10 (Banach-Steinhaus theorem)**.** *Let X be a Banach space and Y a normed space. Let $T_n : X \to Y$ be a sequence of bounded operators. If $(T_n x)$ is a convergent sequence for every $x \in X$, the linear map $T : X \to Y : Tx = \lim_n T_n x$ is bounded and* $\sup_n \|T_n\| < \infty$.

Proof. The theorem follows immediately from the uniform boundedness principle once you observe that a convergent sequence in a Banach space is bounded. □

Remark 6.11. The Banach-Steinhaus theorem does not hold for *nets* of bounded operators converging pointwise. The point is that a convergent net in a normed space is not necessarily bounded.

6.5 Exercises

Exercise 2. A topological space is called *Hausdorff* if there exist for every two points $x \neq y$ in X open neighborhoods $\mathcal{U}$ of x and $\mathcal{V}$ of y such that $\mathcal{U} \cap \mathcal{V} = \emptyset$. A compact Hausdorff space X has the following property that you may use in this exercise without proving it: whenever $\mathcal{U}$ is a nonempty open subset of X, there exists a nonempty open subset $\mathcal{W}$ such that the closure $\overline{\mathcal{W}}$ of $\mathcal{W}$ is contained in $\mathcal{U}$.

Prove now, inspired by the proof of the Baire category theorem, the following variant: if X is a compact Hausdorff space and $(\mathcal{U}_n)$ a sequence of open dense subsets, then the intersection $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ is still dense.

Exercise 3. Let X be a topological space and $f : X \to \mathbb{R}$ a function. Then f is called *lower semicontinuous* if

$$\{x \in X \mid f(x) > \alpha\}$$

is open for all $\alpha \in \mathbb{R}$. The function f is called *upper semicontinuous* if

$$\{x \in X \mid f(x) < \alpha\}$$

is open for all $\alpha \in \mathbb{R}$.

Check that a function $f : X \to \mathbb{R}$ is continuous if and only if f is both lower and upper semicontinuous. The aim of the exercise is to prove the following statement: if X is a compact Hausdorff space and $f : X \to \mathbb{R}$ is a lower semicontinuous function, then

$$\mathcal{C} = \{x \in X \mid \ f \ \text{ is continuous in } \ x\}$$

is a dense subset of X.

1. Replacing f by $\frac{f}{1+|f|}$, we can assume that f is bounded.

2. Define the function

$$g : X \to [0, +\infty) : g(x) = (\limsup_{y \to x} f(y)) - f(x) \; .$$

   Here we used the notation

$$\limsup_{y \to x} f(y) := \inf_{\mathcal{U} \text{ neighborhood of } x} \big(\sup_{y \in \mathcal{U}} f(y)\big) \; .$$

   Prove that g is upper semicontinuous.

3. Define $\mathcal{U}_n = \{x \in X \mid g(x) < 1/n\}$ and prove that $\mathcal{C} = \bigcap_{n \in \mathbb{N}_0} \mathcal{U}_n$.

4. Because of Exercise 3, it remains to prove that $\mathcal{U}_n$ is dense for every n. Suppose that $\mathcal{V}$ is a nonempty open set and $\mathcal{V} \subset X \setminus \mathcal{U}_n$. Take $x_0 \in \mathcal{V}$. Construct inductively a sequence (x_k) in $\mathcal{V}$ with the property that $f(x_{k+1}) \geq f(x_k) + \frac{1}{2n}$. Obtain a contradiction with the assumption that f is bounded.

Exercise 4. Let $X = \ell^1(\mathbb{N})$ with norm $\|\cdot\|_1$. Consider $Y = \ell^1(\mathbb{N})$ with the norm $\|\cdot\|_\infty$. Then, Y is a normed space but not a Banach space. The identity map $\mathrm{id} : X \to Y : x \mapsto x$ is a bounded, bijective linear map. Prove that id is not open.

Exercise 5. Let Y be a Banach space and $\omega : Y \to \mathbb{C}$ an unbounded linear map. Define for $y \in Y$ the new norm $\|y\|_{\text{bizarre}} = \|y\| + |\omega(y)|$. Define X as the normed space $(Y, \|\cdot\|_{\text{bizarre}})$. Prove that the identity map $\mathrm{id} : X \to Y$ is bounded and bijective, but not open.

Use a vector space basis for Y to prove that every infinite dimensional Banach space Y admits unbounded linear maps $\omega : Y \to \mathbb{C}$. The following exercise gives an indication why it is impossible to give an *explicit* unbounded linear functional on an infinite dimensional Banach space.

Exercise 6. Let $x : \mathbb{N} \to \mathbb{C}$ be a sequence such that $xy \in \ell^1(\mathbb{N})$ for all $y \in \ell^2(\mathbb{N})$. Use the closed graph theorem to prove that $x \in \ell^2(\mathbb{N})$. I personally do not know a more elementary proof of this fact.

Exercise 7. Define for every $n \in \mathbb{N}$, the operator $T_n \in \mathrm{B}(\ell^2(\mathbb{N}))$ given by

$$(T_n x)(k) = \begin{cases} k\, x(k) & \text{if } \; k \leq n \; , \\ 0 & \text{if } \; k > n \; . \end{cases}$$

Define

$$X_0 = \{x \in \ell^2(\mathbb{N}) \mid \text{ There exists } \; n_0 \; \text{ such that } \; x(k) = 0 \; \text{ for all } \; k \geq n_0\} \; .$$

Check that $\sup\{\|T_n x\| \mid n \in \mathbb{N}\} < \infty$ for all $x \in X_0$. Deduce that completeness of X is crucial in the formulation of the uniform boundedness principle.

A series of exercises: continuous functions with divergent Fourier series

A function $f : \mathbb{R} \to \mathbb{C}$ is called 2π*-periodical* if $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$. If f is 2π-periodical and integrable on $[0, 2\pi]$, one defines the *Fourier coefficients* of f as

$$\widehat{f}(n) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx} \, dx \quad \text{for all } \; n \in \mathbb{Z} \; .$$

The *Fourier series* of f is given as

$$\sum_{n=-\infty}^{+\infty} \widehat{f}(n)e^{inx} \; .$$

More precisely, we associate to f the sequence of 2π-periodical functions s_n defined by

$$s_n(x) = \sum_{k=-n}^{n} \widehat{f}(k)e^{ikx} \; .$$

An important aspect of Fourier theory is to study under which conditions $s_n(x) \to f(x)$. Dirichlet's theorem says for example the following: if f is continuous in x and admits a left and a right derivative in x, then $s_n(x) \to f(x)$.

Starting from Dirichlet's theorem, it is a natural question whether $s_n(x) \to f(x)$ for every 2π-periodical continuous function f and every $x \in \mathbb{R}$. The answer is no, but it is not entirely trivial to provide an explicit continuous 2π-periodical function f such that, say, $s_n(0) \not\to f(0)$. But, using the Banach-Steinhaus theorem, we will easily prove the existence of such functions f.

Consider the Banach space X of continuous 2π-periodical functions from $\mathbb{R}$ to $\mathbb{C}$ equipped with the supremum norm $\| \cdot \|_\infty$. Check that $|\widehat{f}(n)| \leq \|f\|_\infty$ for all $f \in X$ and all $n \in \mathbb{Z}$. It follows that

$$T_n : X \to \mathbb{C} : T_n(f) = s_n(0)$$

is a sequence of bounded linear maps from X to $\mathbb{C}$. We claim that $\sup_n \|T_n\| = +\infty$. It then follows from the Banach-Steinhaus theorem that $T_n(f)$ is divergent for at least one $f \in X$.

In order to prove the claim, we have to take a closer look at some (elementary) aspects of Fourier theory. We rewrite the Fourier series s_n as a convolution of f and the *Dirichlet kernel:* using the fact that the integral of a 2π-periodical function over $[a, a + 2\pi]$ is independent of a, we get

$$\begin{aligned} s_n(x) &= \sum_{k=-n}^{n} \widehat{f}(k)e^{ikx} = \frac{1}{2\pi} \int_0^{2\pi} f(y)e^{ik(x-y)} \, dy \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y)\Big(\sum_{k=-n}^{n} e^{iky}\Big) \, dy \\ &= \int_{-\pi}^{\pi} f(x-y)D_n(y) \, dy \quad \text{with} \quad D_n(y) = \frac{1}{2\pi} \frac{\sin\big((n+\frac{1}{2})y\big)}{\sin\big(\frac{y}{2}\big)} \; . \end{aligned}$$

Prove that for all $g \in \mathrm{L}^1([a,b],\lambda)$, we have

$$\|g\|_1 = \sup\Big\{ \Big| \int_a^b f(x)g(x) \, dx \Big| \;\Big|\; f \in C([a,b]), \|f\|_\infty \leq 1 \Big\} \; . \tag{6.1}$$

You may use without proving it that $C([a,b])$ is dense in $\mathrm{L}^1([a,b],\lambda)$.

Hint. It is easy to prove that the right hand side is smaller or equal than the left hand side. Check that it is sufficient to prove the converse inequality when $g \in C([a,b])$. Take $g \in C([a,b])$. Define for every $\varepsilon > 0$, the function

$$\varphi_\varepsilon : \mathbb{C} \to \mathbb{C} : \varphi_\varepsilon(z) = \begin{cases} \frac{z}{|z|} & \text{if } |z| \geq \varepsilon \; , \\ \frac{z}{\varepsilon} & \text{if } |z| \leq \varepsilon \; . \end{cases}$$

Prove that for all $\varepsilon > 0$, the function $f_\varepsilon = \overline{\varphi_\varepsilon \circ g}$ is continuous and satisfies $\| f_\varepsilon\|_\infty \leq 1$. Also, $f_\varepsilon(x)g(x) \to |g(x)|$ uniformly in $x \in [a, b]$ when $\varepsilon \to 0$.

Using (6.1) we get that

$$\|T_n\| = \int_{-\pi}^{\pi} |D_n(y)|\, dy = 2\int_0^{\pi} |D_n(y)|\, dy\ .$$

If $y \in \left[\frac{k+\frac{1}{4}}{n+\frac{1}{2}}\pi, \frac{k+\frac{1}{2}}{n+\frac{1}{2}}\pi\right]$ and $k = 0, \ldots, n$, we have

$$\left|\sin\big((n+\frac{1}{2})y\big)\right| \geq \frac{1}{\sqrt{2}} \quad \text{and} \quad 0 < \sin\big(\frac{y}{2}\big) \leq \frac{y}{2} \leq \frac{(k+1)\pi}{2n+1}\ .$$

It then follows that for the same y, $|D_n(y)| \geq \dfrac{n+\frac{1}{2}}{\sqrt{2}\pi^2(k+1)}$. We conclude that

$$\int_0^{\pi} |D_n(y)|\, dy \geq \frac{1}{4\sqrt{2}\pi} \sum_{k=1}^{n+1} \frac{1}{k}\ .$$

Because the harmonic series diverges, it follows that $\|T_n\| \to +\infty$.

Lecture 7

A quick course in topology

From this lecture onwards we deal with the heart of functional analysis. The basic idea is that the same vector space can often be equipped in a natural way with *several topologies.* In the case of a Banach space X, these are for instance the well known norm topology and another, weaker, topology defined by the continuous functionals on X. Only when X is finite-dimensional, both topologies coincide.

In this lecture we remind a number of results in general topology. We strongly emphasize topologies defined by a family of pseudometrics (see Definition 7.2). Intuitively, pseudometric spaces are almost the same as ordinary metric spaces, but they are nevertheless sufficiently general for our purposes.

Definition 7.1. A *topology* $\mathcal{T}$ on a set X is a collection of subsets of X satisfying the following conditions.

- The empty set $\emptyset$ and the whole set X belong to $\mathcal{T}$.
- If $\mathcal{U}, \mathcal{V} \in \mathcal{T}$, then $\mathcal{U} \cap \mathcal{V} \in \mathcal{T}$.
- If $(\mathcal{U}_i)_{i \in I}$ is a family of elements of $\mathcal{T}$, then $\bigcup_{i \in I} \mathcal{U}_i$ is an element of $\mathcal{T}$.

A set X equipped with a topology $\mathcal{T}$ is called a *topological space.* The elements of $\mathcal{T}$ are called *open subsets of* X*,* or simply *opens.* We say that $K \subset X$ is *closed* if $X \setminus K$ is open.

7.1 Metric and pseudometric spaces

If (X, d) is a *metric space,* the associated *metric topology* is defined as follows: a subset $\mathcal{U} \subset X$ is open if and only if there exists for every $x \in \mathcal{U}$, a $\rho > 0$ satisfying $B(x, \rho) \subset \mathcal{U}$. Here, $B(x, \rho)$ denotes the ball with center x and radius ρ, i.e.

$$B(x, \rho) = \{y \in X \mid d(y, x) < \rho\} .$$

In particular, every *normed space* is equipped with its *norm topology.* Even more specifically, $\mathbb{R}^n$ and $\mathbb{C}^n$ get a topology in this way. In whole this course, the usual norm topology on $\mathbb{R}^n$ and $\mathbb{C}^n$ is the only topology that we consider on $\mathbb{R}^n$, $\mathbb{C}^n$. But it is *absolutely crucial that we consider more than just the norm topology on infinite dimensional normed spaces.*

The most general type of topology that we will be interested in, is induced by a family of *pseudometrics.*

Definition 7.2. Let X be a set.

- A map $d : X \times X \to [0, +\infty)$ is called a *pseudometric* if the following conditions hold.
  - $d(x, x) = 0$ for all $x \in X$,
  - $d(x, y) = d(y, x)$ for all $x, y \in X$,
  - $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.
- The pair $(X, \mathcal{D})$ is called a *pseudometric space* if $\mathcal{D}$ is a family of pseudometrics on X with the following property : if $d(x, y) = 0$ for all $d \in \mathcal{D}$, then x must be equal to y.
- If $(X, \mathcal{D})$ is a pseudometric space, the *pseudometric topology* on X is defined as follows: $\mathcal{U} \subset X$ is open if and only if for all $x \in \mathcal{U}$, there exist $\rho > 0$ and $d_1, \ldots, d_n \in \mathcal{D}$ such that

$$B_{d_1}(x, \rho) \cap \cdots \cap B_{d_n}(x, \rho) \subset \mathcal{U} \quad \text{where} \quad B_d(x, \rho) = \{y \in X \mid d(y, x) < \rho\} \; .$$

Example 7.3. Let H be a Hilbert space. Define for every $x \in H$ the pseudometric

$$d_x(y, z) = |\langle x, y - z\rangle| \; .$$

Check that the family $\{d_x \mid x \in H\}$ of pseudometrics turns H into a pseudometric space. The associated pseudometric topology is called the *weak topology* on H.

Example 7.4. Let X be a set and let $K := F(X, \mathbb{C})$ be the space of complex-valued functions. For any $x \in X$ consider the pseudometric $d_x(f, g) = |f(x) - g(x)|$. The associated topology on K is called the topology of pointwise convergence.

7.2 Continuity, convergence, interior, closure and the subspace topology

The motivation to speak abstractly about a topology, is the following: it provides the natural framework to introduce continuity of functions, convergence of sequences (or nets) and other topological notions like closure, compactness, etc.

We recall in this section several of these abstract definitions and, more importantly, say what they mean in the case of interest for us, namely for pseudometric spaces.

Definition 7.5. Let X, Y be topological spaces, $f : X \to Y$ a function and $x \in X$.

- A subset $K \subset X$ is called a *neighborhood* of x if there exists an open subset $\mathcal{U} \subset X$ such that $x \in \mathcal{U}$ and $\mathcal{U} \subset K$.

- We say that f is *continuous in* x if $f^{-1}(\mathcal{V})$ is a neighborhood of x whenever $\mathcal{V}$ is a neighborhood of $f(x)$.
- We say that f is *continuous* if $f^{-1}(\mathcal{V})$ is open whenever $\mathcal{V} \subset Y$ is open.
- We call f a *homeomorphism* if f is bijective and f and f^{-1} are both continuous.
- The *closure* of $K \subset X$ is the smallest closed subset of X containing K.
- The *interior* of $K \subset X$ is the largest open subset of X contained in K.
- The sequence (x_n) is said to *converge to* x if there exists for every neighborhood $\mathcal{U}$ of x an n_0 such that $x_n \in \mathcal{U}$ for all $n \geq n_0$.
- We say that x is an *accumulation point* of $K \subset X$ if there exists for every neighborhood $\mathcal{U}$ of x an element $y \in \mathcal{U} \cap K$ with $y \neq x$.
- Let (x_n) be a sequence in X. We say that $x \in X$ is a *limit point* of the sequence (x_n) if there exists for every neighborhood $\mathcal{U}$ of x and every n_0, an $n \geq n_0$ with $x_n \in \mathcal{U}$.

Exercise 1. Let X, Y be topological spaces.

1. Prove that $f : X \to Y$ is continuous if and only if f is continuous in every $x \in X$.
2. Prove that x belongs to the closure of $K \subset X$ if and only if $\mathcal{U} \cap K$ is nonempty for every neighborhood $\mathcal{U}$ of x.
3. Prove that x belongs to the interior of $K \subset X$ if and only if K is a neighborhood of x.

If X is a topological space and $Y \subset X$ is a subset, the *subspace topology* on Y is defined as follows: a subset $\mathcal{V} \subset Y$ is open if and only if there exists an open subset $\mathcal{U} \subset X$ such that $\mathcal{V} = \mathcal{U} \cap Y$.

If $\mathcal{T}$ and $\mathcal{T}'$ are topologies on the same set X, we say that $\mathcal{T}$ is *weaker* than $\mathcal{T}'$ (or $\mathcal{T}'$ stronger than $\mathcal{T}$) if $\mathcal{T} \subset \mathcal{T}'$. We say that $\mathcal{T}$ is *strictly weaker* if moreover $\mathcal{T} \neq \mathcal{T}'$. In Example 7.3, we have seen that a Hilbert space H admits its so called weak topology. Prove as an exercise that if (e_n) is an orthonormal family in H, then $e_n \to 0$ weakly. But of course, (e_n) does not converge to 0 in the norm topology. So, for infinite dimensional Hilbert spaces, the weak topology is strictly weaker than the norm topology.

7.3 Continuity, convergence, etc. for pseudometric spaces

We specify the notions of interior, closure, convergence, continuity, ... to the special case of pseudometric spaces. The following exercises should make you familiar enough with pseudometric spaces.

Exercise 2. Let $(X, \mathcal{D})$ be a pseudometric space equipped with its pseudometric topology. Prove the following statements.

1. A point $x \in X$ lies in the interior of a subset $Z \subset X$ if and only if there exist $d_1, \dots, d_n \in \mathcal{D}$ and $\varepsilon > 0$ such that
$$B_{d_1}(x,\varepsilon) \cap \cdots \cap B_{d_n}(x,\varepsilon) \subset Z \; .$$

2. A point $x \in X$ lies in the closure of a subset $Z \subset X$ if and only if for all $d_1, \dots, d_n \in \mathcal{D}$ and all $\varepsilon > 0$ there exists $z \in Z$ satisfying $d_i(x,z) < \varepsilon$ for all $i = 1, \dots, n$.

3. A sequence (x_n) in X converges to x if and only if for all $d \in \mathcal{D}$ we have that $d(x_n, x) \to 0$ as $n \to \infty$.

Exercise 3. Let $(X, \mathcal{D})$ and $(X', \mathcal{D}')$ be pseudometric spaces. Let $f : X \to X'$. Prove that f is continuous in $x \in X$ if and only if for all $d' \in \mathcal{D}'$ and all $\varepsilon > 0$, there exist $d_1, \dots, d_n \in \mathcal{D}$ and $\delta > 0$ satisfying $d'(f(x), f(y)) < \varepsilon$ whenever $d_i(x,y) < \delta$ for all $i = 1, \dots, n$.

Exercise 4. Let $(X, \mathcal{D})$ be a pseudometric space equipped with its pseudometric topology. Assume that $Z \subset X$. Prove the following statements.

1. Restricting every $d \in \mathcal{D}$ to Z, the set Z becomes a pseudometric space itself.

2. The pseudometric topology on Z coincides with the subspace topology as a subset of X.

7.4 Be careful with sequences ... and say hello to nets

Without going into details, we point out some *pitfalls when working with sequences in general topological spaces.* Recall the following two properties of a *metric space* (X, d) *:*

- A point x belongs to the closure of a subset $K \subset X$ if and only if there exists a sequence (x_n) in K converging to x.

- A function $f : X \to Y$ is continuous in the point $x \in X$ if and only if $(f(x_n))_{n \in \mathbb{N}}$ converges to $f(x)$ whenever (x_n) is a sequence in X converging to x.

None of both properties holds true in arbitrary topological spaces. And the crucial point is that this is not a pathological phenomenon. In Exercise 12 you see that both properties fail for instance when X is an infinite dimensional Hilbert space equipped with the weak topology of Example 7.3.

The correct generalization of a sequence is the notion of a *net*. We give here the definition and an illustration.

Definition 7.6. We call $(I, \leq)$ a *directed set* if $(I, \leq)$ is a partially ordered set such that for all $i, j \in I$, there exists a $k \in I$ satisfying $i \leq k$ and $j \leq k$.

A *net* $(x_i)_{i \in I}$ in a set X is a map $I \to X : i \mapsto x_i$ from a directed set I to the set X.

Let X be a topological space and $x \in X$. We say that the net $(x_i)_{i \in I}$ *converges to* x if for every neighborhood $\mathcal{U}$ of x there exists an $i_0 \in I$ such that $x_i \in \mathcal{U}$ whenever $i \geq i_0$.

We say that x is a *limit point* of the net $(x_i)_{i \in I}$ if for every neighborhood $\mathcal{U}$ of x and for every $i_0 \in I$ there exists an $i \geq i_0$ such that $x_i \in \mathcal{U}$.

Exercise 5. Let X be a topological space and $x \in X$. Define I as the set of neighborhoods of x and order I by inverse inclusion: $\mathcal{U} \leq \mathcal{V}$ if and only if $\mathcal{U} \supset \mathcal{V}$. Check that I is a directed set.

Use the directed set of neighborhoods of x to prove the following two statements.

1. The point x belongs to the closure of a subset $K \subset X$ if and only if there exists a net $(x_i)_{i \in I}$ in K converging to x.

2. Let Y be a topological space. A function $f : X \to Y$ is continuous in x if and only if the net $(f(x_i))_{i \in I}$ converges to $f(x)$ whenever $(x_i)_{i \in I}$ is a net in X converging to x.

Exercise 6. Let I be a set. Let J be the family of finite subsets of I, ordered by inclusion. Prove that J is a directed set.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in a normed space. We say that the series $\sum_{n=1}^{\infty} x_n$ converges unconditionally to x if $\sum_{n=1}^{\infty} x_{\sigma(n)} = x$ for any bijection $\sigma \colon \mathbb{N} \to \mathbb{N}$. Define a net $(S_F)_{F \in J}$ via $S_F := \sum_{n \in F} x_n$. Show that the series $\sum_{n=1}^{\infty} x_n$ converges unconditionally to x iff the net $(S_F)_{F \in J}$ converges to x.

Exercise 7. Let $(X, \mathcal{D})$ be a pseudometric space equipped with the pseudometric topology. Let $(x_i)_{i \in I}$ be a net in X and $x \in X$. Prove that $x_i \to x$ if and only if $d(x_i, x) \to 0$ for every $d \in \mathcal{D}$.

One can work with nets in essentially the same way as with sequences. The only tricky point is to define the notion of a *subnet*.

Definition 7.7. Let $(x_i)_{i \in I}$ be a net in a set X. A *subnet* of $(x_i)_{i \in I}$ is a net $(y_j)_{j \in J}$ of the form $y_j = x_{h(j)}$ where $h : J \to I$ is a map such that for every $i \in I$ there exists a $j_0 \in J$ satisfying $h(j) \geq i$ for all $j \geq j_0$.

In most cases the map h in the definition of a subnet will be nondecreasing, i.e. $h(j) \leq h(j')$ whenever $j \leq j'$, but it is more convenient not to make this a hypothesis.

Exercise 8. Let $(x_i)_{i \in I}$ be a net in a topological space X. Assume that $(x_i)_{i \in I}$ converges to x. Prove that every subnet of $(x_i)_{i \in I}$ converges to x.

Lemma 7.8. *Let $(x_i)_{i \in I}$ be a net in a topological space X and take $x \in X$. Then x is a limit point of the net $(x_i)_{i \in I}$ if and only if $(x_i)_{i \in I}$ admits a subnet that converges to x.*

Proof. Assume first that $h : J \to I$ such that $y_j := x_{h(j)}$ defines a subnet of $(x_i)_{i \in I}$ that converges to x. We have to prove that x is a limit point of $(x_i)_{i \in I}$. Take a neighborhood $\mathcal{U}$ of x and take $i_0 \in I$. We have to prove the existence of $i \in I$ such that $i \geq i_0$ and $x_i \in \mathcal{U}$. From the definition of a subnet we find $j_0 \in J$ such that $h(j) \geq i_0$ whenever $j \geq j_0$. Since $y_j \to x$ there exists $j_1 \in J$ such that $y_j \in \mathcal{U}$ for all $j \geq j_1$. Take $j \in J$ such that $j \geq j_0$ and $j \geq j_1$. Put $i := h(j)$. It follows that $i \geq i_0$ and $x_i = y_j \in \mathcal{U}$.

Conversely assume that x is a limit point of $(x_i)_{i \in I}$. Denote by $\mathcal{N}$ the set of neighborhoods of x ordered by inverse inclusion. Define on $I \times \mathcal{N}$ the partial order given by $(i, \mathcal{U}) \leq (j, \mathcal{V})$ if and only if $i \leq j$ and $\mathcal{U} \supset \mathcal{V}$. Since x is a limit point of $(x_i)_{i \in I}$ we can choose for every $(i, \mathcal{U}) \in I \times \mathcal{N}$ an element $h(i, \mathcal{U}) \in I$ such that $h(i, \mathcal{U}) \geq i$ and $x_{h(i,\mathcal{U})} \in \mathcal{U}$. Define $y_{(i,\mathcal{U})} := x_{h(i,\mathcal{U})}$. Prove yourself that $(y_{(i,\mathcal{U})})_{(i,\mathcal{U}) \in I \times \mathcal{N}}$ is a subnet of $(x_i)_{i \in I}$ that converges to x. $\square$

7.5 Compactness

Definition 7.9. Let X be a topological space and $K \subset X$. We say that K is *compact* if every open cover of K admits a finite subcover: whenever $\mathcal{O}$ is a family of open subsets of X covering K, i.e. satisfying

$$K \subset \bigcup_{\mathcal{U} \in \mathcal{O}} \mathcal{U},$$

there exists a finite subset $\mathcal{O}_1 \subset \mathcal{O}$ such that still

$$K \subset \bigcup_{\mathcal{U} \in \mathcal{O}_1} \mathcal{U}.$$

Note that compactness is in fact an intrinsic property of K. Indeed, one easily checks that $K \subset X$ is compact if and only if K is compact when we equip K with the subspace topology.

Proposition 7.10. *A family of subsets is said to have the* finite intersection property *if every finite subfamily has a nonempty intersection.*

A topological space K is compact if and only if every family of closed subsets of K having the finite intersection property, actually has a nonempty intersection.

Proof. The proof is almost tautological. Taking complements, families of closed subsets of K are in one-to-one correspondence with families of open subsets. The family of closed subsets has the finite intersection property if and only if the corresponding family of open subsets has no finite subfamily that covers K. On the other hand the whole family of closed subsets has a nonempty intersection if and only if the corresponding family of open subsets does not cover K. □

For metric spaces, several more natural conditions are equivalent with compactness, see Section 3.1. Existence of convergent subsequences is particularly useful. Using nets, we can arrive at an analogous characterization of compactness for general topological spaces.

Proposition 7.11. *Let X be a topological space. Then the following statements are equivalent.*

(i) X is compact.

(ii) Every net in X has a limit point.

(iii) Every net in X admits a convergent subnet.

Proof. Not (ii) $\implies$ not (i). Assume that $(x_i)_{i \in I}$ is a net in X without limit point. For every $x \in X$ choose a neighborhood $\mathcal{U}_x$ of x and an index $i_x \in I$ such that $x_i \notin \mathcal{U}_x$ for all $i \geq i_x$. Then $\{\mathcal{U}_x \mid x \in X\}$ is an open covering of X. Finitely many $\mathcal{U}_{x_1}, \ldots, \mathcal{U}_{x_n}$ never cover the whole of X because we can find $i \in I$ such that $i \geq i_{x_k}$ for all $k = 1, \ldots, n$. Then $x_i \in X$ belongs to none of the sets $\mathcal{U}_{x_k}$, $k = 1, \ldots, n$.

(ii) $\Leftrightarrow$ (iii). This follows immediately from Lemma 7.8.

Not (i) $\implies$ not (ii). Suppose that $\mathcal{U} = (\mathcal{O}_i)_{i \in I}$ is an open cover of X without a finite subcover. Let J be the directed set of finite subsets of I. For any $F \in J$ the subfamily $(\mathcal{O})_{i \in F}$ does not cover X, so there exists $x_F \in X \setminus \bigcup_{i \in F} \mathcal{O}_i$. The net $(x_F)_{F \in J}$ does not have a limit point. Indeed, suppose that x is a limit point. There exists $i \in I$ such that $x \in \mathcal{O}_i$. If we choose $F = \{i\}$ then any $G \in J$ satisfying $F \subset G$ will give an element $x_G \notin \mathcal{O}_i$, so x cannot be a limit point. □

7.6 Infinite products of topological spaces and Tychonoff's theorem

The main aim of this section is to define *infinite cartesian products* and to prove the *Tychonoff theorem* on compactness of an infinite cartesian product of compact spaces.

Definition 7.12. Let I be a set and suppose that for all $i \in I$, we are given a set X_i. Then, the *infinite cartesian product* of the sets X_i is defined and denoted as

$$\prod_{i \in I} X_i := \{(x_i)_{i \in I} \mid x_i \in X_i \text{ for all } i \in I\}.$$

If $X_i = X$ for all X, check that $\prod_{i \in I} X_i$ is in fact the set of all functions from I to X.

Definition 7.13. Let $(X_i)_{i \in I}$ be a family of topological spaces. Set $X = \prod_{i \in I} X_i$. A set $U \subset X$ is a basic open subset if there exist finitely many indices $i_1, \ldots, i_n \in I$ and open sets $U_k \in X_{i_k}$ such that $U = \{(x_i)_{i \in I} : x_{i_k} \in U_k \text{ for } 1 \leqslant k \leqslant n\}$. A set is open if it is equal to a union of basic open subsets.

Remark 7.14. For every $i \in I$ we have map $\pi_i : X \to X_i$ that maps an element of the product to its i-th coordinate. The topology in the product is the smallest topology that makes all these maps continuous.

Exercise 9. Let $(X_i)_{i \in I}$ be a family of topological spaces. Set $X = \prod_{i \in I} X_i$, equipped with the product topology. Let $(x_j)_{j \in J}$ be a sequence in X. Prove that $\lim_{j \in J} x_j = x$ if and only if for every $i \in I$ we have $\lim_{j\ inJ} \pi_i(x_j) = \pi_i(x)$.

For those familiar with the usual definition of the product topology, the following exercise shows that both definitions are equivalent.

Theorem 7.15 (Tychonoff theorem)**.** *Let $(X_i, \mathcal{D}_i)_{i \in I}$ be a family of compact pseudometric spaces. Then, $X = \prod_{i \in I} X_i$ equipped with the product topology, is compact.*

Proof. Let $(x_k)_{k \in K}$ be a net in X. We will show that it admits a limit point. For any $J \subset I$ define the projection map $\pi_J : X \to X_J := \prod_{i \in J} X_i$ by choosing only the coordinates indexed by J. We say that $g_J \in X_J$ is a *partial limit point* of $(x_k)_{k \in K}$ if its a limit point of the net $(\pi_J(x_k))_{k \in K}$. We introduce a partial order on the set of partial limit points in the following way: $g_J \leqslant g_{J'}$ if $J \subset J'$ and $\pi_J(g_{J'}) = g_J$, i.e. $g_{J'}$ is an extension of g_J.

We will show that there exists a maximal partial limit point, using Zorn's lemma. First of all, since each of the spaces X_i is compact, if we take $J = \{i\}$ we can find a partial limit point g_i. Now we have to show that any linearly ordered family of partial limit points admits an upper bound. Suppose that $(g_J)_{J \in Y}$, where $Y \subset \mathcal{P}(I)$, is such a family. Consider $\widetilde{J} := \bigcup_{J \in Y} J$. For any $j \in \widetilde{J}$ there exists $J \in Y$ such that $j \in J$, so we may define $g_{\widetilde{J}}(j) := g_J(j)$; the definition does not depend on the choice of J. It remains to check that $g_{\widetilde{J}}$ is a limit point of $(\pi_{\widetilde{J}}(x_k))_{k \in K}$. Let U be a basic neighborhood around $g_{\widetilde{J}}$ in $X_{\widetilde{J}}$. For any $k \in K$ we would like to find $k' \geqslant k$ such that $\pi_{\widetilde{J}}(x_{k'}) \in U$. Since U is a basic neighbourhood, there are only finitely many indices $j_1, \ldots, j_n$ that we have to care about. We can find $J \in Y$ such that $j_1, \ldots, j_n \in J$ and now use the fact that g_J is a partial limit point to conclude.

Let now g_J be a maximal partial limit point of the net $(x_k)_{k\in K}$. We claim that $J = I$. If not, there exists $i \notin J$ and we will find a partial limit point $g_{J\cup\{i\}}$. Indeed, there exists a subnet of $(\pi_J(x_k))_{k\in K}$ converging to g_J and we can ensure the convergence of $(\pi_i(x_k))_{k\in K}$ after passing to a further subnet, simply by using the compactness of X_i. So if $J \neq I$, then the partial limit point g_J cannot be maximal. Therefore $J = I$ and g_I is a genuine limit point. □

Remark 7.16. The proof above is similar in spirit to the proof of the Hahn-Banach theorem (Theorem 5.1. We have partial limit points and we want to extend them to the full product. What we can do, using compactness, is to extend the partial limit points a single coordinate at a time. Then Zorn's lemma gives extension to the whole space.

7.7 Topological vector spaces

Definition 7.17. A topological space X is said to be *Hausdorff* if for every $x \neq y$ in X there exist neighborhoods $\mathcal{U}$ of x and $\mathcal{V}$ of y that are disjoint, i.e. $\mathcal{U} \cap \mathcal{V} = \emptyset$.

Proposition 7.18. *Let $(X, \mathcal{D})$ be a pseudometric space. The pseudometric topology is Hausdorff.*

Proof. Let $x \neq y$ be two distinct points in X. Take $d \in \mathcal{D}$ such that $d(x, y) \neq 0$. Put $\varepsilon := d(x, y)/2$. Define $\mathcal{U} := B_d(x, \varepsilon)$ and $\mathcal{V} := B_d(y, \varepsilon)$. Then $\mathcal{U}$ and $\mathcal{V}$ are neighborhoods of x and y respectively and $\mathcal{U} \cap \mathcal{V} = \emptyset$. □

When dealing with vector spaces, metrics are usually given by norms on the vector space. Similarly, pseudometrics are usually given by seminorms.

Definition 7.19. Let X be a vector space over $\mathbb{C}$.

- A map $p : X \to [0, +\infty)$ is called a *seminorm* on X if the following conditions hold.
  - $p(\lambda x) = |\lambda|\, p(x)$ for all $x \in X, \lambda \in \mathbb{C}$,
  - $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$.
- The pair $(X, \mathcal{P})$ is called a *seminormed space* if $\mathcal{P}$ is a family of seminorms on X such that $x = 0$ whenever $p(x) = 0$ for all $p \in \mathcal{P}$.
- If $(X, \mathcal{P})$ is a seminormed space, the *seminorm topology* on X is defined as the pseudometric topology on X given by the family of pseudometrics $\{d_p \mid p \in \mathcal{P}\}$ defined by $d_p(x, y) = p(x - y)$.

Definition 7.20. A *topological vector space* is a vector space X equipped with a Hausdorff topology such that the maps

$$X \times X \to X : (x, y) \mapsto x + y \quad \text{and} \quad \mathbb{C} \times X \to X : (\lambda, x) \mapsto \lambda x$$

are continuous when we put on $X \times X$, respectively $\mathbb{C} \times X$, the product topology.

Proposition 7.21. *Let $(X, \mathcal{P})$ be a seminormed vector space. The seminorm topology turns X into a topological vector space.*

Proof. We use nets to check the continuity of the addition. Let $((x_i, y_i))_{i\in I}$ be a net in $X \times X$ converging to (x, y), which means that $\lim_{i\in I} x_i = x$ and $\lim_{i\in I} y_i = y$. We want to check that the net $(x_i + y_i)_{i\in I}$ converges to $x + y$. We have to check that for any seminorm $p \in \mathcal{P}$ we have $\lim_{i\in I} p(x_i + y_i - x - y) = 0$. By the triangle inequality we have $p(x_i + y_i - x - y) \leqslant p(x_i - x) + p(y_i - y)$ and both terms on the right-hand side converge to 0, because $\lim_{i\in I} x_i = x$ and $\lim_{i\in I} y_i = y$.

Prove yourself the continuity of the scalar multiplication. □

Remark 7.22. All topological vector spaces that we encounter in these lecture notes are seminormed spaces. The terminology 'seminormed space' is not a standard one. On page 92, a series of exercises will lead to the following alternative characterization of seminormed spaces: they are exactly the *locally convex topological vector spaces*. So, the class of seminormed spaces (with their seminorm topology) is the same as the class of locally convex spaces. This last terminology is much more standard. The reason for our unorthodox terminology is our belief that seminormed spaces are intuitively much easier to understand than locally convex spaces.

7.8 Exercises

Exercise 10. The norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on the vector space X are called *equivalent* if there exist constants $C_1, C_2 > 0$ such that

$$C_1\|x\|_a \leq \|x\|_b \leq C_2\|x\|_a \quad \text{for all } \; x \in X \ .$$

Prove that equivalent norms induce the same norm topology on X.

Exercise 11. Let H be a Hilbert space with its weak topology defined in Example 7.3. Suppose that the sequence $(x_n)_{n\to\infty}$ converges in the weak topology. Using the uniform boundedness principle, show that this sequence is bounded, i.e. $\sup_{n\ in\mathbb{N}} \|x_n\| < \infty$.

Exercise 12. Let H be an infinite dimensional Hilbert space. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal family in H. Follow the hints below to prove the following statements.

(i) The point 0 belongs to the weak closure of $K := \{\sqrt{n}e_n \mid n \in \mathbb{N}_0\}$.

(ii) There is no sequence in K that converges to 0 weakly.

To prove (i), you have to prove that there exists for all $x_1, \dots, x_k$ and all $\varepsilon > 0$ an $n \in \mathbb{N}_0$ such that

$$|\langle\sqrt{n}e_n, x_i\rangle| < \varepsilon \quad \text{for all } \; i = 1, \dots, k \ .$$

This can be done by contradiction.

To prove (ii) use Exercise 11.

Exercise 13. Let H and K be as in Exercise 12. Define the function

$$f : H \to \mathbb{R} : f(x) = \begin{cases} 1 & \text{if } \; x \in K \ , \\ 0 & \text{if } \; x \notin K \ . \end{cases}$$

Prove that f is not weakly continuous in 0 but that nevertheless $(f(x_n))$ converges to 0 whenever (x_n) is a sequence in H converging weakly to 0.

Lecture 8

Weak topologies and the Banach-Alaoglu theorem

8.1 Examples of topological vector spaces

We will systematically equip one and the same space (typically a Banach space) with different topologies[1] and that is the crucial aspect of the list of examples in this section. It is extremely useful to consider several topologies on the same vector space, motivated by the following simplistic reasoning. In a weaker topology, there are fewer open sets. Therefore, subsets are more often compact. Compactness is very useful, since it allows to take limit points. A typical illustration of this phenomenon is the Banach-Alaoglu theorem below.

Our first example was already given as Example 7.3.

Example 8.1. Let H be a Hilbert space. Define the family $\mathcal{P}$ of seminorms on H given by

$$\mathcal{P} = \{p_y \mid y \in H\} \quad \text{with} \quad p_y(x) = |\langle y, x\rangle| \;\text{ for all }\; x \in H \;.$$

The resulting seminorm topology on H is called the *weak topology* on the Hilbert space H.

In the last paragraph of Section 7.2, it was shown that the weak topology on an infinite dimensional Hilbert space H, is strictly weaker than the norm topology.

Example 8.2. This example generalizes the previous Example 8.1. Let X be a Banach space. Define the family $\mathcal{P}$ of seminorms on X as

$$\mathcal{P} = \{p_\omega \mid \omega \in X^*\} \quad \text{with} \quad p_\omega(x) = |\omega(x)| \;\text{ for all }\; x \in X \;.$$

The seminorm topology defined by $\mathcal{P}$ is called the *weak topology* on the Banach space X.

Since every Hilbert space is a Banach space, Examples 8.1 and 8.2 could lead to a potential problem of terminology, but the Riesz representation theorem 1.10 tells us that this is not the case.

[1] Browsing some books, you will find at least seven seminorm topologies on $B(H)$ that are all different when H is an infinite dimensional Hilbert space. In Example 8.5 and Exercises 5 and 6, we treat five of them.

Example 8.3. Let X be a Banach space and X^* its dual Banach space. Define the family $\mathcal{P}$ of seminorms on X^* as

$$\mathcal{P} = \{p_x \mid x \in X\} \quad \text{where} \quad p_x(\omega) = |\omega(x)| \quad \text{for all} \quad \omega \in X^* \ .$$

The seminorm topology defined by $\mathcal{P}$ is called the *weak* topology* on the dual Banach space X^*.

The Banach-Alaoglu theorem below tells us that the unit ball of X^* is compact in the weak* topology for every Banach space X. Using Theorem 1.10, it follows that the unit ball of a Hilbert space is compact in the weak topology. More generally, the same holds for the unit ball in an arbitrary *reflexive* Banach space, but not in other Banach spaces.

Example 8.4. Let $X = C(\mathbb{R})$ be the space of continuous functions on $\mathbb{R}$. We define two families of seminorms on X: $\mathcal{P}_1 := \{p_K : K \subset \mathbb{R} \text{ is compact}\}$, where $p_K(f) := \sup_{x\in K} |f(x)|$, and $\mathcal{P}_2 := \{p_x : x \in \mathbb{R}\}$, where $p_x(f) := |f(x)|$. The topology induced by $\mathcal{P}_1$ is called the topology of *convergence on compact subsets*, whereas the topology defined by $\mathcal{P}_2$ is called the topology of *pointwise convergence*.

Example 8.5. Let H be a Hilbert space. On the Banach space of bounded operators $\mathrm{B}(H)$, we have the following families of seminorms.

$$\begin{aligned}
&\mathcal{P}_1 \text{ consisting of the seminorms } T \mapsto |\langle Tx, y\rangle| && \text{for all } x, y \in H \ .\\
&\mathcal{P}_2 \text{ consisting of the seminorms } T \mapsto \|Tx\| && \text{for all } x \in H \ .\\
&\mathcal{P}_3 \text{ consisting of the seminorms } T \mapsto \|Tx\| + \|T^*x\| && \text{for all } x \in H \ .
\end{aligned}$$

The family $\mathcal{P}_1$ defines the *weak topology* on $\mathrm{B}(H)$. The family $\mathcal{P}_2$ defines the *strong topology* on $\mathrm{B}(H)$. The family $\mathcal{P}_3$ defines the *strong* topology* on $\mathrm{B}(H)$.

Remark 8.6. It is unfortunate but crucial to observe that the *weak topology on* $\mathrm{B}(H)$ *defined in Example 8.5 does not coincide with the weak topology on* $\mathrm{B}(H)$ *when viewing* $\mathrm{B}(H)$ *as a Banach space.* Whenever we deal with $\mathrm{B}(H)$, the terminology ‘weak topology’ will refer to the terminology introduced in Example 8.5.

8.2 The Banach-Alaoglu theorem

Theorem 8.7 (Banach-Alaoglu theorem)**.** *Let X be a Banach space. The unit ball $(X^*)_1$ of the dual Banach space X^* is compact in the weak* topology defined in 8.3.*

Proof. We will identify $(X^*)_1$ with a closed subset of an infinite product of compact metric spaces. Tychonoff’s theorem 7.15 will then yield the compactness of $(X^*)_1$.

Elements of X^* are, in particular, functions from X to $\mathbb{C}$, so we can identify them with a subset of $\prod_{x\in X} \mathbb{C}$. But this space is not compact, so we have to use something else. Every element $\omega \in (X^*)_1$ satisfies $|\omega(x)| \leqslant \|x\|$, so we may identify $(X^*)_1$ with a subset of the product $K := \prod_{x\in X} \overline{B}_{\|x\|}$, where for $r \geq 0$ we define $\overline{B}_r = \{z \in \mathbb{C} \mid |z| \leq r\}$ – a compact subset of $\mathbb{C}$. By Tychonoff’s theorem 7.15 this product is compact.

More formally, we define a map

$$\theta : (X^*)_1 \to K : \theta(\omega)_x = \omega(x) \ .$$

We have to check two things: that the topology on $\theta((X^*)_1)$ induced from K agrees with the topology of $(X^*)_1$ (so θ is a homeomorphism onto its image) and that $\theta((X^*)_1)$ is closed in K, because this would imply that $\theta((X^*)_1)$ is compact.

The first part is simple: both topologies are given by pointwise convergence on X, so a net $(\omega_i)_{i\in I}$ in $(X^*)_1$ converges iff the net $(\theta(\omega_i))_{i\in I}$ converges.

To check that $\theta((X^*)_1)$ is closed in K, we also use nets. Elements of K are functions f from X to $\mathbb{C}$ that satisfy $|f(x)| \leqslant \|x\|$, while the subset $\theta((X^*)_1)$ consists of linear functions. Prove yourself that linearity is preserved under convergence of nets. The condition $|f(x)| \leqslant \|x\|$ for a linear functional f implies that $\|f\| \leqslant 1$, so any element in the closure of $\theta((X^*)_1)$ actually belongs to $\theta((X^*)_1)$, i.e. this set is closed. □

8.3 Illustration: an invariant mean on the group of integers

Definition 8.8. Let X be a set. A *mean* or *finitely additive probability measure* m on X is a map that assigns to every subset $A \subset X$ a number $m(A) \in [0, 1]$ such that

- $m(\emptyset) = 0$ and $m(X) = 1$,
- $m(A \cup B) = m(A) + m(B)$ whenever A and B are disjoint subsets of X.

Definition 8.9. Let G be a group. An *invariant mean m on G* is a finitely additive probability measure m on G satisfying

$$m(gA) = m(A) \quad \text{for all} \quad g \in G, A \subset G\ .$$

In words, an invariant mean is a translation invariant finitely additive probability measure on the group. A group G that admits an invariant mean, is called an *amenable group*.

Exercise 1. Denote by $|A|$ the number of elements of a finite set A. Let G be a finite group. Prove that

$$m(A) = \frac{|A|}{|G|}$$

is the unique invariant mean on G.

It is far from obvious that certain infinite groups can have an invariant mean. This is already illustrated by the following exercise.

Exercise 2. Let G be an infinite group and m an invariant mean on G. Prove that $m(A) = 0$ for every finite subset $A \subset G$.

Using the Banach-Alaoglu theorem, we can prove the following.

Theorem 8.10. *The group $G = \mathbb{Z}$ has an invariant mean.*

The idea of the proof of Theorem 8.10 is the following. For every $n \in \mathbb{N}$, we define the finitely additive probability measure m_n on $\mathbb{Z}$ as

$$m_n(A) = \frac{|A \cap [-n, n]|}{2n + 1}\ .$$

In a sense to be made precise, the finitely additive probability measure m_n is more and more invariant as $n \to \infty$. The Banach-Alaoglu theorem will allow us to take a limit point of the sequence $(m_n)_{n\in\mathbb{N}}$. This limit point will be an invariant mean.

Proof of Theorem 8.10. Define the Banach space $X = \ell^\infty(\mathbb{Z})$ with the supremum norm $\|\cdot\|_\infty$. Define for every $n \in \mathbb{N}$, the linear maps

$$\omega_n : \ell^\infty(\mathbb{Z}) \to \mathbb{C} : \omega_n(f) = \frac{1}{2n+1} \sum_{i=-n}^{n} f(i) \; .$$

Check that $\|\omega_n\| = 1$ for every n.

Define for every $\omega \in \ell^\infty(\mathbb{Z})^*$ and every $k \in \mathbb{Z}$, the element $k \cdot \omega \in \ell^\infty(\mathbb{Z})^*$ by the formulae

$$(k \cdot \omega)(f) = \omega(f \cdot k) \quad \text{where} \quad (f \cdot k)(n) = f(k+n) \; .$$

We claim that for every $k \in \mathbb{Z}$,

$$\|k \cdot \omega_n - \omega_n\| \to 0 \quad \text{when} \quad n \to \infty \; .$$

You easily check that for all $n > k \geq 1$, we have

$$(k \cdot \omega_n - \omega_n)(f) = \frac{1}{2n+1}\Big(- \sum_{i=-n}^{-n+k-1} f(i) + \sum_{i=n+1}^{n+k} f(i) \Big) \; .$$

Hence, $\|k \cdot \omega_n - \omega_n\| \leq \frac{2k}{2n+1}$ proving the claim.

Combining the Banach-Alaoglu theorem and Proposition 7.11 we find a subnet $(\mu_j)_{j\in J}$ of the sequence $(\omega_n)_{n\in\mathbb{N}}$ such that $\mu_j \to \mu \in \ell^\infty(\mathbb{Z})^*$. Write $\mu_j = \omega_{h(j)}$ where $h : J \to \mathbb{N}$ is a map such that for every $n_0 \in \mathbb{N}$ there exists a $j_0 \in J$ satisfying $h(j) \geq n_0$ for all $j \geq j_0$. The fact that $\mu_j \to \mu$ in the weak* topology means that $\mu_j(f) \to \mu(f)$ for every $f \in \ell^\infty(\mathbb{Z})$. Deduce from this that

1. $k \cdot \mu = \mu$ for all $k \in \mathbb{Z}$,
2. $\mu(f) \in [0,1]$ if $0 \leq f(i) \leq 1$ for all $i \in \mathbb{Z}$,
3. $\mu(1) = 1$.

Conclude that denoting by χ_A the indicator function of a subset $A \subset \mathbb{Z}$, the formula $m(A) = \mu(\chi_A)$ provides us with an invariant mean on $\mathbb{Z}$. □

8.4 Exercises

Exercise 3. Let X be a Banach space and suppose that its norm topology and weak topology coincide. Find $\omega_1, \ldots, \omega_n \in X^*$ such that

$$\|x\| \leq \max_{i=1,\ldots,n} |\omega_i(x)| \quad \text{for all} \;\; x \in X \; .$$

Deduce that the linear map $X \to \mathbb{C}^n : x \mapsto (\omega_1(x), \ldots, \omega_n(x))$ is injective and hence, X is finite dimensional.

Exercise 4. Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded operator. Use the basic neighborhood in the weak topology to prove that T remains continuous, when X and Y are endowed with their respective weak topologies. Try to prove this continuity using nets.

Exercise 5. Consider the different seminorm topologies on $\mathrm{B}(H)$ defined in Example 8.5. Prove that the weak topology is weaker than the strong topology; that the strong topology is weaker than the strong* topology and that the strong* topology is weaker than the norm topology.

Use next the following examples to prove that we can replace 'weaker' by 'strictly weaker' everywhere in the previous paragraph, whenever H is infinite dimensional. Define the sequences of operators.

1. $U_n : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z}) : (U_n x)(k) = x(n+k)$ for all $x \in \ell^2(\mathbb{Z}), k \in \mathbb{Z}$,
2. $V_n : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}) : (V_n x)(k) = x(n+k)$ for all $x \in \ell^2(\mathbb{N}), k \in \mathbb{N}$,
3. $P_n : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}) : (P_n x)(k) = \begin{cases} x(k) & \text{if } k \geq n , \\ 0 & \text{if } k < n . \end{cases}$

Check that $U_n \to 0$ in the weak topology, but not in the strong topology; that $V_n \to 0$ in the strong topology, but not in the strong* topology and that $P_n \to 0$ in the strong* topology, but not in the norm topology.

Exercise 6. In Theorem 3.22, we have seen that $\mathrm{B}(H) \cong \mathcal{TC}(H)^*$. Prove that the weak* topology on $\mathrm{B}(H)$ (defined through the identification of $\mathrm{B}(H)$ and $\mathcal{TC}(H)^*$) and the weak topology on $\mathrm{B}(H)$ defined in Example 8.5, coincide on the unit ball $(\mathrm{B}(H))_1$.

Exercise 7. Prove that the map $\mathrm{B}(H) \to \mathrm{B}(H) : T \mapsto T^*$ is continuous if we equip $\mathrm{B}(H)$ with its weak topology. Is this still true for the strong topology?

Denote by $(\mathrm{B}(H))_1$ the unit ball of $\mathrm{B}(H)$ and prove that the multiplication map

$$(\mathrm{B}(H))_1 \times \mathrm{B}(H) \to \mathrm{B}(H) : (S, T) \mapsto ST$$

is continuous if we equip $(\mathrm{B}(H))_1 \times \mathrm{B}(H)$ with the product of the strong topologies and $\mathrm{B}(H)$ with the strong topology.

Exercise 8. Let H be an infinite dimensional Hilbert space. Prove as follows that the multiplication map

$$m : \mathrm{B}(H) \times \mathrm{B}(H) \to \mathrm{B}(H) : (S, T) \mapsto ST$$

is not continuous for the strong topology. Denote by $\theta_{x,y}$ the rank one operator defined by

$$\theta_{x,y}(z) = \langle z, y \rangle x .$$

Assume that the multiplication map is strongly continuous. It follows that also the map $\mathrm{B}(H) \to \mathrm{B}(H) : T \mapsto T^2$ is strongly continuous. Let $(e_n)_{n \in \mathbb{N}}$ be an orthonormal family in H. Consider the subset $\mathcal{A} \subset \mathrm{B}(H)$ given by

$$\mathcal{A} = \{\sqrt{n}\theta_{e_n,e_n} \mid n \in \mathbb{N}_0\} .$$

Use Exercise 12 in Lecture 7 to prove that 0 belongs to the closure of $\mathcal{A}$ in the strong topology. By continuity of $T \mapsto T^2$, it follows that 0 belongs to the strong closure of

$$\mathcal{B} = \{n\theta_{e_n,e_n} \mid n \in \mathbb{N}_0\} .$$

Find a vector $x \in H$ such that $\|Tx\| = 1$ for all $T \in \mathcal{B}$. Derive a contradiction.

Finally prove as follows that the multiplication map *m does preserve limits of sequences.* Indeed, if $(S_n, T_n) \to (S, T)$ in the product of the strong topologies, it follows that $S_n \to S$ and $T_n \to T$ strongly. Use the Banach-Steinhaus theorem and Exercise 7 to conclude that $S_n T_n \to ST$ strongly.

We conclude that the multiplication map, although preserving limits of sequences, is not a continuous map. We already met this phenomenon in Exercise 13 of Lecture 7, but now it appears in a seemingly innocent case, namely for the map $T \mapsto T^2$ on $\mathrm{B}(H)$ equipped with the strong topology.

Exercise 9. This exercise complements Theorem 5.8 in Lecture 5, where the necessary terminology is introduced. Imitate the proof of Theorem 8.10 to give a different proof for the existence of a Banach limit L on $\ell^\infty(\mathbb{N})$ satisfying $L(x) = \lambda$ whenever the sequence of Cesàro means of $(x(n))_{n\in\mathbb{N}}$ converges to λ.

Lecture 9

The Hahn-Banach separation theorem

Let A and B be disjoint subsets of $\mathbb{R}^3$. Under which assumptions does there exist a plane $P \subset \mathbb{R}^3$ such that A lies entirely on one side of P and B entirely on the other side? Depending on whether we want A and B to lie strictly on both sides of P (i.e. not intersecting P), sufficient conditions vary but all have a common ground: A and B should be convex. If we replace $\mathbb{R}^3$ by $\mathbb{R}^n$, the natural question becomes to separate A and B by a hyperplane, i.e. an affine subspace of dimension $n-1$.

The main subject of this lecture is to prove such a separation-by-a-hyperplane theorem for convex disjoint subsets of infinite dimensional topological vector spaces. At first this sounds as a very abstract, almost nonsensical topic. But as we will see, the so called Hahn-Banach separation theorem reveals deep properties about the weak topology of Banach spaces. This will be applied to prove striking results in group theory.

9.1 The Hahn-Banach separation theorem

Theorem 9.1 (Hahn-Banach separation theorem)**.** *Let X be a topological vector space. If A and B are nonempty disjoint convex subsets of X and if A is open, there exists a continuous linear functional $\omega : X \to \mathbb{C}$ and a number $\alpha \in \mathbb{R}$ such that*

$$\mathrm{Re}(\omega(a)) < \alpha \leq \mathrm{Re}(\omega(b)) \quad \textit{for all } \ a \in A, b \in B \ .$$

The Hahn-Banach separation theorem may seem more innocent than it really is. Given a topological vector space, it is not obvious to find open convex subsets. Indeed, there exist topological vector spaces such that the only open convex subsets are $\emptyset$ and X. For such topological vector spaces, the statement of the Hahn-Banach separation theorem is empty. Remark also that such topological vector spaces have only one continuous functional $\omega : X \to \mathbb{C}$, namely $\omega = 0$. But, if $(X, \mathcal{P})$ is a seminormed space with its seminorm topology, there are plenty of open convex sets, e.g. the sets given by

$$\mathcal{U} = \{x \in X \mid p_1(x) < \varepsilon_1, \dots, p_n(x) < \varepsilon_n\} \tag{9.1}$$

whenever $p_1, \dots, p_n \in \mathcal{P}$ and $\varepsilon_1, \dots, \varepsilon_n > 0$. The relation between open convex sets and seminorms goes much further, see the exercises on page 92.

Before proving the Hahn-Banach separation theorem, we need a number of preliminary results. The main point in the proof will be an application of the Hahn-Banach extension theorem 5.6. So, we will need to construct sublinear maps on X. Such sublinear maps (called Minkowski functionals in [Ped]) are provided by the following result.

Lemma 9.2. *Let X be a topological vector space and $\mathcal{U}$ an open convex neighborhood of 0. Define for every $x \in X$,*

$$m(x) = \inf\{t \mid t \geq 0, t^{-1}x \in \mathcal{U}\} .$$

Then, $m(x) \in [0, +\infty)$ for all $x \in X$. Moreover, m is a sublinear map on X in the sense of Definition 5.5. Finally,

$$\mathcal{U} = \{x \in X \mid m(x) < 1\} .$$

Proof. Prove yourself this lemma according to the following steps. Details can be found in [Ped, 2.4.6].

1. Let $x \in X$ and $t > 0$ such that $t^{-1}x \in \mathcal{U}$. Prove that $s^{-1}x \in \mathcal{U}$ for all $s \geq t$.

2. Let $x \in X$. Prove that $n^{-1}x \to 0$ as $n \to +\infty$. Deduce that $m(x) < \infty$.

3. Prove that $m(sx) = sm(x)$ for all $x \in X$ and $s \geq 0$.

4. For all $s, t > 0$ and $x, y \in X$ one has

$$(s+t)^{-1}(x+y) = \frac{s}{s+t} s^{-1}x + \frac{t}{s+t} t^{-1}y .$$

   Deduce that $m(x+y) \leq m(x) + m(y)$.

5. Let $x \in \mathcal{U}$. Use the fact that $\mathcal{U}$ is open to find $\varepsilon > 0$ satisfying $(1+\varepsilon)x \in \mathcal{U}$. Deduce that $m(x) < 1$.

6. Conversely, prove that if $m(x) < 1$, then $x \in \mathcal{U}$.

□

Example 9.3. Let X be a normed space and let B be the open unit ball, i.e. $B := \{x \in X : \|x\| < 1\|$. Then the associated Minkowski functional is equal to the norm.

We also need the following characterization of continuous functionals on a topological vector space. We formulate the result over the field $\mathbb{R}$, which is the version that we will apply below, but the same result holds of course over the field $\mathbb{C}$.

Lemma 9.4. *Let X be a topological vector space over $\mathbb{R}$ and $\omega : X \to \mathbb{R}$ a linear functional. Then, the following statements are equivalent.*

(i) *ω is continuous.*

(ii) *ω is continuous in 0.*

(iii) *There exists a neighborhood $\mathcal{U}$ of 0 such that $|\omega(x)| \leq 1$ for all $x \in \mathcal{U}$.*

Proof. (i) $\Longrightarrow$ (ii). Obvious.

(ii) $\Longrightarrow$ (iii). Since ω is continuous in 0, since $\omega(0) = 0$ and since the unit interval $[-1, 1]$ is a neighborhood of 0 in $\mathbb{R}$, there exists a neighborhood $\mathcal{U}$ of 0 in X such that $\omega(\mathcal{U}) \subset [-1, 1]$. This means that $|\omega(x)| \leq 1$ for all $x \in \mathcal{U}$.

(iii) $\Longrightarrow$ (i). Take a neighborhood $\mathcal{U}$ of 0 in X such that $|\omega(x)| \leq 1$ for all $x \in \mathcal{U}$. Choose $x_0 \in X$ and $\varepsilon > 0$. Put $\mathcal{V} := x_0 + \varepsilon\mathcal{U}$. Check that $\mathcal{V}$ is a neighborhood of x_0 in X and that $|\omega(x) - \omega(x_0)| \leq \varepsilon$ for all $x \in \mathcal{V}$. So ω is continuous at x_0 for every $x_0 \in X$. □

Lemma 9.5. *Let X be a vector space over $\mathbb{R}$ and $\varphi : X \to \mathbb{R}$ a nonzero linear functional. Let $\mathcal{U} \subset X$ be a nonempty convex subset. Then $\varphi(\mathcal{U})$ is a nonempty interval. If X is a topological vector space and $\mathcal{U}$ is open, then $\varphi(\mathcal{U})$ is an open interval.*

Proof. Also the proof of this lemma is left as an exercise.

(i) Prove that $\varphi(\mathcal{U})$ is convex.

(ii) Prove that intervals (open, closed, half open, etc.) are the only convex subsets of $\mathbb{R}$. Deduce that $\varphi(\mathcal{U})$ is a nonempty interval whenever $\mathcal{U} \subset X$ is nonempty and convex.

(iii) Finally assume that X is topological and that $\mathcal{U}$ is open. Fix $a \in X$ such that $\varphi(a) \neq 0$. Choose $x \in \mathcal{U}$. Prove the existence of $\varepsilon > 0$ such that $x + sa \in \mathcal{U}$ for all $s \in (-\varepsilon, \varepsilon)$. Deduce that $\varphi(x)$ lies in the interior of $\varphi(\mathcal{U})$. Conclude that $\varphi(\mathcal{U})$ is open.

□

Contemplating about the previous lemma it might sound strange that we make a statement about a *topological* vector space without assuming continuity of φ. Actually we only used the following algebraic weakening of $\mathcal{U}$ being open: for every $x \in \mathcal{U}$ and every $a \in X$, there exists an $\varepsilon > 0$ such that $x + sa \in \mathcal{U}$ for all $s \in (-\varepsilon, \varepsilon)$.

We are now ready to prove the Hahn-Banach separation theorem.

Proof of Theorem 9.1. Give a proof yourself according to the following steps. Details can be found in [Ped, 2.4.7].

(i) Fix $a_0 \in A$ and $b_0 \in B$. Put $c_0 := b_0 - a_0$ and $C := A - B + c_0$. Prove that C is convex. Write

$$C = \bigcup_{b \in B} (A - b + c_0)$$

and conclude that C is an open neighborhood of 0. Denote by m the sublinear map associated with C as in Lemma 9.2.

(ii) Check that $m(c_0) \geq 1$.

(iii) Define the linear map $\varphi_0 : \mathbb{R}c_0 \to \mathbb{R} : \varphi_0(sc_0) = s$. Check that $\varphi_0(x) \leq m(x)$ for all $x \in \mathbb{R}c_0$. Apply the Hahn-Banach extension theorem 5.6 to get a linear map $\varphi : X \to \mathbb{R}$ satisfying $\varphi(x) \leq m(x)$ for all $x \in X$ and $\varphi(c_0) = 1$.

(iv) Prove that $|\varphi(x)| \leq 1$ whenever $x \in C \cap (-C)$. Use Lemma 9.4 to deduce that φ is continuous.

(v) Use the fact that $m(x) < 1$ for all $x \in C$ to deduce that $\varphi(a) < \varphi(b)$ for all $a \in A, b \in B$.

(vi) Use Lemma 9.5 to find $\alpha \in \mathbb{R}$ such that $\varphi(a) < \alpha \leq \varphi(b)$ for all $a \in A, b \in B$.

(vii) Define $\omega : X \to \mathbb{C} : \omega(x) = \varphi(x) - i\varphi(ix)$ to conclude the proof of the theorem.

□

Corollary 9.6. *Let X be a seminormed space with its seminorm topology. Suppose that A and B are nonempty disjoint convex subsets of X such that A is compact and B is closed. Then, there exists a continuous linear map $\omega : X \to \mathbb{C}$ and numbers $\alpha_1, \alpha_2 \in \mathbb{R}$ such that*

$$\mathrm{Re}(\omega(a)) \leq \alpha_1 < \alpha_2 \leq \mathrm{Re}(\omega(b)) \quad \textit{for all} \;\; a \in A, b \in B \; .$$

Proof. We claim that there exists an open convex subset $A_1 \subset X$ such that $A \subset A_1$ and $A_1 \cap B = \emptyset$. Indeed, for every $a \in A$, we find by (9.1), an open convex neighborhood $\mathcal{U}_a$ of 0 such that $(a + \mathcal{U}_a + \mathcal{U}_a) \cap B = \emptyset$. Take $a_1, \ldots, a_n$ such that $A \subset \bigcup_{i=1}^n (a_i + \mathcal{U}_{a_i})$. Define $\mathcal{U} = \mathcal{U}_{a_1} \cap \cdots \cap \mathcal{U}_{a_n}$. Then, $\mathcal{U}$ is an open convex neighborhood of 0 and $(A + \mathcal{U}) \cap B = \emptyset$. Check this and check that we can take $A_1 = A + \mathcal{U}$.

By Theorem 9.1, we find a continuous linear map $\omega : X \to \mathbb{C}$ and a real number α_2 such that

$$\mathrm{Re}(\omega(a)) < \alpha_2 \leq \mathrm{Re}(\omega(b)) \quad \text{for all} \;\; a \in A_1, b \in B \; .$$

Set $\alpha_1 = \sup\{\mathrm{Re}(\omega(a)) \mid a \in A\}$. Since A is compact and ω continuous, the supremum α_1 is attained in some $a \in A$. So, $\alpha_1 < \alpha_2$ and we are done. □

Our final corollary to the Hahn-Banach extension theorem may seem very abstract, but it has far reaching consequences. The idea is the following: if in a Banach space X, we can approximate an element x *in the weak topology* by elements of some subset $A \subset X$, then we can approximate x *in norm* by elements of $\mathrm{conv}(A)$, where

$$\mathrm{conv}(A) = \left\{ \sum_{i=1}^n t_i a_i \;\middle|\; n \in \mathbb{N}_0, t_i \in [0,1] \text{ and } a_i \in A \;\text{ for all }\; i = 1, \ldots, n \; , \; \sum_{i=1}^n t_i = 1 \right\} .$$

In words, $\mathrm{conv}(A)$ is the smallest convex subset of X that contains A.

Corollary 9.7. *Let X be a Banach space and $A \subset X$ a convex subset that is closed in the norm topology. Then, A is closed in the weak topology.*

Also, if $A \subset X$ and if x belongs to the weak closure of A, then x belongs to the norm closure of $\mathrm{conv}(A)$.

Proof. Take $x \in X \setminus A$. We have to find a weak neighborhood $\mathcal{U}$ of x such that $\mathcal{U} \cap A = \emptyset$. Apply Corollary 9.6 to the Banach space X equipped with the norm topology. Since $\{x\}$ is compact and convex, while A is convex and closed, we can take $\omega \in X^*$ and $\alpha \in \mathbb{R}$ such that

$$\mathrm{Re}(\omega(x)) < \alpha \leq \mathrm{Re}(\omega(a)) \quad \text{for all} \;\; a \in A \; .$$

Defining $\mathcal{U} = \{y \in X \mid \mathrm{Re}(\omega(y)) < \alpha\}$, we are done. □

9.2 Linear functionals that are continuous for a weak topology

The Hahn-Banach separation theorem can only be interpreted in a concrete situation if we know which are the continuous functionals on X. In this course, all concrete topological vector spaces are seminormed spaces and most of the time, they are even defined as follows.

Let X be a vector space and $\mathcal{F}$ a *faithful family of linear maps* $\omega : X \to \mathbb{C}$, meaning that $\bigcap_{\omega \in \mathcal{F}} \operatorname{Ker} \omega = \{0\}$. Define for every $\omega \in \mathcal{F}$ the seminorm p_ω on X as $p_\omega(x) = |\omega(x)|$ for all $x \in X$. Set $\mathcal{P} = \{p_\omega \mid \omega \in \mathcal{F}\}$. Then, $(X, \mathcal{P})$ is a seminormed space and hence, a topological vector space when equipped with the seminorm topology.

This brings us to the following natural question: describe the linear maps $X \to \mathbb{C}$ that are continuous for the seminorm topology described above. The answer to this problem is surprisingly simple: these are exactly the linear maps belonging to $\operatorname{span} \mathcal{F}$. This is the contents of the next proposition and is concretely interpreted below in Example 9.9.

Proposition 9.8. *Let X be a vector space, $\mathcal{F}$ a faithful family of linear functionals from X to $\mathbb{C}$ and $\mathcal{P}$ the associated family of seminorms on X. If we equip X with the seminorm topology given by $\mathcal{P}$, a linear functional $\omega : X \to \mathbb{C}$ is continuous if and only if $\omega \in \operatorname{span} \mathcal{F}$.*

Proof. It is straightforward to check that functionals in $\operatorname{span} \mathcal{F}$ are continuous. Suppose conversely that $\omega : X \to \mathbb{C}$ is continuous. Take $\omega_1, \ldots, \omega_n \in \mathcal{F}$ and $\varepsilon > 0$ such that $|\omega(x)| \leq 1$ whenever $|\omega_i(x)| < \varepsilon$ for all $i = 1, \ldots, n$. Use the linearity of $\omega, \omega_1, \ldots, \omega_n$ to conclude that $\omega(x) = 0$ whenever $\omega_i(x) = 0$ for all $i = 1, \ldots, n$. Define the linear map

$$\pi : X \to \mathbb{C}^n : \pi(x) = (\omega_1(x), \ldots, \omega_n(x)) .$$

Set $K = \pi(X)$. Then, K is a vector subspace of $\mathbb{C}^n$. Also, $\operatorname{Ker} \pi \subset \operatorname{Ker} \omega$, allowing to define the linear map $\mu : K \to \mathbb{C}$ such that $\omega(x) = \mu(\pi(x))$ for all $x \in X$. Extending μ to a linear functional on the whole of $\mathbb{C}^n$, we find $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that $\mu(z_1, \ldots, z_n) = \sum_{i=1}^n \lambda_i z_i$ for all $(z_1, \ldots, z_n) \in K$. But this means that $\omega = \sum_{i=1}^n \lambda_i \omega_i$, proving that $\omega \in \operatorname{span} \mathcal{F}$. □

Example 9.9. We apply Proposition 9.8 to the following more concrete cases.

- If X is a Banach space, a linear functional $X \to \mathbb{C}$ is continuous for the weak topology if and only if it is continuous for the norm topology.
- If X^* is the dual of a Banach space X, a linear functional $\theta : X^* \to \mathbb{C}$ is continuous for the weak* topology if and only if there exists $x \in X$ with $\theta(\omega) = \omega(x)$ for all $\omega \in X^*$.

9.3 Exercises

Exercise 1. Use the hints below to prove the following theorem (Goldstine's theorem).

Let X be a Banach space and $i : X \to X^{**}$ the isometry given by $i(x)(\omega) = \omega(x)$ introduced in Corollary 5.3. Then, the image $i((X)_1)$ of the unit ball $(X)_1$ in X is weak dense in the unit ball $(X^{**})_1$ of X^{**}.*

Hints. Equip X^{**} with its weak* topology. Denote by K the weak* closure of $i((X)_1)$. Suppose that $\theta \in X^{**} \setminus K$. You have to prove that $\|\theta\| > 1$. Apply Corollary 9.6 to the closed convex subset K and the compact convex subset $\{\theta\}$. Use Example 9.9 to prove that $\|\theta\| > 1$.

Exercise 2. We know that in reflexive Banach spaces the unit ball is weakly compact (by the Banach-Alaoglu theorem). Prove the converse, according to the following steps:

(i) Check that the embedding $\iota\colon X \to X^{**}$ is continuous when X is endowed with its weak topology and $X^{**} = (X^*)^*$ with its weak* topology.

(ii) Deduce that the image of the unit ball $\iota((X)_1)$ is compact in the weak* topology, hence closed.

(iii) Use Exercise 1 to show that $\iota((X)_1) = (X^{**})_1$.

(iv) Conclude that $\iota(X) = X^{**}$.

The following is a series of exercises that leads to a proof of the equivalence of the following two statements about a topological vector space X:

- the topological vector space X is a seminormed space; this means that there exists a family $\mathcal{P}$ of seminorms on X such that the topology on X coincides with the seminorm topology defined by $\mathcal{P}$;
- the topological vector space X is locally convex; this means[1] that every neighborhood $\mathcal{U}$ of 0 contains a convex neighborhood of 0.

Exercise 3. Let $(X, \mathcal{P})$ be a seminormed space and consider on X the seminorm topology. Let $\mathcal{U}$ be a neighborhood of 0. Prove that there exist $p_1, \ldots, p_n \in \mathcal{P}$ and an $\varepsilon > 0$ such that

$$\{x \in X \mid p_i(x) < \varepsilon \ \text{ for all } \ i = 1, \ldots, n\}$$

is a convex neighborhood of 0 contained in $\mathcal{U}$.

So, you have shown that every seminormed space is locally convex.

Exercise 4. Let X be a topological vector space and $\mathcal{U}$ an open convex neighborhood of 0. Define the associated sublinear map m as in Lemma 9.2. Prove that m is a seminorm if and only if $\lambda\mathcal{U} = \mathcal{U}$ for all $\lambda \in \mathbb{C}$ with $|\lambda| = 1$.

We call a convex neighborhood $\mathcal{U}$ of 0 *balanced* if $\lambda\mathcal{U} = \mathcal{U}$ for all $\lambda \in \mathbb{C}$ with $|\lambda| = 1$.

Exercise 5. Let X be a topological vector space.

1. Use the compactness of the circle $S^1 := \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$ to prove that

$$\mathcal{V} = \bigcap_{\lambda \in S^1} \lambda\mathcal{U}$$

is a balanced convex neighborhood of 0 whenever $\mathcal{U}$ is a convex neighborhood of 0.

2. Prove that the interior of a balanced convex neighborhood of 0 is still a balanced convex neighborhood of 0.

[1] In the terminology of general topology, this means that the topology on X has a basis consisting of convex open sets.

3. Assume that X is locally convex. Prove that every neighborhood of 0 contains an open balanced convex neighborhood of 0.

Exercise 6. Let X be a locally convex topological vector space. Define $\mathcal{O}$ as the set of all open balanced convex neighborhoods of 0. Define for every $\mathcal{U} \in \mathcal{O}$ the map

$$p_{\mathcal{U}}(x) = \inf\{t \in [0, +\infty) \mid t^{-1}x \in \mathcal{U}\} .$$

By Exercise 4, $\mathcal{P} = \{p_{\mathcal{U}} \mid \mathcal{U} \in \mathcal{O}\}$ is a family of seminorms on X. Use Exercise 5 to prove that the seminorm topology on X defined by $\mathcal{P}$ coincides with the original topology on X.

Lecture 10

The Krein-Milman theorem

In this lecture we make a careful study of *compact convex sets in a seminormed space.* Keeping in mind the intuition about convex sets in the plane, the notion of an *extreme point* is probably not so surprising: we say that a point in a convex subset K of a vector space X is extreme if the point does not lie on an interval contained in K. The Krein-Milman theorem says that a compact convex set K has "enough" extreme points, in the sense that K can be retrieved as the closure of the convex hull of its extreme points.

As an application we prove a result in the representation theory of groups: all groups have "enough" irreducible representations.

10.1 The Krein-Milman theorem

Definition 10.1. Let X be a vector space and $K \subset X$ a convex subset. We say that $x \in K$ is an *extreme point of* K if the following holds: whenever $x = ty + (1-t)z$ with $y, z \in K$ and $t \in (0, 1)$, we have $x = y = z$.

The set of extreme points of K is denoted as $\operatorname{ext} K$.

A non-empty subset $F \subset K$ is called a *face* if the following condition holds: whenever $ty + (1-t)z \in F$ with $y, z \in K$ and $t \in (0, 1)$ then $y, z \in F$. We usually assume that the faces are convex; if not mentioned otherwise, the word face will always mean a convex face.

Remark 10.2. An extreme point is a face that consists of a single point.

If you want to find a maximum of convex function defined on a convex set, knowing the extreme points might be very helpful.

Exercise 1. Let X be a vector space and let $K \subset X$ be a convex subset. Suppose that $f : K \to \mathbb{R}$ is a convex function that attains a maximum. Show that the set of points, where f attains a maximum is a (not necessarily convex) face of K. In particular, if the maximum is unique, it is attained at an extreme point.

Remark 10.3. In many cases (see the proof of Theorem 10.4) one can find an extreme point inside a face, so even in the case of a non-unique maximum, one can often conclude that there is an extreme point at which the maximum is attained.

The central question of this section is the following: do convex subsets always have extreme points? If we think about convex subsets of the plane, a few natural conditions pop up: an open convex set never has extreme points and also a line in the plane has no extreme points. But a convex subset K of the plane seems to have extreme points whenever it is closed and bounded, i.e. compact. Moreover, in that case, you can recover K by taking convex combinations of its extreme points.

Such a geometric intuition holds in arbitrary seminormed spaces and that is the content of the Krein-Milman theorem. The way that we find extreme points in the proof of the Krein-Milman theorem is by proving that minimal closed faces consist of a single point. In order to do that, we need a bit more information about faces.

Exercise 2. Let X be a vector space and $K \subset X$ a convex subset. Assume that $F_1 \subset K$ is a face and that $F_2 \subset F_1$ is a face (of F_1). Prove that F_2 is a face of K.

Exercise 3. Let X be a seminormed space and let K be a convex compact subset. Let $\varphi\colon X \to \mathbb{R}$ be a bounded linear functional. Prove that $F_\varphi := \{x \in K : \varphi(x) = \sup_{y \in K} \varphi(y)\}$ is a closed face of K.

Theorem 10.4 (Krein-Milman Theorem)**.** *Let X be a seminormed space and $K \subset X$ a nonempty compact convex subset. Then, K equals the closure of* $\operatorname{conv}(\operatorname{ext} K)$.

Proof. The proof will consist of two parts. In the first one we will show that extreme points exist. In the second one we will upgrade this statement to $K = \overline{\operatorname{conv}}(\operatorname{ext} K)$.

Let $\mathcal{F}$ denote the set of closed faces of K, ordered by reverse inclusion. We claim that there exists a maximal element of this set, a minimal closed face. In order to apply Zorn's lemma, we have to show that any decreasing family of faces has a lower bound. By compactness, the intersection will be non-empty and you should check that it is a face. Therefore we get a minimal face F and our aim is to show that it consists of a single point. If there are two distinct points $x, y \in F$, then we can find a functional $\varphi\colon X \to \mathbb{R}$ such that $\varphi(x) \neq \varphi(y)$, using the Hahn-Banach theorem. Define now $F_1 := \{x \in F : \varphi(x) = \sup_{y \in F} \varphi(y)\}$. By Exercise 3 F_1 is a face of F, hence a face of K by Exercise 2. But $F_1 \neq F$, because at most one of the points x and y belongs to F_1. It contradicts the minimality of F, so $F = \{x\}$ for some point $x \in K$; we thus have an extreme point.

Let us consider the set $K_1 := \overline{\operatorname{conv}}(\operatorname{ext} K)$. It is a compact convex set. If $K_1 \neq K$ then there exists a point $y \in K \setminus K_1$. Using the Hahn-Banach separation theorem 9.6, we can find a functional $\varphi\colon X \to \mathbb{R}$ such that $\sup_{x \in K_1} \varphi(x) < \varphi(y)$. Now define a face F_φ as in Exercise 3. By the first part of the proof, we can find an extreme point belonging to F_φ. But $F_\varphi \cap K_1 = \emptyset$, so K_1 did not contain all of the extreme points of K; this is a contradiction. □

In applications, as for instance in the proof of the Ryll-Nardzewski fixed point theorem below, it is often not sufficient just to know that a compact convex set has a lot of extreme points, but it is often also needed to locate the extreme points. The following theorem provides a tool for that.

Theorem 10.5. *Let X be a seminormed space and $K \subset X$ a nonempty convex compact subset of X. If $F \subset K$ is closed and if K is equal to the closure of* $\operatorname{conv}(F)$*, then* $\operatorname{ext} K \subset F$.

Proof. Suppose by contradiction that $x \in \operatorname{ext}(K) \setminus F$. Since F is closed, we can take a neighborhood V of 0 in X such that V is closed convex and $x \notin F + V$. Since F is compact, take $y_1, \ldots, y_n \in F$ such that

$$F \subset \bigcup_{k=1}^{n} (y_k + V)\,.$$

Define K_k as the closure of $\operatorname{conv}(F \cap (y_k + V))$. Since $y_k + V$ is closed and convex, it follows that $K_k \subset y_k + V$ for all k. Moreover, $F \subset \bigcup_{k=1}^n K_k$, implying that K is the closure of $\operatorname{conv}(K_1 \cup \cdots \cup K_n)$. Because the K_n are compact convex, the set $\operatorname{conv}(K_1 \cup \cdots \cup K_n)$ is already closed and we conclude that

$$K = \operatorname{conv}(K_1 \cup \cdots \cup K_n) \ .$$

Write $x = \sum_{k=1}^n t_k x_k$ with $x_k \in K_k$, $t_k \in [0,1]$ and $\sum_{k=1}^n t_k = 1$. Since x is an extreme point of K, there exists a k such that $x = x_k$. It follows that $x \in K_k \subset y_k + V \subset F + V$. This yields the required contradiction. □

10.2 Irreducible representations of groups

Definition 10.6. A *unitary representation* of a group G on a Hilbert space H is a homomorphism of G to the group $\mathcal{U}(H)$ of unitary operators on H. In other words, it is a map $\pi : G \to \mathcal{U}(H)$ satisfying $\pi(e) = 1$ and $\pi(gh) = \pi(g)\pi(h)$ for all $g, h \in G$.

Example 10.7. Every group has at least the following two unitary representations.

- The *trivial representation* on the one dimensional Hilbert space $H = \mathbb{C}$ given by $\pi(g) = 1$ for all $g \in G$.
- The *regular representation* on the Hilbert space $H = \ell^2(G)$ given by $\pi(g)\xi = \xi \cdot g^{-1}$, i.e. $(\pi(g)\xi)(h) = \xi(g^{-1}h)$ for all $g, h \in G$, $\xi \in \ell^2(G)$.

Often unitary representations are built up out of subrepresentations. More precisely, assume that $\pi_i : G \to \mathcal{U}(H_i)$ are two unitary representations of G. Then we can construct the unitary representation $\pi_1 \oplus \pi_2$ of G on the Hilbert space $H_1 \oplus H_2$ given by $(\pi_1 \oplus \pi_2)(g) : \xi_1 \oplus \xi_2 \mapsto \pi_1(g)\xi_1 \oplus \pi_2(g)\xi_2$.

Unitary representations that cannot be broken up into a direct sum of two subrepresentations, are called *irreducible*. The precise definition goes as follows.

Definition 10.8. Let $\pi : G \to \mathcal{U}(H)$ be a unitary representation of the group G.

- A vector subspace $K \subset H$ is called a π*-invariant subspace* if $\pi(g)K = K$ for all $g \in G$.
- The representation π is called *irreducible* if $\{0\}$ and H are the only π-invariant closed vector subspaces of H.

Obviously the trivial representation is irreducible, since $\{0\}$ and $\mathbb{C}$ simply are the only vector subspaces of $\mathbb{C}$. A priori it is not clear that a group admits other irreducible representations. The fact that they always do, is a consequence of the Krein-Milman theorem, as we shall see below.

Exercise 4. Prove that the direct sum of two unitary representations π_1, π_2 (with $H_i \neq \{0\}$) is never irreducible.

Conversely prove that if $\pi : G \to \mathcal{U}(H)$ is a unitary representation and $K \subset H$ a closed π-invariant vector subspace, then π can be written as the direct sum of two unitary representations, on K and $K^\perp$ respectively.

10.3 Unitary representations and positive definite functions

Let G be a group and $\pi : G \to \mathcal{U}(H)$ a unitary representation. Whenever $\xi \in H$, the function

$$\varphi_\xi : G \to \mathbb{C} : \varphi_\xi(g) := \langle \pi(g)\xi, \xi \rangle$$

has a very special property. Indeed, whenever $g_i \in G$ and $\lambda_i \in \mathbb{C}$, we have

$$\begin{aligned}\sum_{i,j=1}^{n} \lambda_i \overline{\lambda_j}\, \varphi_\xi(g_j^{-1} g_i) &= \sum_{i,j=1}^{n} \lambda_i \overline{\lambda_j}\, \langle \pi(g_j^{-1} g_i)\xi, \xi \rangle \\ &= \sum_{i,j=1}^{n} \lambda_i \overline{\lambda_j}\, \langle \pi(g_j)^* \pi(g_i)\xi, \xi \rangle \\ &= \langle \sum_{i=1}^{n} \lambda_i \pi(g_i)\xi, \sum_{j=1}^{n} \lambda_j \pi(g_j)\xi \rangle \\ &= \Big\| \sum_{i=1}^{n} \lambda_i \pi(g_i)\xi \Big\|^2 \\ &\geq 0 \; .\end{aligned}$$

Functions from G to $\mathbb{C}$ satisfying this property are called *positive definite.*

Definition 10.9. Let G be a group. A function $\varphi : G \to \mathbb{C}$ is called *positive-definite* if for all $n \in \mathbb{N}$, all $g_1, \ldots, g_n \in G$ and all $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ we have

$$\sum_{i,j=1}^{n} \overline{\lambda_j} \lambda_i \varphi(g_j^{-1} g_i) \geq 0 \; . \tag{10.1}$$

This exactly means that the matrix $(\varphi(g_j^{-1} g_i))_{i,j}$ is positive (semi)definite.

We have seen above that unitary representations give rise to positive definite functions φ_ξ. We now prove that also the converse holds. Before we can have any chance to prove this statement, we have to check that every positive definite function φ on G satisfies $\varphi(g^{-1}) = \overline{\varphi(g)}$ for all $g \in G$. Indeed, the positive definite functions of the form φ_ξ satisfy

$$\varphi_\xi(g^{-1}) = \langle \pi(g^{-1})\xi, \xi \rangle = \langle \pi(g)^*\xi, \xi \rangle = \langle \xi, \pi(g)\xi \rangle = \overline{\langle \pi(g)\xi, \xi \rangle} = \overline{\varphi_\xi(g)} \; .$$

Lemma 10.10. *Let $\varphi : G \to \mathbb{C}$ be a positive definite function on the group G. Then for all $g \in G$ we have*

- $\varphi(g^{-1}) = \overline{\varphi(g)}$,
- $\varphi(e) \geq 0$ *and* $|\varphi(g)| \leq \varphi(e)$.

Proof. Choose $g_1 = e$ and $g_2 = g$. As φ is positive definite, the matrix $\begin{bmatrix} \varphi(e) & \varphi(g^{-1}) \\ \varphi(g) & \varphi(e) \end{bmatrix}$ is positive semidefinite. It follows that the diagonal entries are non-negative, i.e. $\varphi(e) \leqslant 0$, that the matrix is Hermitian, i.e. $\varphi(g^{-1}) = \overline{\varphi(g)}$, and that the determinant is non-negative, i.e. $|\varphi(g)|^2 = \varphi(g)\varphi(g^{-1}) \leqslant \varphi(e)^2$. □

Proposition 10.11. *Let $\varphi : G \to \mathbb{C}$ be a positive-definite function on a countable group G. There exists a Hilbert space H_φ, a unitary representation $\pi_\varphi : G \to \mathcal{U}(H_\varphi)$ and a vector $\xi_\varphi \in H_\varphi$ such that*

$$\varphi(g) = \langle \pi_\varphi(g)\xi_\varphi, \xi_\varphi \rangle \ \textit{for all} \ g \in G \quad \textit{and} \quad \operatorname{span}\{\pi_\varphi(g)\xi_\varphi \mid g \in G\} \ \textit{is dense in} \ H_\varphi. \tag{10.2}$$

Proof. Denote by $\mathrm{Fun}_0(G)$ the vector space of finitely supported functions from G to $\mathbb{C}$. For all $\xi, \eta \in \mathrm{Fun}_0(G)$, we define

$$\langle \xi, \eta \rangle = \sum_{g,h \in G} \overline{\eta(h)}\, \xi(g)\, \varphi(h^{-1}g) \ .$$

By (10.1) we have that $\langle \xi, \xi \rangle \geq 0$ for all $\xi \in \mathrm{Fun}_0(G)$. By Lemma 10.10 we get that $\langle \eta, \xi \rangle = \overline{\langle \xi, \eta \rangle}$. So we have defined a positive Hermitian form on $\mathrm{Fun}_0(G)$.

Denote by $\mathrm{Fun}_{00}(G) \subset \mathrm{Fun}_0(G)$ the vector subspace consisting of the functions $\xi \in \mathrm{Fun}_0(G)$ satisfying $\langle \xi, \xi \rangle = 0$. The same formula as above defines a positive-definite Hermitian form on the quotient $\mathrm{Fun}_0(G)/\mathrm{Fun}_{00}(G)$. Denote by $\| \cdot \|$ the associated norm. Define H_φ to be the completion of $\mathrm{Fun}_0(G)/\mathrm{Fun}_{00}(G)$ with respect to the norm $\| \cdot \|$. Define ξ_φ to be the vector in H_φ that corresponds to the function in $\mathrm{Fun}_0(G)$ that equals 1 on e and 0 elsewhere.

For every $g \in G$ and $\xi \in \mathrm{Fun}_0(G)$ define $\xi \cdot g \in \mathrm{Fun}_0(G)$ as the translated function $(\xi \cdot g)(h) = \xi(gh)$. One checks easily that $\langle \xi \cdot g, \xi \cdot g \rangle = \langle \xi, \eta \rangle$ for all $\xi, \eta \in \mathrm{Fun}_0(G)$ and $g \in G$. Therefore the formula

$$\pi_\varphi(g)(\xi + \mathrm{Fun}_{00}(G)) = \xi \cdot g^{-1} + \mathrm{Fun}_{00}(G)$$

provides a well defined linear operator $\pi_\varphi(g)$ on the vector space $\mathrm{Fun}_0(G)/\mathrm{Fun}_{00}(G)$. Moreover $\pi_\varphi(g)$ preserves the norm $\| \cdot \|$ so that $\pi_\varphi(g)$ extends in a unique way to an isometric operator $\pi_\varphi(g) : H_\varphi \to H_\varphi$. By construction π_φ is a unitary representation satisfying (10.2). □

10.4 Application of the Krein-Milman theorem: all groups admit many irreducible representations

We denote by $\mathcal{PD}_1(G)$ the set of all positive-definite functions φ on G satisfying $\varphi(e) = 1$. Note that $\mathcal{PD}_1(G)$ is a convex subset of the vector space $\mathrm{Fun}(G)$.

Proposition 10.12. *Let G be a group and $\varphi : G \to \mathbb{C}$ a positive-definite function satisfying $\varphi(e) = 1$. Assume that φ is an extreme point of $\mathcal{PD}_1(G)$. Then the unitary representation π_φ given by Proposition 10.11 is irreducible.*

Proof. Assume that π_φ is reducible. We have to show that φ is not an extreme point of $\mathcal{PD}_1(G)$. Take a closed vector subspace $K \subset H_\varphi$ such that $\{0\} \neq K \neq H_\varphi$ and $\pi_\varphi(g)K = K$ for all $g \in G$. Denote by p_K the orthogonal projection of H_φ onto K. Check that $\pi_\varphi(g)p_K = p_K\pi_\varphi(g)$ for all $g \in G$. Define the positive-definite functions φ_1 and φ_2 by the formulae

$$\varphi_1(g) = \langle \pi_\varphi(g)p_K\xi_\varphi, \xi_\varphi \rangle \quad \text{and} \quad \varphi_2(g) = \langle \pi_\varphi(g)(1 - p_K)\xi_\varphi, \xi_\varphi \rangle \ .$$

By construction $\varphi(g) = \varphi_1(g) + \varphi_2(g)$. Put $\alpha = \varphi_1(e)$. Note that $1 - \alpha = \varphi_2(e)$.

We prove that $\varphi_1 \neq \alpha\varphi$. Assume the contrary. It follows that for all $g, h \in G$ we have

$$\langle \alpha 1\, \pi_\varphi(g)\xi_\varphi, \pi_\varphi(h)\xi_\varphi\rangle = \alpha\varphi(h^{-1}g) = \varphi_1(h^{-1}g) = \langle p_K\, \pi_\varphi(g)\xi_\varphi, \pi_\varphi(h)\xi_\varphi\rangle \ .$$

The second condition in (10.2) now implies that $p_K = \alpha 1$. This means that either $K = \{0\}$, $\alpha = 0$ or $K = H_\varphi$, $\alpha = 1$. Both are absurd and it follows that $\varphi_1 \neq \alpha\varphi$. We analogously get that $\varphi_2 \neq (1-\alpha)\varphi$.

Note that it also follows that $0 < \alpha < 1$. Indeed, if $\alpha = 0$ we have $\varphi_1(e) = 0$ and Lemma 10.10 implies that $\varphi_1 = 0$. Hence $\varphi = \varphi_2$, which we contradicted above. So, $\alpha \neq 0$. We similarly prove that $\alpha \neq 1$.

Finally the formula

$$\varphi = \alpha\, \frac{1}{\alpha}\varphi_1 \ + \ (1-\alpha)\, \frac{1}{1-\alpha}\varphi_2$$

shows that φ is not an extreme point of $\mathcal{PD}_1(G)$. □

The converse of Proposition 10.12 also holds but is not needed to prove Theorem 10.13 below: if π_φ is irreducible, then φ is an extreme point of $\mathcal{PD}_1(G)$. To prove this converse one needs to prove first Schur's lemma: if π_φ is irreducible and $T \in \mathrm{B}(H_\varphi)$ is an operator satisfying $\pi_\varphi(g)T = T\pi_\varphi(g)$ for all $g \in G$, then T must be a multiple of the identity operator 1.

Theorem 10.13. *Let G be a countable group and $g, h \in G$ two distinct elements in G. There exists an irreducible unitary representation π of G satisfying $\pi(g) \neq \pi(h)$.*

In other words: a countable group admits sufficiently many irreducible representations to distinguish the group elements.

Proof. For every $g \in G$ define the seminorm p_g on $\mathrm{Fun}(G)$ given by $p_g(F) = |F(g)|$. The family of seminorms $\{p_g \mid g \in G\}$ turns $\mathrm{Fun}(G)$ into a seminormed space. We claim that $\mathcal{PD}_1(G) \subset \mathrm{Fun}(G)$ is a compact subset. By Lemma 10.10 we know that $|\varphi(g)| \leq 1$ for all $\varphi \in \mathcal{PD}_1(G)$ and all $g \in G$. Imitating the proof of the Banach-Alaoglu theorem 8.7, one identifies $\mathcal{PD}_1(G)$ with a closed subset of $\prod_{g\in G} D$, where $D \subset \mathbb{C}$ denotes the unit disc. The compactness of $\mathcal{PD}_1(G)$ then follows from Tychonoff's theorem.

We claim that $\mathcal{PD}_1(G)$ admits an extreme point $\varphi \in \mathcal{PD}_1(G)$ satisfying $\varphi(g) \neq 1$. Assuming the contrary the Krein-Milman theorem 10.4 implies that $\varphi(g) = 1$ for all $\varphi \in \mathcal{PD}_1(G)$. Denote by $\lambda : G \to \mathcal{U}(\ell^2(G))$ the regular representation of G as defined in Example 10.7. Let δ_e be the function that is equal to 1 in e and equal to 0 elsewhere. Then the function $\psi(g) = \langle\lambda(g)\delta_e, \delta_e\rangle$ belongs to $\mathcal{PD}_1(G)$ and satisfies $\psi(g) = 0$, contradiction.

So we can take an extreme point φ of $\mathcal{PD}_1(G)$ satisfying $\varphi(g) \neq 1$. Proposition 10.11 provides a unitary representation π_φ of G satisfying (10.2). Proposition 10.12 says that π_φ is irreducible. Since

$$\langle\pi_\varphi(g)\xi_\varphi, \xi_\varphi\rangle = \varphi(g) \neq 1 = \varphi(e) = \langle\xi_\varphi, \xi_\varphi\rangle \ ,$$

it follows that $\pi_\varphi(g) \neq 1$. This proves the theorem. □

10.5 Exercises

Exercise 5. Let H be a Hilbert space and $K = \{x \in H \mid \|x\| \leq 1\}$. Prove that $x \in \operatorname{ext} K$ if and only if $\|x\| = 1$.

Exercise 6. Prove that the unit balls of $L^1([0,1])$ (with $\|\cdot\|_1$) and $c_0(\mathbb{N}) = \{x : \mathbb{N} \to \mathbb{C} \mid \lim_n x(n) = 0\}$ (with $\|\cdot\|_\infty$) have no extreme points (see [Con, V.7.2,V.7.7]).

Lecture 11

Applications to amenability of groups

Recall the following definition of amenability of a countable group G.

Definition 11.1. Let G be a group. An *invariant mean m on G* is a finitely additive probability measure m on G satisfying

$$m(gA) = m(A) \quad \text{for all} \quad g \in G, A \subset G \ .$$

In words, an invariant mean is a translation invariant finitely additive probability measure on the group. A group G that admits an invariant mean is called an *amenable group.*

In this lecture we show the strength of the abstract theorems proven in the previous lectures and obtain a number of deep results on amenability of groups.

11.1 The Markov-Kakutani fixed point theorem

There are quite a few fixed point theorems in mathematics that have remarkable consequences. In *Analysis I and II,* you met the theorem that says that a (strict) contraction on a complete metric space has a unique fixed point. That theorem is used in the Picard iteration method to prove existence and uniqueness of solutions of certain differential equations. But it is also used to prove the inverse function theorem.

In this section, we prove a fixed point theorem of a different nature. As an application, we will study invariant means on abelian groups. In the final "extra" lecture we will prove the more powerful Ryll-Nardzewski fixed point theorem.

Definition 11.2. Let X and Y be vector spaces and $K \subset X$ a convex subset. A map $T : K \to Y$ is called *affine* if

$$T(\alpha x + (1-\alpha)y) = \alpha T(x) + (1-\alpha)T(y) \quad \text{for all } \ \alpha \in [0,1] \ , \ x, y \in K \ .$$

Check that an affine map $T : K \to Y$ satisfies

$$T\Big(\sum_{i=1}^{n} \alpha_i x_i\Big) = \sum_{i=1}^{n} \alpha_i T(x_i)$$

whenever $n \in \mathbb{N}_0$, $\alpha_i \in [0,1]$, $x_i \in K$ for all $i = 1,\ldots,n$ and $\sum_{i=1}^n \alpha_i = 1$.

Obvious examples of affine maps are of course restrictions to K of linear maps from X to Y.

Theorem 11.3 (Markov-Kakutani fixed point theorem)**.** *Let X be a seminormed space and $K \subset X$ a compact convex subset. Let $\mathcal{F}$ be a family of continuous affine maps from K to K. Assume that $ST = TS$ for all $S, T \in \mathcal{F}$. Then, there exists an $x_0 \in K$ such that $T(x_0) = x_0$ for all $T \in \mathcal{F}$.*

Proof. Prove yourself this theorem according to the following steps. Details can be found in [Con, V.10.1].

Whenever $T \in \mathcal{F}$ and $n \in \mathbb{N}_0$, we write $T^{(n)} : K \to K$ given by

$$T^{(n)}(x) := \frac{1}{n}\sum_{k=0}^{n-1} T^k(x) ,$$

where T^k denotes the k-fold composition $T^k = T \circ \cdots \circ T$ and where by convention $T^0(x) = x$ for all $x \in K$.

1. Check that $T^{(n)}$ indeed maps K into K. Check that $T^{(n)}S^{(m)} = S^{(m)}T^{(n)}$ for all $S, T \in \mathcal{F}$ and $n, m \in \mathbb{N}_0$.

2. Consider the family of sets $\mathcal{K} := \{T^{(n)}(K) \mid T \in \mathcal{F}, n \in \mathbb{N}_0\}$. Prove that each set in $\mathcal{K}$ is compact.

3. Observe that for all $T_1, \ldots, T_p \in \mathcal{F}$ and $n_1, \ldots, n_p \in \mathbb{N}_0$ we have that

$$T_1^{(n_1)} \cdots T_p^{(n_p)}(K) \subset \bigcap_{k=1}^{p} T_k^{(n_k)}(K) .$$

   Deduce that the family $\mathcal{K}$ has the finite intersection property. Deduce the existence of $x_0 \in K$ such that $x_0 \in T^{(n)}(K)$ for all $T \in \mathcal{F}$ and all $n \in \mathbb{N}_0$.

4. It remains to prove that $T(x_0) = x_0$ for all $T \in \mathcal{F}$. So fix $T \in \mathcal{F}$ and $n \in \mathbb{N}_0$. Write $x_0 = T^{(n)}(x)$ for some $x \in K$ and prove that

$$T(x_0) - x_0 = \frac{1}{n}(T^n(x) - x) .$$

   Deduce that $T(x_0) - x_0 \in \frac{1}{n}(K - K)$ for all $n \in \mathbb{N}_0$.

5. Use the continuity of the map $X \times X \to X : (x, y) \mapsto x - y$ to prove that $K - K$ is a compact subset of X.

6. Prove as follows that $T(x_0) - x_0 \in \mathcal{U}$ for every $T \in \mathcal{F}$ and every convex open neighborhood of 0. Fix such T and $\mathcal{U}$. Prove that $(n\mathcal{U})_{n \in \mathbb{N}_0}$ is an open covering of X. Deduce the existence of $n \in \mathbb{N}_0$ such that $\frac{1}{n}(K - K) \subset \mathcal{U}$. Conclude that $T(x_0) - x_0 \in \mathcal{U}$.

7. By definition a topological vector space is Hausdorff. Hence the statement above implies that $T(x_0) = x_0$ for all $T \in \mathcal{F}$.

□

11.2 Abelian groups are amenable

If G is a countable group and $\xi : G \to \mathbb{C}$ a function, we define for every $g \in G$, the translated function

$$\xi \cdot g : G \to \mathbb{C} : (\xi \cdot g)(h) = \xi(gh) \; .$$

Recall that we have proven in Theorem 8.10 that the group $\mathbb{Z}$ is amenable. As an application of the Markov-Kakutani fixed point theorem, we can prove that every abelian group is amenable.

Theorem 11.4. *Every abelian group is amenable.*

Proof. Define

$$K = \{\Psi \in \ell^\infty(G)^* \mid \|\Psi\| \leq 1 \; , \; \Psi(1) = 1 \; , \; \Psi(F) \geq 0 \;\text{ whenever }\; F(g) \geq 0 \text{ for all } \; g \in G \; \} \; .$$

We equip $\ell^\infty(G)^*$ with the weak* topology.

1. Prove that K is weak* compact and convex.
2. Define for every $g \in G$ the map

   $$T_g : K \to K : (T_g \Psi)(F) = \Psi(F \cdot g) \; .$$

   Prove that every T_g, $g \in G$, is weak* continuous and affine.
3. Since G is a commutative group, we can apply the Markov-Kakutani fixed point theorem to the family $\{T_g \mid g \in G\}$ of affine maps from K to K. So, we get $\Psi \in K$ such that $T_g \Psi = \Psi$ for all $g \in G$. Define $m(A) = \Psi(\chi_A)$ whenever $A \subset G$ and check that m is an invariant mean on G.

□

11.3 The compact space of means

Let G be a countable set. Denote by $\mathcal{P}(G)$ the power set of G, i.e. the set of all subsets of G. Recall that a mean m on the set G is a finitely additive probability measure on G, i.e. a map $m : \mathcal{P}(G) \to [0, 1]$ satisfying $m(\emptyset) = 0$, $m(G) = 1$ and $m(A \cup B) = m(A) + m(B)$ whenever A and B are disjoint subsets of G.

Then define

$$\begin{aligned} \mathcal{M}(G) &:= \text{the set of means on } G \\ &:= \{m : \mathcal{P}(G) \to [0, 1] \mid \; m \text{ is a mean } \} \; . \end{aligned}$$

Whenever $A \subset G$ the formula $d_A(m, m') := |m(A) - m'(A)|$ defines a pseudometric on $\mathcal{M}(G)$. The family $\mathcal{D} := \{d_A \mid A \subset G\}$ turns $\mathcal{M}(G)$ into a pseudometric space.

Proposition 11.5. *The space $\mathcal{M}(G)$ of means on a countable set G equipped with the pseudometric topology given by $\mathcal{D}$ is compact.*

Proof. Prove this proposition yourself. The proof is almost identical to the proof of the Banach-Alaoglu theorem 8.7, identifying $\mathcal{M}(G)$ with a closed subset of the infinite Cartesian product $\prod_{A\subset G}[0,1]$. □

The most obvious means on G are given as follows. Assume that $\xi : G \to [0,1]$ is a finitely supported function with $\sum_{g\in G}\xi(g) = 1$. Define

$$m_\xi : \mathcal{P}(G) \to [0,1] : m_\xi(A) = \sum_{g\in A}\xi(g)\ .$$

Check that m_ξ is indeed a mean on G. We denote by $\mathcal{S}(G)$ the set of these easy means :

$$\mathcal{S}(G) := \{m_\xi \mid \xi : G \to [0,1] \ \text{ is finitely supported and } \ \sum_{g\in G}\xi(g) = 1\}\ . \tag{11.1}$$

Proposition 11.6. *Equip the space of means $\mathcal{M}(G)$ with the pseudometric topology given by $\mathcal{D}$. Then $\mathcal{S}(G)$ is a dense subset of $\mathcal{M}(G)$.*

Proof. Fix $m \in \mathcal{M}(G)$. We have to prove that m lies in the closure of $\mathcal{S}(G)$. By definition we have to prove that for all $A_1,\dots,A_n \subset G$ and every $\varepsilon > 0$ there exists $\xi \in \mathcal{S}(G)$ such that $d_{A_i}(m, m_\xi) < \varepsilon$ for all $i = 1,\dots,n$. We will actually do better and prove that we can find $\xi \in \mathcal{S}(G)$ such that $m(A_i) = m_\xi(A_i)$ for all $i = 1,\dots,n$.

1. Take intersections of A_i's and complements $G \setminus A_j$ to find a partition of G into nonempty *disjoint* subsets $D_1,\dots,D_m \subset G$ with the property that every A_i can be written as the union of some D_j's.

2. Choose points $g_i \in D_i$. Define

$$\xi : G \to [0,1] : \begin{cases} \xi(g_i) = m(D_i) & \text{for all } \ i = 1,\dots,m\ , \\ \xi(g) = 0 & \text{if } \ g \notin \{g_1,\dots,g_m\}\ . \end{cases}$$

   Prove that $\xi \in \mathcal{S}(G)$ and that $m(D_i) = m_\xi(D_i)$ for all $i = 1,\dots,m$.

3. Deduce that $m(A_i) = m_\xi(A_i)$ for all $i = 1,\dots,n$.

□

11.4 A first characterization of amenability: approximately invariant functions

We already saw in Theorem 11.4 that every abelian group is amenable. Actually many more groups are amenable as we shall see below. However once the group is infinite, it is impossible to give a concrete formula for an invariant mean. When proving the amenability of the group $\mathbb{Z}$ (Theorem 8.10) we nevertheless started with very concrete "approximate" invariant means

$$m_n : \ell^\infty(\mathbb{Z}) \to \mathbb{C} : m_n(F) = \frac{1}{2n+1}\sum_{k=-n}^{n} F(k)\ . \tag{11.2}$$

The really invariant mean then arises as a (highly nonconcrete and nonunique) weak* limit point of the sequence $(m_n)_{n\in\mathbb{N}}$.

The first result that we prove in this section is the following: any invariant mean on a countable group G arises as the weak* limit point of a concrete sequence of "approximate" invariant means.

We first introduce the following notation. Throughout G denotes a countable group. Whenever $1 \le p < \infty$ and $\xi : G \to \mathbb{C}$, we put

$$\|\xi\|_p = \Big(\sum_{g\in G} |\xi(g)|^p\Big)^{1/p} .$$

We define $\ell^p(G) := \{\xi : G \to \mathbb{C} \mid \|\xi\|_p < \infty\}$. By Proposition 0.10 we know that $\ell^p(G)$, equipped with $\|\cdot\|_p$, is a Banach space.

Also recall that the following concrete formula provides an isometric embedding of $\ell^1(G)$ into the dual of $\ell^\infty(G)$. Whenever $\xi \in \ell^1(G)$, define

$$m_\xi : \ell^\infty(G) \to \mathbb{C} : m_\xi(F) = \sum_{g\in G} \xi(g)\, F(g) . \tag{11.3}$$

Then $\|m_\xi\| = \|\xi\|_1$.

In the case $G = \mathbb{Z}$, we define $\xi_n := \frac{1}{2n+1}\chi_{[-n,n]}$ and observe that the approximate invariant means m_n in (11.2) are precisely equal to $m_n = m_{\xi_n}$. One checks that for all $g \in \mathbb{Z}$ we have $\|\xi_n \cdot g - \xi_n\|_1 \to 0$ as $n \to \infty$. All this motivates the following theorem.

Theorem 11.7. *Let G be a countable group. Then, the following statements are equivalent.*

1. *The group G is amenable.*
2. *There exists a sequence $\xi_n : G \to [0, +\infty)$ of finitely supported functions such that*
$$\sum_{g\in G} \xi_n(g) = 1 \;\; \text{for all} \;\; n \in \mathbb{N} \;\; \text{and} \;\; \lim_{n\to\infty} \|\xi_n \cdot g - \xi_n\|_1 = 0 \;\; \text{for all} \;\; g \in G .$$
3. *There exists a sequence $\xi_n \in \ell^2(G)$ satisfying*
$$\|\xi_n\|_2 = 1 \;\; \text{for all} \;\; n \in \mathbb{N} \;\; \text{and} \;\; \lim_{n\to\infty} \|\xi_n \cdot g - \xi_n\|_2 = 0 \;\; \text{for all} \;\; g \in G .$$

Before proving the theorem we need the following easy lemma.

Lemma 11.8. *Let G be a countable set. The finite rank functions from G to $\mathbb{C}$*

$$\operatorname{span}\{\chi_A \mid A \subset G\}$$

form a $\|\cdot\|_\infty$-norm dense subspace of $\ell^\infty(G)$.

Proof. Fix $F \in \ell^\infty(G)$ and $\varepsilon > 0$. Take finitely many balls $B(z_i, \varepsilon)$, $i = 1, \dots, n$, of radius ε that cover $\{z \in \mathbb{C} \mid |z| \le \|F\|_\infty\}$. Put $B_i := F^{-1}(B(z_i, \varepsilon))$. Observe that $\bigcup_{i=1}^n B_i = G$. Write $A_1 := B_1$ and $A_i := B_i \setminus (B_1 \cup \cdots \cup B_{i-1})$ for $i = 2, \dots, n$. Define the function

$$\xi := \sum_{i=1}^n z_i \chi_{A_i} .$$

By construction $\|F - \xi\|_\infty \le \varepsilon$. □

Proof of Theorem 11.7. $1 \implies 2$ (due to Day). Let $m : \mathcal{P}(G) \to [0,1]$ be an invariant mean on G. Using the notation in (11.1) and Proposition 11.6 we can take a net $(\xi_i)_{i\in I}$ in $\mathcal{S}(G)$ such that $m_{\xi_i} \to m$ in the pseudometric topology on $\mathcal{M}(G)$. This concretely means that for all $A \subset G$ we have that $\lim_{i\in I} m_{\xi_i}(A) = m(A)$. Check that $m_{\xi_i}(gA) = m_{\xi_i\cdot g}(A)$ for all $g \in G$ and $A \subset G$. The invariance of m implies that

$$\lim_{i\in I}(m_{\xi_i}(A) - m_{\xi_i\cdot g}(A)) = 0 \quad \text{for all } \; g \in G \; , \; A \subset G \; .$$

Whenever $\xi : G \to \mathbb{C}$ is a finitely supported function, we view m_ξ as an element of $\ell^\infty(G)^*$ as in formula (11.3). Because of Lemma 11.8 it follows that

$$\lim_{i\in I} m_{\xi_i - \xi_i\cdot g}(F) = 0 \quad \text{for all } \; g \in G \; , \; F \in \ell^\infty(G) \; .$$

This means that for all $g \in G$ we have that $\xi_i - \xi_i \cdot g$ converges to zero weakly in $\ell^1(G)$. We want to turn this weak convergence into norm convergence.

Fix a finite subset $\mathcal{F} \subset G$ and an $\varepsilon > 0$. We will prove the existence of $\xi \in \mathcal{S}(G)$ such that $\|\xi - \xi \cdot g\|_1 < \varepsilon$ for all $g \in \mathcal{F}$. Define the vector space

$$X = \bigoplus_{g\in\mathcal{F}} \ell^1(G)$$

and turn X into a Banach space using the norm $\|\cdot\|_{\max}$ of Exercise 6 in Lecture 0. Observe that X is defined as a direct sum of finitely many Banach spaces. Define the convex subset $K \subset X$ as

$$K = \left\{ \bigoplus_{g\in\mathcal{F}} (\xi \cdot g - \xi) \;\middle|\; \xi \in \mathcal{S}(G) \right\} .$$

Since the net $\left(\bigoplus_{g\in\mathcal{F}}(\xi_i \cdot g - \xi_i)\right)_{i\in I}$ converges weakly to 0, it follows that 0 belongs to the weak closure of K. But then Corollary 9.7 guarantees that 0 belongs to the norm closure of K. This in turn means that there exists a $\xi \in \mathcal{S}(G)$ such that $\|\xi - \xi \cdot g\|_1 < \varepsilon$ for all $g \in \mathcal{F}$.

Let $G = \{g_0, g_1, \dots\}$ be an enumeration of the elements of G. Because of the previous paragraph we can choose for every $n \in \mathbb{N}$ an element $\xi_n \in \mathcal{S}(G)$ such that $\|\xi_n - \xi_n \cdot g\|_1 < 1/n$ for all $g \in \{g_0, \dots, g_n\}$. By construction we have that $\lim_{n\to\infty} \|\xi_n \cdot g - \xi_n\|_1 = 0$ for every $g \in G$.

$2 \implies 3$. Given $\xi_n \in \mathcal{S}(G)$ satisfying $\lim_{n\to\infty} \|\xi_n \cdot g - \xi_n\|_1 = 0$ for every $g \in G$, we put

$$\eta_n : G \to \mathbb{C} : \eta_n(g) = \sqrt{\xi_n(g)} \; .$$

One checks easily that $\eta_n \in \ell^2(G)$ with $\|\eta_n\|_2 = 1$ for all $n \in \mathbb{N}$. It remains to prove that $\lim_{n\to\infty} \|\eta_n - \eta_n \cdot g\|_2 = 0$ for all $g \in G$. To prove this statement, first observe that for all *positive* real numbers $a, b \in [0, +\infty)$ we have

$$|\sqrt{a} - \sqrt{b}|^2 \leq |\sqrt{a} - \sqrt{b}|\,|\sqrt{a} + \sqrt{b}| = |a - b| \; .$$

It follows that $\|\eta_n - \eta_n \cdot g\|_2^2 \leq \|\xi_n - \xi_n \cdot g\|_1 \to 0$. □

11.5 A second characterization of amenability: positive definite functions

A related characterization can be obtained using positive definite functions.

Proposition 11.9. *Let G be a countable group. Then G is a amenable if and only if there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of finitely supported positive definite functions that converges pointwise to* 1*, i.e.* $\lim_{n\to\infty}\varphi_n(g) = 1$ *for any $g \in G$.*

Proof. Suppose that G is amenable. By theorem 11.7 we get a sequence $(\xi_n)_{n\in\mathbb{N}}$ of finitely supported unit vectors in $\ell^2(G)$ which are approximately invariant. We can define the positive definite functions $\varphi_n(g) := \langle\lambda(g)\xi_n, \xi_n\rangle$. They are finitely supported, because ξ_n's are finitely supported and they converge to 1 pointwise, because $\|\lambda_g(\xi_n) - \xi_n\| \to 1$.

Suppose now that we have a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of positive definite functions with the desired properties. We may assume that $\varphi_n(e) = 1$, because $\varphi_n(e)$ converges to 1 and we can replace the function φ_n by $\widetilde{\varphi}_n := \frac{\varphi_n}{\varphi_n(e)}$ if needed. We will now construct the approximately invariant vectors in $\ell^2(G)$. We could have applied Proposition 10.11 in order to get some unitary representations and unit vectors, but we have to insist on finding those vectors inside $\ell^2(G)$ and using the λ as our representation, so we need to find another way. Our aim is to construct a positive operator $T_n\colon \ell_2(G) \to \ell_2(G)$ that commutes with $\lambda(g)$ and $T_n\delta_e = \varphi_n$. Indeed, if we have that, we may define $\xi_n := \sqrt{T_n}\delta$, and then we have

$$\begin{aligned}\varphi_n(g) &= \langle\varphi_n, \delta_g\rangle = \langle\varphi_n, \lambda(g)\delta_e\rangle\\ &= \langle T_n\delta_e, \lambda(g)\delta_e\rangle = \langle\sqrt{T_n}\delta_e, \sqrt{T_n}\lambda(g)\delta_e\rangle\\ &= \langle\xi_n, \lambda(g)\sqrt{T_n}\delta_e\rangle = \langle\xi_n, \lambda(g)\xi_n\rangle\end{aligned}$$

How do we construct the operators T_n? We want them to commute with $\lambda(g)$, which multiply from the left, so they should be built from operators of multiplication from the right. We declare

$$(T_n f)(g) := (f * \varphi_n)(g) = \sum_{h\in G} f(gh^{-1})\varphi_n(h).$$

Because φ_n is finitely supported, the sum above is always finite and it is easy to see that T_n is a bounded operator. Show that T_n is positive, using the fact that φ_n is positive definite. We clearly have $T_n\delta_e = \varphi_n$, so we get we wanted: unit vectors ξ_n such that $\varphi_n(g) = \langle\xi_n, \lambda(g)\xi_n\rangle$. Let us now check that the condition $\lim_{n\to\infty}\|\lambda(g)\xi_n - \xi_n\| = 0$ follows from the condition $\lim_{n\to\infty}\varphi_n(g) = 1$. Indeed, we have

$$\begin{aligned}\|\lambda(g)\xi_n - \xi_n\|^2 &= \|\lambda(g)\xi_n\|^2 - 2\operatorname{Re}\langle\xi_n, \lambda(g)\xi_n\rangle + \|\xi_n\|^2\\ &= 2(1 - \operatorname{Re}\langle\xi_n, \lambda(g)\xi_n\rangle) = 2(1 - \operatorname{Re}\varphi_n(g)),\end{aligned}$$

which tends to 0. ∎

11.6 A third characterization of amenability: existence of fixed points

In Theorem 11.4 we proved that every abelian group is amenable. The method of the proof went as follows: we defined the convex, weak* compact subset of $\ell^\infty(G)^*$ given by

$$K := \{\Psi \in \ell^\infty(G)^* \mid \|\Psi\| \leq 1\ ,\ \Psi(1) = 1\ ,\ \Psi(F) \geq 0 \text{ if } F(g) \geq 0 \text{ for all } g \in G\}$$

and we considered the action of G on K by translation. The Markov-Kakutani fixed point theorem provided us with the invariant mean.

We now prove that conversely amenability can be characterized by a fixed point property: a group is amenable if and only if every "affine action" on a convex compact set has a fixed point. We first recall some terminology.

- We call $(\alpha_g)_{g\in G}$ an *action* of the group G on a set K if every α_g, $g \in G$, is a permutation of the set K and if the following two conditions hold: $\alpha_e(x) = x$ for all $x \in K$ and $\alpha_g(\alpha_h(x)) = \alpha_{gh}(x)$ for all $g, h \in G$, $x \in X$. In other words, an action of G on K is nothing else than a homomorphism from G to the permutation group of K.
- If X is a vector space and $K \subset X$ a convex subset, we say that $(\alpha_g)_{g\in G}$ is an *affine action* of G on K if every α_g, $g \in G$, is an affine map from K to K.
- We call $(\alpha_g)_{g\in G}$ an *action by homeomorphisms* if K is a topological space and every $\alpha_g : K \to K$ is a homeomorphism.
- We call $x \in K$ a *fixed point* for the action $(\alpha_g)_{g\in G}$ of G on K if $\alpha_g(x) = x$ for all $g \in G$.

Theorem 11.10. *A countable group G is amenable if and only if every action of G by affine homeomorphisms of a nonempty compact convex subset $K \subset X$ of a seminormed space X, admits a fixed point.*

Proof. Assume first that every action of G by affine homeomorphisms of a nonempty compact convex subset $K \subset X$ of a seminormed space X, admits a fixed point. We can repeat the proof of Theorem 11.4 to find an invariant mean on G.

Conversely assume that G is amenable. Let X be a seminormed space, $K \subset X$ a nonempty compact convex subset and $(\alpha_g)_{g\in G}$ an action of G on K by affine homeomorphisms. Fix an arbitrary point $x_0 \in K$. By Theorem 11.7 and using notation (11.1) take a sequence (ξ_n) in $\mathcal{S}(G)$ such that $\lim_{n\to\infty} \|\xi_n - \xi_n \cdot g\|_1 = 0$ for all $g \in G$. Since $\xi_n(g) \in [0, 1]$ and $\sum_{g\in G} \xi_n(g) = 1$ and since K is convex, we can define the sequence

$$x_n := \sum_{g\in G} \xi_n(g)\alpha_g(x_0)\ .$$

Check that for all $h \in G$ and $n \in \mathbb{N}$ we have

$$\alpha_h(x_n) = \sum_{g\in G} (\xi_n \cdot h^{-1})(g)\alpha_g(x_0)\ .$$

Denote by $\mathcal{P}$ the family of seminorms on X that define its topology. Check that for all $p \in \mathcal{P}$ and all $h \in G$, $n \in \mathbb{N}$, we have that

$$p(x_n - \alpha_h(x_n)) \leq (\sup_{k \in K} p(k)) \|\xi_n - \xi_n \cdot h\|_1 .$$

Since K is compact we can take a limit point $x \in K$ of the sequence $(x_n)_{n \in \mathbb{N}}$. Check that $p(x - \alpha_h(x)) = 0$ for all $p \in \mathcal{P}$. Deduce that x is a fixed point for $(\alpha_g)_{g \in G}$. □

11.7 A large class of amenable groups

The following theorem shows that amenability is preserved under several constructions of groups. In combination with the fact that all finite groups and all abelian groups are amenable, it follows that the class of amenable groups is really large.

Theorem 11.11. *The following properties hold.*

1. *If the countable group G is amenable, then all its subgroups are amenable.*
2. *Let G be a countable group and $N \lhd G$ a normal subgroup. Then G is amenable if and only if both N and G/N are amenable.*
3. *Let G be a countable group and (G_n) an increasing sequence of subgroups of G. If $G = \bigcup_n G_n$ and if all the G_n are amenable, then also G is amenable.*

Proof. 1. Let $K \subset G$ be a subgroup. Take a sequence of positive definite functions $\varphi_n : G \to \mathbb{C}$ like in Proposition 11.9. Then the sequence of restrictions $\psi_n := (\varphi_n)_{|K}$ shows that K is amenable.

2. First assume that G is amenable with invariant mean m. By 1 above, the subgroup $N \subset G$ is amenable. Denote by $\pi : G \to G/N$ the quotient map and check that $n(A) = m(\pi^{-1}(A))$ defines an invariant mean on G/N.

To prove the converse we use Theorem 11.10. So assume that N and G/N are amenable and let $(\alpha_g)_{g \in G}$ be an action of G by affine homeomorphisms of a nonempty compact convex subset $K \subset X$ of a seminormed space X. By Theorem 11.10 and because N is amenable, the restriction of this action to an action of N admits a fixed point. This means that the set

$$K' := \{x \in K \mid \alpha_h(x) = x \quad \text{for all} \quad h \in N\}$$

is nonempty. Check that K' is convex and compact. Use the fact that N is normal in G to prove that $\alpha_g(K') = K'$ for every $g \in G$. Also check that the formula

$$\beta_{gN}(x) := \alpha_g(x)$$

provides a well defined action $(\beta_{gN})_{gN \in G/N}$ of G/N on K' by affine homeomorphisms. By Theorem 11.10 and because G/N is amenable, this action admits a fixed point $x \in K'$. By construction x is a fixed point for the original action $(\alpha_g)_{g \in G}$. So every action of G by affine homeomorphisms of a nonempty compact convex subset $K \subset X$ of a seminormed space X, admits a fixed point. Again by Theorem 11.10, it follows that G is amenable.

3. Let $\mathcal{F} \subset G$ be a finite subset and $\varepsilon > 0$. We have to find a function $\xi \in \mathcal{S}(G)$ such that $\|\xi \cdot g - \xi\|_1 < \varepsilon$ for all $g \in \mathcal{F}$. Since $\mathcal{F}$ is finite, we find n such that $\mathcal{F} \subset G_n$. Because G_n is amenable, we can take a function $\xi_0 \in \mathcal{S}(G_n)$ such that $\|\xi_0 \cdot g - \xi_0\|_1 < \varepsilon$ for all $g \in \mathcal{F}$. It suffices to define the function $\xi \in \mathcal{S}(G)$ as

$$\xi(g) = \begin{cases} \xi_0(g) & \text{if } g \in G_n \,, \\ 0 & \text{if } g \notin G_n \,. \end{cases}$$

□

Example 11.12. Theorem 11.11 implies that the following groups are amenable.

1. Define the group S_∞ as follows.

   $$S_\infty = \{\sigma : \mathbb{N} \to \mathbb{N} \mid \sigma \text{ is a bijection, } \exists N \in \mathbb{N} \text{ such that } \sigma(n) = n \text{ for all } n \geq N\} \,.$$

   The group S_∞ contains the usual permutation groups $S_n = \mathrm{Perm}\{0, 1, \ldots, n-1\}$ as an increasing sequence of finite subgroups with $S_\infty = \bigcup_n S_n$.

2. If G is a group, one defines the *derived group* $[G, G]$ as the subgroup of G generated by all the *commutators* $\{ghg^{-1}h^{-1} \mid g, h \in G\}$. Check that $[G, G]$ is a normal subgroup of G and that the quotient group $G/[G, G]$ is commutative. Then, define by induction $G^{(0)} = G$ and $G^{(n+1)} = [G^{(n)}, G^{(n)}]$. The group G is said to be *solvable* if there exists an n such that $G^{(n)} = \{e\}$.

   Prove that every solvable group is amenable. Prove that for all $n \in \mathbb{N}$, $n \geq 2$, the group

   $$G = \{A \in \mathrm{GL}_n(\mathbb{Z}) \mid A_{ij} = 0 \text{ if } i > j\}$$

   is solvable.

3. Fix an invertible matrix $A \in \mathrm{GL}_n(\mathbb{Z})$. Define the group G on the set $\mathbb{Z}^n \times \mathbb{Z}$ with product

   $$(x, k) \cdot (y, l) = (x + A^k y, k + l) \,.$$

   Here, we regard A as a group homomorphism from $\mathbb{Z}^n$ to $\mathbb{Z}^n$. Check that G is a group and that this group is amenable.

11.8 Nonamenable groups

We shall prove that for $n \geq 2$, the group

$$\mathrm{SL}_n(\mathbb{Z}) = \{A \in \mathrm{GL}_n(\mathbb{Z}) \mid \det A = 1\}$$

is nonamenable. It is not hard to see that for $n \geq 3$, the group $\mathrm{SL}_2(\mathbb{Z})$ can be realized as a subgroup of $\mathrm{SL}_n(\mathbb{Z})$. By Theorem 11.11, it is therefore sufficient to prove that $\mathrm{SL}_2(\mathbb{Z})$ is nonamenable. And even this will not be proven in a direct way: we rather show that $\mathrm{SL}_2(\mathbb{Z})$ contains the *free group on two generators* $\mathbb{F}_2$ as a subgroup and that $\mathbb{F}_2$ is nonamenable. Of course, we first have to explain what is $\mathbb{F}_2$.

Definition 11.13. Let G be a group and $a, b \in G \setminus \{e\}$. We say that a and b are *free* if the following condition holds: every product where the factors are alternatingly a nonzero power of a and a nonzero power of b, is different from e. So, whenever $k \geq 1$ and $n_i, m_i \in \mathbb{Z} \setminus \{0\}$, we have

$$\begin{aligned} a^{n_1} b^{m_1} \cdots a^{n_k} b^{m_k} &\neq e \, , \\ a^{n_1} b^{m_1} \cdots b^{m_{k-1}} a^{n_k} &\neq e \, , \\ b^{m_1} a^{n_1} \cdots b^{m_k} a^{n_k} &\neq e \, , \\ b^{m_1} a^{n_1} \cdots a^{n_{k-1}} b^{m_k} &\neq e \, . \end{aligned} \tag{11.4}$$

All the products in G where the factors are alternatingly a nonzero power of a and a nonzero power of b, are called *alternating products of powers of a and powers of b*.

Definition 11.14. Let G be a group. We call G the *free group on two generators* if G is generated by two elements $a, b \in G \setminus \{e\}$ that are free in the sense of Definition 11.13. We denote $G = \mathbb{F}_2$.

Does the free group exist and is it uniquely determined up to isomorphism? The answer to both questions is yes. First of all the usual construction of the free group freely generated by a and b goes as follows. First consider the set of words W in the alphabet $\{a, a^{-1}, b, b^{-1}\}$, including the empty word that we denote by e. Elements of W look like e, $ababab^{-1}$, $aaab^{-1}bbba^{-1}$, etc. A word is called reduced if there are "no obvious simplifications" : the letter a is never followed or preceded by the letter a^{-1}, the letter b is never followed or preceded by the letter b^{-1}. Also there is an "obvious" reduction procedure, reducing arbitrary words to reduced words by canceling all occurrences of aa^{-1}, $a^{-1}a$, bb^{-1}, $b^{-1}b$. The set of reduced words can be turned into a group. The group law consists of concatenating two words and then reducing the concatenated word:

- $a^{-1}a^{-1}b \cdot bbb = a^{-1}a^{-1}bbbb$,
- $a^{-1}a^{-1}b \cdot b^{-1}bb = a^{-1}a^{-1}bb$,
- $a^{-1}a^{-1}b \cdot b^{-1}aaab^{-1} = ab^{-1}$.

It is painful to prove that this group law is well defined and associative. We therefore proceed in a different way. We first show that $\mathbb{F}_2$ is unique up to isomorphism, if it exists. We then exhibit a concrete group that is freely generated by two elements.

Proposition 11.15. *Let G be a group that is generated by two elements $a, b \in G$ that are free in the sense of Definition 11.13. Then the following universal property holds: whenever H is a group and $x, y \in H$, there exists a unique group homomorphism $\pi : G \to H$ satisfying $\pi(a) = x$ and $\pi(b) = y$.*

Proof. Define

$$\pi(a^{n_1} b^{m_1} \cdots a^{n_k} b^{m_k}) = x^{n_1} y^{m_1} \cdots x^{n_k} y^{m_k}$$

and define π analogously on the other alternating products of powers of a and b given in (11.4). Because a and b are free, two alternating products of powers of a and b can only be equal in the group G if they are factorwise equal. Therefore, π is a well defined map from G to H. We need to prove that π is a group homomorphism. By definition $\pi(e) = e$. It also follows almost directly from the definition of π that $\pi(g^{-1}) = \pi(g)^{-1}$ for all $g \in G$.

One then checks that the set $\{g \in G \mid \pi(gh) = \pi(g)\pi(h) \text{ for all } h \in G\}$ is a subgroup of G. By the definition of π this subgroup contains a and b. Hence it is the whole of G.

The uniqueness of π follows because a and b generate G. □

Corollary 11.16. *If it exists, the free group on two generators is uniquely defined up to a unique isomorphism.*

Proof. Assume that G and H are freely generated by $a, b \in G$, respectively $x, y \in H$. By Proposition 11.15 we find group homomorphisms $\pi : G \to H$ and $\rho : H \to G$ satisfying $\pi(a) = x$, $\pi(b) = y$, $\rho(x) = a$ and $\rho(y) = b$. But then $\pi \circ \rho$ and $\rho \circ \pi$ are the identity homomorphism, because they are the identity on the generators. □

It remains to prove the existence of a group with two free generators a and b. In fact, it suffices to provide an example of a group G with elements $a, b \in G \setminus \{e\}$ being free in the sense of Definition 11.13, because we can then take the subgroup generated by a and b. Given a concrete group G and elements $a, b \in G$, it is most of the time not so easy to prove that a and b are free (if they are). An example is provided by Proposition 11.18. Most of the examples in the literature are based on the following extremely nice lemma.

The lemma uses the notion of a *group action:* an action of a group G on a set X is a map

$$G \times X \to X : (g, x) \mapsto g \cdot x$$

satisfying $e \cdot x = x$ for all $x \in X$ and $g \cdot (h \cdot x) = (gh) \cdot x$ for all $g, h \in G$, $x \in X$. Check that an action of G on X is nothing else than a group homomorphism $G \to \mathrm{Perm}(X)$ from G to the group of all permutations of X, mapping an element $g \in G$ to the permutation $x \mapsto g \cdot x$.

Lemma 11.17 (Serre's Lemma, also called ping-pong lemma)**.** *Let G be a group acting on a set X. Suppose that $S, T \in G$ and that X is the disjoint union of nonempty subsets X_1 and X_2 satisfying*

$$S^n \cdot X_1 \subset X_2 \quad \text{and} \quad T^n \cdot X_2 \subset X_1 \quad \text{for all } n \in \mathbb{Z} \setminus \{0\} .$$

Then, S and T are free in the sense of Definition 11.13.

Proof. Two elements x and y in a group G are called *conjugate* if there exists a $g \in G$ such that $gyg^{-1} = x$. Conjugating an arbitrary alternating product of powers of S and T with a sufficiently large power of S, it is sufficient to prove that every alternating product of powers of S and T starting and ending with a power of S, is different from the neutral element e :

$$S^{n_0} T^{m_1} S^{n_1} \cdots T^{m_k} S^{n_k} \neq e$$

whenever $k \in \mathbb{N}$ and $n_i, m_i \in \mathbb{Z} \setminus \{0\}$. Denote the left-hand side of the above expression by g. The assumptions of the lemma imply that $g \cdot X_1 \subset X_2$. Since X_1, X_2 are disjoint and nonempty, it follows that $g \neq e$. □

Proposition 11.18. *The elements*

$$S = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

of the group $\mathrm{SL}_2(\mathbb{Z})$ are free.

Proof. Set $X = \mathbb{R} \setminus \mathbb{Q}$, the set of irrational numbers. Define the action of $\mathrm{SL}_2(\mathbb{Z})$ on X by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x = \frac{ax+b}{cx+d} .$$

If we set $X_1 = X \cap [-1, 1]$ and $X_2 = X \setminus X_1$, you can easily check that S, T, X_1 and X_2 satisfy the conditions of Serre's Lemma. Therefore, S and T are free. $\square$

The above proposition ends the proof of the existence of the free group on two generators.

Proposition 11.19. *Let $G = \mathbb{F}_2$ be the free group on two generators. Then, G is nonamenable.*

Proof. Suppose that m is an invariant mean on G. Denote by G_a the set of all alternating powers of a and b that *start with a power of a*. Add the neutral element e to G_a. Denote by G_b the set of all alternating powers of a and b that *start with a power of b*. It is clear that G is the disjoint union of G_a and G_b.

Because bG_a and b^2G_a are disjoint subsets of G_b, it follows that

$$2m(G_a) = m(bG_a) + m(b^2G_a) \leq m(G_b) .$$

On the other hand, $aG_b \subset G_a$, implying that $m(G_b) = m(aG_b) \leq m(G_a)$. We conclude that $2m(G_a) \leq m(G_b) \leq m(G_a)$. Therefore $m(G_a) = m(G_b) = 0$. But then $m(G) = m(G_a) + m(G_b) = 0$, a contradiction. $\square$

Corollary 11.20. *The group $\mathrm{SL}_n(\mathbb{Z})$, $n \geq 2$, is nonamenable*

Proof. The group $\mathrm{SL}(n, \mathbb{Z})$ admits a copy of $\mathrm{SL}(2, \mathbb{Z})$ and hence a copy of $\mathbb{F}_2$ as a subgroup. Since $\mathbb{F}_2$ is nonamenable, Theorem 11.11 implies that also $\mathrm{SL}(n, \mathbb{Z})$ is nonamenable. $\square$

11.9 Concluding remarks on amenability of groups

Theorem 11.11 and Proposition 11.19 imply that a group G is nonamenable if G contains a subgroup isomorphic with $\mathbb{F}_2$. In 1929, John von Neumann posed the problem whether the converse holds: does every nonamenable group have a subgroup isomorphic with $\mathbb{F}_2$? Ol'shanskii proved in 1980 that the answer is no.

But this does not mean that all mysteries about amenability of groups disappeared. Motivated by Theorem 11.11, one defines the class $\mathcal{EG}$ of groups as the smallest class of groups satisfying the following properties.

- The class $\mathcal{EG}$ contains all finite groups and all commutative groups.
- If G belongs to $\mathcal{EG}$, then all subgroups and all quotients of G belong to $\mathcal{EG}$.
- If G has a normal subgroup N such that both N and G/N belong to $\mathcal{EG}$, then G belongs to $\mathcal{EG}$.
- If G has an increasing sequence of subgroups G_n that all belong to $\mathcal{EG}$ and satisfy $G = \bigcup_n G_n$, then G belongs to $\mathcal{EG}$.

Theorem 11.11 implies that all groups in the class $\mathcal{EG}$ are amenable. In 1957, Day asked if $\mathcal{EG}$ equals the class of amenable groups. But in 1984, Grigorchuk proved that also the answer to Day's question is no.

Finally consider again the free group $\mathbb{F}_2$ on two free generators a and b. Denote by N the smallest normal subgroup of $\mathbb{F}_2$ containing the elements $[ab^{-1}, a^{-1}ba]$ and $[ab^{-1}, a^{-2}ba^2]$, where we use the notation $[g, h] = ghg^{-1}h^{-1}$ to denote the *commutator* of g and h. Define $F = \mathbb{F}_2/N$. We call F *Thompson's group.* It is still an *open problem* to decide whether Thompson's group F is amenable. It is known though that F does not contain a subgroup isomorphic with $\mathbb{F}_2$ and it is also known that F does not belong to $\mathcal{EG}$.

At first sight, the construction of F may seem a little strange. This method of constructing finitely generated groups is called the *method of generators and relations:* F 'is' the group with generators a and b subject to the relations

- the element ab^{-1} commutes with the element $a^{-1}ba$,
- the element ab^{-1} commutes with the element $a^{-2}ba^2$.

If you then define $x_0 = a$, $x_1 = b$ and $x_{n+1} = x_0^{-n} x_1 x_0^n$ for all $n \geq 1$, Thompson's group F can also be defined as the group generated by $x_0, x_1, \ldots$ subject to the relations

$$x_k^{-1} x_n x_k = x_{n+1} \quad \text{whenever} \;\; 0 \leq k < n \; .$$

11.10 Exercises

Exercise 1. Let G be a countable set. Whenever $\varphi \in \ell^\infty(G)^*$ satisfies $\|\varphi\| = 1$, $\varphi(1) = 1$ and $\varphi(F) \geq 0$ whenever $F(g) \geq 0$ for all $g \in G$, we can define a mean m on G by the formula $m(A) := \varphi(\chi_A)$. The aim of this exercise is to perform a converse construction: whenever m is a mean on G, there is a unique $\varphi \in \ell^\infty(G)^*$ such that $m(A) = \varphi(\chi_A)$ for all $A \subset G$.

1. Use Lemma 11.8 to prove the uniqueness of φ.
2. Check the result in the special case where $m \in \mathcal{S}(G)$.
3. Use Proposition 11.6 to deduce the general case.

Intuitively a mean is a finitely additive probability measure, while $\varphi \in \ell^\infty(G)^*$ is a "finitely additive" integral. So we have shown that there is a one-to-one correspondence between finitely additive measures and integrals.

Exercise 2. Consider the group $G = \mathbb{Z}$ and denote $A_n \subset G$ given by $A_n = [-n, n]$. The subsets $A_n \subset G$ are finite and more and more translation invariant in the following sense.

Definition 11.21. Let G be a countable group. A sequence of nonempty finite subsets $A_n \subset G$ is called a *Følner sequence* if

$$\lim_{n \to \infty} \frac{|gA_n \triangle A_n|}{|A_n|} = 0 \quad \text{for all} \;\; g \in G \; ,$$

where $\triangle$ denotes the symmetric difference of two sets.

The aim of this exercise is to prove the following theorem due to Namioka: a countable group G is amenable if and only if G admits a Følner sequence.

First assume that $A_n \subset G$ is a Følner sequence. Check that $\xi_n := |A_n|^{-1}\chi_{A_n}$ satisfies the second assumption in Theorem 11.7. So G is amenable.

Conversely assume that G is amenable.

1. Observe that it suffices to prove the following: for every finite subset $\mathcal{F} \subset G$ and every $\varepsilon > 0$, there exists a nonempty subset $A \subset G$ such that $|gA \triangle A|/|A| < \varepsilon$ for all $g \in \mathcal{F}$. So fix a finite subset $\mathcal{F} \subset G$ and fix $\varepsilon > 0$.

2. We use Namioka's trick. For every $r \in [0, 1]$ we denote by E_r the function that is equal to 1 on $(r, 1]$ and equal to 0 elsewhere. Prove that for every $a, b \in [0, 1]$ we have
$$|a - b| = \int_0^1 |E_r(a) - E_r(b)|\, dr \quad \text{and} \quad a = \int_0^1 E_r(a)\, dr \; .$$

3. Whenever $\xi : G \to [0, 1]$ and $r \in [0, 1]$, we denote $\xi^r := E_r \circ \xi$. Note that $\xi^r = \chi_{A_r}$ where $A_r = \{g \in G \mid \xi(g) \in (r, 1]\}$. Prove that for all finitely supported functions $\xi, \eta : G \to [0, 1]$ we have
$$\|\xi - \eta\|_1 = \int_0^1 \|\xi^r - \eta^r\|_1\, dr \quad \text{and} \quad \|\xi\|_1 = \int_0^1 \|\xi^r\|_1\, dr \; .$$

4. Since G is amenable Theorem 11.7 provides us with a finitely supported function $\xi : G \to [0, 1]$ such that $\|\xi\|_1 = 1$ and
$$\sum_{g \in \mathcal{F}} \|\xi \cdot g^{-1} - \xi\|_1 < \varepsilon \; .$$
Deduce that
$$\int_0^1 \sum_{g \in \mathcal{F}} \|\xi^r \cdot g^{-1} - \xi^r\|_1\, dr < \int_0^1 \varepsilon\, \|\xi^r\|_1\, dr \; .$$
Conclude that we can fix an $r \in [0, 1]$ such that
$$\sum_{g \in \mathcal{F}} \|\xi^r \cdot g^{-1} - \xi^r\|_1 < \varepsilon \|\xi^r\|_1 \; .$$

5. Prove that $\xi^r \neq 0$ and that $A := \{g \in G \mid \xi(g) \in (r, 1]\}$ is a nonempty finite subset of G satisfying
$$\sum_{g \in \mathcal{F}} \frac{|gA \triangle A|}{|A|} < \varepsilon \; .$$
This ends the proof of the theorem.

Dessert

The Ryll-Nardzewski fixed point theorem

We treat a very powerful and deep fixed point theorem due to Ryll-Nardzewski. In the formulation and the proof of the theorem, one plays all the time back and forth between two topologies on the same vector space. In order to avoid confusion, we introduce the following terminology.

Terminology 12.1 (Only relevant in 12.2 and 12.3)**.** We work with a seminormed space X and refer to the seminorm topology on X as the *strong topology* on X. At the same time, we consider the *weak topology* on X defined by the seminorms $p_\omega(x) = |\omega(x)|$ where $\omega : X \to \mathbb{C}$ runs through the strongly continuous linear functionals on X.

We distinguish between two typical cases.

- We start with a Banach space X. Then, the strong topology is the norm topology and the weak topology is the weak topology of Banach spaces introduced in Example 8.2.
- We start with a dual Banach space X^* equipped with the weak* topology. Because of Proposition 9.8, both the strong and the weak topology are given by the weak* topology.

Let $\mathcal{G}$ be a family of maps from a set K to itself. We call $\mathcal{G}$ a *semigroup* if the identity map belongs to $\mathcal{G}$ and if $S \circ T \in \mathcal{G}$ whenever $S, T \in \mathcal{G}$.

Theorem 12.2 (Ryll-Nardzewski fixed point theorem)**.** *Let X be a seminormed space and K a nonempty, convex, weakly compact subset of X. Let $\mathcal{G}$ be a semigroup of weakly continuous affine maps from K to K satisfying the following property.*

If $x, y \in K$, $x \neq y$, then 0 does not belong to the strong closure of $\{Tx - Ty \mid T \in \mathcal{G}\}$.

Then there exists an element $x \in K$ such that $T(x) = x$ for all $T \in \mathcal{G}$.

Proof. Define for every $T \in \mathcal{G}$ the weakly closed subset of K defined by

$$\mathrm{Fix}_T = \{x \in K \mid Tx = x\} .$$

If we can prove that $\mathrm{Fix}_{T_1} \cap \cdots \cap \mathrm{Fix}_{T_n} \neq \emptyset$ for all $n \in \mathbb{N}_0$ and $T_1, \ldots, T_n \in \mathcal{G}$, the weak compactness of K implies the existence of $x \in K$ such that $Tx = x$ for all $T \in \mathcal{G}$. Considering the semigroup generated by $T_1, \ldots, T_n$ we therefore may suppose that $\mathcal{G}$ is countable.

By Zorn's Lemma, we can take a minimal weakly closed, convex, nonempty, $\mathcal{G}$-invariant subset of K. So, we may assume from the beginning that this minimal element is K itself. In other words: if $L \subset K$ is weakly closed, convex, nonempty and $\mathcal{G}$-invariant, then $L = K$.

Applying once more Zorn's Lemma, we find a minimal weakly closed, nonempty, $\mathcal{G}$-invariant subset $C \subset K$. (Observe that convexity disappeared from our list of conditions.) We will prove that C is a singleton and hence provides the required fixed point. Suppose that $x, y \in C$ and $x \neq y$. By Theorem 10.4, we can take an extreme point $z \in \operatorname{ext} K$. Define

$$F = \text{weak closure of } \left\{T\left(\frac{x+y}{2}\right) \;\middle|\; T \in \mathcal{G}\right\}.$$

Since F is a nonempty, weakly closed and $\mathcal{G}$-invariant subset of K, our minimality condition on K implies that K equals the weak closure of $\operatorname{conv}(F)$. By Theorem 10.5, we have $z \in F$.

Take a net $(T_i)_{i \in I}$ in $\mathcal{G}$ such that $T_i\left(\frac{x+y}{2}\right) \to z$ weakly. Since K is weakly compact, we may assume that $T_i(x) \to u$ and $T_i(y) \to v$ weakly. It follows that $z = \frac{u+v}{2}$ and the extremality of z implies that $u = z = v$. It follows that $T_i(x) - T_i(y) \to 0$ weakly and hence, 0 belongs to the weak closure of $\{Tx - Ty \mid T \in \mathcal{G}\}$. This yields a contradiction with Lemma 12.3. □

Lemma 12.3. *Let X be a seminormed space and C a nonempty, weakly compact subset of X. Let $\mathcal{G}$ be a countable semigroup of weakly continuous maps from C to C satisfying the following minimality condition.*

For all $x \in C$, the set $\{Tx \mid T \in \mathcal{G}\}$ is weakly dense in C.

If $x, y \in C$ are such that 0 does not belong to the strong closure of $\{Tx - Ty \mid T \in \mathcal{G}\}$, then 0 does not belong either to the weak closure of $\{Tx - Ty \mid T \in \mathcal{G}\}$.

Proof. Let $a \in C$ be arbitrary. Define D as the strong closure of $\operatorname{conv}\{Ta \mid T \in \mathcal{G}\}$. By construction and because $\mathcal{G}$ is countable, D is strongly separable, meaning that D has a countable subset that is strongly dense. Since D is convex, D is also weakly closed. The minimality assumption on C, says that the set $\{Ta \mid T \in \mathcal{G}\}$ is weakly dense in C. Hence, $C \subset D$.

Take $x, y \in C$ and assume that 0 does not belong to the strong closure of $\{Tx - Ty \mid T \in \mathcal{G}\}$. Take a strong neighborhood V of 0 in X such that $V \cap \{Tx - Ty \mid T \in \mathcal{G}\} = \emptyset$. Take a strong neighborhood W of 0 in X such that W is strongly closed and convex, and satisfies $W - W \subset V$. Since W is convex, W is also weakly closed.

In the first paragraph of the proof, we included C in a strongly separable set D. Therefore, we can take a sequence $(x_k)_{k \in \mathbb{N}}$ in X such that $C \subset \bigcup_{k=0}^{\infty}(x_k + W)$. But then

$$C = \bigcup_{k \in \mathbb{N}} \big(C \cap (x_k + W)\big)$$

is a covering of the weakly compact set C by a sequence of weakly closed subsets. A variant of Baire's theorem (Exercise 3 in Lecture 6) implies the existence of $k \in \mathbb{N}$ such that $C \cap (x_k + W)$ has a nonempty weak interior with respect to the weak subspace topology on C. So, we can take a weakly open subset $\mathcal{U} \subset X$ such that

$$\emptyset \neq \mathcal{U} \cap C \subset C \cap (x_k + W).$$

But then, $C \setminus \left(\bigcup_{T\in\mathcal{G}} T^{-1}(\mathcal{U})\right)$ is a weakly closed, $\mathcal{G}$-invariant subset of C that is different from C. So, it must be the empty set, yielding $C = \bigcup_{T\in\mathcal{G}} T^{-1}(\mathcal{U})$.

Assume now that 0 belongs to the weak closure of $\{Tx - Ty \mid T \in \mathcal{G}\}$. We derive a contradiction. By weak compactness of C, we can take a net $(T_i)_{i\in I}$ such that $T_i x \to z$ and $T_i y \to z$ for some $z \in C$. Take $S \in \mathcal{G}$ such that $z \in S^{-1}(\mathcal{U})$. Take i sufficiently large such that $T_i x$ and $T_i y$ belong both to $S^{-1}(\mathcal{U})$. It follows that $ST_i x$ and $ST_i y$ belong both to $\mathcal{U} \cap C \subset x_k + W$. Hence,

$$ST_i x - ST_i y \in W - W \subset V \; .$$

This is a contradiction with the choice of the neighborhood V in the beginning of the proof. □

DEPARTMENT OF MATHEMATICS
Celestijnenlaan 200B – box 2400
BE-3001 Heverlee, Belgium
tel. + 32 16 32 70 54
stefaan.vaes@kuleuven.be and mateusz.wasilewski@kuleuven.be