# FUNCTIONAL ANALYSIS

NICOLAS MÜNCHINGER

Abstract. These notes are a typed version of Professor Burger's lecture notes, whereby some explanations and proofs were added.

Contents



---

*Date*: September 15, 2024.

# Chapter 1. Banach Spaces, bounded linear Maps: first properties and examples

## 1.1. Normed Spaces, Banach Spaces, Examples

In this course all vector spaces will be over the field $\mathbf{K}$ where $\mathbf{K} = \mathbf{R}, \mathbf{C}$. These are normed fields where

- for $x \in \mathbf{R}$, $|x| := \max(x, -x)$
- for $z = x + iy \in \mathbf{C}$, $|z| = \sqrt{x^2 + y^2}$

Now, let $V$ be a $\mathbf{K}$-vector space.

**Definition 1.1.** A *norm* on $V$ is a map $V \to \mathbf{R}, v \mapsto \|v\|$ satisfying

(1) $\|v\| \geqslant 0$ for all $v \in V$ and $\|v\| = 0$ iff $v = 0$
(2) $\|v + w\| \leqslant \|v\| + \|w\|$ for all $v, w \in V$
(3) $\|\alpha v\| = |\alpha| \|v\|$ for all $\alpha \in \mathbf{K}$ and $v \in V$

**Definition 1.2.** A normed space is a $\mathbf{K}$-vector space together with a norm $\| \cdot \|$; it is often denoted by $(V, \| \cdot \|)$.

Define $d(v, w) := \|v - w\|$. Then properties (1) + (2) are equivalent to the distance axioms (of a metric). In addition:

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

Thus a normed space $(V, \| \cdot \|)$ has a natural distance and in particular is a topological space. By definition of the topology, a basis of open sets is given by[1]

$$\{B_{<r}(x) \mid x \in V,\, r \geqslant 0\}.$$

The $\mathbf{K}$-vector space structure and the topology on $V$ are then compatible in the following sense:

**Lemma 1.3.** The maps

(1) $\mathbf{K} \times V \to V$, $(\alpha, v) \mapsto \alpha v$
(2) $V \times V \to V$, $(v_1, v_2) \mapsto v_1 + v_2$

are continuous.

**Proof.** Note that we interpret $\mathbf{K} \times V$ and $V \times V$ to be endowed with the product topology.

(1) Let $\varepsilon > 0$ and $(\alpha, v) \in \mathbf{K} \times V$ be arbitrary; write $p$ for the function in (1). Next, we want to find some open $U \subset \mathbf{K} \times V$ s.t. $p(U) \subset B_{<\varepsilon}(\alpha v)$. One finds

$$p(B_{<\delta}(\alpha) \times B_{<\delta}(v)) \subset B_{<\delta\|v\| + \delta(|\alpha| + \delta)}(\alpha v) \subset B_{<\varepsilon}(\alpha v)$$

for $\delta$ suitably small. This holds, since for $\beta \in B_{<\delta}(0)$ we have

$$(\alpha + \beta) B_{<\delta}(v) = B_{<\delta|\alpha + \beta|}((\alpha + \beta) v).$$

[1] We will write $B_r(x)$ to denote the open ball of radius $r$ centered at $x$. For closed balls we will write $B_{\leqslant r}(x)$.

(2) Again, for $\varepsilon > 0$ and $(v_1, v_2) \in V \times V$ arbitrary and write $a$ for the function in (2). Then

$$p(B_{<\delta}(v_1) \times B_{<\delta}(v_2)) = B_{<\delta}(v_1) + B_{<\delta}(v_2) \subset B_{<2\delta}(v_1 + v_2)$$

so choosing $2\delta < \varepsilon$ does the job. ■

Later in the course we will have to focus on more general objects than normed spaces, namely:

**Definition 1.4.** A topological vector space is a vector space $V$ endowed with a topology for which the maps (1) + (2) in Lemma 1.3 are continuous.

Clearly all concepts pertaining to the theory of metric spaces make sense for normed spaces. The most imporant one:

**Definition 1.5.** A normed space $(V, \|\cdot\|)$ is called a *Banach* space if the underlying metric space $(V, d)$ is complete.

And:

**Definition 1.6.** A normed space is $(V, \|\cdot\|)$ is separable if the underlying metric space $(V, d)$ is.

We now turn to examples.

**Example 1.7.** Let $V$ be a $\mathbf{K}$-vector space with an inner product

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbf{K}.$$

Then $\|x\| := \sqrt{\langle x, x\rangle}$ defines a norm on $V$ since

$$\begin{aligned}\|x+y\|^2 &= \langle x, x\rangle + 2\operatorname{Re}\langle x, y\rangle + \langle y, y\rangle \\ &\leqslant \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2\end{aligned}$$

using Cauchy-Schwarz. An inner product space $(V, \langle \cdot, \cdot \rangle)$ is called a *Hilbert space* if $(V, \|\cdot\|)$ is complete, that is a Banach space.

Inner product spaces can be characerised among normed spaces as those whose norm satisfies the parallelogramm law:

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

For an exposition cf. Iacobelli, Analysis 4, section 1.1.

**Example 1.8.** is a $\sigma$-algebra of subsets of $\Omega$ called measurable sets and $\mu \colon \mathcal{F} \to [0, +\infty]$ is a $\sigma$-additive measure. For $1 \leqslant p < +\infty$ let

$$L^p(\Omega, \mathbf{K}) := \left\{ f \colon \Omega \to \mathbf{K} \text{ measurable} \mid \|f\|_{L^p(\Omega)} := \int_\Omega |f(x)|^p \, d\mu(x) < +\infty \right\} / \sim$$

whith $f \sim g$ if $f(x) = g(x)$ almost everywhere (with respect to the measure $\mu$). Then $\|\cdot\|_{L^p(\Omega)}$ satisfies all properties of a norm. The triangle inequality follows from the convexity of $x \mapsto x^p$ when $p \geqslant 1$ or can also be deduced using Hölder's inequality[1], which states that for $f, g\colon \Omega \to \mathbf{R}$ we have

$$\|fg\|_{L^1(\Omega)} \leqslant \|f\|_{L^p(\Omega)}\|g\|_{L^q(\Omega)}$$

with $1 \leqslant p < +\infty$ and $q$ the corresponding *conjugate exponent*, i.e.

$$\frac{1}{p} + \frac{1}{q} = 1.$$

For $p = +\infty$ one defines

$$L^\infty(\Omega, \mathbf{K}) := \{f\colon X \to \mathbf{K} \text{ measurable} \mid \exists M > 0\colon |f(x)| \leqslant M \text{ a.e.}\}$$

with norm

$$\|f\|_{L^\infty(\Omega)} = \inf\{M > 0\colon |f(x)| \leqslant M \text{ a.e.}\}.$$

**Special case:** If $\mathcal{F} = 2^\Omega$ and $\mu$ is the counting measure, the corresponding $L^p$ space is denoted by $\ell^p(X, \mathbf{K})$.

**Theorem 1.9.** Let $1 \leqslant p \leqslant \infty$ and $(f_n)_{n\geq 1}$ a Cauchy-Sequence in $L^p(\Omega, \mu, \mathbf{K})$. Then tehre is a susbequence $(f_{n_k})_{k\geqslant 1}$ converging almost everywhere to a measurable function $f\colon \Omega \to \mathbf{K}$. In addition, $f \in L^p(\Omega, \mu, \mathbf{K})$ and

$$\lim_{n\to\infty} \|f - f_n\|_{L^p(\Omega)} = 0.$$

The case $p = 2$ is special as $\|\cdot\|_{L^2(\Omega)}$ is induced by the inner product

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

**Example 1.10.** Let $X$ be a topological space and

$$C_b(X) := \{f\colon X \to \mathbf{R} \text{ continuous and bounded}\}.$$

For $f \in C_b(X)$ we let

$$\|f\|_b := \sup_{x\in X} |f(x)|$$

which makes is a Banach space (this is readily checked since $\|\cdot\|_b$ is the uniform norm and continuity respectively boundedness are properties which are preserved by uniform convergence).

[1]For the proof of this inequality we can exploit the convexity of $x \mapsto e^x$: First, since the inequality is multiplicatively symmetric we can assume that $\|f\|_{L^p(\Omega)} = \|g\|_{L^q(\Omega)} = 1$. Next, by convexity, for $a, b > 0$ one has

$$\begin{aligned} ab &= \exp(\ln(a) + \ln(b)) \\ &= \exp\Big(\frac{\ln(a^p)}{p} + \frac{\ln(b^q)}{q}\Big) \\ &\leqslant \frac{\exp(\ln(a^p))}{p} + \frac{\exp(\ln(b^q))}{q} = \frac{a^p}{p} + \frac{b^q}{q}. \end{aligned}$$

Setting $a = f$, $b = g$ and integrating both sides yields the desired bound of 1.

**Example 1.11.** Let $\alpha > 0$; then $\Lambda^\alpha(\mathbf{R})$ is the space of all bounded continuous functions $f\colon \mathbf{R} \to \mathbf{R}$ satisfying

$$\sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|^\alpha} < +\infty.$$

Then for $f \in \Lambda^\alpha(\mathbf{R})$

$$\|f\|_{\Lambda^\alpha} := \sup_{x\in\mathbf{R}} |f(x)| + \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|^\alpha}$$

gives rise to a norm on $\Lambda^\alpha(\mathbf{R})$ for which it is a Banach space. If $\alpha > 1$ then any $f \in \Lambda^\alpha(\mathbf{R})$ is constant.

**Example 1.12.** The following family of functions spaces on $\mathbf{R}^d$, called Sobolov spaces, are fundamental in the study of partial differential equations. First a definition: a function $f \in L^p(\mathbf{R}^d)$ (where the underlying measure is the Lebesgue measure on $\mathbf{R}^d$) is said to have weak derivatives in $L^p$ up to order $k \in \mathbf{N}$ if for every $(\alpha_1, \ldots, \alpha_d) \in \mathbf{N}^d$ with $|\alpha| := \alpha_1 + \cdots + \alpha_d \leqslant k$, there is some $g_\alpha \in L^p(\mathbf{R}^d)$ with

$$\int_{\mathbf{R}^d} g_\alpha(x)\varphi(x)\,dx = (-1)^{|\alpha|} \int_{\mathbf{R}^d} f(x)\partial_x^\alpha \varphi(x)\,dx$$

for all $\varphi \in C_c^\infty(\mathbf{R}^d)$. Observe that if $f \in C^\infty(\mathbf{R}^d)$ then

$$g_\alpha(x) = \partial_x^\alpha f(x)$$

as can be seen by repeated integration by parts. In general, if $f \in L^p(\mathbf{R}^d)$ has weak derivatives in $L^p$ up to order $k$, we write by abuse of terminology,

$$g_\alpha = \partial_x^\alpha f$$

and denote by $L_k^p(\mathbf{R}^d)$ this function space. Then

$$\|f\|_{L_k^p(\mathbf{R}^d)} := \sum_{|\alpha|\leqslant k} \|\partial_x^\alpha f\|_{L^p(\mathbf{R}^d)}$$

turns $L_k^p(\mathbf{R}^d)$ into a Banach space. A version of the Sobolov embedding theorem says that if $m > \frac{d}{2}$ and $f \in L_m^2(\mathbf{R}^d)$ then $f$ can be corrected on a set of measure zero to become $C^k$ for $k < m - \frac{d}{2}$.

We now turn to properties of finite dimensional normed spaces. The following concept of equivalence for norms will prove useful.

**Definition 1.13.** Two norms $\|\cdot\|_1$, $\|\cdot\|_2$ on a vector space $V$ are called equivalent if there exists some $C > 0$ s.t.

$$\frac{1}{C}\|x\|_1 \leqslant \|x\|_2 \leqslant C\|x\|_1$$

for all $x \in V$.

Clearly, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent norms then $(V, \|\cdot\|_1)$ is a Banach space iff $(V, \|\cdot\|_2)$ is.

**Example 1.14.** Consider $c_{00}(\mathbf{Z})$, the space of all continuous functions $f\colon \mathbf{Z} \to \mathbf{K}$ having finite support. Then for $1 \leqslant p_1, p_2 < +\infty$ the norms $\|\cdot\|_{p_1}$ and $\|\cdot\|_{p_2}$ on $c_{00}(\mathbf{Z})$ are equivalent iff $p_1 = p_2$.

For finite dimensional vector spaces we have the following:

**Proposition 1.15.** On a finite dimensional $\mathbf{K}$-vector space $V$ all norms are equivalent.

**Proof.** Since any finite dimensional vector space $V$ is isomorphic to $\mathbf{K}^d$ it suffices to show the result for $\mathbf{K}^d$. Let $\|\cdot\|$ be any norm on $\mathbf{K}^d$ and $\|\cdot\|_2$ the Euclidean norm. Now, if $(e_1, \ldots, e_n)$ be the canonical basis, we have that for all $x = \sum_{k=1}^n \alpha_k e_k$

$$\begin{aligned}\|x\| = \left\|\sum_{k=1}^n \alpha_k e_k\right\| &\leqslant \sum_{k=1}^n |\alpha_k| \|e_k\| \\ &\leqslant Mn \max_{1\leqslant k\leqslant n} |\alpha_k| \leqslant Mn\left(\sum_{k=1}^n |\alpha_k|^2\right)^{1/2} = Mn\|x\|_2\end{aligned}$$

for $M = \max\{\|e_1\|, \ldots, \|e_n\|\}$.

For the other direction, note that $\|\cdot\|$ is a continuous function so it attains its Minimum on the compact set $S := \{x \in V\colon \|x\|_2 = 1\}$, let it be $C$. Then for all $x \in S$ we have $\|x\| \geqslant C = C\|x\|_2$.

Since norm-equivalence is a transitive relation we are done. ■

We can deduce the following fact:

**Corollary 1.16.** In a normed space, any finite dimensional subspace is closed.

**Proof.** Since all norms are equivalent we find that such a subspace is complete and hence closed as a subspace of any normed space. ■

Even if norms are equivalent, their respective unit balls can have very different geometric properties, e.g. the unit ball with respect to the euclidean norm is a literal ball, whereas it is a square for the maximum norm (aka infinity norm).

## 1.2. Continuous Linear Maps

Having defined the objects of the category of normed spaces, we need the morphisms. These turn out to be continuous linear maps and admit various characterisations. In a normed space $(V, \|\cdot\|)$ we say that a subset $B \subset V$ is bounded if there is some $0 \leqslant R < +\infty$ s.t. $B \subset B_{\leqslant R}(0)$.

**Definition 1.17.** A linear map $T\colon V \to W$ between normed spaces $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ is bounded if $T(B)$ is bounded whenever $B \subset V$ is bounded.

Observe that the property for $T$ to be bounded is equivalent to

$$\|T\| := \sup\{\|T(x)\|_W : \|x\|_V \leqslant 1,\, x \in V\} < +\infty$$

meaning that $T(B_{\leqslant 1}(0))$ is bounded, whereby $\|T\|$ is called the *operator norm*.

**Theorem 1.18.** Let $T\colon V \to W$ be a linear map of normed spaces $V$, $W$. The following are equivalent:

(1) $T$ is continuous in $0 \in V$
(2) $T$ is continuous on $V$
(3) $T$ is bounded
(4) $T$ is Lipschitz continuous with Lipschitz constant $\|T\|$.

**Proof.** (1) $\Longrightarrow$ (2): From Lemma 1.3 we know that for all $v \in V$ the map $L_v\colon V \to V$, $x \mapsto x + v$ is continuous. The additivity of $T$ can be expressed by the commutativity of the below diagram

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & W \\ {\scriptstyle L_{-v}}\downarrow & & \uparrow{\scriptstyle L_{T(v)}} \\ V & \xrightarrow[\quad T \quad]{} & W \end{array}$$

Thus $T = L_{T(v)} \circ T \circ L_{-v}$. If now $T$ is continuous in 0 this implies that $L_{T(v)} \circ T \circ L_{-v}$ is continuous at 0 and hence , using $L_{-v}(v) = 0$, $T$ is continuous at $v$.

(2) $\Longrightarrow$ (3): Since $T$ is continuous at zero there exists some $\varepsilon > 0$ s.t. $T(B^V_{\leqslant\varepsilon}(0)) \subset B^W_{\leqslant 1}(0)$ so in particular $T(B^V_{\leqslant 1}(0)) \subset B^W_{\leqslant 1/\varepsilon}(0)$.

(3) $\Longrightarrow$ (4): $T$ being bounded implies the existence of some $C \geqslant 0$ s.t. $\|T(x)\|_W \leqslant C$ for all $x \in B^V_{<1}(0)$. In particular we have $\|T(x)\|_W \leqslant C\|x\|_V$ for arbitrary $x$ and hence

$$\|T(x) - T(y)\|_W = \|T(x-y)\|_W \leqslant C\|x-y\|_V$$

for all $x, y \in V$.

(4) $\Longrightarrow$ (1): This direction is clear. ∎

From now on, given normed spaces $V$ and $W$ we will denote

$$B(V,W) := \{T\colon V \to W : T \text{ is linear and bounded}\}.$$

Next, let us consider a property of the operator norm, namely *submultiplicativity.*

**Proposition 1.19.** Let $U \xrightarrow{T} V \xrightarrow{S} W$ be bounded linear maps between normed spaces, then $\|S \circ T\| \leqslant \|S\| \cdot \|T\|$.

**Proof.** For any $x \in B^U_{\leqslant 1}(0)$ we compute

$$\|S(Tx)\| \leqslant \|S\|\|Tx\| \leqslant \|S\|\|T\|\|x\| \leqslant \|S\|\|T\|$$

so taking the supremum over $B^V_{\leqslant 1}(0)$ the result follows. ■

**Remark 1.20.** This simple observation is quite fundamental as it endows $B(V, V)$ with teh structure of a Banach Algebra, which will be topic of Functional Analysis II.

We observe that if $V$ and $W$ are normed spaces then $T \mapsto \|T\|$ gives rise to a norm on $B(V, W)$.

**Definition 1.21.** $B(V, \mathbf{K})$ is called the *dual space* of $V$ and denoted $V^*$.

**Proposition 1.22.** If $W$ is a Banach space then $B(V, W)$ equipped with the operator norm is a Banach space. In particular for any normed space $V$ its dual space $V^*$ is a Banach space with respect to the operator norm.

**Proof.** Let $(A_n)_{n \geqslant 1}$ be a Cauchy sequence in $B(V, W)$. Since for any $x \in V$ we have

$$\|A_n(x) - A_m(x)\| \leqslant \|A_n - A_m\| \|x\|$$

the sequence $(A_n(x))_{n \geqslant 1}$ is Cauchy in $W$ and hence converges. This means the function

$$A\colon V \to W, \quad x \mapsto \lim_{n \to \infty} A_n(x)$$

is well defined. Linearity of $A$ follows from properties of the limit and linearity of the $A_n$; next we show boundedness. One has

$$\begin{aligned} \|A(x)\| &= \|A(x) - A_n(x) + A_n(x)\| \\ &\leqslant \|A(x) - A_n(x)\| + \|A_n(x)\| \leqslant \|A(x) - A_n(x)\| + \|A_n\| \|x\| \end{aligned}$$

whereby the first term goes to zero as $n \to \infty$ and the second term converges since

$$|\|A_n\| - \|A_m\|| \leqslant \|A_n - A_m\|$$

so that $\|A\| \leqslant \lim_{n \to \infty} \|A_n\|$. It remains to show that $A_n \to A$ with respect to the operator norm (so far we have only established pointwise convergence). By continuity of the norm we find

$$\|A_n(x) - A(x)\| = \lim_{m \to \infty} \|A_n(x) - A_m(x)\| \leqslant \limsup_{m \to \infty} \|A_n - A_m\| \|x\|$$

so in particular

$$\|A - A_n\| \leqslant \limsup_{m \to \infty} \|A_n - A_m\|$$

which lets us deduce

$$\limsup_{n \to \infty} \|A_n - A\| \leqslant \limsup_{n \to \infty} \limsup_{m \to \infty} \|A_n - A_m\| = 0$$

since $(A_n)_{n \geqslant 1}$ is Cauchy, which implies $\lim_{n \to \infty} \|A_n - A\| = 0$. ■

In certain situations linear maps are automaically continuous.

**Proposition 1.23.** Let $V$ and $W$ be normed spaces and $T\colon V \to W$ a linear map. Assume $V$ is finite dimensional, then $T$ is bounded.

**Proof.** Since for finite dimensional vector spaces the unit sphere is compact[1], we can argue by contradiction. If there existed a sequence $(v_n)_{n\geqslant 1} \subset B^V_{<1}(0)$ with $\|Tv_n\|_W \geqslant n$. By compactness there exists a convergent subsequence $(v_{n_k})_{k\geqslant 1}$ converging to $v$, contradicting that $Tv_{n_k}$ is unbounded. ■

Before we move to examples of bounded linear maps, we deifne the operation of adjunction which in a sense generalises the transpoition of a matrix.

Recall that if $V$ is a normed $\mathbf{K}$-vector space, its dual $V^*$ is defined as: $V^* = B(V, \mathbf{K})$. Given now $A \in B(V, W)$ a bounded linear map of normed $\mathbf{K}$-vector spaces, we have that $\forall \lambda \in W^*$, the composition

$$\lambda \circ A \colon V \xrightarrow{A} W \xrightarrow{\lambda} \mathbf{K}$$

defines an element in $V^*$ denotes $A^*\lambda$. This way we obtain a linear map

$$A^* \colon W^* \to V^*$$

called the adjoint of $A$. Let us show that $A^*$ is bounded:

$$|(A^*\lambda)(x)| = |\lambda(Ax)| \leqslant \|\lambda\|\,\|A\|\,\|x\|$$

which implies

$$\|A^*\lambda\| \leqslant \|A\|\|\lambda\|$$

so taking the supremum over all $\|\lambda\| \leqslant 1$ we find $\|A^*\| \leqslant \|A\|$. Later, we will see that this inequality is in fact an equality, a consequence of the Hahn-Banach theorem.

Laslty, let us quickly look at the special case of $A^*$ when the underlying space is a Hilbert space (i.e. an inner product space that is additionally complete). Assume $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces; we know that the map

$$i_1 \colon \mathcal{H}_1 \to \mathcal{H}_1^*, \quad i_1(v)(x) = \langle x, v\rangle$$

is a bijection (Riesz representation theorem). Now, letting $T \in B(\mathcal{H}_1, \mathcal{H}_2)$ and $T^* \colon \mathcal{H}_2^* \to \mathcal{H}_1^*$ we define

[1] By Heinel-Borel we know that $\{|x| \leqslant 1 \colon x \in \mathbf{R}^d\} \subset \mathbf{R}^d$ is compact, being closed and bounded. If $V$ is now an arbitrary finite dimensional space of dimension $d$, there exists an isomorphism $\Phi \colon \mathbf{R}^d \to V$ and the function $\|\cdot\|_\Phi \colon \mathbf{R}^d \to \mathbf{R}$, $x \mapsto \|\Phi(x)\|_V$ defines a norm on $\mathbf{R}^d$. By norm equivalence we conclude that $B^{\mathbf{R}^d}_{<1}(0, \|\cdot\|_\Phi)$ is compact. If now $(v_n)_{n\geqslant 1} \subset B^V_{<1}(0)$ then the sequence $(\Phi^{-1}(v_n))_{n\geqslant 1} \subset B^{\mathbf{R}^d}_{<1}(0, \|\cdot\|_\Phi)$ has a convergent subsequence $(\Phi^{-1}(v_{n_k}))_{k\geqslant 1}$ so that by definition of $\|\cdot\|_\Phi$ the sequence $(v_{n_k})_{k\geqslant 1}$ converges w.r.t. $\|\cdot\|_V$.

$$\begin{array}{ccc} \mathcal{H}_2 & \xrightarrow{\;i_2\;} & \mathcal{H}_2^* \\ {\scriptstyle T'}\big\downarrow & & \big\downarrow{\scriptstyle T^*} \\ \mathcal{H}_1 & \xrightarrow[\;i_1\;]{} & \mathcal{H}_1^* \end{array}$$

i.e. $T' := i_1^{-1} \circ T^* \circ i_2$. Then $T' \in B(\mathcal{H}_2, \mathcal{H}_1)$ and satisfies

$$\begin{aligned} \langle v, T'w \rangle &= i_1(T'w)(v) \\ &= i_1((i_1^{-1} \circ T^* \circ i_2)(w))(v) \\ &= i_1(i_1^{-1} \circ T^*[h_2 \mapsto \langle h_2, w \rangle])(v) \\ &= i_1(i_1^{-1}([h_1 \mapsto \langle Th_1, w \rangle]))(v) \\ &= [h_1 \mapsto \langle Th_1, w \rangle](v) = \langle Tv, w \rangle \end{aligned}$$

In the case of Hilbert spaces we will denote $T'$ by $T^*$ (by abuse of notation).

**Definition 1.24.** Let $\mathcal{H}$ be a Hilbert space; a bounded linear operator $T\colon \mathcal{H} \to \mathcal{H}$ is called

(a) Self-adjoint if $T^* = T$
(b) Unitary if $T^*T = TT^* = \mathrm{id}_{\mathcal{H}}$

**Remark 1.25.** A unitary operator has in particular has in particular the property that $\|Tv\|^2 = \|v\|^2$ for all $v \in \mathcal{H}$.

More generally:

**Definition 1.26.** A bounded operator $T\colon V \to W$ of normed spaces is an isometry if

$$\|Tv\|_W = \|v\|_V \qquad \forall v \in V$$

**Example 1.27** (Multiplicative Operator)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\varphi \in L^\infty(\Omega)$. Then if $f \in L^p(\Omega)$ then

$$|f(x)\varphi(x)| \leqslant \|\varphi\|_{L^\infty(\Omega)}|f(x)|$$

and hence $f\varphi \in L^p(\Omega)$. We deduce that the linear operator

$$M_\varphi\colon L^p(\Omega) \to L^p(\Omega), \quad f \mapsto f\varphi$$

is bounded with $\|M_\varphi\| \leqslant \|\varphi\|_{L^\infty(\Omega)}$. In fact, this inequality can be turned into an equality: For arbitrary $\varepsilon > 0$ there exists a set $E_\varepsilon$ of positive measure s.t.

$$|\varphi(x)| \geqslant (1-\varepsilon)\|\varphi\|_{L^\infty(\Omega)}.$$

Now we consider the function $f_\varepsilon := \mathbf{1}_{E_\varepsilon}\frac{1}{\mu(E_\varepsilon)^{1/p}}$ which clearly has unit $L^p$-norm and satisfies

$$\int_\Omega |f_\varepsilon \varphi|^p \, d\mu = \frac{1}{\mu(E_\varepsilon)} \int_{E_\varepsilon} |\varphi|^p \, d\mu \geqslant (1-\varepsilon)^p \|\varphi\|^p_{L^\infty(\Omega)}.$$

Since $\varepsilon$ can be made arbitrarily small the desired result follows.

This example is absolutely fundamental: Let $V$ be a finite dimensional **R**-inner product space (finite dimensional real Hilbert space) and $T\colon V \to V$ a self-adjoint map, $n = \dim(V)$ and $X = \{1, \ldots, n\}$ with $\mu$ the counting measure. Then there exists a Hilbert space isomorphism

$$V \xrightarrow{\delta} \ell^2(X)$$

and $\varphi \in L^\infty(X)$ s.t.

$$\begin{array}{ccc} V & \xrightarrow{\quad\delta\quad} & \ell^2(X) \\ {\scriptstyle T}\big\downarrow & & \big\downarrow{\scriptstyle M_\varphi} \\ V & \xrightarrow[\quad\delta\quad]{} & \ell^2(X) \end{array}$$

commutes. This is a reformulation of the theorem that a real symmetric matrix admits an orthonormal set of eigenvectors with real eigenvalues.

**Example 1.28** (unitary representation)**.** Let $\Gamma$ be a group which we consider as a measure space with counting measure. In this example $\ell^2(\Gamma) = \ell^2(\Gamma, \mathbf{C})$. For $f \in \ell^2(\Gamma)$ and $\gamma \in \Gamma$ define

$$(*) \qquad (\lambda(\gamma)f)(\eta) := f(\gamma^{-1}\eta).$$

Then

$$\langle \lambda(\gamma)f, g\rangle = \sum_{\eta\in\Gamma} f(\gamma^{-1}\eta)\overline{g(\eta)} = \sum_{\eta\in\Gamma} f(\eta)\overline{g(\gamma\eta)} = \langle f, \lambda(\gamma^{-1})g\rangle$$

from which we deduce

$$\lambda(\gamma)^* = \lambda(\gamma^{-1}).$$

In addition the definition in $(*)$ with the inverse of $\gamma$ is chosen s.t.

$$\lambda(\gamma_1\gamma_2)f = [x \mapsto f((\gamma_1\gamma_2)^{-1}x)] = [x \mapsto f(\gamma_2^{-1}\gamma_1^{-1}x)] = \lambda(\gamma_1)\lambda(\gamma_2)f.$$

In particular:

$$\begin{aligned} \lambda(\gamma)^*\lambda(\gamma) &= \lambda(\gamma^{-1})\lambda(\gamma) = \lambda(e) = \mathrm{id} \\ \lambda(\gamma)\lambda(\gamma)^* &= \lambda(\gamma)\lambda(\gamma^{-1}) = \lambda(e) = \mathrm{id} \end{aligned}$$

meaning $\lambda\colon \Gamma \to U(\ell^2(\Gamma))$ is a homomorphism into the group $U(\ell^2(\Gamma))$ of unitary operators of $\ell^2(\Gamma)$.

**Fact:** For $\gamma \in \Gamma$, $\lambda(\gamma)$ has an eigenvector in $\ell^2(\Gamma) \iff \gamma$ is of finite order in $\Gamma$.

For the necessary condition, if $\gamma$ had infinite order and $f$ was an eigenvector of $\lambda(\gamma)$ then there would exist some $\alpha \in \mathbf{C}$ s.t. $\lambda(\gamma)f = \alpha f$. In particular, given any $\eta \in \Gamma$ with[1] $f(\eta) \neq 0$ we would have $f(\gamma^{-1}\eta) = \alpha f(\eta)$. This

[1]Such an $\eta$ must exist since, by definition of eigenvectors, $f$ has to be nonzero.

means $\alpha \neq 0$ since otherwise $f(\eta) = 0$ and iterating the identity we obtain $f(\gamma^{-m}\eta) = \alpha^m f(\eta)$ for any $m \in \mathbf{Z}$. However, this contradicts $f \in \ell^2(\Gamma)$.

Next the sufficient condition: We construct an eigenvector $f$ by setting $f(\eta) = 1$ for all $\eta \in \langle\gamma\rangle$ and let it be zero for all other group members. Since for $\eta \notin \langle\gamma\rangle$ we have $\gamma^{-1}\eta \notin \langle\gamma\rangle$ this is indeed an eigenvector (with corresponding eigenvalue $\alpha = 1$) and $f \in \ell^2(\Gamma)$ since

$$\|f\|^2_{\ell^2(\Gamma)} = \sum_{\eta\in\langle\gamma\rangle} |f(\eta)|^2 = \operatorname{ord}(\gamma) < +\infty$$

**Example 1.29** (Integral Operators)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a $\sigma$-finite measure space. This means that $\Omega$ is a countable union of measurable sets of finite measure and hence Fubini's theorem holds. Let $K \in L^2(\Omega \times \Omega, \mathbf{K})$ with $\mathbf{K}$ being $\mathbf{R}$ or $\mathbf{C}$. Then[1]

$$\int_{\Omega\times\Omega} |K|^2 \, d(\mu \times \mu) = \int_\Omega \int_\Omega |K(x, y)|^2 \, d\mu(x) \, d\mu(y) < +\infty$$

and hence by Fubini's Theorme we have that for almost every $x \in \Omega$

$$\int_\Omega |K(x, y)|^2 \, d\mu(y) < +\infty$$

meaning $y \mapsto K(x, y)$ is in $L^2(\Omega)$ so that for all $f \in L^2(\Omega)$

$$T_K f(x) = \int_\Omega f(y) K(x, y) \, d\mu(y)$$

is well defined a.e. Writing $K_x(y) = K(x, y)$ we estimate

$$\begin{aligned} \|T_K f\|^2_{L^2(\Omega)} &= \int_\Omega \left|\left\langle K_x, \overline{f}\right\rangle\right|^2 d\mu(x) \\ &\leqslant \int_\Omega \|K_x\|^2_{L^2(\Omega)} \|f\|^2_{L^2(\Omega)} \, d\mu(x) = \|K\|^2_{L^2(\Omega)} \|f\|^2_{L^2(\Omega)} \end{aligned}$$

which shows that $T_K$ defines a bounded operator on $L^2(\Omega)$ with norm

$$\|T_k\| \leqslant \|K\|_{L^2(\Omega)}.$$

Let us also compute the adjoint of $T_K$:

$$\begin{aligned} \langle T_K f, g\rangle &= \int_\Omega T_K f(x) \overline{g(x)} \, d\mu(x) \\ &= \int_\Omega \int_\Omega \overline{g(x)} f(y) K(x, y) \, d\mu(y) \, d\mu(x) \\ &= \int_\Omega f(y) \overline{\int_\Omega g(x) \overline{K(x, y)} \, d\mu(x)} \, d\mu y = \langle f, T_{K^*} g\rangle \end{aligned}$$

with $K^*(x, y) = \overline{K(y, x)}$. In particular, $T_K$ is self adjoint if and only if

$$K(x, y) = \overline{K(y, x)} \qquad \forall x, y \in \Omega$$

[1]The equality is due to Tonelli's theorem

## 2.1. Hahn-Banach, Analytic Form

Hahn-Banach refers to a set of results that assert the existence of continuous linear forms with addional properties. Whether hidden or explicit, convexity plays always a fundemental role. We begin with a very general result called "Hahn-Banach, Analytic Form" that concerns $\mathbf{R}$-vector spaces.

**Definition 2.1.** A *gauge*[1] on a $\mathbf{R}$-vector space $V$ is a funciton $p\colon V \to \mathbf{R}$ such that

(1) $p(\alpha x) = \alpha p(x)$, $\forall \alpha > 0$
(2) $p(x+y) \leqslant p(x) + p(y)$, $\forall x, y \in V$

**Remark 2.2.** Observe that $\forall r \in \mathbf{R}$ the sublevels $\{v \in V : p(v) < r\}$ and $\{v \in V : p(v) \leqslant r\}$ are convex.

**Remark 2.3.** Let $C \subset V$ be a convex set with the property that $\forall v \in V\, \exists \alpha > 0$ s.t. $v \in \alpha C$. Then

$$p(v) := \inf\{\alpha > 0 \colon v \in \alpha C\}$$

is a gauge function on $V$. In addition

$$\{v \colon p(v) < 1\} \subset C \subset \{v \colon p(v) \leqslant 1\}.$$

**Theorem 2.4.** Let $V$ be an $\mathbf{R}$-vector space, $p\colon V \to \mathbf{R}$ a gauge, $M \subset V$ a vector subspace and $f\colon M \to \mathbf{R}$ a linear form with $f(v) \leqslant p(v)$ for all $v \in M$. Then there exists a linear extension[2] $F\colon V \to \mathbf{R}$ of $f$ with $F(v) \leqslant p(v)$ for all $v \in V$.

First, some informal explanation why this theorem intuitively ought to be true: Starting from the given functional $f$ we pick some $x_0 \in V \setminus M$ and construct a new functional $\widetilde{f}\colon M + \mathbf{R}x_0 \to \mathbf{R}$ via

$$\widetilde{f}(v + tx_0) = f(v) + t\alpha$$

whereby we will show that there exists an $\alpha \in \mathbf{R}$ so that $\widetilde{f} \leqslant p$ on $M + \mathbf{R}x_0$. Intuitively we would want to iterate this construction until our functional is defined on all of $V$, but this idea already hints that the proof will somehow involve the Axiom of Choice (or one of its equivalent formulations) since we would need some kind of transfinite induction. The proof below will not proceed by transfinite induction, but the idea is still the same.

We will make use of Zorn's lemma which is a statement about (partially) ordered sets which we recall now.

Let $\mathcal{P}$ be a set with a partial order $\leqslant$. A subset $Q \subset \mathcal{P}$ is totally ordered if $\forall a, b \in Q$ either $a \leqslant b$ or $b \leqslant a$. We say that $c \in \mathcal{P}$ is an *upper bound* for a subset $Q \subset \mathcal{P}$ if $a \leqslant c$ for all $a \in Q$. We say that $m \in \mathcal{P}$ is *maximal*

[1] Also called a *sublinear* map.

[2] The term *linear extension* already encompasses that $F|_M = f$.

if $m \leqslant x \implies x = x$. Finally we say that $\mathcal{P}$ is *inductive* if every totally ordered subset $Q \subset \mathcal{P}$ has an upper bound. The following is the Zorn's lemma which is equivalent to the Axiom of Choice:

**Lemma 2.5** (Zorn's Lemma)**.** Let $\mathcal{P} \neq \varnothing$ be ordered and inductive. Then $\mathcal{P}$ admits a maximal element.

With this we turn to the proof of Theorem 2.4.

**Proof.** Let

$$\begin{aligned} \mathcal{P} = \{(h, D)\colon D \subset V &\text{ is an } \mathbf{R}\text{-linear subspace} \\ &h\colon D \to \mathbf{R} \text{ is linear, } M \subset D, \\ &h|_M = f,\ h(v) \leqslant p(v) \quad \forall v \in D\}. \end{aligned}$$

We order $\mathcal{P}$ in the following way:

$$(h_1, D_1) \leqslant (h_2, D_2) \iff D_1 \subset D_2 \text{ and } h_2|_{D_1} = h_1.$$

Clearly $\mathcal{P} \neq \varnothing$ since $(m, f) \in \mathcal{P}$.

Let us verify the hypothesis of Zorn's lemma. For any totally ordered subset $Q \subset \mathcal{P}$ define

$$E := \bigcup_{(h,D)\in Q} D.$$

Since $Q$ is totally ordered, $E$ is a $\mathbf{R}$-vector subspace of $V$. Define $j\colon E \to \mathbf{R}$ by $j|_D = h$ whenever $(h, D) \in Q$. Now one verifies easily that $(j, E) \in \mathcal{P}$; it is clearly an upper bound for $Q$. By Zorn's lemma there is a maximal element $(F, E) \in \mathcal{P}$. All we have to show is that $E = V$, which we will do by contradiction.

Suppose $E \neq V$ and let $x_0 \in V \setminus E$ and $D := E + \mathbf{R}x_0$. Define $h\colon D \to \mathbf{R}$ by

$$h(v + tx_0) = F(v) + t\alpha \qquad \forall v \in E,\ t \in \mathbf{R}$$

whereby $\alpha = F(x_0)$ is a constant yet to be determined so that $h \leqslant p$ on $D$. We need $\alpha \in \mathbf{R}$ s.t.

$$F(v) + t\alpha \leqslant p(v + tx_0) \qquad \forall v \in E,\ t \in \mathbf{R}.$$

Using the homogeneity of $p$ (and linearity of $F$), this amounts to the following inequalities:

(1) $F(x) + \alpha \leqslant p(x + x_0) \qquad \forall x \in E$
(2) $F(x) - \alpha \leqslant p(x - x_0) \qquad \forall x \in E.$

Combining these we want to show that

$$\sup_{y\in E}(F(y) - p(y - x_0)) \leqslant \inf_{x\in E}(p(x + x_0) - F(x))$$

i.e. that for all $x, y \in E$:

$$F(y) - p(y - x_0) \leqslant p(x + x_0) - F(x)$$

$$\iff F(x) + F(y) \leqslant p(x + x_0) + p(y - x_0).$$

Now

$$\begin{aligned} F(x) + F(y) = F(x + y) \leqslant p(x + y) &= p((x + x_0) + (y - x_0)) \\ &\leqslant p(x + x_0) + p(y - x_0) \end{aligned}$$

which yields the desired inequality and concludes the proof. ■

The geometric form of Hahn-Banach for real vector spaces will be used in the theory of topological vector spaces, in particular to establish the Krein-Milman theorem.

For many applications to dual spaces of normed **K**-vector spaces, where $\mathbf{K} = \mathbf{R}, \mathbf{C}$, the notion of seminorms, a bit more restrictive than a gauge, will suffice.

**Definition 2.6.** A seminorm on a **K**-vector space $V$ is a function $p\colon V \to [0, +\infty)$ s.t.

(1) $p(\alpha v) = |\alpha| p(v) \qquad \forall \alpha \in \mathbf{K}, \forall v \in V$
(2) $p(v_1 + v_2) \leqslant p(v_1) + p(v_2) \qquad \forall v_1, v_2 \in V.$

We have the following form of Hahn-Banach valid for **K**-vector spaces.

**Theorem 2.7.** Let $V$ be a **K**-vector space, $p\colon V \to [0, +\infty)$ a seminorm, $M \subset V$ a **K**-vector subspace and $f\colon M \to \mathbf{K}$ a linear form with $|f(v)| \leqslant p(v)$ for $\forall v \in M$. Then there exists a **K**-linear extension $F\colon V \to \mathbf{K}$ with $|F(v)| \leqslant p(v)$ for $\forall v \in V$.

**Proof.** We can assume that $\mathbf{K} = \mathbf{C}$ since for $\mathbf{K} = \mathbf{R}$ the theorem holds by invoking Theorem 2.4, since $F(v) \leqslant p(v)$ and due to absolute homogeneity we have $-F(v) = F(-v) \leqslant p(-v) = p(v)$.

Now, by writing

$$f(v) = f_1(v) + i f_2(v) \qquad \forall v \in V$$

for $f_1, f_2\colon M \to \mathbf{R}$, we not only see that $f_1$ and $f_2$ are **R**-linear, but **C**-linearity of $f$ also enforces

$$f_1(iv) + i f_2(iv) = f(iv) = i f(v) = i f_1(v) - f_2(v)$$

meaning $f_2(v) = -f_1(iv)$. Hence we can solely focus on $f_1$ which satisfies

$$|f_1(v)| \leqslant \sqrt{(\operatorname{Re} f_1(v))^2 + (\operatorname{Im} f_2(v))^2} = |f(v)| \leqslant p(v)$$

so there exists a linear extension $F_1\colon V \to \mathbf{R}$ with

$$|F_1(v)| \leqslant p(v) \qquad \forall v \in V.$$

We now set $F(v) := F_1(v) - i F_1(iv)$ which is indeed **C**-linear and extends $f$ (due to our above observation). To estimate $|F|$ we do a little trick: Pick $\theta \in [0, 2\pi]$ so that $e^{i\theta} F(v) = |F(v)|$, i.e. we rotate the complex number $F(v)$ onto the real-axis. This gives us

$$|F(v)| = e^{i\theta} F(v) = F(e^{i\theta} v) \overset{(*)}{=} F_1(e^{i\theta} v) \leqslant p(e^{i\theta} v) = |e^{i\theta}| p(v) = p(v)$$

whereby in $(*)$ we used that $F(e^{i\theta}v)$ is purely real. ■

We draw some immediate consequences which can loosely be summarised by saying tha a normed $\mathbf{K}$-vector space has enough continuous linear functinoals.

**Corollary 2.8.** Let $(V, \|\cdot\|)$ be a normed $\mathbf{K}$-vector space, $M \subset V$ a subspace and $f\colon M \to \mathbf{K}$ continuous linear. Then there is $F\colon V \to \mathbf{K}$ continuous linear with $F|_M = f$ and $\|F\| = \|f\|$.

**Proof.** By hypothesis $|f(v)| \leqslant \|f\|\|v\|$ for all $v \in M$ whereby $\|f\|$ is the operator norm of $f$ constraint to $M$, i.e.

$$\|f\| = \sup_{\substack{x\in M\\ \|x\|\leqslant 1}} \|f(x)\|.$$

Let $p\colon V \to [0, +\infty)$ be defined by

$$p(v) := \|f\|\|v\|.$$

Then $p$ is a norm, in particular sublinear, so that Theorem 2.7 applies to obtain a $\mathbf{K}$-linear form $F\colon V \to \mathbf{K}$ with $F|_M = f$ and $|F(v)| \leqslant p(v) = \|f\|\|v\|$ for all $v \in V$. This implies

$$\|f\| = \sup_{\substack{v\in M\\ \|v\|\leqslant 1}} |f(v)| = \sup_{\substack{v\in M\\ \|v\|\leqslant 1}} |F(v)| \leqslant \sup_{\substack{v\in V\\ \|v\|\leqslant 1}} |F(v)| \leqslant \|f\|$$

demonstrating $\|F\| = \|f\|$. ■

**Corollary 2.9.** Let $V$ be a normed space and $v_0 \in V$. Then there is $f_0 \in V^*$ with

(1) $\|f_0\| = 1$
(2) $f_0(x_0) = \|x_0\|$.

**Proof.** Let $M = \mathbf{K}x_0$ and $f\colon M \to \mathbf{K}$, $f(tx_0) = t\|x_0\|$. By Corollary 2.8 there exists $f_0 \in V^*$ with $f_0|_M = f$, in particular $f_0(x_0) = \|x_0\|$ and

$$\|f_0\| = \|f\| = \sup_{|t|\leqslant \frac{1}{\|x_0\|}} |f(tx_0)| = 1$$

■

The next statement is an immediate consequence of 2.9:

**Corollary 2.10.** Let $V$ be a normed space, then for all $v \in V$:

$$\|v\| = \sup\{|f(v)|\colon f \in V^*,\ \|f\| \leqslant 1\} = \max\{|f(v)|\colon f \in V^*,\ \|f\| \leqslant 1\}$$

This corollary allows us now to compute hte norm of hte adjoint $T^*\colon W^* \to V^*$ of a bounded linear map $T\colon V \to W$.

**Corollary 2.11.** We have $\|T^*\| = \|T\|$.

**Proof.** Plugging in the definitions we compute[1]

$$\begin{aligned}\|T^*\| &= \sup_{\substack{\lambda\in W^*\\ \|\lambda\|\leqslant 1}} \|T^*(\lambda)\| \\ &= \sup_{\|\lambda\|\leqslant 1}\sup_{\|v\|\leqslant 1} |T^*(\lambda)(v)| \\ &= \sup_{\|v\|\leqslant 1}\sup_{\|\lambda\|\leqslant 1} |\lambda(Tv)|\end{aligned}$$

By Corollary 2.10 we have

$$\sup_{\|\lambda\|\leqslant 1} |\lambda(Tv)| = \|Tv\|$$

which together with the above computation implies

$$\|T^*\| = \sup_{\|v\|\leqslant 1} \|T(v)\| = \|T\|.$$

∎

Another application of 2.10 is to the bidual of a normed space $V$: this is by definition $V^{**} = B(V^*, \mathbf{K})$ and the point is that we have a canonical map

$$J\colon V \to V^{**}, \quad J(v)(\lambda) := \lambda(v)$$

**Proposition 2.12.** The map $J\colon V \to V^{**}$ is a $\mathbf{K}$-linear isometry into the Banach space $V^{**}$.

**Proof.** First, we have that

$$|J(v)(\lambda)| = |\lambda(v)| \leqslant \|\lambda\|\|v\|$$

which shows that $J(v) \in (V^*)^*$. Next we have by 2.10:

$$\|J(v)\| = \sup_{\|\lambda\|\leqslant 1} |\lambda(v)| = \|v\|.$$

∎

Spaces for which $J$ is surjective are called *reflexive*, they are automatically Banach spaces.

Now we turn to some important examples of dual spaces.

---

[1] Suprema are interchangable: Let $f\colon A\times B \to \mathbf{R}$ for sets $A, B$; then

$$\sup_{a\in A}\sup_{b\in B} f(a,b) = \sup_{b\in B}\sup_{a\in A} f(a,b).$$

To see this, note that for any $a \in A$ we have

$$f(a,b) \leqslant \sup_{a\in A} f(a,b) \implies \sup_{b\in B} f(a,b) \leqslant \sup_{b\in B}\sup_{a\in A} f(a,b)$$

and taking $\sup_{a\in A}$ on both sides we obtain the first inequality. The other one can be obtained by applying this argument to $b \in B$.

**Example 2.13** (The dual of $L^p(\Omega)$, $1 \leqslant p < +\infty$)**.** Let $1 \leqslant p < +\infty$ and $q$ the conjugate exponent, that is

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then Hölder's inequality shows that every $g \in L^q(\Omega)$ gives rise to a continuous linear function on $L^p(\Omega)$ by

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

with $\|\ell\| \leqslant \|g\|_{L^q(\Omega)}$.

In fact:

**Theorem 2.14.** For $1 \leqslant p < +\infty$ the map

$$L^q(\Omega) \to (L^p(\Omega))^*, \quad g \mapsto \ell_g$$

is an isometric isomorphism.

**Proof.** Cf. Stein-Shakarchi [**SS11**], section 1.4. ∎

**Corollary 2.15.** For $1 < p < +\infty$, $L^p(\Omega)$ is reflexive.

We will see later on that for Banach spaces there is a relation between uniform convexity and reflexivity.

**Example 2.16.** Let $X$ be a locally compact Hausdorff space. A continuous function $f\colon X \to \mathbf{R}$ is said to vanish at infinity if for all $\varepsilon > 0$ there exists some compact set $K \subset X$ s.t.

$$|f(x)| < \varepsilon \qquad \forall x \in X \setminus K.$$

Let $C_0(X, \mathbf{C})$ be the space of continuous $\mathbf{C}$-valued functions that vanish at $+\infty$. Endowed with the norm

$$\|f\|_b := \sup_{x \in X} |f(x)|$$

$C_0(X)$ becomes a Banach space.

The dual space $C_0(X)^*$ is described by the space of complex measures: a complex measure is a set function

$$\mu\colon \mathcal{B}_X \to \mathbf{C}$$

defined on the $\sigma$-algebra $\mathcal{B}_X$ of Borel sets s.t. for all $E \in \mathcal{B}_X$ and any countable partition $E = \bigsqcup_{n \in \mathbf{N}} E_n$ with $E_n \in \mathcal{B}_X$ we have

$$\mu(E) = \sum_{n=1}^{\infty} \mu(E_n).$$

Since this series is assumed to converge for any permutation of the summands, it converges absolutely. One defines then the total variation measure of $\mu$ as

$$|\mu|(E) = \sup\left\{\sum_{n=1}^{\infty} |\mu(E_n)| \colon E = \bigsqcup_{n\in\mathbf{N}} E_n,\, E_n \in \mathcal{B}_X\right\}$$

and shows that $|\mu|$ is a positiv measure on $\mathcal{B}_X$ with $|\mu|(X) < +\infty$.

In order to define the integral of say a bounded Borel function $f \colon X \to \mathbf{C}$ w.r.t. $\mu$, one reduces oneself to the case of positive measures (where the Lebesgue integral is available) in the following way. First one can evidently decompose $\mu$ as

$$\mu = \mu_1 + i\mu_2$$

where $\mu_1, \mu_2$ are both complex measures with values in $\mathbf{R}$: such measures are called *signed* measures. Given a signed measure $\nu \colon \mathcal{B}_X \to \mathbf{R}$ define then

$$\nu^+ = \frac{1}{2}(|\nu| + \nu)$$
$$\nu^- = \frac{1}{2}(|\nu| - \nu).$$

Then $\nu^+, \nu^-$ are positive measures with $\nu^+(X), \nu^-(X) < +\infty$ and $\nu = \nu^+ - \nu^-$. Thus, given a complex measure $\mu$, we can decompose it as follows into a combination of positive measures:

$$\mu = (\mu_1^+ - \mu_1^-) + i(\mu_2^+ - \mu_2^-)$$

and hence $\int_\Omega f\, d\mu$ makes sense for say any bounded Borel function. Finally, we say that $\mu$ is regular if its total variation measure $|\mu|$ is a regular Borel measure. Then:

**Theorem 2.17.** (Riesz Representation) For every bonded linear map $\Phi \colon C_0(X, \mathbf{C}) \to \mathbf{C}$ there is a unique complex regular measure $\mu$ defined on Borel sets such that

$$\Phi(f) = \int_X f\, d\mu \qquad \forall f \in C_0(X, \mathbf{C}).$$

In addition $\|\Phi\| = |\mu|(X)$.

## 2.2. The Problem of Measure

The Hahn-Banach theorem can be used to show that there is a finitely additive set funciton defined on all subsets of $\mathbf{R}^d$ that agrees with Lebesgue measure on measurable sets and is translation invariant. However, this set function cannot be $\sigma$-additive and this is connected to the existence of non-measurable sets.

A deeper fact is that it is not possible to extend the Lebesgue measure on $\mathbf{R}^d$ ($d \geqslant 3$) to a finitely-additive measure on all subsets of $\mathbf{R}^d$ so that it is both translation and rotation invariant.

Here we are going to treat the case $d = 1$ which proceeds in two steps, the first of which contains the main idea based on the use of Hahn-Banach; the second step, more formal can be read in Stein-Shakarchi (chapter 5.4 of [**SS11**]).

Let $\mathbf{R}/\mathbf{Z}$ be the group of real numbers mod 1, that is, the quotient of the abelian group $\mathbf{R}$ by the subgroup $\mathbf{Z}$ and $\pi\colon \mathbf{R} \to \mathbf{R}/\mathbf{Z}$ the canonical projection. Let

$$C^{\infty}(\mathbf{R}/\mathbf{Z}) = \{f\colon \mathbf{R}/\mathbf{Z} \to \mathbf{R}\colon f \text{ is bounded}\}.$$

We say that $f \in \ell^{\infty}(\mathbf{R}/\mathbf{Z})$ is measurable if $f \circ \pi\colon \mathbf{R} \to \mathbf{R}$ is $\mathcal{L}^1$-measurable w.r.t. the Lebesgue measure $\mathcal{L}^1$ so that $\mathcal{L}^1([0,1]) = 1$. For $f \in \ell^{\infty}(\mathbf{R}/\mathbf{Z})$ measurable we define

$$\int_{\mathbf{R}/\mathbf{Z}} f \, d\mathcal{L}^1 := \int_0^1 (f \circ \pi)(x) \, d\mathcal{L}^1(x)$$

which exists since $f \circ \pi$ is bounded and measurable.

Next we have an action of $\mathbf{R}$ by translation $\ell^{\infty}(\mathbf{R}/\mathbf{Z})$ defined as follows: observe that for $x \in \mathbf{R}/\mathbf{Z}$ and $h \in \mathbf{R}$, $x + h \in \mathbf{R}/\mathbf{Z}$ is well defined. Then for $f \in \ell^{\infty}(\mathbf{R}/\mathbf{Z})$, set $f_h(x) = f(x + h)$.

**Theorem 2.18.** There is a linear map

$$I\colon \ell^{\infty}(\mathbf{R}/\mathbf{Z}) \to \mathbf{R}$$

s.t.

(1) $I(f) \geqslant 0$ if $f \geqslant 0$
(2) $I(f) = \int_{\mathbf{R}/\mathbf{Z}} f \, d\mathcal{L}^1$ whenever $f$ is measurable
(3) $I(f_h) = I(f)$ for all $h \in \mathbf{R}$ and $f \in \ell^{\infty}(\mathbf{R}/\mathbf{Z})$.

**Proof.** This is going to be an application of Hahn-Banach (Theorem 2.4) with $V = \ell^{\infty}(\mathbf{R}/\mathbf{Z})$ and $M = \{f \in \ell^{\infty}(\mathbf{R}/\mathbf{Z})\colon f \text{ measurable}\}$ with the linear form

$$I_0\colon M \to \mathbf{R}, \quad I_0(f) := \int_{\mathbf{R}/\mathbf{Z}} f \, d\mathcal{L}^1.$$

The key now is to find the appropriate gauge function $p\colon V \to \mathbf{R}$ s.t. $I_0(f) \leqslant p(f)$ for all $f \in M$. Banach's ingenious construction goes as follows: For every pair $(A, \alpha)$ consisting of a finite set $A$ and a function $\alpha\colon A \to \mathbf{R}$ define

$$M_{(A,\alpha)}(f) := \sup_{x \in \mathbf{R}/\mathbf{Z}} \frac{1}{|A|} \Big( \sum_{a \in A} f(x + \alpha(a)) \Big)$$

for $f \in \ell^{\infty}(\mathbf{R}/\mathbf{Z})$ where $|A|$ is the cardinality of $A$. Define then

$$p(f) := \inf\{M_{(A,\alpha)}(f)\colon A \text{ finite}, \ \alpha\colon A \to \mathbf{R}\}.$$

Observe that since

$$-\|f\|_{\infty} \leqslant M_{A,\alpha}(f) \leqslant \|f\|_{\infty},$$

$p(f)$ is well defined.

To establish that $p$ is a guage it will be convenient to define for $f \in \ell\infty(\mathbf{R}/\mathbf{Z})$,

$$S(f) := \sup_{x\in\mathbf{R}/\mathbf{Z}} f(x) \in \mathbf{R}.$$

Then $S$ satisfies the following properties:

(1) $S(cf) = cS(f)$ if $c \geqslant 0$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$.
(2) $S(f_1 + f_2) \leqslant S(f_1) + S(f_2)$ for all $f_1, f_2 \in \ell^\infty(\mathbf{R}/\mathbf{Z})$
(3) $S(f_h) = S(f)$ for all $h \in \mathbf{R}$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$.

It is then convenient to rewrite $M_{A,\alpha}(f)$ as $S(\frac{1}{|A|}\sum_{a\in A} f_{\alpha(a)})$. From this we deduce

(1) $M_{A,\alpha}(cf) = cM_{A,\alpha}(f)$ for all $c \geqslant 0$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$
(2) $M_{A,\alpha}(f_1 + f_2) \leqslant M_{A,\alpha}(f_1) + M_{A,\alpha}(f_2)$ for all $f_1, f_2 \in \ell^\infty(\mathbf{R}/\mathbf{Z})$
(3) $M_{A,\alpha}(f_h) = M_{A,\alpha}(f)$ for all $h \in \mathbf{R}$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$.

Property (1) implies immediately that for all $c \geqslant 0$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$

$$p(cf) = cp(f).$$

Concerning the second defining property of a gauge we make the following observation: let $(A, \alpha), (B, \beta)$ be maps from finite sets to $\mathbf{R}$. Define

$$\alpha + \beta\colon A \times B \to \mathbf{R}, \quad (a, b) \mapsto \alpha(a) + \beta(b).$$

Then for all $g \in \ell^\infty(\mathbf{R}/\mathbf{Z})$ we have

(1) $M_{A\times B,\alpha+\beta}(g) \leqslant M_{A,\alpha}(g)$
(2) $M_{A\times B,\alpha+\beta}(g) \leqslant M_{B,\beta}(g)$.

We show (1), as (2) follows from (1) by interchaning the roles of $(A, \alpha)$ and $(B, \beta)$. To this end we compute

$$\begin{aligned}
M_{A\times B,\alpha+\beta}(g) &= S\Big(\frac{1}{|A||B|}\sum_{\substack{a\in A\\ b\in B}} g_{\alpha(a)+\beta(b)}\Big)\\
&= S\Big(\frac{1}{|B|}\sum_{b\in B}\Big(\frac{1}{|A|}\sum_{a\in A} g_{\alpha(a)}\Big)_{\beta(b)}\Big)\\
&\leqslant \frac{1}{|B|}\sum_{b\in B} S\Big(\Big(\frac{1}{|A|}\sum_{a\in A} g_{\alpha(a)}\Big)_{\beta(b)}\Big)\\
&= \frac{1}{|B|}\sum_{b\in B} S\Big(\frac{1}{|A|}\sum_{a\in A} g_{\alpha(a)}\Big) = M_{A,\alpha}(g)
\end{aligned}$$

which shows (1).

Let now $f_1, f_2 \in \ell^\infty(\mathbf{R}/\mathbf{Z})$, $\varepsilon > 0$ and $(A, \alpha), (B, \beta)$ s.t.

$$M_{A,\alpha}(f_1) < p(f_1) + \varepsilon$$
$$M_{B,\beta}(f_2) < p(f_2) + \varepsilon.$$

Then

$$p(f_1 + f_2) \leqslant M_{A\times B,\alpha+\beta}(f_1 + f_2)$$

$$\begin{aligned}
&\leqslant M_{A\times B,\alpha+\beta}(f_1) + M_{A\times B,\alpha+\beta}(f_2)\\
&\leqslant M_{A,\alpha}(f_1) + M_{B,\beta}(f_2)\\
&< p(f_1) + \varepsilon + p(f_2) + \varepsilon
\end{aligned}$$

which implies $p(f_1 + f_2) \leqslant p(f_1) + p(f_2)$ and shows that $p$ is a gauge.

Next we observe that for all $h \in \mathbf{R}$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$ measurable, we have

$$\begin{aligned}
I_0(f) &= \frac{1}{|A|}\int_{\mathbf{R}/\mathbf{Z}} \sum_{a\in A} f_{\alpha(a)}\, d\mathcal{L}^1\\
&\leqslant \int_{\mathbf{R}/\mathbf{Z}} S\Big(\frac{1}{|A|}\sum_{a\in A} f_{\alpha(a)}\Big)\, d\mathcal{L}^1 = M_{A,\alpha}(f)
\end{aligned}$$

which by taking the infimum over all $(A,\alpha)$ implies $I_0(f) \leqslant p(f)$ for all $f \in M$.

Let $I\colon \ell^\infty(\mathbf{R}/\mathbf{Z}) \to \mathbf{R}$ be the linear form extending $I_0$ and satisfying

$$I(f) \leqslant p(f)$$

for all $f \in \ell^\infty(\mathbf{RZ})$ given by Theorem 2.4.

Now we show that $I$ satisfies properties (1), (2) and (3) of Theorem 2.18. Clearly, if $f(x) \leqslant 0$ for all $x \in \mathbf{R}/\mathbf{Z}$ then $M_{A,\alpha}(f) \leqslant 0$ and hence $p(f) \leqslant 0$. Thus $I(f) \leqslant p(f) \leqslant 0$. If now $f(x) \geqslant 0$ for all $x \in \mathbf{R}/\mathbf{Z}$ we have $-f(x) \leqslant 0$ for all $x \in \mathbf{R}/\mathbf{Z}$, hence $I(-f) \leqslant 0$ and by linearity of $I$, $I(f) \geqslant 0$ which proves (1). Property (2) is immediate since $I$ extends $I_0$. ■

For (3) we claim that $p(f - f_h) \leqslant 0$ for all $\ell^\infty(\mathbf{R}/\mathbf{Z})$ and $h \in \mathbf{R}$. Indeed, let $N \geqslant 1$ in $\mathbf{N}$ arbitrary, $A_N = \{1,\dots,N\}$ and $\alpha_N(j) = jN$. Then the sum entering the definition of $M_{A_N,\alpha_N}(f)$ is

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N}(f - f_h)(x + jh) &= \frac{1}{N}\sum_{j=1}^{N}(f(x+jh) - f(x+(j+1)h))\\
&= \frac{1}{N}(f(x+h) - f(x+(N+1)h)).
\end{aligned}$$

And hence

$$M_{A_N,\alpha_N}(f - f_h) \leqslant \frac{2\|f\|_\infty}{N} \xrightarrow{N\to\infty} 0$$

which implies $p(f - f_h) \leqslant 0$. Thus $I(f - f_h) \leqslant 0$. Replacing $f$ by $f_{-h}$ and then $-h$ by $h$ we get $I(f_h - f) \leqslant 0$ and by linearity $I(f_h) = I(f)$ for all $h \in \mathbf{R}$ and $f \in \ell^\infty(\mathbf{R}/\mathbf{Z})$.

For $E \subset \mathbf{R}/\mathbf{Z}$, we say that $E$ is measurable if $\mathbf{1}_E \in \ell^\infty(\mathbf{R}/\mathbf{Z})$ is and define its Lebesgue measure

$$\mathcal{L}^1(E) := \int_{\mathbf{R}/\mathbf{Z}} \mathbf{1}_E\, d\mathcal{L}^1.$$

Then we have the follwing immediate corollary from Theorem 2.18.

**Corollary 2.19.** There is a non-negative set function $\widehat{\lambda}$ defined on all subsets of $\mathbf{R}/\mathbf{Z}$ s.t.

(1) $\widehat{\lambda}(E_1 \cup E_2) = \widehat{\lambda}(E_1) + \widehat{\lambda}(E_2)$ for all disjoint subsets $E_1, E_2$
(2) $\widehat{\lambda}(E) = \mathcal{L}^1(E)$ if $E$ is measurable
(3) $\widehat{\lambda}(E + h) = \widehat{\lambda}(E)$ for all $h \in \mathbf{R}$ and $E \subset \mathbf{R}/\mathbf{Z}$.

From this it is not difficult to deduce

**Theorem 2.20.** There is a function $\widehat{\lambda} \colon \mathcal{P}(\mathbf{R}) \to [0, +\infty]$ with the following properties

(1) $\widehat{\lambda}(E_1 \cup E_2) = \widehat{\lambda}(E_1) + \widehat{\lambda}(E_2)$ whenever $E_1, E_2$ are disjoint
(2) $\widehat{\lambda}(E) = \mathcal{L}^1(E)$ whenever $E$ is Lebesgue measurable.
(3) $\widehat{\lambda}(E + h) = \widehat{\lambda}(E)$ for all $h \in \mathbf{R}$ and $E \subset \mathbf{R}$.

Corollary 2.19 can obviously be rephrased in terms of the existence of a finitely additive set function on $S^1$ that is SO(2)-invariant measure on $S^1$.

In contrast to the action of SO(3) on $S^2$ one has a paradoxical decomposition as was shown by Banach-Tarski.

**Theorem 2.21.** There is a countable subset $E \subset S^2$, a partition

$$S^2 \setminus E = A_1 \cup A_2 \cup A_3 \cup A_4$$

and two rotations $a, b \in \mathrm{SO}(3)$ such that

$$a(A_2) = A_2 \cup A_3 \cup A_4$$
$$b(A_4) = A_1 \cup A_2 \cup A_4.$$

**Corollary 2.22.** There is no SO(3)-invariant additive set function on $S^2$ extending the Lebesgue measure.

**Proof.** If $\widehat{\lambda} \colon \mathcal{P}(S^2) \to [0, \infty)$ were such a set function we would first have $\widehat{\lambda}(E) = 0$ since $E$ is countable. Then

$$\widehat{\lambda}(A_2) = \widehat{\lambda}(aA_2) = \widehat{\lambda}(A_2) + \widehat{\lambda}(A_3) + \widehat{\lambda}(A_4)$$

which implies $\widehat{\lambda}(A_3) = \widehat{\lambda}(A_4) = 0$ and similarly,

$$\widehat{\lambda}(A_4) = \widehat{\lambda}(bA_4) = \widehat{\lambda}(A_1) + \widehat{\lambda}(A_2) + \widehat{\lambda}(A_4)$$

implying $\widehat{\lambda}(A_1) = \widehat{\lambda}(A_2) = 0$. Thus

$$\widehat{\lambda}(S^2) = \widehat{\lambda}(E) + \widehat{\lambda}(A_1) + \widehat{\lambda}(A_2) + \widehat{\lambda}(A_3) + \widehat{\lambda}(A_4)$$

which implies $\widehat{\lambda}(E) = 0$ for all $E \subset S^2$. ■

## Chapter 3. Compact Operators, Spectral Theorem

The main result of this chapter is the spectral theorem for self-adjoint compact operators on a Hilbert space. In fact many operators arising " in nature" are compact, examples will arise in the first section of this chapter, while the second is devoted to the proof of the spectral theorem.

### 3.1. Compact operators and Hilbert-Schmidt operators

Certain natural classes of operators between Banach spaces have much stronger properties than being bounded.

**Definition 3.1.** A (bounded) operator $T : V \to W$ between Banach spaces is said to be *compact* if $\overline{T(B_{\leqslant 1}(0))}$ is a compact subset of $W$.

This is equivalent to requiring that $\overline{T(B)} \subset W$ is compact for whenever $B \subset V$ is bounded[1].

The fundamental examples is:

**Example 3.2.** If $T\colon V \to W$ has finite rank then $T$ is compact. Indeed, $R(T) := \operatorname{im}(T)$ is finite dimensional and $T(B_{\leqslant 1}(0))$ is closed and bounded; it is compact by Heine-Borel.

Let $V, W$ be Banach spaces and $K(V, W) \subset B(V, W)$ the subset consisting of compact operators. Then:

**Proposition 3.3.** (1) $K(V, W)$ is a subvector space of $B(V, W)$.
(2) If $A \in B(V, V)$, $T \in K(V, W)$ and $S \in B(W, W)$ then $STA \in K(V, W)$
(3) $K(V, W)$ is closed in $B(V, W)$ for the operator norm.

Before we get to the proof, let us recall a characterisation of compactness particularly well suited for complete metric spaces:

**Proposition 3.4.** A metric space $(X, d)$ is compact if and only if it is complete and totally bounded. $X$ is totally bounded if for all $\varepsilon > 0$ there exists some finite subset $A \subset X$ s.t.

$$X = \bigcup_{a \in A} B_{\leqslant \varepsilon}(a)$$

i.e. $X$ is the union of finitely many balls of radius $\varepsilon$.

For a proof see 8.1. Now lets turn to the proof of Proposition 3.3:

**Proof.** (1) is follows from continuity of scalar multiplication and addition. For (2) note that $TA$ is again a compact operator since for any bounded set $E \subset V$ also $A(E)$ is bounded and hence $\overline{T(A(E))}$ is compact. Moreover, if

[1] This equivalence follows from the identities

$$T(B_{\leqslant r}(0)) = T(rB_{\leqslant 1}(0)) = rT(B_{\leqslant 1}(0))$$

and $\overline{rX} = r\overline{X}$ whereby $rX := \{rx\colon x \in X\}$.

$K\colon V \to W$ is any compact operator then $SK$ is also a compact operator since

$$\overline{S(K(B_{\leqslant 1}(0)))} \subset S\big(\overline{K(B_{\leqslant 1}(0))}\big)$$

whereby the set on the right side is compact, being the image of a compact set under a continuous transformation. Lastly, since closed subsets of compact sets are again compact in a Hausdorff space this concludes the proof of (2).

(3) Let $T = \lim_{n\to\infty} T_n$ with $T_n$ compact for $n \geqslant 1$. We show that $T(B_{\leqslant 1}(0))$ is totally bounded. Let $\varepsilon > 0$ and $n$ s.t. $\|T_n - T\| \leqslant \varepsilon$. For every $x, y \in B_{\leqslant 1}(0)$ we have then:

$$\begin{aligned}\|T(x) - T(y)\| &\leqslant \|T(x) - T_n(x)\| + \|T_n(x) - T_n(y)\| + \|T_n(y) - T(y)\| \\ &\leqslant 2\|T - T_n\| + \|T_n(x) - T_n(y)\|.\end{aligned}$$

Now $T_n(B_{\leqslant 1}(0))$ is totally bounded, hence $\exists F \subset B_{\leqslant 1}(0)$ finite s.t. for all $y \in B_{\leqslant 1}(0)$ there exists some $x \in F$ s.t.

$$\|T_n(x) - T_n(y)\| \leqslant \varepsilon$$

which implies that for all $y \in B_{\leqslant 1}(0)$ there exists some $x \in F$ s.t.

$$\|T(x) - T(y)\| \leqslant 3\varepsilon$$

and shows that $T(B_{\leqslant 1}(0))$ is totally bounded. ■

**Corollary 3.5.** If $T \in B(V, W)$ is the limit of a sequence $(T_n)_{n\geqslant 1}$ where each $T_n$ has finite rank, then $T \in K(V, W)$ is compact.

**Example 3.5.**$^{(*)}$ Let $\mathcal{H}$ be a separable Hilbert space with orthonormal basis $\{e_k\colon k \geqslant 1\}$ and define

$$T\colon \bigoplus_{k\geqslant 1} \mathbf{C}e_k \to \bigoplus_{k\geqslant 1} \mathbf{C}e_k, \quad T(e_k) = \lambda_k e_k$$

with $(\lambda_n)_{n\geqslant 1} \subset \mathbf{C}$. Then $T$ extends to a bounded operator $\mathcal{H} \to \mathcal{H}$ iff

$$\sup_{k\geqslant 1} |\lambda_k| < +\infty$$

which then coincides with $\|T\|$. Furthermore, we have that $T$ is compact iff $\lim_{n\to\infty} \lambda_n = 0$.

For diagonal operators in a Hilbert space as in the above example, bounded operators correspond to bounded sequences, and compact operators to sequences vanishing at infinity. We are going to define a class of operators which in the diagonal case would correspond to the condition $\sum_{n=1}^{\infty} |\lambda_n|^2 < +\infty$. These are the *Hilbert-Schmidt* operators.

**Definition 3.6.** Let $\mathcal{H}$ be a separable Hilbert space with orthonormal basis $\{e_k\colon k \in \mathbf{N}\}$. Then $T \in B(\mathcal{H}, \mathcal{H})$ is called Hilbert-Schmidt if

$$\sum_{n,m=1}^{\infty} |\langle Te_n, e_m\rangle|^2 < +\infty.$$

**Lemma 3.7.** If $\{f_k \colon k \in \mathbf{N}\}$ is another orthonormal basis we have

$$\sum_{n,m=1}^{\infty} |\langle Te_n, e_m\rangle|^2 = \sum_{n,m=1}^{\infty} |\langle Tf_n, f_m\rangle|^2.$$

**Proof.**

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |\langle Te_n, e_m\rangle|^2 &= \sum_{n=1}^{\infty} \|Te_n\|^2 \\ &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |\langle Te_n, f_m\rangle|^2 \\ &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} |\langle e_n, T^* f_m\rangle|^2 = \sum_{m=1}^{\infty} \|T^* f_m\|^2\end{aligned}$$

and the last term can now be expanded in terms of the basis $\{f_k\}_{k\in\mathbf{N}}$ to conclude. ■

**Definition 3.8.** If $T \colon \mathcal{H} \to \mathcal{H}$ is Hilbert-Schmidt we define its Hilbert-Schmidt norm by

$$\|T\|_2 := \Big(\sum_{n,m\geqslant 1} |\langle Te_n, e_m\rangle|^2\Big)^{1/2}.$$

**Corollary 3.9.** If $T \colon \mathcal{H} \to \mathcal{H}$ is Hilbert-Schmidt, so is $T^*$ and $\|T\|_2 = \|T^*\|_2$.

As one can guess from Example 3.5.$^{(*)}$ the operator norm and the Hilbert-Schmidt norm are quite different. However, we always have the following inequality.

**Lemma 3.10.** If $T \in B(\mathcal{H}, \mathcal{H})$ is Hilbert-Schmidt then $\|T\| \leqslant \|T\|_2$.

**Proof.** For $x \in \mathcal{H}$ we find (using the triangle inequality)

$$\|Tx\| \leqslant \sum_{n=1}^{\infty} |\langle x, e_n\rangle| \|Te_n\| \leqslant \Big(\sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2\Big)^{1/2} \Big(\sum_{n=1}^{\infty} \|Te_n\|^2\Big)^{1/2} = \|x\|\|T\|_2$$

■

We conclude the following:

**Proposition 3.11.** If $T \colon \mathcal{H} \to \mathcal{H}$ is a Hilbert-Schmidt operator, $T$ is compact.

**Proof.** Let $\{e_n \colon n \in \mathbf{N}\}$ be an ONB. Define $T_n \colon B(\mathcal{H}, \mathcal{H})$ by

$$T_n(e_k) = \begin{cases} e_k, & 1 \leqslant k \leqslant n \\ 0, & k \geqslant n+1 \end{cases}$$

and notice that it has finite rank. In addition $T_n - T$ is Hilbert-Schmidt with

$$\|T_n - T\|_2^2 = \sum_{k=n+1}^{\infty} \|T(e_k)\|^2 \xrightarrow{n\to\infty} 0$$

since by definition $\sum_{n=1}^{\infty} \|T(e_n)\|^2 < +\infty$. Using the previous Lemma we conclude that $\|T_n - T\| \to 0$ for $n \to \infty$ proving that $T$ is compact (cf. (3) of Proposition 3.3). ∎

From this we are going to get a large class of compact operators.

**Proposition 3.12.** Let $(\Omega, \mathcal{F}, \mu)$ be a $\sigma$-finite measure space, $K \in L^2(\Omega \times \Omega, \mu \times \mu)$ and $T_K \colon L^2(\Omega) \to L^2(\Omega)$ the corresponding bounded operator we have already discussed in Example 1.29. Then $T_K$ is Hilbert-Schmidt, in particular compact and $\|T_k\|_2 = \|K\|_{L^2(\Omega\times\Omega)}$.

**Proof.** We assume that $\Omega$ is s.t. $L^2(\Omega)$ is separable, so let $\{f_n \colon n \in \mathbf{N}\}$ be an ONB of $L^2(\Omega)$. Recall that by Fubini, for almost every $x \in \Omega$, $K_x(y) := K(x, y)$ is in $L^2(\Omega)$. We compute (employing monotone convergence)

$$\begin{aligned}
\sum_{n=1}^{\infty} \|T_K f_n\|_{L^2(\Omega)}^2 &= \sum_{n=1}^{\infty} \int_\Omega |(T_K f_n)(x)|^2 \, d\mu(x) \\
&= \sum_{n=1}^{\infty} \int_\Omega \left|\left\langle K_x, \overline{f}_n \right\rangle\right|^2 d\mu(x) \\
&= \int_\Omega \sum_{n=1}^{\infty} \left|\left\langle K_x, \overline{f}_n \right\rangle\right|^2 d\mu(x) \\
&= \int_\Omega \sum_{n=1}^{\infty} \left|\left\langle \overline{K_x}, f_n \right\rangle\right|^2 d\mu(x) \\
&= \int_\Omega \left\|\overline{K_x}\right\|_{L^2(\Omega)}^2 d\mu(x) = \|K\|_{L^2(\Omega\times\Omega)}
\end{aligned}$$

∎

## 3.2. Spectral theorem for compact self-adjoint operators

If $\mathcal{H}$ is a $\mathbf{K}$-Hilbert space and $T \in B(\mathcal{H}, \mathcal{H})$ is self-adjoint, if $\dim(\mathcal{H}) < +\infty$ we know that all eigenvalues of $T$ are real and there is an ONB of $\mathcal{H}$ consisting of eigenvectors of $T$. We are going to generalise this result by replacing the hypothesis $\dim(\mathcal{H}) < +\infty$ by the hypothesis that $T$ is compact.

Fo simplicity of notation we will assume that all our Hilbert spaces are $\mathbf{C}$-vector spaces. Analogous results hold over $\mathbf{R}$.

Let $V$ be a Banach space and $T \in B(V, V)$. For $\lambda \in \mathbf{C}$ let

$$V_\lambda := \{v \in V \colon T(v) = \lambda v\}.$$

Then $V_\lambda$ is clearly a closed subspace of $V$ (since $T$ is continuous). Recall that $\lambda$ is an eigenvalue of $T$ if $V_\lambda \neq \{0\}$ in which case the elements of $V_\lambda$ are called eigenvectors of $T$ corresponding to the eigenvalue $\lambda$.

**Example 3.13.** Let $\mathcal{H} = L^2([0,1], \mathbf{C})$ where we take the usual Lebesgue measure $\lambda$ on $[0,1]$. Define $Tf(x) = xf(x)$ for $f \in \mathcal{H}$. Then $T$ is clearly bounded and self-adjoint since for $f, g \in \mathcal{H}$ we have

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

However, if $T$ had an eigenvalue $\alpha \in \mathbf{C}$ it would satisfy $\alpha f(x) = xf(x)$ for a.e. $x \in [0,1]$ so that $(\alpha - x)f(x) = 0$ a.e. implying $f(x) = 0$ a.e. because $(\alpha - x) \neq 0$ a.e.

**Theorem 3.14** (Spectral Theorem)**.** Let $T \in B(\mathcal{H}, \mathcal{H})$ be compact self-adjoint where $\mathcal{H}$ is a Hilbert space. Then $\mathcal{H}$ has an ONB consisting of eigenvectors of $T$. In addition: $\dim(\mathcal{H}_\lambda) < +\infty$ for all $\lambda \neq 0$ and

$$\{\lambda \in \mathbf{C} \colon |\lambda| \geqslant \varepsilon,\ \dim(\mathcal{H}_\lambda) > 0\}$$

is finite for all $\varepsilon > 0$.

The proof is based on two lemmas, one of which is a verification whereas the second one is trickier.

**Lemma 3.15.** Let $T \in B(\mathcal{H}, \mathcal{H})$ where $\mathcal{H}$ is a Hilbert space.

(1) If $T = T^*$ and $W \subset \mathcal{H}$ is a $T$-invariant subspace, so is $W^\perp$.
(2) If $T = T^*$ then $\langle Tv, v\rangle \in \mathbf{R}$ for all $v \in \mathcal{H}$; in particular all eigenvalues of $T$ are real.
(3) $\|T\| = \sup\{|\langle T(v), w\rangle| \colon \|v\| \leqslant 1,\ \|w\| \leqslant 1\}$.
(4) If $T = T^*$ and $\lambda \neq \alpha$ then $\mathcal{H}_\lambda$ and $\mathcal{H}_\alpha$ are orthogonal.

**Proof.** (1) For $u \in W^\perp$ we find

$$\langle Tu, w\rangle = \langle u, T^*w\rangle = \langle u, Tw\rangle = 0 \qquad \forall w \in W$$

since $Tw \in W$.

(2) Indeed:

$$\overline{\langle Tv, v\rangle} = \langle v, Tv\rangle = \langle v, T^*v\rangle = \langle Tv, v\rangle$$

(3) Using Corollary 2.10 we compute

$$\|T\| = \sup_{\|v\|\leqslant 1} \|Tv\| = \sup_{\|v\|\leqslant 1} \sup\{|f(v)| \colon f \in \mathcal{H}^*,\ \|f\| \leqslant 1\}.$$

Now by Riesz representation theorem all linear forms in $\mathcal{H}^*$ are of the form $x \mapsto \langle x, y\rangle$ for some $y \in \mathcal{H}$ reducing the above to[1]

$$\|T\| = \sup_{\|v\|\leqslant 1} \sup_{\|w\|\leqslant 1} |\langle Tv, w\rangle|.$$

(4) For all $v \in \mathcal{H}_\lambda$ and $w \in \mathcal{H}_\alpha$:

$$\lambda\langle v, w\rangle = \langle Tv, w\rangle = \langle v, Tw\rangle = \overline{\alpha}\langle v, w\rangle = \alpha\langle v, w\rangle$$

but since $\lambda \neq \alpha$ we must have $\langle v, w\rangle = 0$. ■

The next lemma gives the key to the whole theorem.

**Lemma 3.16.** Let $T \in B(\mathcal{H})$ with $T = T^*$. Then

$$\|T\| = \sup\{|\langle Tv, v\rangle| \colon v \in \mathcal{H}\}.$$

**Proof.** Let $s := \sup\{|\langle Tv, v\rangle| \colon v \in \mathcal{H}\}$; then clearly $s \leqslant \|T\|$. We want to show that

$$|\langle Tv, w\rangle| \leqslant s\|v\|\|w\|$$

which by (3) of the previous lemma implies the desired equality. Since multiplying $w$ by some $\alpha \in \mathbf{C}$ with $|\alpha| = 1$ does not affect the above inequality, we may assume that $\langle Tv, w\rangle \in \mathbf{R}$. Now, from $T = T^*$ and $\langle Tv, w\rangle \in \mathbf{R}$ we deduce

$$\langle T(v+w), v+w\rangle = \langle Tv, v\rangle + 2\langle Tv, w\rangle + \langle Tw, w\rangle$$
$$\langle T(v-w), v-w\rangle = \langle Tv, v\rangle - 2\langle Tv, w\rangle + \langle Tw, w\rangle$$

which combined yields

$$4\langle Tv, w\rangle = \langle T(v+w), v+w\rangle - \langle T(v-w), v-w\rangle.$$

Hence

$$|\langle Tv, w\rangle| \leqslant \frac{s}{4}(\|v+w\|^2 + \|v-w\|^2) = \frac{s}{2}(\|v\|^2 + \|w\|^2).$$

Lastly, to turn this sum into our desired product we will again exploit some symmetry, namely that replacing $v$ and $w$ simultaneously by $\sqrt{a}v$ and $\frac{w}{\sqrt{a}}$ for some $a > 0$ we find:

$$|\langle Tv, w\rangle| \leqslant \frac{s}{2}\Big(a\|v\|^2 + \frac{1}{a}\|w\|^2\Big).$$

Since we may assume $v \neq 0$ we can set $a = \frac{\|w\|}{\|v\|}$ and get

$$|\langle Tv, w\rangle| \leqslant s\|v\|\|w\|$$

■

Now, let us turn to the proof of Theorem **??**.

[1] Recall that for all $y \in \mathcal{H}$

$$\|[x \mapsto \langle x, y\rangle]\| = \sup_{\|x\|\leqslant 1} |\langle x, y\rangle| \leqslant \|x\|\|y\|$$

and for $x = y$ this holds with equality, i.e. $\|[x \mapsto \langle x, y\rangle]\| = \|y\|$.

**Proof.** (1) We claim that either $\|T\|$ or $-\|T\|$ is an eigenvalue: We may assume $T \neq 0$; let $(v_n)_{n\geqslant 1}$ be a sequence with $\|v_n\| = 1$ for all $n \geqslant 1$ and $\lim_{n\to\infty} |\langle Tv_n, v_n\rangle| = \|T\|$. We may assume, passing to a subsequence if necessary, that $\lim_{n\to\infty}\langle Tv_n, v_n\rangle = \lambda \in \mathbf{R}$ (it is a real number by (2) of Lemma 3.15) and proceed to show that $\lambda$ is an eigenvalue. Clearly $\lambda = \|T\|$ or $\lambda = -\|T\|$. Since $T$ is compact, again modulo passing to a subsequence[1], we may assume $\lim_{n\to\infty} Tv_n = w$. Since $\lambda \neq 0$ we get that $w \neq 0$. Next we compute

$$\begin{aligned}\|Tv_n - \lambda v_n\|^2 &= \|Tv_n\|^2 - 2\lambda\langle Tv_n, v_n\rangle + \lambda^2\|v_n\|^2 \\ &\leqslant 2\|T\|^2 - 2\lambda\langle Tv_n, v_n\rangle\end{aligned}$$

which together with $\lim_{n\to\infty}\langle Tv_n, v_n\rangle = \lambda$ and $\lambda^2 = \|T\|^2$ implies

$$\lim_{n\to\infty} \|Tv_n - \lambda v_n\| = 0.$$

Combining this with $\lim_{n\to\infty} Tv_n = w$ we obtain

$$\lim_{n\to\infty} \lambda v_n = w \iff \lim_{n\to\infty} v_n = \frac{w}{\lambda}$$

and hence $T(w) = \lambda w$.

(2) By Zorn's Lemma we can choose an orthonormal set $\mathcal{A} \subset \mathcal{H}$ of eigenvectors which is maximal among all orthonormal sets of eigenvectors. Let $\langle\mathcal{A}\rangle$ be the $\mathbf{C}$-vector subspace of $\mathcal{H}$ spanned by these vectors and $W := \overline{\langle A\rangle}$ its closure. We want to show that $W = \mathcal{H}$. Indeed, if this was not the case we would have $W^\perp \neq \{0\}$ and since $T(W) \subset W$ we have $T(W^\perp) \subset W^\perp$ (by (2) of Lemma 3.15). In addition $W^\perp$ is a Hilbert space and $T|_{W^\perp} : W^\perp \to W^\perp$ is self-adjoint and compact. Hence, by (1) $T|_{W^\perp}$ admits an eigenvector, contradicting the maximality of $\mathcal{A}$.

(3) Let $\varepsilon > 0$ and define

$$W := \overline{\bigoplus_{|\lambda|\geqslant\varepsilon} \mathcal{H}_\lambda}.$$

Observe that the sum $\bigoplus_{|\lambda|\geqslant\varepsilon} \mathcal{H}_\lambda$ is direct since for all $\alpha \neq \beta$ we have $\mathcal{H}_\alpha \perp \mathcal{H}_\beta$ (by (4) of Lemma 3.15). We are going to show that $\dim(W) < +\infty$ by showing the inclusion

$$T(B^W_{\leqslant 1}(0)) \supset B^W_{\leqslant\varepsilon}(0)$$

which implies that $B^W_{\leqslant\varepsilon}(0)$ is compact, so $\dim(W) < +\infty$.

Since $\mathcal{H}_\lambda \subset \mathcal{H}$ is a closed subspace (by continuity of $T$), for $\lambda \in \mathbf{R}$ let $P_\lambda \colon \mathcal{H} \to \mathcal{H}_\lambda$ be the orthogonal projection onto $\mathcal{H}_\lambda$. Let $v \in B^W_{\leqslant\varepsilon}(0)$; then $v = \sum_{|\lambda|\geqslant\varepsilon} P_\lambda(v)$ with

$$\|v\|^2 = \sum_{|\lambda|\geqslant\varepsilon} \|P_\lambda(v)\|^2.$$

[1] Note that $(Tv_n)_{n\geqslant 1} \subset \overline{T(B_{\leqslant 1}(0))}$ and by compactness of $T$ the set $\overline{T(B_{\leqslant 1}(0))}$ is compact.

Set $w := \sum_{|\lambda|\geqslant\varepsilon} \frac{1}{\lambda} P_\lambda(v)$ which exists since

$$\|w\|^2 = \sum_{|\lambda|\geqslant\varepsilon} \frac{1}{\lambda^2} \|P_\lambda(v)\|^2 \leqslant \frac{1}{\varepsilon^2} \sum_{|\lambda|\geqslant\varepsilon} \|P_\lambda(v)\|^2 = \frac{\|v\|^2}{\varepsilon^2} \leqslant 1$$

and

$$Tw = T\Big(\sum_{|\lambda|\geqslant\varepsilon} \frac{1}{\lambda} P_\lambda(v)\Big) = \sum_{|\lambda|\geqslant\varepsilon} \frac{1}{\lambda} T(P_\lambda(v)) = \sum_{|\lambda|\geqslant\varepsilon} P_\lambda(v) = v$$

This shows $T(B^W_{\leqslant 1}(0)) \supset B^W_{\leqslant\varepsilon}(0)$. ■

**Example 3.17** (Unitary representations of compact groups)**.** This examples is meant to give a glimpse into the field of unitary representations and more specifically in the problem of decomposing them into irreducible ones.

We assume that $(X, d)$ is a compact metric space on which a group $G$ acts by isometries (distance preserving bijections):

$$G \times X \to X, \quad (g, x) \mapsto gx$$

with $d(gx, gy) = d(x, y)$ for all $g \in G$ and $x, y \in X$. We assume in addition that $G$ preserves a regular positive Borel measure $\mu$ on $X$.

Fundamental example of such a situation is: $X = \mathbf{S}^2$ with $d$ being the angular distance on $\mathbf{S}^2$, $\mathcal{L}$ the Lebesgue measure[1] on $\mathbf{S}^2$ and $G = \mathrm{SO}(3)$.

Now back to the general setting, for $g \in G$ and $f \in L^2(X, \mu)$ define

$$\pi(g)f(x) = f(g^{-1}x).$$

As we have already seen in Example 1.28, $\pi(g)$ is an unitary operator of $L^2(X)$, since (cf. Theorem 8.3)

$$\begin{aligned} \|\pi(g)f\|^2_{L^2(X)} &= \int_X |f(g^{-1}x)|^2 \, d\mu(x) \\ &= \int_X |f(x)|^2 \, dg^{-1}_*\mu(x) \\ &= \int_X |f(x)|^2 \, d\mu(x) = \|f\|_{L^2(X)} \end{aligned}$$

where by used that $g$ preserves the measure $\mu$, i.e.

$$g^{-1}_*\mu(E) = \mu(g(E)) = \mu(E).$$

Hence

$$\pi\colon G \to U(L^2(X))$$

is a group homomorphism.

**Task:** Decompose $L^2(X)$ into an orthogonal sum of closed subspaces that are invariant under $\pi(g)$ for all $g \in G$ and "minimal" in a reasonable sense.

---

[1]The Lebesgue measure on $\mathbf{S}^n$ can be defined by

$$\mathcal{L}(E) = \lambda^{n+1}(\{tx\colon x \in E,\, 0 \leqslant t \leqslant 1\}) \qquad \forall E \subset \mathbf{S}^n.$$

Let now $K \in C(X \times X)$ be a continuous kernel such that $K(gx, gy) = K(x, y)$ for all $g \in G$ and $x, y \in X$. We claim now that

$$\pi(g)T_K = T_K\pi(g).$$

Indeed, a direct computation yields

$$\begin{aligned}(\pi(g)T_K f)(x) &= T_K f(g^{-1}x) \\ &= \int_X K(g^{-1}x, y)f(y)\, d\mu(y) \\ &= \int_X K(x, gy)f(y)\, d\mu(y) \\ &= \int_X K(x, y)f(g^{-1}y)\, d\mu(y) = (T_K\pi(g))f(x).\end{aligned}$$

This has the following remarkable consequences: If $K(x, y) = \overline{K(y, x)}$ for all $x, y \in X$; then $T_K\colon L^2(X) \to L^2(X)$ is a compact self-adjoint operator. For every eigenvalue $\lambda \neq 0$ of $T_K$, the corresponding finite dimensional eigenspace $\mathcal{H}_\lambda \subset L^2(X)$ of $T_K$ is invariant under $\pi(g)$, $g \in G$.

In fact in our situation there is a pletora of such kernels, namely if $k\colon [0, \infty) \to \mathbf{R}$ is continuous then $K(x, y) := k(d(x, y))$ is such a valid kernel.

This leads to the following theorem.

**Theorem 3.18.** $L^2(X)$ is a direct orthogonal sum of $\pi(G)$-invariant (irreducible) finite dimensional subspaces.

In the case of SO(3) acting on $\mathbf{S}^2$ this decomposition takes the following concrete form: Recall that a polynomial $P \in \mathbf{R}[x, y, z]$ is *harmonic* if $\Delta P = 0$ where

$$\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$$

is the *Laplace operator*. Let then

$$\mathcal{H}_n = \{\, P|_{\mathbf{S}^2} : P\colon \mathbf{R}^3 \to \mathbf{R} \text{ is homogeneous of degree } n \text{ and harmonic}\}.$$

Then:

$$L^2(\mathbf{S}^2) = \overline{\bigoplus_{n \geqslant 0} \mathcal{H}_n}$$

and the action of SO(3) in $\mathcal{H}_n$ is irreducible.

## 3.3. Mercer's Theorem

We begin with the current terminology belonging to this context. A kernel on a set $X$ is a function $K\colon X \times X \to \mathbf{R}$; it is symmetric if $K(x, y) = K(y, x)$ for all $x, y \in X$.

**Definition 3.19.** A symmetric kernel $K$ on a set $X$ is positive semidefinite if for all $n \geqslant 1$ and $x_1, \dots, x_n \in X$, the symmetric matrix $(K(x_k, x_j))_{k,j}$ is positive semidefinite. That is, for all $c_1, \dots, c_n \in \mathbf{R}$,

$$\sum_{k,j=1}^{n} c_k c_j K(x_k, x_j) \geqslant 0.$$

The case $n = 1$ implies $K(x, x) \geqslant 0$ for all $x \in X$ (3.20).

**Example 3.21.** If $\mathcal{H}$ is a $\mathbf{R}$-Hilbert space and $\varphi\colon X \to \mathcal{H}$ is any map, then

$$K(x, y) := \langle \varphi(x), \varphi(y) \rangle$$

is a positive semidefinite kernel on $X$.

In our context we will take $(X, d)$ to be a compact metric space endowed with a regular Borel probability measure $\mu \in M^1(X)$. Given $K \in C(X \times X)$ continuous we know from Proposition 3.12 that the operator

$$T_K\colon L^2(X, \mu) \to L^2(X, \mu), \quad T_K f(x) = \int_X K(x, y) f(y)\, d\mu(y)$$

is Hilbert-Schmidt and hence compact. If in addition $K$ is a symmetric kernel, $T_K$ is self-adjoint and the spectral theorem (Theorem 3.14) applies. Observe that our hypothesis on $X$ and $\mu$ guarantees that $L^2(X)$ is separable.

**Theorem 3.22** (Mercer)**.** Let $(X, d)$ be a compact metric space, $\mu \in M^1(X)$ Borel regular such that for all $U \subset X$ open non-empty $\mu(U) > 0$. Let $K \in C(X \times X)$ be a continuous positive semi-definite kernel on $X$. Then there is an ONB $\{\varphi_n\}_{n \geqslant 1}$ of $\ker(T_K)^{\perp}$ consisting of continuous eigenfunctions of $T_K$ and if $\lambda_k$ is the eigenvalue corresponding to $\varphi_k$ then $\lambda_k > 0$ for all $k \geqslant 1$. In addition,

$$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \varphi_n(y)$$

the sum being absolutely and uniformly convergent.

Observe that $T_K$ being Hilbert-Schmidt gives us

$$\sum_{n=1}^{\infty} \lambda_n^2 = \sum_{n,m=1}^{\infty} |\langle T_K \varphi_n, \varphi_m \rangle|^2 < +\infty$$

**Corollary 3.23.** In the situation of Theorem 3.22 we have

$$K(x, x) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x)^2,$$

with uniform convergence on the right hand side. In particular

$$\sum_{n=1}^{\infty} \lambda_n = \int_X K(x, x)\, d\mu(x) < +\infty.$$

We proceed with the proof of Mercer's theorem by first showing the following lemma whose proof is left as an exercise.

**Lemma 3.24.** Let $(X, d)$ and $\mu \in M^1(X)$ be as in Mercer's theorem. In particular, $\mu(U) > 0$ for nonempty open $U \subset X$. Given a symmetric $K \in C(X \times X)$, the following are equivalent:

(1) $K$ is positive semidefinite.
(2) For all $f \in C(X)$ we have
$$\int_X \int_X f(x)f(y)K(x,y)\,d\mu(x)\,d\mu(y) \geqslant 0.$$
(3) $\langle T_K f, f\rangle \geqslant 0$ for all $f \in L^2(X)$.

**Proof.** (1) $\implies$ (2) Let $f \in C(X)$ and $\varepsilon > 0$ be arbitrary. By uniform continuity of $K$ and $f$ we can find $\delta > 0$ s.t. for all $x_1, x_2, y_1, y_2 \in X$, $(x_2, y_2) \in B_{<\delta}(x_1) \times B_{<\delta}(y_1)$ implies that $|K(x_1, y_1) - K(x_2, y_2)| < \varepsilon$ as well as $|f(x_2) - f(x_1)| < \varepsilon$. Pick $x_1, \ldots, x_n \in X$ s.t. $\bigcup_{i=1}^n B_{<\delta}(x_i) = X$ (compactness) and turn this into a disjoint partition of Borel sets by setting $M_1 := B_{<\delta}(x_1)$ and $M_k := B_{<\delta}(x_k) \setminus \bigcup_{1 \leqslant i < k} B_{<\delta}(x_i)$. Then we have
$$\begin{aligned}
&\int_X \int_X f(x)f(y)K(x,y)\,d\mu(x)\,d\mu(y)\\
&= \sum_{i,j=1}^n \int_{M_j} \int_{M_i} f(x)f(y)K(x,y)\,d\mu(x)\,d\mu(y)\\
&= \sum_{i,j=1}^n \int_{M_j} \int_{M_i} (f(x_i) + \varepsilon_i)(f(x_j) + \varepsilon_j)(K(x_i, x_j) + \varepsilon_{ij})\,d\mu(x)\,d\mu(y)\\
&= \underbrace{\sum_{i,j=1}^n f(x_i)f(x_j)K(x_i, x_j)}_{\geqslant 0} + O(\varepsilon) \geqslant O(\varepsilon)
\end{aligned}$$
with $\varepsilon_i, \varepsilon_{jk} \in (-\varepsilon, \varepsilon)$ for all $1 \leqslant i, j, k \leqslant n$. We used that due to continuity one has $|f| \leqslant C_1$ and $|K| \leqslant C_2$ on $X$ respectively $X \times X$. Letting $\varepsilon \to 0$ yields the desired inequality.

(2) $\implies$ (3) We employ density of $C(X)$ in $L^2(X)$ (note that due to compactness $C(X) \subset L^2(X)$ coincides with the continuous compactly supported functions on $X$). Let $f \in L^2(X)$ and $\varepsilon > 0$ be arbitrary; let $\varphi_\varepsilon \in C(X)$ s.t. $\|\varphi_\varepsilon - f\|_{L^2(X)} < \varepsilon$. Then
$$\begin{aligned}
\langle T_K f, f\rangle &= \langle T_K(f - \varphi_\varepsilon + \varphi_\varepsilon), f - \varphi_\varepsilon + \varphi_\varepsilon\rangle\\
&= \langle T_K(f - \varphi_\varepsilon), (f - \varphi_\varepsilon)\rangle + 2\langle T_K(f - \varphi_\varepsilon), \varphi_\varepsilon\rangle + \underbrace{\langle T_K \varphi_\varepsilon, \varphi_\varepsilon\rangle}_{\geqslant 0}.
\end{aligned}$$

Now, by Cauchy-Schwarz,
$$|\langle T_K(f - \varphi_\varepsilon), \varphi_\varepsilon\rangle| \leqslant \|T_K(f - \varphi_\varepsilon)\|_{L^2(X)} \|\varphi_\varepsilon\|_{L^2(X)}$$

$$\leqslant \|K\|_{L^\infty(X\times X)}\|f-\varphi_\varepsilon\|_{L^2(X)}\|\varphi_\varepsilon\|_{L^2(X)}$$
$$\leqslant \varepsilon\|K\|_{L^\infty(X\times X)}\|\varphi_\varepsilon\|_{L^2(X)}$$

which goes to 0 as $\varepsilon \to 0$. We used that $\|K\|_{L^\infty(X\times X)} < +\infty$ and $\|\varphi_\varepsilon\|_{L^2(X)} \to \|f\|_{L^2(X)} < +\infty$ as $\varepsilon \to 0$. A similar argument applies to the first summand.

(3) $\implies$ (1) Let $c_1,\dots,c_n \in \mathbf{R}$ and $x_1,\dots,x_n \in X$ be arbitrary. Given $\varepsilon > 0$ we choose, by uniform continuity of $K$, some $\delta > 0$ s.t. for any $x_1, x_2, y_1, y_2 \in \mathbf{R}$,

$$d(x_1,x_2)+d(y_1,y_2) < \delta \implies |K(x_1,y_1)-K(x_2,y_2)| < \varepsilon.$$

Now we set

$$f_\varepsilon = \sum_{k=1}^n c_k \frac{1}{\mu(B_{<\delta}(x_k))}\mathbf{1}_{B_{<\delta}(x_k)}$$

which is well defined (since $B_{<\delta}(x_k)$ is open and nonempty) and in $L^2(X)$. We find

$$\begin{aligned}
0 &\leqslant \langle T_K f_\varepsilon, f_\varepsilon\rangle \\
&= \sum_{k,j=1}^n \frac{c_k c_j}{\mu(B_{<\delta}(x_k))\mu(B_{<\delta}(x_j))}\int_{B_{<\delta}(x_j)}\int_{B_{<\delta}(x_k)} K(x,y)\,d\mu(x)\,d\mu(y) \\
&= \sum_{k,j=1}^n \frac{c_k c_j}{\mu(B_{<\delta}(x_k))\mu(B_{<\delta}(x_j))}\int_{B_{<\delta}(x_k)}\int_{B_{<\delta}(x_j)} (K(x_k,x_j)+O(\varepsilon))\,d\mu(x)\,d\mu(y) \\
&= \sum_{k,j=1}^n c_k c_j K(x_k,x_j) + O(\varepsilon)\sum_{k,j=1}^n c_k c_j
\end{aligned}$$

so letting $\varepsilon \to 0$ yields the desired inequality.

■

We now come to the proof of Mercer's Theorem.

**Proof.** (1) We start by observing that for all $f \in L^2(X)$,

$$T_K f(x) = \int_X K(x,y) f(y)\,d\mu(y)$$

is well defined for all $x \in X$ and continuous. Indeed,

$$\begin{aligned}
|T_K f(x_1) - T_K f(x_2)| &\leqslant \int_X |K(x_1,y)-K(x_2,y)||f(y)|\,d\mu(y) \\
&\leqslant \Big(\int_X |K(x_1,y)-K(x_2,y)|^2\,d\mu(y)\Big)^{1/2}\|f\|_{L^2(X)}.
\end{aligned}$$

Now $K\colon X\times X \to \mathbf{R}$ being continuous on the compact metric space implies uniform continuity; that is, for all $\varepsilon > 0$ there exists a $\delta > 0$ s.t.

$$d(x_1,x_2)+d(y_1,y_2) < \delta \implies |K(x_1,y_1)-K(x_2,y_2)| < \varepsilon.$$

In particular,

$$d(x_1,x_2)<\delta \implies |K(x_1,y)-K(x_2,y)|<\varepsilon$$

for all $y \in X$ which implies

$$|T_K f(x_1) - T_K f(x_2)| \leqslant \varepsilon \|f\|_{L^2(X)}.$$

(2) By the spectral theorem let $\{f_n\}_{n\geqslant 1}$ be an ONB of $\ker(T_K)^\perp$ consisting of eigenvectors of $T_K$ and write $\lambda_n$ for the eigenvalue corresponding to $f_n$. Then, by Lemma 3.24, $0 \leqslant \langle T_K f_n, f_n\rangle = \lambda_n$ and since $f_n \notin \ker(T_K)$ we get $\lambda_n > 0$ for all $n \geqslant 1$. Thus, we have $f_n = \frac{1}{\lambda_n} T_K f_n$ a.e. By (1) we know that $T_K f_n$ is continuous, hence we define $\varphi_n(x) = \frac{1}{\lambda_n} T_K f_n(x)$ for all $x \in X$. Then $\varphi_n \in C(X)$ and $\varphi_n = f_n$ a.e. This proves the first part of the theorem.

(3) Define

$$K_n(x,y) := K(x,y) - \sum_{k=1}^{n} \lambda_k \varphi_k(x)\varphi_k(y).$$

Then $K_n \in C(X\times X)$ and it is symmetric. We claim that $K_n$ is psd. Indeed, for $f \in L^2(X)$ we have

$$\langle T_{K_n} f, f\rangle = \langle T_K f, f\rangle - \sum_{k=1}^{n} \lambda_k \langle f, \varphi_k\rangle^2.$$

Now we expand $f$ as

$$f = \sum_{k=1}^{\infty} \langle f, \varphi_k\rangle \varphi_k + g$$

where $g$ is the orthogonal projection of $f$ onto $\ker(T_k)$. Then

$$\langle T_K f, f\rangle = \sum_{k=1}^{\infty} \lambda_k \langle f, \varphi_k\rangle^2$$

and hence

$$\langle T_{K_n} f, f\rangle = \sum_{k=n+1}^{\infty} \lambda_k \langle f, \varphi_k\rangle^2 \geqslant 0$$

so that psd. follows from Lemma 3.24. In particular

$$K(x,x) - \sum_{k=1}^{n} \lambda_k \varphi_k(x)^2 = K_n(x,x) \geqslant 0$$

so that

$$\sum_{k=1}^{n} \lambda_k \varphi_k(x)^2 \leqslant K(x,x)$$

with the left hand side (absolutely) convergent for every $x \in X$.

(4) We deduce for $1 \leqslant N \leqslant M$:

$$\sum_{k=N}^{M} \lambda_k |\varphi_k(x)||\varphi_k(y)| \leqslant \Big(\sum_{k=N}^{M} \lambda_k \varphi_k(x^2)\Big)^{1/2} \Big(\underbrace{\sum_{k=N}^{M} \lambda_k \varphi_k(y)^2}_{\leqslant K(y,y)}\Big)^{1/2}$$

$$\leqslant \Big(\sum_{k=N}^{M} \lambda_k \varphi_k(x)^2\Big)^{1/2} \|K\|_b^{1/2}.$$

This implies that for all $x \in X$, $\sum_{k=1}^{\infty} \lambda_k \varphi_k(x)\varphi_k(y)$ converges absolutely and uniformly in $y$ and by symmetry for all $y \in X$ converges absolutely and uniformly in $x$.

(5) Let now $K_x(y) := K(x, y)$, take any ONB $\{\psi_n\}_{n \geqslant 1}$ of $\ker(T_K)$ and expand $K_x \in L^2(X)$ in the ONB $\{\varphi_n, \psi_n\}_{n \geqslant 1}$:

$$K_x = \sum_{k=1}^{\infty} \underbrace{\langle K_x, \varphi_k \rangle}_{\lambda_k \varphi_k(x)} \varphi_k + \underbrace{\sum_{k=1}^{\infty} \langle K_x, \psi_k \rangle \psi_k}_{=0}$$

so that $K_x = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x) \varphi_k$. This means

$$K_x = \lim_{n \to \infty} \sum_{k=1}^{n} \lambda_k \varphi_k(x) \varphi_k$$

in $L^2(X)$. Thus, there exists a subsequence $(n_\ell)_{\ell \geqslant 1}$ s.t. $\sum_{k=1}^{n_\ell} \lambda_k \varphi_k(x)\varphi_k(y)$ converges pointwise a.e. in $y$ to $K_x(y)$. Now this implies that for all $x \in X$ the continuous functions $y \mapsto \sum_{k=1}^{\infty} \lambda_k \varphi_k(x)\varphi_k(y)$ and $y \mapsto K(x, y)$ coincide a.e. Since $\mu(U) > 0$ for nonempty open $U \subset X$ we deduce that they coincide everywhere (employing continuity). Hence

$$K(x, y) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x) \varphi_k(y)$$

for all $x, y \in X$ and in particular

$$K(x, x) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(x)^2$$

for all $x \in X$.

(6) Now we show that the convergence of $\sum_{k=1}^{n} \lambda_k \varphi_k(x)^2$ to $K(x, x)$ is uniform. Fix $\varepsilon > 0$ and let

$$V_n^{\varepsilon} = \Big\{ x \in X \colon \sum_{k=1}^{n} \lambda_k \varphi_k(x)^2 > K(x, x) - \varepsilon \Big\}.$$

Then $V_n^{\varepsilon}$ is open for all $n \geqslant 1$, $V_n^{\varepsilon} \subset V_{n+1}^{\varepsilon}$ and by pointwise convergence

$$\bigcup_{n \geqslant 1} V_n^{\varepsilon} = X.$$

Since $X$ is compact there exists a finite subcover and due to the inclusions there must exist some $n(\varepsilon) \geqslant 1$ s.t. $V^{\varepsilon}_{n(\varepsilon)} = X$ which shows that the convergence is uniform.

(7) Going back to the inequality in (4):

$$\sum_{k=N}^{M} \lambda_k |\varphi_k(x)||\varphi_k(y)| \leqslant \Big(\sum_{k=N}^{M} \lambda_k \varphi_k(x)^2\Big)^{1/2} \Big(\sum_{k=N}^{M} \lambda_k \varphi_k(y)^2\Big)^{1/2}$$

we deduce that $\sum_{k=1}^{\infty} \lambda_k \varphi_k(x)\varphi_k(y)$ converges absolutely and uniformly in $X \times X$ with sum $K(x, y)$. ■

**Chapter** 4. **Baire Category and its consequences**

This chapter is devoted to some of the major theorems in functional analysis. They are all consequences of a result in point-set topology, that is the Baire category theorem. This theorem is a topological analogue of the fact from measure theory that a set of positive measure cannot be countable union of sets of measure zero.

## 4.1. Baire Category

The idea of category of a set in a metric space is to describe "smallness" resp. "generosity" in purely topological terms. Its origins lie in the thesis of Baire who answered the following question: given a sequence $f_n\colon \mathbf{R}\to\mathbf{R}$ of continuous functions converging *pointwise* to a function $f\colon \mathbf{R}\to\mathbf{R}$, that is

$$f(x)=\lim_{n\to\infty} f_n(x) \qquad \forall x\in\mathbf{R}$$

what can one say about the subset of points in $\mathbf{R}$ at which $f$ is continuous? We will see that this set is "big" in a precise way. We now turn to the relevant definitions: let $X$ be a topological space and $S\subset X$ a subset. We recall that the interior $S^\circ$ of $S$ is the union of all open subsets of $X$ contained in $S$.

**Definition 4.1.** A subset $S\subset X$ is *nowhere dense* if its closure $\overline{S}$ has empty interior, that is $(\overline{S})^\circ=\varnothing$.

Note that $S$ is nowhere dense iff it is not dense in any open ball $B_{<r}(x)\subset X$. Indeed, if $(\overline{S})^\circ=\varnothing$ and $S$ was dense in some $B_<(x)$ then its closure would contain $B_{<r}(x)$, a contradiction. Conversely, if $(\overline{S})^\circ\neq\varnothing$ then $\overline{S}$ contains some open ball $B_{<r}(x)$, so in particular $S$ is dense in $B_{<r}(x)$.

**Example 4.2.** (a) A point in $\mathbf{R}^d$ is nowhere dense ($n\geqslant 1$).
(b) The Cantor set in $[0,1]$ is (closed and) nowhere dense.
(c) $\mathbf{Q}\subset\mathbf{R}$ is not nowhere dense since $\overline{\mathbf{Q}}=\mathbf{R}$.

However:

(d) $\{(x,0)\colon x\in\mathbf{Q}\}$ is nowhere dense in $\mathbf{R}^2$.
(e) Let $f\colon\mathbf{R}^d\to\mathbf{R}$ be smooth and assume $y\in f(\mathbf{R}^d)$ is a regular value. Then $f^{-1}(y)\subset\mathbf{R}^d$ is nowhere dense in $\mathbf{R}^d$.

**Definition 4.3.** (1) A set $S\subset X$ is of *first category* in $X$ if it is a countable union of nowhere dense subsets of $X$; a subset $S\subset X$ that is not of the first category is of the *second category*.
(2) A subset $S\subset X$ is *generic* if its complement is of the first category.

**Example 4.4.** $\mathbf{Q}$, while being dense in $\mathbf{R}$, is however of first category and hence $\mathbf{R}\setminus\mathbf{Q}$, while being dense in $\mathbf{R}$ as well, is generic.

The main result of Baiare is that $\mathbf{R}$ is of second category in itself. This actually holds for complete metric spaces as the following shows.

**Theorem 4.5.** Let $(X, d)$ be a complete metric space with $X \neq \varnothing$. Then the following assertions hold:

(1) Let $U_n \subset X$, $n \in \mathbf{N}$, be open and dense subsets. Then $U := \bigcap_{n\in\mathbf{N}} U_n$ is dense in $X$.
(2) Let $F_n \subset X$, $n \in \mathbf{N}$, be a family of closed subsets of $X$ s.t. $(\bigcup_{n\in\mathbf{N}} F_n)^\circ \neq \varnothing$. Then there is $n_0 \in \mathbf{N}$ with $F_{n_0}^\circ \neq \varnothing$.
(3) Let $X = \bigcup_{n\in\mathbf{N}} F_n$ with $F_n$ closed for all $n \in \mathbf{N}$. Then there exists $n_0 \in \mathbf{N}$ with $F_{n_0}^\circ \neq \varnothing$.
(4) If $X = \bigcup_{n\in\mathbf{N}} A_n$ then there exists $A_n$ which is dense in some $B_{<r}(x) \subset X$.

We begin with

**Lemma 4.6.** For $U \subset X$ and $F = X \setminus U$ the following are equivalent:

(1) $U$ is open and dense in $X$.
(2) $F$ is closed and nowhere dense in $X$.

**Proof.** (1) $\implies$ (2): If there existed $V \subset F$ open we would also have $V \cap U \neq \varnothing$ due to density of $U$ in $X$, which is a contradiction.

(2) $\implies$ (1): Let $x \in X$ and $U_x$ be some neighbourhood of $x$. Then $U_x \cap U \neq \varnothing$ since otherwise $U_x \subset X \setminus U$ which would contradict nowhere density of $X \setminus U$. Thus $U$ is dense in $X$. ■

Now let us proof Theorem 4.5.

**Proof.** (1) Let $V \subset X$ be an arbitrary open set. Then by density of $U_1$ there exists some $x_1 \in U_1 \cap V$ and some $0 < r_1 < 1$ s.t. $B_{\leqslant r_1}(x_1) \subset U_1 \cap V$. This follows from the fact that $U_1 \cap V$ is open so we can find an open ball around $x_1$ contained in $U_1 \cap V$ and in turn some smaller open ball stricly contained in the former so that its closure will also be contained in $U_1 \cap V$. We now iterate this, so in the next step with $B_{r_1}(x_1)$ in the role of $V$ and $0 < r_2 < \frac{1}{2}$.

By means of this we construct the sequence $(x_n, r_n)$ recursively[1] s.t. $0 < r_n < \frac{1}{n}$ and

$$B_{\leqslant r_n}(x_n) \subset U_n \cap B_{r_{n-1}}(x_{n-1}) \subset \bigcap_{k=1}^{n} U_k \cap V$$

whereby we use that the finite intersection of open sets is open. From the construction it becomes clear that $(x_n)_{n\geqslant 1}$ is a Cauchy sequence since for $n, m \geqslant N$ we have $d(x_n, x_m) \leqslant \frac{1}{N}$, so by completeness of $X$ it converges to some $x \in X$. Lastly, for any $n \geqslant 1$ we have that $x \in B_{\leqslant r_n}(x_n)$ since $(x_m)_{m\geqslant n}$ is a convergent sequence contained in the closed set $B_{\leqslant r_n}(x_n)$, letting us conclude $x \in U_n$ for all $n \geqslant 1$, as desired.

[1]Note that for justifying this construction we require the axiom of dependent choice. One can in fact show that the axiom of dependent choice is equivalent to the Baire Category Theorem.

(1) $\Longrightarrow$ (3): If $F_n^\circ = \varnothing$ for all $n \in \mathbf{N}$ then

$$\varnothing = X \setminus \bigcup_{n\in\mathbf{N}} F_n = \bigcup_{n\in\mathbf{N}} (X \setminus F_n)$$

which contradicts (1) since $X \setminus F_n$ is dense in $X$ for all $n \in \mathbf{N}$ using Lemma 4.6.

(1) $\Longrightarrow$ (4): This is a direction consequence of (3), which guarantees the existence of some $A_n$ s.t. $\overline{A}_n$ has non-empty interior. ■

In order to prove (2) we need

**Lemma 4.7.** Let $\varnothing \neq Y \subset X$ be open in a complete metric space $(X, d)$. Then $Y$ satisfies properties (1) and (3) in Theorem 4.5.

**Proof.** $\overline{Y}$ is again complete, being the closed subset of a complete metric space. Now if $U_n$ is dense in $Y$ it is also dense in $\overline{Y}$ so we can apply (1) of Theorem 4.5 to deduce that $\bigcap_{n\geqslant 1} U_n$ is dense in $\overline{Y}$ and thus also in $Y$. (3) is implied by (1) (note that (3) does not require the metric space to be complete but merely to satisfy (1)). ■

Now let us proof (2) of Theorem 4.5.

**Proof.** Let $F_n$ be closed in $X$ and $U := (\bigcup_{n\geqslant 1} F_n)^\circ$ non empty. Then $F_n \cap U$ is closed in $U$ and clearly $U = \bigcup_{n\geqslant 1}(F_n \cap U)$. Hence by (3) of Lemma 4.7 there is a $n_0$ s.t. $U \cap F_{n_0}$ contains a non-empty subset $W$ that is open in $U$ hence in $X$. Thus $0 \neq W \subset F_{n_0}$ which implies $F_{n_0}^\circ \neq \varnothing$. ■

We can rephrase a consequence of Theorem 4.5 as follows:

**Corollary 4.8.** A complete non-empty metric space is of second category, as is any of its non-empty open subsets.

**Corollary 4.9.** Any generic subset of a non-empty complete metric space is dense, as is any generic subset of an open subset of such a metric space.

**Remark 4.10.** There is little relation between being generic and bein of positive Lebesgue measure, as the following examples shows:

(1) $\lambda([0, 1]) = 1$ but $[0, 1]$ is not generic, say in $\mathbf{R}$.
(2) Let $\mathbf{N} \to \mathbf{Q}$, $n \mapsto q_n$ be a bijection and for every $n \geqslant 1$,

$$U_n := \bigcup_{k\in\mathbf{N}} (q_k - 2^{-(n+k+1)}, q_k + 2^{-(n+k+1)}).$$

Then

$$\lambda(U_n) \leqslant \sum_{k=0}^{\infty} 2^{-(n+k)} = 2^{-(n-1)}.$$

However, $U_n$ is open and dense in $\mathbf{R}$, hence $\bigcap_{n\geqslant 1} U_n \subset \mathbf{R}$ is generic but

$$\lambda\Big(\bigcap_{n\geqslant 1} U_n\Big) = \lim_{n\to\infty} \lambda(U_n) = 0$$

## 4.2. Some applications

Next we present two applications of the Baire category theorem, the first is due to Baire:

**Theorem 4.11.** Let $f_n \colon X \to \mathbf{C}$ be a sequence of continuous functions on a complete metric space $X$ such that

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists for all $x \in X$. Then the set of points where $f$ is continuous is generic in $X$. For a proof cf. [**SS11**] Chapter 4, section 1.1.

It is well know that in $\mathbf{R}$ there are continuous functinos that are nowhere differentiable, e.g.

$$f(x) = \sum_{n=0}^{\infty} 2^{-n\alpha} e^{i2^n x}, \quad 0 < \alpha \leqslant 1$$

and the question is: how common is this phenomenon? In fact, let $C([0,1])$ be the Banach space of continuous functions with sup norm $\|f\|_b = \sup_{x\in[0,1]} |f(x)|$. Then:

**Theorem 4.12.** The set of functinos in $C([0,1])$ that are nowhere differentiable is generic.

Again, a proof can be found in [**SS11**]: Chapter 4, section 1.2.

In fact, while both theorems use the Baire Category Theorem, the proofs are rather tricky.

We close this subsection with an application of Baire category which will have far reaching far reaching consequences in Function Analysis.

**Proposition 4.13** (Principle of uniform boundedness)**.** Let $(X,d)$ be a complete metric space and $f_\lambda \colon X \to \mathbf{R}$, $\lambda \in \Lambda$ a family of continuous functions such that

$$\sup_{\lambda\in\Lambda} |f_\lambda(x)| < +\infty \qquad \forall x \in X.$$

Then there is an open ball $B_{<r}(y)$ $(r > 0)$ s.t.

$$\sup_{\lambda\in\Lambda} \sup_{x\in B} |f_\lambda(x)| < +\infty.$$

**Proof.** For every $n \in \mathbf{N}$ consider the closed subset

$$A_n := \{x \in X \colon |f_\lambda(x)| \leqslant n \quad \forall \lambda \in \Lambda\}$$

$$= \bigcap_{\lambda \in \Lambda} \{x \in X : |f_\lambda(x)| \leqslant n\}.$$

Then by hypothesis $X = \bigcup_{n\in\mathbf{N}} A_n$ and by (3) of Theorem 4.5 there exists $n_0 \in \mathbf{N}$ with $A_{n_0}^\circ \neq \varnothing$. Now take $y \in A_{n_0}^\circ$ and $r > 0$ with $B_{<r}(y) \subset A_{n_0}$. ■

## 4.3. The uniform boundedness principle

The combination of Proposition 4.13 with the linear structure of a vector space has the following consequences:

**Theorem 4.14** (Banach-Steinhaus)**.** Let $(V, \|\cdot\|_V)$ be a Banach space, $(W, \|\cdot\|_W)$ a normed space and $T_\lambda \in B(V,W)$, $\lambda \in \Lambda$, a family of bounded linear operators with

$$\sup_{\lambda\in\Lambda} \|T_\lambda(v)\|_W < +\infty \qquad \forall v \in V.$$

Then $\sup_{\lambda\in\Lambda} \|T_\lambda\| < +\infty$.

**Proof.** Proposition 4.13 gives us an open ball $B_{<r}(x) \subset V$ s.t.

$$\sup_{\lambda\in\Lambda} \sup_{v\in B_{<r}(x)} \|T_\lambda(v)\|_W \leqslant C < +\infty$$

so for all $v \in B_{<1}(0) = \frac{1}{r}B_{<r}(x) - \frac{x}{r}$ we can write $v = \frac{1}{r}u_v - \frac{x}{r}$ for $u_v \in B_{<r}(x)$ to find

$$\|T_\lambda(v)\|_W = \left\|\frac{1}{r}T_\lambda(u_v) - \frac{1}{r}T_\lambda(x)\right\| \leqslant \frac{2C}{r} \qquad \forall v \in B_{<1}(0), \forall \lambda \in \Lambda$$

yielding $\sup_{\lambda\in\Lambda} \|T_\lambda\| < +\infty$. ■

Our first application is to what one can say about a sequence $T_n : V \to W$ of bounded operators converging pointwise.

**Corollary 4.15.** Let $T_n \in B(V,W)$ where $V$ is a Banach space and assume

$$T(x) = \lim_{n\to\infty} T_n(x)$$

exists for for all $x \in V$. Then

(1) $\sup_{n\geqslant 1} \|T_n\| < +\infty$
(2) $T \in B(V,W)$
(3) $\|T\| \leqslant \liminf_{n\to\infty} \|T_n\|$.

**Remark 4.16.** The theorem does **not** say that $\|T_n - T\| \to 0$. Indeed, consider

$$T_n : \ell^2(\mathbf{N}) \to \ell^2(\mathbf{N}), \quad x = \sum_{k=1}^\infty x_k\delta_k \mapsto x_n\delta_n.$$

Then $\|T_n x\| = |x_n| \to 0$ for all $n \in \mathbf{N}$ so $\lim_{n\to\infty} T_n x = 0$. However, $\|T_n\| = 1$ for all $n \in \mathbf{N}$.

**Proof.** (1) Since $(T_n x)_{n\geqslant 1}$ converges for all $x$, we have $\sup_{n\geqslant 1}\|T_n x\| < +\infty$ for all $x$ so, by Theorem 4.14 , $\sup_{n\geqslant 1}\|T_n\| < +\infty$.

(2)+(3) By linearity of the limit $T$ is linear. Let $l := \liminf_{n\to\infty}\|T_n\|$ and $(T_{n_k})_{k\geqslant 1}$ be a subsequence s.t. $l = \lim_{k\to\infty}\|T_{n_k}\|$. Now, for all $x \in V$ we have $Tx = \lim_{k\to\infty} T_{n_k}(x)$ and hence

$$\|Tx\| = \lim_{k\to\infty}\|T_{n_k}x\| \leqslant \lim_{k\to\infty}\|T_{n_k}\|\|x\|$$

implying $\|T\| \leqslant l$. ■

Next we deduce two corollaries that are useful to detect bounded subsets in Banach spaces.

**Corollary 4.17.** Let $E$ be a normed space and $B \subset E$ a subset such that for all $f \in E^*$, $f(B) \subset \mathbf{K}$ is bounded. Then $B \subset E$ is bounded.

**Proof.** We apply Theorem 4.14 with $V = E^*$, $W = \mathbf{K}$ and $\Lambda = B$. Define for all $v \in B$,

$$T_v\colon E^* \to \mathbf{K}, \quad f \mapsto f(v).$$

Then $\sup_{v\in B}|T_v f| < +\infty$ for all $f \in E^*$. Using Theorem 4.14 we deduce that

$$\sup_{v\in B}\|T_v\| < +\infty$$

and by Corollary 2.10 we have $\|T_v\| = \|v\|$ concluding the proof. ■

We also have the analogous dual statement:

**Corollary 4.18.** Assume $E$ is a Banach space and $B^* \subset E^*$ is a subset such that $\{f(x)\colon f \in B^*\} \subset \mathbf{K}$ is bounded for all $x \in E$. Then $B^* \subset E^*$ is bounded.

**Proof.** Similarly to the previous proof we define for all $f \in B^*$

$$T_f\colon E \to \mathbf{K}, \quad x \mapsto f(x).$$

By assumption we have $\sup_{f\in B^*}\|T_f(x)\| < +\infty$ for all $x \in E$, hence invoking Theorem 4.14 (with $V = E$, $W = \mathbf{K}$ and $\Lambda = B^*$) yet again yields

$$\sup_{f\in B^*}\|T_f\| < +\infty.$$

Together with

$$\sup_{f\in B^*}\|T_f\| = \sup_{f\in B^*}\sup_{x\in E}|f(x)| = \sup_{f\in B^*}\|f\|$$

this concludes the proof. ■

## 4.4. The open mapping theorem and the closed graph theorem

A general question, once one has a category, is whether bijective morphisms are automatically isomorphisms. In our case this translates to the question whether a bounded linear operator between normed spaces that is bijective has a bounded inverse. In general the answer to this question is no. But if both spaces are Banach, the answer is yes will follow from the more general theorem we present now.

**Theorem 4.19** (open mapping theorem)**.** Let $X, Y$ be a Banach space and $T\colon X \to Y$ a bounded operator. Then the following are equivalent:

(1) $T$ is surjective.
(2) $T$ is open.
(3) (Qualitative solvability) For every $y \in Y$ there exists a solution $u \in X$ to the equation $Tu = y$.
(4) (Quantitative solvability) There exists a constant $C > 0$ such that for every $y \in Y$ there exists a solution $u \in X$ to the equation $Tu = y$ which obeys the bound $\|u\|_X \leqslant C\|y\|_Y$.

**Proof.** It is clear that (1) and (3) are equivalent and that (4) implies (3). We will first show that (2) and (4) are equivalent. If $f$ is open there exists $\varepsilon > 0$ s.t. $B_{<\varepsilon}(0) \subset T(B_{<1}(0))$. Now, for arbitrary $y \in Y \setminus \{0\}$ we have $\frac{\varepsilon}{2}\frac{y}{\|y\|_Y} \in B_{<\varepsilon}(0)$ so there exists $x \in B_{<1}(0)$ with $Tx = \frac{\varepsilon}{2}\frac{y}{\|y\|_Y}$. Then, for $u := \frac{2\|y\|_Y}{\varepsilon}x$,

$$T(u) = T\Big(\frac{2\|y\|_Y}{\varepsilon}x\Big) = \frac{2\|y\|_Y}{\varepsilon}T(x) = y$$

and $\|u\| \leqslant C\|y\|_Y$ for $C := \frac{2}{\varepsilon}$.

Conversely, given (4) holds, for every $y \in T(B_{<1}(0))$ we want to find some $\delta > 0$ s.t. $B_{<\delta}(y) \subset T(B_{<1}(0))$. Let $u_y \in B_{<1}(0)$ s.t. $Tu_y = y$. Any $w \in B_{<\delta}(y)$ can be written as $y + a$ with $a \in B_{<\delta}(0)$ and there exists some $u_a \in B_{<1}(0)$ s.t. $\|u_a\|_X \leqslant C\|a\|_Y < C\delta$. Now we have

$$T(u_y + u_a) = T(u_y) + T(u_a) = y + a = w$$

and $\|u_y + u_a\|_X < \|u_y\|_X + C\delta$, hence choose $\delta > 0$ s.t. $\|u_y\|_X + C\delta < 1$ (note that $\|u_y\|_X < 1$).

It remains to show the main direction, namely (3) $\implies$ (4). The proof is taken from [**Tao10**] (page 99-100). For every $n \geqslant 1$, let $E_n \subset Y$ be the set of all $y \in Y$ for which there exists a solution to $Tu = y$ with $\|u\|_X \leqslant n\|y\|_Y$. From the hypothesis of (3), we see that $\bigcup_{n\geqslant 1} E_n = Y$ so by (4) of Theorem 4.5 there exists some $n_0$ s.t. $E_{n_0}$ is dense in some ball $B_{<r}(y_0)$. In other words, the problem $Tu = y$ is approximately quantitatively solvable in the ball $B_{<r}(y_0)$ in the sense that for every $\varepsilon > 0$ and every $y \in B_{<r}(y_0)$ there exists an approximate solution $u \in E_{n_0}$ with $\|Tu - y\|_Y \leqslant \varepsilon$ and $\|u\|_X \leqslant n_0\|Tu\|_Y$, and thus $\|u\|_X \leqslant n_0 r + n_0\varepsilon$.

By subtracting two such approximate solutions, we conclude that for any $y \in B_{<2r}(0)$ and $\varepsilon > 0$, there exists $u \in X$ with $\|Tu - y\|_Y \leqslant 2\varepsilon$ and $\|u\|_X \leqslant 2n_0 r + 2n_0\varepsilon$.

Since $T$ is homogeneous, we can rescale and conclude that for any $y \in Y$ and any $\varepsilon > 0$ there exists $u \in X$ with $\|Tu - y\|_Y \leqslant 2\varepsilon$ and $\|u\|_X \leqslant 2n_0\|y\|_Y + 2n_0\varepsilon$.

In particular, setting $\varepsilon = \frac{1}{4}\|y\|_Y$ (treating t he case $y = 0$ separately), we conclude that for any $y \in Y$ we may write $y = Tu_1 + y_1$, where $\|y_1\|_Y \leqslant \frac{1}{2}\|y\|_Y$ and $\|u_1\|_X \leqslant \frac{5}{2}n_0\|y\|_Y$.

Iterating this procedure (in the second step with $y_1$ in the role of $y$), we find that in the $n$th step there exist $u_1, \dots, u_n$ with $\|u_k\|_X \leqslant \frac{5}{2^k}n_0\|y\|_Y$ s.t.

$$y = \sum_{k=1}^{n} Tu_k + y_n = T\Big(\sum_{k=1}^{n} u_k\Big) + y_n$$

with $\|y_n\|_Y \leqslant \frac{1}{2^n}\|y\|_Y$.

Taking limits we see that $\sum_{n=1}^{\infty} u_n = u$ for some $u \in X$ since $\sum_{n=1}^{\infty} \|u_n\|_X$ converges and $X$ is complete; $u$ is a solution to $Tu = y$ with $\|u\|_X \leqslant 5n_0\|y\|_Y$, so the claim follows. ■

**Corollary 4.20.** Let $T\colon V \to W$ be a bounded linear operator between Banach spaces that is bijective. Then $T^{-1}\colon W \to V$ is bounded.

**Proof.** $T^{-1}\colon W \to V$ is well defined and by Theorem 4.19 $T$ is open, hence $T^{-1}$ is continuous. ■

**Corollary 4.21.** Assume $V$ is a vector space endowed with two norms $\|\cdot\|_1$, $\|\cdot\|_2$ s.t. $(V, \|\cdot\|_1)$ and $(V, \|\cdot\|_2)$ are Banach. Assume there exists $c > 0$ s.t. $\|v\|_2 \leqslant c\|v\|_1$ for all $v \in V$. Then there is $C > 0$ s.t.

$$\|v\|_1 \leqslant C\|v\|_2 \qquad \forall v \in V.$$

**Proof.** The identity map $(V, \|\cdot\|_1) \to (V, \|\cdot\|_2)$, $v \mapsto v$ is a bijective bounded operator, hence the claim follows from (4) of Theorem 4.19. ■

Next we turn to a rather astonishing consequence of consequence of Corollary 4.20.

**Theorem 4.22** (closed graph theorem)**.** Let $T\colon V \to W$ be a linear map between Banach spaces $V, W$. Assume that

$$\Gamma := \{(v, Tv)\colon v \in V\}$$

is closed in $V \times W$. Then $T$ is bounded.

**Remark 4.23.** (1) The converse holds since $W$ is Hausdorff.

(2) $f\colon \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \begin{cases} \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

has a closed graph, but is not continuous.

**Proof.** $V \times W$ with the norm $\|(v,w)\| = \|v\|_V + \|w\|_W$ is a Banach space and so is $\Gamma$, being the closed subspace of a Banach space. Observe that the projections $\pi_V\colon V\times W \to V$ and $\pi_W\colon V\times W\to W$ are continuous and linear. Now

$$\pi_V\big|_\Gamma : \Gamma \to V$$

is a continuous linear bijection, hence by Corollary 4.20 its inverse

$$(\pi_V\big|_\Gamma)^{-1}\colon V \to \Gamma$$

is continuous as well. Since $T = \pi_W \circ (\pi_V\big|_\Gamma)^{-1}$, it is continuous. ■

**Remark 4.24.** Let $C([0,1])$ and $C^1([0,1])$ both be endowed with the sup norm $\|\cdot\|_b$. The derivative

$$C^1([0,1]) \to C([0,1]), \quad f \mapsto f'$$

is a linear map and its graph is closed in $C^1([0,1]) \times C([0,1])$ (this is a formulation of the fact that if a sequence of functions $(f_n)_{n\geqslant 1}$ converges uniformly to $f$ and also $(f_n')_{n\geqslant 1}$ converges uniformly to some $g$, then $f' \equiv g$). However, the derivative operator is not bounded. The closed graph theorem was not applicable because $C^1([0,1])$ is not complete w.r.t. $\|\cdot\|_b$.

## 4.5. Grothendieck's theorem on closed subspaces of $L^p$

Here we present a quite non-trivial application of the closed graph theorem, namely

**Theorem 4.25.** Let $(X,\mathcal{A},\mu)$ be a finite measure space, that is $\mu(X) < +\infty$. Suppose that

(1) $E$ is a closed subspace of $L^p(X)$ for some $1 \leqslant p < +\infty$
(2) $E \subset L^\infty(X)$.

Then $E$ is finite dimensional.

**Proof.** Equipped with the $L^p$-norm, $E$ is a Banach space. Let

$$I\colon E \to L^\infty(X)$$

be the identity map, $I(f) = f$ for all $f \in E$. We claim that the graph of $I$ is closed: indeed assume $f_n \to f$ in $L^p$ and $f_n \to g$ in $L^\infty$. There exists a subsequence $(f_{n_k})_{k\geqslant 1}$ that converges a.e. to $f$ and together with the convergence in $L^\infty$ we can conclude that $f = g$ a.e. By the closed graph theorem there is $M > 0$ s.t. $\|f\|_{L^\infty(X)} \leqslant M\|f\|_{L^p(X)}$ for all $f \in E$.

**Claim.** There exists $A > 0$ s.t. $\|f\|_{L^\infty(X)} \leqslant A\|f\|_{L^2(X)}$ for all $f \in E$.

If $1 \leqslant p \leqslant 2$ this follows from Hölder's inequality,

$$\int_X |f|^p\,d\mu \leqslant \left(\int_X |f|^2\,d\mu\right)^{\frac{p}{2}} \left(\int_X 1\,d\mu\right)^{\frac{2-p}{2}} = \|f\|^p_{L^2(X)}\mu(X)^{\frac{2-p}{2}}$$

so in particular $\|f\|_{L^p(X)} \leqslant \|f\|_{L^2(X)}\mu(X)^{\frac{2-p}{2p}}$.

Assume $2 < p < +\infty$ and notice that

$$|f(x)|^p \leqslant \|f\|_{L^\infty(X)}^{p-2} |f(x)|^2$$

and integrating this inequality gives

$$\|f\|_{L^p(X)}^p \leqslant \|f\|_{L^\infty(X)}^{p-2} \|f\|_{L^2(X)}^2.$$

We now use $\|f\|_{L^\infty(X)} \leqslant M \|f\|_{L^p(X)}$ for all $f \in E$ and deduce

$$\|f\|_{L^p(X)}^p \leqslant M^{p-2} \|f\|_{L^p(X)}^{p-2} \|f\|_{L^2(X)}^2$$

from which $\|f\|_{L^p(X)} \leqslant M^{\frac{p-2}{2}} \|f\|_{L^2(X)}$ for all $f \in E$ follows.

Now we return to the proof of Theorem 5.25. Let $f_1, \dots, f_n$ be an orthonormal set in $E$. If $\dim(E) \geqslant n$ such a set can be obtained by the Gram-Schmidt orthogonalisation procedure. Let

$$B = \left\{ s = (s_1, \dots, s_n) \in \mathbf{C}^n \colon \sum_{j=1}^n |s_j|^2 \leqslant 1 \right\}.$$

be the unit ball in $\mathbf{C}^n$ and for every $s \in B$, let $f_s(x) = \sum_{j=1}^n s_j f_j(x)$. Then $\|f_s\|_{L^2(X)} \leqslant 1$ and by the claim we deduce $\|f_s\|_{L^\infty(X)} \leqslant A$ for all $s \in B$. So for every $s \in B$ there exists a measurable subset $X_s \subset X$ with $\mu(X_s) = \mu(X)$ s.t. $|f_s(x)| \leqslant A$ for all $x \in X_s$. Let now $\{s_j \colon j \geqslant 1\} \subset B$ be a countable dense subset of $B$ and $S := \bigcap_{j \geqslant 1} X_{s_j}$. Then $|f_{s_j}(x)| \leqslant A$ for all $x \in S$ and $j \in \mathbf{N}$, and $\mu(S) = \mu(X)$. But observe that for all $x \in S$, $s \mapsto f_s(x)$ is continuous, and hence $|f_s(x)| \leqslant A$ for all $x \in S$ and $s \in B$. From this, we claim that

$$(*) \qquad \sum_{j=1}^n |f_j(x)|^2 \leqslant A^2$$

for $x \in S$. Indeed, we may assume that the left hand side is non-zero; then if we let $\sigma := (\sum_{j=1}^n |f_j(x)|^2)^{\frac{1}{2}}$ and set $s_j := \overline{f_j(x)}/\sigma$, $|f_s(x)| \leqslant A$ implies

$$\frac{1}{\sigma} \sum_{j=1}^n |f_j(x)|^2 \leqslant A$$

as we claimed. Finally integrating $(*)$ over $X$ we find $n \leqslant A^2 \mu(X)$. ■

## 4.6. Complementary subspaces and a counterexample

First we show some geometric properties of closed subspaces in a Banach space that follow from the open mapping theorem and then present and elementary proof of the fact that $c_0(\mathbf{N})$ does not admit a closed complement in $\ell^\infty(\mathbf{N})$.

**Proposition 4.26.** Let $V$ be a Banach space and $E, F$ two closed subspaces of $V$ s.t. $E + F$ is closed. Then there exists $C > 0$ such that every $z \in E + F$ admits a decomposition $z = e + f$ for $e \in E$, $f \in F$ with $\|e\|_V \leqslant C \|z\|_V$ and $\|f\|_V \leqslant C \|z\|_V$.

**Proof.** $E \times F$ with the norm $\|(e,f)\|_{E\times V} = \|e\|_V + \|f\|_V$ is a Banach space, as well is $E + F$ (with $\|\cdot\|_V$), being a closed subspace of a Banach space. The map

$$E \times F \to E + F, \quad (e,f) \mapsto e + v$$

is a surjective bounded linear operator. Hence, by (4) of Theorem 4.19, for every $z \in E + F$ there exists some $(e,f) \in E \times F$ s.t.

$$\|e\|_V + \|f\|_V = \|(e,f)\|_{E\times F} \leqslant C\|e + f\|_V = C\|z\|_V.$$

■

Let $V$ be a Banach space and $E \subset V$ a closed subspace. We say that $E$ admits a closed complement if there is a closed subspace $F \subset V$ s.t.

(1) $E + F = V$
(2) $E \cap F = \{0\}$.

We know that if $V$ is a Hilbert space then $E^\perp$ is a closed complement of $E$. A remarkable result of Lindenstrauss and Tzafriri says that if a Banach space $V$ has the property that every closed subspace admits a closed complement then there is an equivalent norm on $V$ coming from a scalar product.

Here We are going to limit ourselves to giving a concrete example. As usual, let

$$\ell^\infty(\mathbf{N}) = \left\{ f \colon \mathbf{N} \to \mathbf{C} \colon \|f\|_{\ell^\infty(\mathbf{N})} := \sup_{n\in\mathbf{N}} |f(n)| < +\infty \right\}$$

be the Banach space of bounded sequences and

$$c_0(\mathbf{N}) = \left\{ f \colon \mathbf{N} \to \mathbf{C} \colon \lim_{n\to\infty} f(n) = 0 \right\}$$

the closed subspace of those converging to 0. Our objective is to show:

**Theorem 4.27.** $c_0(\mathbf{N})$ does not admit a closed complement in $\ell^\infty(\mathbf{N})$.

This elementary (but tricky) proof is due to R. Whitley [**Whi66**].

The strategy of the proof is the following: Assume $R \subset \ell^\infty(\mathbf{N})$ is a closed complement of $c_0(\mathbf{N})$ in $\ell^\infty(\mathbf{N})$. Then the canonical projection $\pi \colon \ell^\infty(\mathbf{N}) \to \ell^\infty(\mathbf{N})/c_0(\mathbf{N})$ restricted to $R$, i.e. $\pi|_R$, is a bijective bounded operator of Banach spaces, hence $(\pi|_R)^{-1}$ is bounded bijective. We are going to establish two properties which will lead to a contradcition:

(1) There is $\mathcal{D} \subset R^*$ countable s.t. $\bigcap_{f\in\mathcal{D}} \ker f = \{0\}$.
(2) If $\mathcal{D} \subset (\ell^\infty(\mathbf{N})/c_0(\mathbf{N}))^*$ is any countable subset, $\bigcap_{f\in\mathcal{D}} \ker f \neq \{0\}$.

If now $T \colon \ell^\infty(\mathbf{N})/c_0(\mathbf{N}) \to R$ were any bounded bijective operator, and $\mathcal{D} \subset R^*$ countable with $\bigcap_{f\in\mathcal{D}} \ker f = \{0\}$, then

$$\{f \circ T \colon f \in \mathcal{D}\} \subset (\ell^\infty(\mathbf{N})/c_0(\mathbf{N}))^*$$

would be a countable subset with $\bigcap_{f\in\mathcal{D}} \ker(f \circ T) = \{0\}$, contradicting (2).

The proof of (2) is based on the following counter-intuitive set theoretic fact:

**Lemma 4.28.** Let $S$ be an infinite countable set. Then there is a family $\{U_a \colon a \in A\}$ of subsets of $S$ s.t.

(1) $U_a$ is infinite for all $a \in A$.
(2) $U_a \cap U_b$ is finite for all $a \neq b$.
(3) $A$ is uncountable.

**Proof.** We identify $S$ with $\mathbf{Q} \cap (0,1)$ and $A := (0,1) \setminus \mathbf{Q}$. For every $a \in A$ choose a sequence $(x_n)_{n\geqslant 1}$ in $S$ with $\lim_{n\to\infty} x_n = a$ and set $U_a = \{x_n \colon n \geqslant 1\} \subset S$. Then $U_a$ is infinite since $a \notin \mathbf{Q}$; if $a \neq b \in S$ and $x_n \to a$, $y_n \to b$ we have for $N$ large enough that $x_n \neq y_m$ for all $n, m \geqslant N$ so that $U_a \cap U_b$ is finite. Since $A$ is uncountable this concludes the proof. ■

We can now prove Theorem 4.27.

**Proof.** Let us show properties (1) and (2).

(1) For all $n \in \mathbf{N}$, define $f_n \colon \ell^\infty(\mathbf{N}) \to \mathbf{C}$, $g \mapsto g(n)$. Then

$$|f_n(g)| = |g(n)| \leqslant \|g\|_\infty.$$

Futhermore, if $g \in \bigcap_{n\geqslant 1} \ker f_n$ then $g(n) = 0$ for all $n \in \mathbf{N}$. Hence

$$\mathcal{D} := \{\, f_n|_R : n \in \mathbf{N}\} \subset R^*$$

is finite and $\bigcap_{n\geqslant 1} \ker f_n|_R = \{0\}$.

(2) Apply Lemma 4.28 to $S = \mathbf{N}$ and let $\{U_a \colon a \in A\}$ be a family of subsets of $\mathbf{N}$ as in the lemma. For all $a \in A$ define

$$f_a := \mathbf{1}_{U_a} + c_0(\mathbf{N}) \in \ell^\infty(\mathbf{N})/c_0(\mathbf{N}).$$

Now, for $\lambda \in (\ell^\infty(\mathbf{N})/c_0(\mathbf{N}))^*$ we claim that the set $\{a \in A \colon \lambda(f_a) \neq 0\}$ is countable; it suffices to show that the set

$$C(n) := \left\{ a \in A \colon |\lambda(f_a)| \geqslant \frac{1}{n} \right\}$$

is countable for every $n \in \mathbf{N}$. Choose $f_1, \dots, f_m \in C(n)$ and let

$$b_k = \operatorname{sgn}(\lambda(f_k)) = \frac{\overline{\lambda(f_k)}}{|\lambda(f_k)|}.$$

We will now proceed to show that $f = \sum_{k=1}^m b_k f_k$ has unit norm. Of course,

$$\widetilde{f} := \sum_{k=1}^m b_k \mathbf{1}_{U_{a_k}} \in \ell^\infty(\mathbf{N})$$

is a representative of the coset $f$. Letting $F := \bigcup_{k\neq j} (U_{a_k} \cap U_{a_j})$ we know that $F$ is finite so $\widetilde{f}\mathbf{1}_F \in c_0(\mathbf{N})$ and $\widetilde{f} - \widetilde{f}\mathbf{1}_F$ represents $f$. Thus $\|f\| \leqslant \big\|\widetilde{f} - \widetilde{f}\mathbf{1}_F\big\|_\infty$ and since

$$\big(\widetilde{f} - \widetilde{f}\mathbf{1}_F\big)(x) = \begin{cases} b_k, & x \in U_{a_k} \setminus F \\ 0, & x \notin \bigcup_{k=1}^m (U_{a_k} \setminus F) \end{cases}$$

it follows that $\left\|\widetilde{f} - \widetilde{f}\mathbf{1}_F\right\|_\infty = 1$ so $\|f\| \leqslant 1$. From this we deduce

$$\|\lambda\| \geqslant |\lambda(f)| = \sum_{k=1}^{m} |\lambda(f_k)| \geqslant \frac{m}{n}$$

which shows that $C(n)$ is finite. We have thus shown that $\{a \in A\colon \lambda(f_a) \neq 0\}$ is countable. If now $\mathcal{D} = \{\lambda_n\colon n \in \mathbf{N}\} \subset (\ell^\infty(\mathbf{N})/c_0(\mathbf{N}))^*$ is a countable family then

$$\bigcup_{n \geqslant 1} \{a \in A\colon \lambda_n(f_a) \neq 0\} \subset A$$

is countable as well and hence $\bigcap_{n\geqslant 1} \ker \lambda_n \neq \{0\}$. ■

# Chapter 5. Topological vector spaces, weak topologies, and the Banach-Alaoglu theorem

In Analysis one encounters function spaces with a natural topology that however cannot be described by a single norm: for example the space of continuous functions on $\mathbf{R}$ with the topology of uniform convergence on compact sets. Another problem one encounters is the fact that the unit ball in an infinite dimensional Banach space is never compact. To remediate these problems we are going to study topological vector spaces whose topology is given by a family of seminorms. On one hand this allows us to study natural function spaces with tools of functional analysis; on the other hand this will lead to weaker topologies on Banach spaces, thereby restoring compactness in certain situations.

## 5.1. Basic Definitions and Examples

We begin by recalling Definition 1.4:

**Definition 5.1.** A topological vector space is a $\mathbf{K}$-vector space $V$ endowed with a topology such that the maps

(1) $\mathbf{K} \times V \to V$, $(\lambda, v) \mapsto \lambda v$
(2) $V \times V \to V$, $(v, w) \mapsto v + w$

are continuous.

We draw the following useful conclusion:

**Lemma 5.2.** The two maps

$$M_\lambda \colon V \to V,\ v \mapsto \lambda v$$
$$L_v \colon V \to V,\ w \mapsto v + w$$

are homeomorphisms.

**Proof.** $M_\lambda$ is continuous with continuous left and right inverse $M_{\lambda^{-1}}$; $L_v$ is continuous with continuous left and right inverse $L_{-v}$. ■

We now turn to describe the topology on a $\mathbf{K}$-vector space $V$ generated by a family of seminorms (see Definition 2.6).

Let $V$ be a $\mathbf{K}$-vector space and $\{\|\cdot\|_\alpha \colon \alpha \in A\}$ a family of seminorms

$$\|\cdot\|_\alpha \colon V \to [0, +\infty)$$

on $V$. There is a priori no restriction on the cardinality of $A$. For all $v \in V$, $F \subset A$ finite and $r > 0$ we let

$$N(v; F; r) := \{w \in V \colon \|w - v\|_\alpha < r \text{ for all } \alpha \in F\}.$$

**Definition 5.3.** Define $U \subset V$ to be *open* if for all $u \in U$ there exists a finite $F \subset A$ and $r > 0$ s.t. $N(v; F; r) \subset U$.

Clearly, $\varnothing$ and $V$ are open w.r.t. to this definition, as well is the arbitrary union of open sets. For finite intersections, note that

$$N(v; F_1; r_1) \cap N(v; F_2; r_2) = N(v; F_1 \cup F_2; \min\{r_1, r_2\})$$

is open. From this we conclude that finite intersections of open sets are open.

**Definition 5.4.** The topology on $V$ generated by the family of seminorms $\{\|\cdot\|_\alpha \colon \alpha \in A\}$ is the topology whose open subsets are given by Definition 5.3.

**Lemma 5.5.** The topology on $V$ generated by the family of seminorms $\{\|\cdot\|_\alpha \colon \alpha \in A\}$ endows $V$ with the structure of a topological vector space.

**Proof.** This follows from

$$\lambda N(v; F; r) = N(\lambda v; F; |\lambda| r)$$

and

$$N(v_1; F_1; r_1) + N(v_2; F_2; r_2) \subset N(v_1 + v_2; F_1 \cap F_2; r_1 + r_2)$$

with a similar argument as given in Lemma 1.3. ■

Of course, if $A = \varnothing$ then the topology on $V$ has exactly two open sets, namely $\varnothing$ and $V$. The following property keeps degenerate cases away:

**Definition 5.6.** A family $A$ of seminorms is *sufficient* if for all $v \in V \setminus \{0\}$ there exists some $\alpha \in A$ s.t. $\|v\|_\alpha \neq 0$.

**Lemma 5.7.** If $A$ is sufficient then the topology generated by $A$ is Hausdorff.

**Proof.** Indeed, if $v_1 \neq v_2 \in V$ then by sufficiency $d := \|v_1 - v_2\|_\alpha \neq 0$ for some $\alpha \in A$. Then $N(v_1; \alpha; \frac{d}{2})$ and $N(v_2; \alpha; \frac{d}{2})$ are two disjoint neighbourhoods of $v_1$ and $v_2$ resp. ■

A particularly important case is when we have a countable sufficient family of seminorms.

**Proposition 5.8.** If $\{\|\cdot\|_n \colon n \in \mathbf{N}\}$ is a sufficient countable family of seminorms on $V$, the generated topology is metrisable.

**Proof.** Since the family is sufficient, it is straightforward to verify that

$$d(v, w) := \sum_{n=1}^{\infty} \frac{1}{2^n} \Big( \frac{\|v - w\|_n}{1 + \|v - w\|_n} \Big)$$

define a distance on $V$. Then observe that for all $n, \ell \geqslant 1$:

$$\frac{1}{2^n} \Big( \frac{\|v - w\|_n}{1 + \|v - w\|_n} \Big) \leqslant d(v, w) \leqslant \sum_{k=1}^{\ell} \frac{1}{2^k} \Big( \frac{\|v - w\|_k}{1 + \|v - w\|_k} \Big) + \frac{1}{2^\ell}.$$

From the second inequality we deduce that if $2^{-\ell} < \varepsilon$, then

$$N(v; \{\|\cdot\|_1, \dots \|\cdot\|_\ell; \varepsilon\}) \subset B_{<2\varepsilon}(v)$$

(here $B_{<r}(v)$ refers to an open ball w.r.t. the metric $d$). Moreover, for $\varepsilon \geqslant 1$ and $d(v,w) < \frac{\varepsilon}{2^{n+1}}$ the first inequality implies

$$\frac{1}{2^n}\Big(\frac{\|v-w\|_n}{1+\|v-w\|_n}\Big) < \frac{\varepsilon}{2^{n+1}}$$

so $(2-\varepsilon)\|v-w\|_n < \varepsilon$ and hence $\|v-w\|_n < \varepsilon$. Thus, for all $\varepsilon \geqslant 1$ we have

$$B_{<\varepsilon 2^{-(n+1)}}(v) \subset N(v; \|\cdot\|_n; \varepsilon)$$

■

**Example 5.9.** Let $X$ be a locally compact Hausdorff space. For every compact subset $K \subset X$, define for all $f \in C(X)$

$$\|f\|_K := \sup_{x\in K} |f(x)|.$$

Then $\{\|\cdot\|_K : K \subset X, \text{ compact}\}$ is a sufficient family of seminorms. We have that $f_n \to f$ in this topology iff $f_n$ converges to $f$ uniformly on every compact subset. Observe that if $X$ is a metric space s.t. $B_{\leqslant n}(x_0)$ is compact for all $n \geqslant 0$, then setting $K_n := B_{\leqslant n}(x_0)$, the countable family $\{\|\cdot\|_{K_n}\}_{n\geqslant 0}$ of seminorms induces the same topology on $C(X)$, which is therefore metrisable.

**Example 5.10.** Let $(X, \mathcal{F}, \mu)$ be a triple consisting of a locally compact Hausdorff space $X$, $\mu$ a positive Borel regular measure on $X$ and $\mathcal{F}$ the $\sigma$-algebra of $\mu$-measurable sets. Define

$$L^p_{\text{loc}}(X) := \{f : X \to \mathbf{C} \text{ measurable} : \forall K \subset X \text{ compact}, f\mathbf{1}_K \in L^p(X)\}.$$

Then $\|f\|_{L^p(X),K} := \|f\mathbf{1}_K\|_{L^p(X)}$, for $K \subset X$ compact, defines a sufficient family of seminorms. Observe that if $X$ is a countable union of compact sets, the topology on $L^p_{\text{loc}}(X)$ is metrisable.

There are many more examples of space of functions on $\mathbf{R}^n$ where the seminorms take into account some local boundedness or local integrability on derivatives.

Let $V, W$ be topological vector spaces defined by families of seminorms

$$\{\|\cdot\|^V_\alpha : \alpha \in A\}, \quad \{\|\cdot\|^W_\beta : \beta \in B\}$$

respectively. Analogous to Theorem 1.18 we have a characterisation of continuous linear maps $T : V \to W$.

**Proposition 5.11.** For a linear maps $T : V \to W$ the following are equivalent:

(1) $T$ is continuous
(2) $T$ is continuous at 0
(3) For all finite $F \subset B$ there exists some finite $G \subset A$ s.t. for all $\beta \in F$:

$$\sup\Big\{\|T(x)\|^W_\beta : \max_{\alpha\in G} \|x\|^V_\alpha \leqslant 1\Big\} < +\infty$$

**Proof.** To see why (2) implies (1) we can apply the same argument as in the proof of Theorem 1.18 which is valid for all topological vector spaces.

(2) $\implies$ (3) Let $F \subset B$ be finite. Since $N(0;F;1)$ is an open neighbourhood of $T(0)$ we can find some finite set $G \subset A$ and $\varepsilon > 0$ s.t. $T(N(0;G;\varepsilon)) \subset N(0;F;1)$ so that

$$\begin{aligned} T(N(0;G;1)) &= T\Big(\frac{1}{\varepsilon}N(0;G;\varepsilon)\Big) \\ &= \frac{1}{\varepsilon}T(N(0;G;\varepsilon)) \\ &\subset \frac{1}{\varepsilon}N(0;F;1) = N(0;F;1/\varepsilon). \end{aligned}$$

Taking closures the statement follows.

(3) $\implies$ (2) Let $N(0;F;r)$ with $F \subset B$ finite and $r > 0$ be arbitrary. Moreover, let $G \subset A$ be s.t.

$$C := \sup\Big\{\|T(x)\|_\beta^W \colon \max_{\alpha\in G} \|x\|_\alpha^V \leqslant 1\Big\} < +\infty$$

i.e. $T(N(0;G;1)) \subset N(0;F;C)$. Then $T(N(0;G;r/C)) \subset N(0;F;r)$ as desired. ■

**Corollary 5.12.** A linear form $f\colon V \to \mathbf{K}$ is continuous iff there exists $G \subset A$ finite such that:

$$\sup\Big\{|f(x)|\colon \max_{\alpha\in G} \|x\|_\alpha^V \leqslant 1\Big\} < +\infty$$

**Proof.** This follows immediately from the previous proposition applied to $W = \mathbf{K}$ and the topology on $\mathbf{K}$ generated by the seminorm (actually a norm) $|\cdot|$ which coincides with the topology induced by the norm $|\cdot|$. ■

Given a topological vector space $V$ (abbreviated TVS from now on), we denote by $V^*$ the vector space of all continuous linear forms $V \to \mathbf{K}$. We then have:

**Theorem 5.13** (Hahn-Banach)**.** Let $V$ be a TVS given by a sufficient family $\{\|\cdot\|_\alpha \colon \alpha \in A\}$ of seminorms. Then for all $v \in V \setminus \{0\}$ there exists some $F \in V^*$ with $F(v) \neq 0$.

**Proof.** Let $v \in V \setminus \{0\}$ and $\alpha \in A$ with $\|v\|_\alpha \neq 0$. We apply Theorem 2.7 to the seminorm $\|\cdot\|_\alpha$, the subspace $M = \mathbf{K}v$ and the linear form

$$f\colon \mathbf{K}v \to \mathbf{K}, \quad f(\lambda v) = \lambda\|v\|_\alpha$$

to obtain a linear extension $F\colon V \to \mathbf{K}$ satisfying $|F(w)| \leqslant \|w\|_\alpha$ for all $w \in V$. By Corollary 5.12, $F \in V^*$ and by construction $F(v) = f(v) = \|v\|_\alpha$. ■

## 5.2. Weak Topologies

When $(V, \|\cdot\|_V)$ is a normed space and $(V^*, \|\cdot\|_{V^*})$ its dual we will use the tools from Section 5.1 and use suitable families of seminorms to define new TVS structures on $V$ and $V^*$ that have ”fewer” open sets than the corresponding norm topologies. In fact, we will apply this construction to any TVS $V$ and its dual $V^*$, and characterise the resulting topologies as initial topologies.

**Definition 5.14.** Let $V$ be a TVS.

(1) The $\sigma(V, V^*)$-topology on $V$ is the topology defined by the family of seminorms $\{\|\cdot\|_\lambda \colon \lambda \in V^*\}$ where $\|v\|_\lambda := |\lambda(v)|$ for all $v \in V$. It is often referred to as the *weak topology* on $V$.
(2) The $\sigma(V^*, V)$-topology on $V^*$ is the topology defined by the family of seminorms $\{\|\cdot\|_v \colon v \in V\}$ where $\|f\|_v := |f(v)|$ for all $v \in V^*$. It is often referred to as the *weak*$^*$*topology* on $V^*$.

**Lemma 5.15.** (1) The family of seminorms defining the $\sigma(V^*, V)$ topology on $V^*$ is sufficient.
(2) If $V$ is a TVS defined by a sufficient family of seminorms then the family of seminorms defining the $\sigma(V, V^*)$ topology on $V$ is sufficient.

Thus, the weak* topology on $V^*$ is always Hausdorff, and if $V$ has a sufficient family of seminorms. the weak topology on $V$ is Hausdorff.

**Proof.** (1) If $f \in V^* \setminus \{0\}$ then there is some $v \in V$ with $f(v) \neq 0$, hence $\|f\|_v \neq 0$.

(2) This follows from Theorem 5.13 which guarantees the existence of some $f \in V^*$ with $f(v) \neq 0$.

Lastly, according to Lemma 5.7, topologies induced by a sufficient family of seminorms are Hausdorff. ■

We now turn to a very useful way of characterising weak and weak* toplogies by putting them into the larger framework of *initial topology* , a concept of wide ranging applications.

Let $X$ be a set and $\mathcal{F} = \{(\varphi_j, Y_j) \colon j \in J\}$ a set of pairs $(\varphi_j, Y_j)$ where $Y_j$ is a topological space and $\varphi_j \colon X \to Y_j$ a map. The task is now to find the most "economical" topology on $X$ making all these maps continuous. Of course the discrete topology on $X$ would do the job, but we want the one with the "least number" of open subsets. Let then $\mathcal{T} \subset 2^X$ be a topology for which the above maps are continuous. Then for all $j \in J$, if $U_j \subset Y_j$ is open, we must have $\varphi_j^{-1}(U_j) \in \mathcal{T}$. Let then

$$S_1 := \{\varphi_j^{-1}(U_j) \colon U_j \subset Y_j \text{ open}, j \in J\}.$$

This is not necessarily a topology as it does not necessarily contain all the finite intersections or arbitrary unions of members of $S_1$. Let $S_2$ be the set containing all finite intersections of elements in $\mathcal{T}_1$ and let $\mathcal{T}_\mathcal{F}$ be the set of arbitrary unions of elements in $U_2$. Then $\mathcal{T}_\mathcal{F} \subset \mathcal{T}$ and

**Lemma 5.16.** $\mathcal{T}_\mathcal{F}$ is a topology.

**Proof.** It is clear that $\varnothing$ and $X$ are in $\mathcal{T}_{\mathcal{F}}$ and that it is stable under arbitrary unions. For finite intersections, let $U_1, U_2 \in \mathcal{T}_{\mathcal{F}}$ and write

$$U_1 = \bigcup_{\alpha \in A} V_\alpha, \quad U_2 = \bigcup_{\beta \in B} W_\beta$$

where $V_\alpha$ and $W_\beta$ are in $S_2$. Then, by definition of $S_2$, we can write

$$V_\alpha = \bigcap_{j=1}^{n_\alpha} \varphi_j^{-1}(V_{j,\alpha}), \quad W_\beta = \bigcap_{k=1}^{n_\beta} \varphi_k^{-1}(W_{k,\beta})$$

for subsets $V_{j,\alpha}$ and $W_{k,\beta}$ open in $Y_j$ respectively $Y_k$. Now

$$U_1 \cap U_2 = \Big( \bigcup_{\alpha \in A} V_\alpha \Big) \cap \Big( \bigcup_{\beta \in B} W_\beta \Big) = \bigcup_{\substack{\alpha \in A \\ \beta \in B}} (V_\alpha \cap W_\beta)$$

and since $V_\alpha$ and $W_\beta$ are in $S_2$, so is their intersection. By definition of $\mathcal{T}_{\mathcal{F}}$, the arbitrary union of sets in $S_2$ is in $\mathcal{T}_{\mathcal{F}}$, concluding the proof. ■

**Definition 5.17.** $\mathcal{T}_{\mathcal{F}}$ is called the *initial topology* defined by the family $\mathcal{F} = \{(\varphi_j, Y_j) \colon j \in J\}$.

**Example 5.18.** Let $\{Y_j\}_{j \in J}$ be a family of topological spaces, $X = \prod_{j \in J} Y_j$ the (set theoretic) cartesian product and $\pi_j \colon X \to Y_j$ the projection onto the $j$th coordinate. Then the product topology on $X$ is the initial topology w.r.t. the family $\{(\pi_j, Y_j)\}_{j \in J}$.

Now, let $\mathcal{T}_{\mathcal{F}}$ be the initial topology on $X$ given by a family $\mathcal{F} = \{(\varphi_j, Y_j)\}_{j \in J}$. The following two lemmas are quite useful:

**Lemma 5.19.** Let $Z$ be a topological space. A map $\psi \colon Z \to X$ is continuous iff for all $j \in J$ the map $\varphi_j \circ \psi \colon Z \to Y_j$ is continuous.

**Proof.** If $\psi$ is continuous then so is $\varphi_j \circ \psi$, being the composition of continuous functions. Conversely, if $\varphi_j \circ \psi$ is continuous for all $j \in J$ we know that

$$(\varphi_j \circ \psi)^{-1}(U_j) = \psi^{-1}(\varphi_j^{-1}(U_j))$$

is open for all open sets $U_j \subset Y_j$. Since the $\varphi_j^{-1}(U_j)$ are somewhat the atoms of $\mathcal{T}_{\mathcal{F}}$, we can use the definition of $\mathcal{T}_{\mathcal{F}}$ together with the fact that $\varphi^{-1} \colon 2^X \to 2^Z$ commutes with unions and intersections to conclude that $\psi^{-1}(U)$ is open for arbitrary open sets $U \in \mathcal{T}_{\mathcal{F}}$. ■

**Lemma 5.20.** A net $(x_\alpha)_{\alpha \in A}$ in $X$ converges to $x \in X$ iff $(\varphi_j(x_\alpha))_{\alpha \in A}$ converges to $\varphi_j(x)$ for all $j \in J$.

**Proof.** The necessary direction follows from the fact that $\varphi_j$ is continuous for all $j \in J$. For the converse, suppose that for a given net $(x_\alpha)_{\alpha \in A}$ there exists some $x \in X$ s.t. $\varphi_j(x_\alpha) \to \varphi_j(x)$ for all $j \in J$. Let $U \subset X$ be some arbitrary open neighbourhood of $x$. By definition of $\mathcal{T}_{\mathcal{F}}$ there exist $W_1, \dots, W_m$ with $W_k \subset Y_k$ s.t.

$$x \in \bigcap_{k=1}^{m} \varphi_k^{-1}(W_k) \subset U.$$

For every $1 \leqslant k \leqslant m$ we find a corresponding $\beta_k$ s.t. $\varphi_k(x_\alpha) \in W_k$ for all $\alpha \geqslant \beta_k$. Since $A$ is a directed set, every two elements have an upper bound so inductively every finite subset has an upper bound. Hence we set $\beta$ to be some upper bound of $\{\beta_1, \dots, \beta_k\}$ so that for all $\alpha \geqslant \beta$ we have for all $1 \leqslant k \leqslant m$ that $\varphi_k(x_\alpha) \in W_k$, in particular $x_\alpha \in \varphi_k^{-1}(W_k)$. We conclude that $x_\alpha \in \bigcap_{k=1}^{m} \varphi_k^{-1}(W_k) \subset U$ for $\alpha \geqslant \beta$. ∎

Now let us return to a vector space $V$ over $\mathbf{K}$ and a family of seminorms $\{\|\cdot\|_\alpha\}_{\alpha\in A}$. Let $\mathcal{T}$ be the topology on $V$ generated by this family of seminorms and $\mathcal{T}_{\mathcal{F}}$ the initial topology on $V$ associated to $\mathcal{F} = \{(x \mapsto \|x-v\|_\alpha, \mathbf{K})\colon v \in V,\ \alpha \in A\}$.

**Proposition 5.21.** $\mathcal{T} = \mathcal{T}_{\mathcal{F}}$.

**Proof.** By construction of $\mathcal{T}$, all seminorms $\|\cdot\|_\alpha\colon V \to \mathbf{K}$ are continuous. Moreover, since $(V, \mathcal{T})$ is a TVS we also know that addition of vectors is continuous, so in particular $x \mapsto \|x - v\|_\alpha$ is continuous for all $v \in V$ and all $\alpha \in A$. Hence $\mathcal{T} \supset \mathcal{T}_{\mathcal{F}}$.

For the other inclusion, let $v \in V$, $\alpha \in A$ and $r > 0$. Then

$$N(v;\alpha;r) = \{x \in V\colon \|x - v\|_\alpha < r\} = [x \mapsto \|x - v\|_\alpha]^{-1}((-r, r))$$

so $N(v;\alpha;r)$ is open with respect to $\mathcal{T}_{\mathcal{F}}$, concluding the proof. ∎

**Corollary 5.22.** Let $V$ be a TVS generated by a family of seminorms and $V^*$ its dual.

(1) The $\sigma(V, V^*)$ topology on $V$ is the initial topology for the family $\mathcal{F} = \{(\lambda, \mathbf{K})\}_{\lambda\in V^*}$.

(2) The $\sigma(V^*, V)$ topology is the initial topology for the family $\{(J(v), \mathbf{K})\}_{v\in V}$ where $J(v)\colon V^* \to V$, $\lambda \mapsto \lambda(v)$.

**Proof.** (1) We already know that every $f \in V^*$ is continuous w.r.t. $\sigma(V, V^*)$. For the converse, using the previous proposition we know that $\sigma(V, V^*)$ is equal to the topology generated by

$$\{(x \mapsto \|x - v\|_\lambda, \mathbf{K})\colon v \in V,\ \lambda \in V^*\} = \{(x \mapsto |\lambda(x) - \lambda(v)|, \mathbf{K})\colon v \in V,\ \lambda \in V^*\}.$$

Now for all $v \in V$ and all $\lambda \in V^*$ the function $V \to \mathbf{R}$, $x \mapsto |\lambda(x) - \lambda(v)|$ is continuous w.r.t. the initial topology generated by $\{(\lambda, \mathbf{K})\}_{\lambda\in V^*}$, being the composition of continuous functions. This shows the other inclusion.

(2) Analogous argument. ∎

We deduce from Lemma 5.19 and 5.20 in combination with Corollary 5.22:

**Corollary 5.23.** (1) A map $\psi\colon Z \to V$ is continuous for the $\sigma(V, V^*)$ topology iff $f \circ \psi\colon Z \to \mathbf{K}$ is continuous for all $f \in V^*$.

(2) A map $\psi\colon Z \to V^*$ is continuous for the $\sigma(V^*, V)$ topology iff $z \mapsto \psi(z)(v)$ is continuous for all $v \in V$.
(3) A sequence $(x_n)_{n\geqslant 1}$ converges in $\sigma(V, V^*)$ to $x$ iff $f(x_n) \to f(x)$ for all $f \in V^*$.
(4) A sequence $(f_n)_{n\geqslant 1}$ converges in $\sigma(V^*, V)$ to $f$ iff $f_n(v) \to f(v)$ for all $v \in V$.

## 5.3. Normed spaces and the Banach Alaoglu Theorem

Let us now turn to a normed space $(V, \|\cdot\|_V)$ and recall that its dual $(V^*, \|\cdot\|_{V^*})$ is a Banach space (cf. Proposition 1.22). We will often refer to the norm topologies as *strong* topologies, to the $\sigma(V, V^*)$ topology on $V$ as the weak topology and the $\sigma(V^*, V)$ topology on $V^*$ as the weak$^*$ topology.

(1) and (2) of Corollary 5.23 have the following consequences:

**Proposition 5.24.** Let $T\colon V \to W$ be a bounded linear operator of normed spaces $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ and $T^*\colon W^* \to V^*$ its adjoint.

(1) $T$ is continuous for the weak topologies on $V$ and $W$.
(2) $T^*$ is continuous for the weak$^*$ topologies on $W^*$ and $V^*$.

As a consequence of the closed graph theorem we have the following converse to (1) of Proposition 5.24:

**Proposition 5.24$^*$.** Let $T\colon V \to W$ be a linear map between Banach spaces $V$ and $W$. Assume $T$ is continuous for the weak topologies on $V$ and $W$. Then $T$ is bounded.

**Proof.** Since graph$(T)$ is weakly closed, it is strongly closed, hence the conclusion follows from the closed graph theorem (cf. Theorem 4.22). ■

Now let $(V, \|\cdot\|_V)$ be a normed space and $(V^*, \|\cdot\|_{V^*})$ its dual. Of course every weakly open set is strongly open and the same applies to $V^*$ with its weak$^*$ topology. Now if $F \subset V^*$ is finite and $\varepsilon > 0$ then

$$N(0; F; r) = \{w \in V\colon |f(w)| < c \text{ for all } f \in F\}$$

contains the subspace $\bigcap_{f\in F} \ker(f)$ which is of finite codimension in $V$. Thus, if $V$ is infinite dimensional, the strong open ball $B_{<\varepsilon}(0)$ is **not** weakly open and the same observation applies to $V^*$.

We have, however, the following:

**Proposition 5.25.** (1) $B_{\leqslant r}(0; V)$ is weakly closed.
(2) $B_{\leqslant r}(0; V^*)$ is weakly$^*$ closed.

**Proof.** (1) Recall that for all $v \in V$ Corollary 2.10 tells us

$$\|v\|_V = \sup\{|f(v)|\colon \|f\| \leqslant 1\}$$

and hence $\|v\|_V \leqslant r$ iff $|f(v)| \leqslant r$ for all $f \in B_{\leqslant 1}(0; V)$. Thus,

$$B_{\leqslant r}(0; V) = \bigcap_{f\in B_{\leqslant 1}(0; V^*)} \{v \in V\colon |f(v)| \leqslant r\}$$

and hence is weakly closed.

(2) Analogous argument. ■

**Example 5.26.** Let $\mathcal{H}$ be a separable infinite dimensional Hilbert space and $\{e_n\}_{n\geqslant 1}$ an orthonormal basis. Then $\lim_{n\to\infty} e_n = 0$ is the weak topology.

Indeed, recall that every $f \in \mathcal{H}^*$ is given by $f(v) = \langle v, w\rangle$ for some $w \in \mathcal{H}$ (Riesz Representation Theorem). We apply the convergence criterion of (3) of Proposition 5.23: Let $x = \sum_{n=1}^{\infty} x_n e_n$ with $\sum_{n=1}^{\infty} |x_n|^2 < +\infty$. Then $\langle e_n, x\rangle = x_n$ and hence $\lim_{n\to\infty}\langle e_n, x\rangle = 0$. Thus, we see that while $B_{\leqslant 1}(0;\mathcal{H})$ is weakly closed, the unit sphere

$$\mathbf{S}^1 = \{v \in \mathcal{H} \colon \|v\|_{\mathcal{H}} = 1\}$$

is not.

**Example 5.27** (Compare with 2.16)**.** Let $X$ be a compact Hausdorff space and $C(X,\mathbf{R})$ the Banach space of continuous functions $f\colon X \to \mathbf{R}$ with the norm $\|f\|_b := \sup_{x\in X} |f(x)|$.

Let $M(X,\mathbf{R})$ be the space of signed regular Borel measures on $X$. Then the $\mathbf{R}$-version of the Riesz representation theorem gives a bijection

$$R\colon M(X,\mathbf{R}) \to C(X,\mathbf{R})^*, \quad \mu \to \Phi_\mu$$

where

$$\Phi_\mu(f) = \int_X f\,d\mu$$

and $\|\Phi_\mu\| = |\mu|(X)$; here $|\mu|$ is the total variation measure of $\mu$. Thus $R$ is a bijective isometry between the Banach space $(M(X,\mathbf{R}), \|\cdot\|)$ (whereby $\|\mu\| := |\mu|(X)$) and the Banach space $C(X,K)^*$, the dual of $C(X,\mathbf{R})$. The weak* topology on $C(X,\mathbf{R})^*$ gives via $R$ a topology on $M(X,\mathbf{R})$ which coincides with the initial topology associated to $\mathcal{F} = \{(J(f),\mathbf{R})\colon f \in C(X,\mathbf{R})\}$, whereby

$$J(f)(\mu) = \int_X f\,d\mu.$$

The unit ball in $M(X,\mathbf{R})$ is then given by

$$M_{\leqslant 1} := \left\{\mu \in M(X,\mathbf{R})\colon \left|\int_X f\,d\mu\right| \leqslant 1 \text{ for all } f \in C(X,\mathbf{R}) \text{ with } \|f\|_b \leqslant 1\right\}.$$

In it there is a particularly interesting subset, namely the space of probability measures on $X$,

$$\begin{aligned} M^1(X) &:= \{\mu\colon \mu \text{ is a positive regular Borel measure on } X,\ \mu(X) = 1\} \\ &= \left\{\mu \in M(X,\mathbf{R})\colon \int_X f\,d\mu \geqslant 0 \text{ whenever } f \geqslant 0 \text{ and } \int_X \mathbf{1}\,d\mu = 1\right\}. \end{aligned}$$

It is thus a convex weakly* closed subset of the unit ball $M_{\leqslant 1}$.

**Example 5.28.** Let $X = [0,1]$, $\lambda$ the Lebesgue measure such that $\lambda([0,1]) = 1$ and $\delta_0$ the Dirac measure at 0. Then $\delta_0 \in M^1(X)$ and $\mu_n := n\lambda\mathbf{1}_{[0,1/n]} \in M^1(X)$ for all $n \geqslant 1$. For every $f \in C([0,1],\mathbf{R})$ we have

$$\int_X f\,d\mu_n = n\int_0^{1/n} f(x)\,d\lambda(x) \to f(0) = \int_X f\,d\delta_0$$

since

$$\begin{aligned}\left|\int_0^{1/n} n(f(x)-f(0))\,d\lambda(x)\right| &\leqslant n\int_0^{1/n} \sup_{y\in[0,1/n]} |f(y)-f(0)|\,d\lambda(x) \\ &= \sup_{x\in[0,1/n]} |f(y)-f(0)| \to 0\end{aligned}$$

by continuity of $f$. Thus, $\mu_n \rightharpoonup \delta_0$ weakly*.

Now we turn to the central result of this chapter:

**Theorem 5.29** (Banach-Alaoglu)**.** Let $V$ be a normed space. Then the unit ball $B_{\leqslant 1}(0;V^*)$ in $V^*$ is weakly* compact.

**Proof.** For the ease of notation, let us write $B^* := B_{\leqslant 1}(0;V^*)$ and $B := B_{\leqslant 1}(0;V)$. Note that for any $\lambda \in B^*$ we have $\lambda(B) \subset \{z \in \mathbf{K} : |z| \leqslant 1\} =: D$ so one can identify[1] $B^*$ with a subset of $D^B = \prod_{v\in B} D$ which is the space of functions from $B$ to $D$ (we will also write $B^*$ for this identification). Note that by Corollary 5.22 we know that the weak* topology on $B^*$ is the initial topology w.r.t. the family $\{([\lambda \mapsto \lambda(v)],\mathbf{K})\}_{v\in B}$ which we can also write as $\{(\pi_v|_{B^*}, D)\}_{v\in B}$; here $\pi_v : D^B \to D$ is the projection onto the "$v$th coordinate". But note that the product topology on $D^B$ is the initial topology w.r.t. the family $\{(\pi_v, D)\}_{v\in B}$, so we find that the weak* topology on $B^*$ is just the product topology on $D^B$ restricted to $B^*$.

Due to Tychonoff's Theorem we know that $D^B$ is compact w.r.t. the product topology, so all that remains is to show that $B^*$ is closed in $D^B$. Let $(\lambda_\alpha)_{\alpha\in A} \subset B^*$ be a net converging to $f \in D^B$; we have to show that $f$ is linear and thus again an element of $B^*$. For arbitrary $v, w \in B$ we have $\lambda_\alpha(v+w) = \lambda_\alpha(v) + \lambda_\alpha(w)$ so by continuity of addition in $\mathbf{K}$ we get that $\lambda_\alpha(v) + \lambda_\alpha(w)$ converges to $f(v) + f(w)$, as desired. Same applies to $\lambda_\alpha(cv)$ for some scalar $c \in \mathbf{K}$.

■

**Remark 5.30.** Even if $V$ is a Banach space, the closed unit ball $B_{\leqslant 1}(0;V)$ is not necessarily compact in the weak topology. In fact a theorem of Kakutani asserts that $B_{\leqslant}(0;V)$ is weakly compact iff $V$ is reflexive, that is iff the linear isometry $J : V \to V^{**}$ from Proposition 2.12 is surjective.

[1] This can be achieved by considering the canonical injection

$$\Psi : B^* \hookrightarrow D^B, \quad \lambda \mapsto (\lambda(v))_{v\in B}$$

which can be turned into a homeomorphism (w.r.t. the weak* topology) via $\widetilde{\Psi} : B^* \to \Psi(B^*)$, $\lambda \mapsto \Psi(\lambda)$.

One of the important consequences of Banach-Alaoglu in the context of Example 5.27 is the following:

**Corollary 5.31.** Let $X$ be a compact Hausdorff space. Then the space $M^1(X)$ of probability measures on $X$ is weakly* compact.

**Proof.** Is it a weakly* closed subset of the unit ball $M_{\leqslant 1}$. ∎

**Remark 5.32.** An equivalent formulation is: the space $M^1(X)$ equipped with the initial topology associated to the family $\mathcal{F} = \{(f, \mathbf{R})\colon f \in C(X \to \mathbf{R})\}$ is compact.

We end this chapter with a construction that will bear its fruits in the next chapter.

Let $X$ be a compact Hausdorff space and $\psi\colon X \to X$ a homeomorphism. Then $\psi$ gives rise to a linear map:

$$\lambda(\psi)\colon C(X) \to C(X), \quad f \mapsto f \circ \psi^{-1}.$$

Then the following two properties are immediate:

(1) $\|\lambda(\psi)(f)\|_b = \|f\|_b$ for all $f \in C(X)$
(2) $\lambda(\psi_1 \circ \psi_2) = \lambda(\psi_1) \circ \lambda(\psi_2)$.

In particular $\lambda(\psi)$ is a bijective isometry of $C(X)$ with inverse $\lambda(\psi^{-1})$.

Moreover, its adjoint $\lambda(\psi)^*\colon C(X)^* \to C(X)^*$ is a bijective isometry (exercise 9 and (2) of Proposition 5.24 imply that $\lambda(\psi)^*$ is weakly* continuous).

Now by (2) and properties of adjunction we have:

$$\lambda(\psi_1 \circ \psi_2)^* = (\lambda(\psi_1)\lambda(\psi_2))^* = \lambda(\psi_2)^*\lambda(\psi_1)^*$$

which leads us to define:

$$\lambda^*(\psi) := (\lambda(\psi)^*)^{-1}.$$

This way we recover:

$$\lambda^*(\psi_1 \circ \psi_2) = \lambda^*(\psi_1)\lambda^*(\psi_2).$$

Now, coming back to Example 5.27, let us compute $\lambda^*(\psi)$ under the identification:

$$M(X) \to C(X)^*, \quad \mu \to \Phi_\mu.$$

We have $\lambda^*(\psi)(\Phi_m) = \Phi_\nu$ and proceed to compute $\nu$. One has

$$\begin{aligned}
\lambda^*(\psi)(\Phi_\mu)(f) &= \underbrace{(\lambda(\psi)^*)^{-1}}_{\lambda(\psi^{-1})^*}(\Phi_\mu)(f) \\
&= \Phi_\mu(\lambda(\psi^{-1})(f)) \\
&= \Phi_\mu(f \circ \psi) = \int_X f \circ \psi(x)\, d\mu(x)
\end{aligned}$$

and $\Phi_\nu(f) = \int_X f(y)\, d\mu(y)$.

Thus we have that for all $f \in C(X)$

$$\int_X f(y)\, d\nu(y) = \int_X f \circ \psi(x)\, d\mu(x).$$

Now let for every Borel set $E \subset X$,

$$\nu'(E) = \mu(\psi^{-1}(E)).$$

Then $\nu'$ is a signed regular Borel measure denoted $\psi_* \mu$ and called the push-forward of the measure $\mu$ by $\psi$. Clearly,

$$\nu'(E) = \int_X \mathbf{1}_{\psi^{-1}(E)}\, d\mu(x) = \int_\Omega \mathbf{1}_E \circ \psi(x)\, d\mu(x)$$

which by using step functions and the monotone convergence theorem implies (cf. Theorem 8.3)

$$\int_X f(y)\, d\nu'(y) = \int_X f(\psi(x))\, d\mu(x)$$

for all $f \in C(X)$. Hence $v' = v$, i.e.

$$\lambda^*(\psi)(\Phi_\mu) = \Phi_{\psi_x(\mu)}.$$

Finally we observe from this that for all homeomorphisms $\psi\colon X \to X$,

$$\lambda^*(\psi)(M'(X)) = M'(X).$$

An interesting point is, that this construction can be generalised in the following way: let $\psi\colon X \to Y$ be a continuous map of compact Hausdorff spaces and $\mu \in M(X)$. Then, for Borel sets $E \subset Y$,

$$(\psi_* \mu)(E) := \mu(\psi^{-1}(E))$$

defines an element $\psi_* \mu \in M(Y)$ and

(1) $M(X) \to M(Y), \quad \mu \to \psi_* \mu$ is weakly$^*$ continuous.
(2) $\psi_*(M'(X)) \subset M'(Y)$.

# Chapter 6. Convexity; the Kakutani-Markov Fixed Point Theorem, and Krain-Milman

This chapter has three interrelated Thema. First, we will exploit the analytic form of Hahn-Banach to establish separation properties of convex sets in TVS whose topology in generated by a family of seminorms. Then we establish a general fixed point theorem (Kakutani-Markov) that has far reaching consequences; for instance, it implies that any homeomorphism of a compact Hausdorff space has an invariant probability measure. Finally, we establish a very geometric result on compact convex subsets of a TVS generated by a sufficient family of seminorms that says one can recover it from its subset of extreme points. Applied to a homeomorphism of a compact Hausdorff space it implies the existence of an ergodic invariant probability measure. Unless otherwise specified, all vector spaces in this chapter are over $\mathbf{R}$.

## 6.1. Convexity

Let $V$ be an $\mathbf{R}$-vector space.

**Definition 6.1.** A subset $K \subset V$ is convex if for all $v, w \in K$ the vector $(1-t)v + tw$ is again in $K$ for all $t \in [0,1]$.

**Example 6.2.** Let $p\colon V \to \mathbf{R}$ be a gauge. Recall that this means
(1) $p(\lambda v) = \lambda p(v)$ for all $\lambda > 0$ and all $v \in V$.
(2) $p(v_1 + v_2) \leqslant p(v_1) + p(v_2)$ for all $v_1, v_2 \in V$.

Then for all $r \in \mathbf{R}$,

$$P_{<r} := \{v \in V : p(v) < r\}$$

is convex. Indeed, for $v_1, v_2 \in P_{<r}$ and $t \in (0,1)$ we have

$$\begin{aligned} p((1-t)v_1 + tv_2) &\leqslant p((1-t)v_1) + p(tv_2) \\ &= tp(v_1) + (1-t)p(v_2) \\ &< tr + (1-t)r = r. \end{aligned}$$

This clearly extends to $t = 1$ and $t = 0$. The same argument shows that if one replaces $<$ in the definition of $P_{<r}$ by $\leqslant$ the corresponding subset is as well convex.

This process can be reversed for convex subsets with additional properties.

**Definition 6.3.** A subset $A \subset V$ is *absorbent* if for all $v \in V$ there exists $\alpha > 0$ s.t. for all $|\lambda| \geqslant \alpha$ we have $v \in \lambda A$.

**Example 6.4.** $B_{\leqslant 1}(0) \subset \mathbf{R}^2$ is absorbent.

**Example 6.5.** Let $V$ be a TVS and $U$ an open subset containing 0. Then $U$ is absorbent. Indeed, for arbitrary $v \in V$ the map $\mathbf{R} \to V$, $t \mapsto tv$ is continuous, in particular at $t = 0$. Hence there exists $\varepsilon > 0$ s.t. $tv \in U$ for all $|t| \leqslant \varepsilon$. Thus $v \in \lambda U$ for all $|\lambda| \geqslant 1/\varepsilon$.

The relationship between convex subsets and gauges is given by the following proposition.

**Proposition 6.6.** Let $V$ be an $\mathbf{R}$-vector space and $A \subset V$ s.t.

(1) $A$ convex
(2) $0 \in A$
(3) $A$ is absorbent.

Then

$$p_A(v) := \inf\{\alpha > 0 \colon v \in \alpha A\}$$

is a gauge on $V$. In addition:

(1) $\{x \colon p_A(x) < 1\} \subset A \subset \{x \colon p_A(x) \leqslant 1\}$
(2) If $A \subset B$ and $B$ satisfies (1), (2) and (3) then $p_B \leqslant p_A$.

**Proof.** Since $A$ is absorbent, $p_A$ is well defined. First, for $\lambda > 0$ we have

$$\begin{aligned} p_A(\lambda v) &= \inf\{\alpha > 0 \colon \lambda v \in \alpha A\} \\ &= \inf\Big\{\alpha > 0 \colon v \in \frac{\alpha}{\lambda} A\Big\} \\ &= \inf\{\lambda\alpha > 0 \colon v \in \alpha A\} \\ &= \lambda \inf\{\alpha > 0 \colon v \in \alpha A\} = \lambda p_A(v). \end{aligned}$$

For subadditivity, let $v, w \in V$. For any $\alpha, \beta > 0$ s.t. $v \in \alpha A$ and $w \in \beta A$ we have

$$\frac{v}{\alpha+\beta} + \frac{w}{\alpha+\beta} = \frac{\alpha}{\alpha+\beta}\frac{v}{\alpha} + \frac{\beta}{\alpha+\beta}\frac{w}{\beta} \in A$$

employing convexity. Hence $v + w \in (\alpha+\beta)A$, demonstrating subadditivity.

For (1), it is clear that $\{x \colon p_A(x) < 1\} \subset A$ due to convexity, since if $x \in \alpha A$ for $\alpha < 1$ we also have $x \in A$. Similarly, $x \in A$ immediately implies $p_A(x) \leqslant 1$, showing the other inclusion. Property (2) is clear.

■

**Example 6.7.** Let $V$ be a TVS defined by a family $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ of seminorms, and $U = N(0, F, \varepsilon)$ where $\varepsilon > 0$ and $F \subset A$ is finite. Let us compute $p_A$: we have for $v \in V$ and $\lambda > 0$,

$$\begin{aligned} v \in \lambda U &\iff v \in \lambda N(0; F; \varepsilon) \\ &\iff \max_{\alpha \in F} \Big\|\frac{v}{\lambda}\Big\|_\alpha < \varepsilon \\ &\iff \max_{\alpha \in F} \|v\|_\alpha < \lambda\varepsilon \end{aligned}$$

and hence $p_A(v) = \frac{1}{\varepsilon} \max_{\alpha \in F} \|v\|_\alpha$.

With these tools at hand we can now prove:

**Theorem 6.8.** Let $V$ be a TVS (over $\mathbf{R}$) defined by a family of seminorms $\{\|\cdot\|_\alpha\}_{\alpha \in A}$. Let $A \subset V$ be a nonempty open convex subset and $v \notin A$. Then there exists $F \in V^*$ with $F(a) < F(v)$ for all $a \in A$.

**Proof.** Pick $a_0 \in A$; then $A' := A - a_0$ is open, convex and $0 \in A'$. Let $p_{A'}\colon V \to \mathbf{R}$ be the associated gauge as in Proposition 6.6. Define $M := \mathbf{R}(v - a_0)$ and

$$f\colon M \to \mathbf{R}, \quad \lambda(v - a_0) \mapsto \lambda.$$

Since $v - a_0 \notin A'$ we have by Proposition 6.6 that $p_{A'}(v - a_0) \geqslant 1$. By homogeneity of $p_{A'}$ and linearity of $f$ this implies $f(w) \leqslant p_{A'}(w)$ for all $w \in M$. By Theorem 2.4 there exists a linear extension $F\colon V \to \mathbf{R}$ of $f$ with $F(w) \leqslant p_{A'}(w)$ for all $w \in V$.

Let us observe that since $A'$ is open we have $A' = \{w \in V\colon p_{A'}(w) < 1\}$. Indeed, the $\supset$ inclusion follows from Proposition 6.6. For the reverse inclusion let $w \in A'$. Since $\mathbf{R} \to V$, $t \to tw$ is continuous and $A'$ open, there exists an open neighbourhood $U \subset A'$ of $w$ and $\varepsilon > 0$ s.t. $(1 - \varepsilon, 1 + \varepsilon)w \subset U \subset A'$ and hence $(1 + \varepsilon/2)w \in A'$.

Thus $F(a - a_0) \leqslant p_{A'}(a - a_0) < 1$ for all $a \in A$ and $F(v - a_0) = f(v - a_0) = 1$. ■

With this at hand we can now show that one can separate points from closed convex subsets.

**Corollary 6.9.** Let $V$ be as in Theorem 6.8, $A \subset V$ closed convex and $x \notin A$. Then there is $\alpha \in \mathbf{R}$ and $F \in V^*$ s.t. $F(a) < \alpha < F(x)$ for all $a \in A$.

**Proof.** Since $A$ is closed and $x \notin A$ we can find an open neighbourhood $U$ of $0$ with $(x + U) \cap A = \varnothing$. Let $J$ be finite and $\varepsilon > 0$ s.t. $N := N(0; J; \varepsilon) \subset U$. Since $N = -N$ we conclude from $(x - N) \cap A = \varnothing$ that $x \notin A + N$. Now observe that

$$A + N = \bigcup_{a \in A} (a + N)$$

which is therefore open; it is also convex. By Theorem 6.8 there is $F \in V^*$ s.t. for all $a \in A$ and $u \in N$ we have $F(a + u) < F(x)$. Since $F \neq 0$, there exists $v_0 \in V$ with $F(v_0) \neq 0$ and since $N$ is absorbent there is $\lambda \neq 0$ with $u_0 := \lambda v_0 \in N$. Thus, $F(u_0) \neq 0$ and exchanging $u_0$ with $-u_0$ if necessary, we may assume $F(u_0) > 0$. Hence $F(a) + F(u_0) < F(x)$ so with $\alpha := F(x) - F(u_0)$ we get $F(a) < \alpha < F(x)$ for all $a \in A$. ■

The simple example of $A = B_{<1}(0; \mathbf{R}^2)$ and $x \in \mathbf{R}^2$ with $\|x\| = 1$ shows that the condition that $A$ is closed in the previous corollary, is important.

It is now time to turn to the concept of convex hull of a subset $A \subset V$ in an $\mathbf{R}$-vector space.

**Definition 6.10.** The convex hull $\operatorname{conv}(A)$ of a subset $A \subset E$ is the intersection of all convex subsets containing $A$.

**Example 6.11.** A triangle.

One shows by recurrence that if $v_1, \dots, v_n$ belong to a convex set $C$ then $\sum_{k=1}^n \lambda_k v_k \in C$ whenever $\lambda_1, \dots, \lambda_n \in \mathbf{R}_{\geqslant 0}$ with $\sum_{k=1}^n \lambda_k = 1$.

This leads to the following formula for the convex hull of a subset $A \subset V$: Let

$$\mathbf{R}_{\geqslant 0}^{(S)} := \{\lambda\colon S \to \mathbf{R}_{\geqslant 0}\colon \lambda \text{ has finite support}\}.$$

Then

$$\operatorname{conv}(A) = \left\{\sum_{a\in A} \lambda(a)a\colon \lambda \in \mathbf{R}_{\geqslant 0}^{(A)}, \sum_{a\in A} \lambda(a) = 1\right\}.$$

Indeed, that $\operatorname{conv}(A)$ contains the right hand side follows from the above remark while the reversed inclusion follows from the fact that the right hand side is convex, as the following computation shows: For all $\lambda, \gamma \in \mathbf{R}_{\geqslant 0}^{(A)}$ and $t \in [0,1]$ we have

$$t\sum_{a\in A} \lambda(a)a + (1-t)\sum_{a\in A} \gamma(a)a = \sum_{a\in A}(t\lambda(a) + (1-t)\gamma(a))a$$

and

$$\sum_{a\in A}(t\lambda(a) + (1-t)\gamma(a)) = t + (1-t) = 1.$$

The following generalises Proposition 5.25.

**Proposition 6.12.** Let $(V, \|\cdot\|_V)$ be a normed vector space and $C \subset V$ convex. Then $C$ is strongly closed iff it is weakly closed.

**Proof.** If $C$ is weakly open it is strongly open, hence weakly closed implies strongly closed.

For the converse, assume that $C \subset V$ is strongly closed; let us show that $V \setminus C$ is weakly open. Let $x_0 \notin C$; by Corollary 6.9 there is $\alpha \in \mathbf{R}$ and $F \in V^*$ with

$$F(c) < \alpha < F(x_0)$$

for all $c \in C$. Thus $\{v \in V\colon F(v) > \alpha\}$ is weakly open, contains $x_0$ and is disjoint from $C$. ■

To proceed further we consider the following lemma.

**Lemma 6.13.** Let $V$ be a TVS. Then the closure $\overline{C}$ of a convex subset $C \subset E$ is convex.

**Proof.** For $t \in [0,1]$ we have

$$(1-t)\overline{C} + t\overline{C} = \overline{(1-t)C} + \overline{tC} \subset \overline{((1-t)C + tC)} \subset \overline{C}$$

using convexity of $C$. We have used that for $X, Y \subset V$ and $t \in \mathbf{K}$, $V$ a TVS over $\mathbf{K}$, one has $t\overline{X} = \overline{tX}$ and $\overline{X} + \overline{Y} \subset \overline{X + Y}$ which follows immediately from continuity of multiplication by a scalar and vector addition.

For example, if $x \in \overline{X}$ and $y \in \overline{Y}$ are limit points of $X$ respectively $Y$, there exist nets $(x_\alpha)_{\alpha\in A} \subset X$ and $(y_\beta)_{\beta\in B} \subset Y$ converging to $x$ resp. $y$. By continuity of addition we have that $x_\alpha + y_\beta$ converges to $x + y$ since the sequence $((x_\alpha, y_\beta))_{(\alpha,\beta)\in A\times B}$ converges to $(x, y)$ (here $A \times B$ is endowed with

$(\alpha,\beta) \leqslant (\alpha',\beta') \iff \alpha \leqslant \alpha' \wedge \beta \leqslant \beta'$ which naturally makes it a directed set). ■

From the preceding lemma we deduce the following proposition.

**Proposition 6.14.** Let $(V, \|\cdot\|_V)$ be a normed vector space and $C \subset V$ a convex subset. Then its closure $\overline{C}$ for the strong topology coincides with its closure $\overline{C}^w$ for the weak topology.

**Proof.** Since $\overline{C}^w$ is weakly closed it is also strongly closed and since $\overline{C}^w \supset C$ this implies $\overline{C}^w \supset \overline{C}$. Conversely, by Lemma 6.13, $\overline{C}$ is convex, meaning since it is strongly closed it is also weakly closed (cf. Proposition 6.12) and hence $\overline{C} \supset \overline{C}^w$. ■

**Corollary 6.15.** Let $(V, \|\cdot\|_V)$ be a normed space and $(v_n)_{n\geqslant 1}$ a sequence s.t. $v_n \rightharpoonup v$ weakly. Then there is a sequence $w_n \in \operatorname{conv}(\{v_n\}_{n\geqslant 1})$ s.t. $w_n \to v$ strongly.

**Proof.** By Proposition 6.14, $\overline{\operatorname{conv}(\{v_n\}_{n\geqslant 1})}$ and $\overline{\operatorname{conv}(\{v_n\}_{n\geqslant 1})}^w$ coincide. ■

Another astonishing fact follows from the closed graph theorem and Proposition 6.12.

**Proposition 6.16.** Let $X, Y$ be Banach spaces and $T\colon X \to Y$ a linear map that is continuous for the weak topologies on $X$ and $Y$. Then $T$ is bounded and the converse also holds.

**Proof.** ($\Longleftarrow$) Follows from (1) of Proposition 5.24.
($\Longrightarrow$) We use the closed graph theorem: $\operatorname{graph}(T) \subset X \times Y$ is weakly closed and clearly convex. Hence it is strongly closed which implies $T$ is bounded. ■

## 6.2. The Markov-Kakutani fixed point theorem

Let $V$ be a topological vector space.

**Definition 6.17.** An automorphism of $V$ is a bijective continuous linear map $T\colon V \to V$ whose inverse $T^{-1}\colon V \to V$ is continuous.

Then the set $\operatorname{Aut}(X)$ of automorphisms of $X$ forms a group under composition.

**Example 6.18.** Let $X$ be a compact Hausdorff space, $V = M(X)$ the space of signed regular Borel measures with weak*-topology. In Chapter 5 we constructed a group homomorphism $\lambda^*\colon \operatorname{Homeo}(X) \to \operatorname{Aut}(E)$ which takes the concrete form

$$\lambda^*(\psi)(\mu) = \psi_*(\mu).$$

We observed that the weakly*-compact convex subset $M^1(X)$ of probability measures is invariant under $\lambda^*(\psi)$ for all $\psi \in \operatorname{Homeo}(X)$.

**Example 6.19.** Let $G$ be a group with discrete topology. A mean on $G$ is a continuous linear form $\gamma \in \ell^\infty(G)^*$ s.t.

(1) $\gamma(\mathbf{1}_G) = 1$
(2) $\gamma(f) \geqslant 0$ for all $f \in \ell^\infty(G)$ with $f \geqslant 0$.

Then the set

$$m(G) = \{\gamma \in \ell^\infty(G)^* \colon \gamma \text{ is a mean}\}$$

is a convex weakly* closed subset of the unit ball of $\ell^\infty(G)^*$ and hence compact.

Now for all $g \in G$ and $\gamma \in \ell^\infty(G)$, let

$$\lambda(g)\gamma(x) := \gamma(g^{-1}x).$$

Then

(1) $\|\lambda(g)\gamma\|_{\ell^\infty(G)} = \|\gamma\|_{\ell^\infty(G)}$ for all $g \in G$ and $\gamma \in \ell^\infty(G)$
(2) $\lambda(g_1 g_2) = \lambda(g_1)\lambda(g_2)$.

Hence $\lambda(g) \colon \ell^\infty(G)$ is a bijective isometry. Its adjoint $\lambda(g)^* \colon \ell^\infty(G)^* \to \ell^\infty(G)^*$ is therefore weakly* continuous. Setting $\lambda^*(g) := (\lambda(g)^*)^{-1}$ we obtain a group homomorphism $\lambda^* \colon G \to \mathrm{Aut}(\ell^\infty(G)^*)$. If $m \in \ell^\infty(G)$ then

$$(\lambda^*(g)m)(\gamma) = m(\lambda(g^{-1})\gamma).$$

Clearly, if $m$ is a mean, $\lambda^*(g)m$ is a mean for all $g \in G$. Thus the compact convex subset $m(G)$ is invariant under $\lambda^*(g)$ for all $g \in G$.

Observe that if $m \in m(G)$ we can define a set function $\mu \colon \mathcal{P}(G) \to \mathbf{R}_{\geqslant 0}$ by $\mu(E) := m(\mathbf{1}_E)$. This set function has then the following properties:

(1) $\mu(G) = 1$
(2) $\mu$ is finitely additive.

**Theorem 6.20** (Markov-Kakutani)**.** Let $V$ be a TVS generated by a sufficient family of seminorms, $G$ an *abelian* group and $\pi \colon G \to \mathrm{Aut}(V)$ a homomorphism. Assume that $A \subset V$ is compact, convex, nonempty and $G$-invariant, that is $\pi(g)(A) \subset A$ for all $g \in G$. Then there exists a point in $A$ that is fixed by $\pi(g)$ for all $g \in G$.

**Proof.** For every $g \in G$ and $n \geqslant 1$, define

$$M_{n,g} \colon V \to V, \quad v \mapsto \frac{1}{n}\sum_{k=0}^{n-1} \pi(g^k)(v).$$

Then $M_{n,g}$ is a continuous linear map and since $A$ is convex and $\pi(g^k)(A) \subset A$ we have $M_{n,g}(A) \subset A$. Let

$$G^* := \{M_{n_1,g_1} \circ \cdots \circ M_{n_\ell,g_\ell} \colon \ell \geqslant 1,\, n_1, \dots, n_\ell \in \mathbf{N},\, g_1, \dots, g_\ell \in G\}.$$

This is a family of continuous linear maps $V \to V$ with the following properties:

(1) If $T, S \in G^*$ then $T \circ S \in G^*$.
(2) $T(A) \subset A$ for all $T \in G^*$.

(3) If $T, S \in G^*$ then $T \circ S = S \circ T$ .

To see (3) it suffices to show that $M_{n,a} \circ M_{m,b} = M_{m,b} \circ M_{n,a}$ which follows from a direct computation (using that $G$ is abelian).

**Claim:** $\bigcap_{T \in G^*} T(A) \neq \varnothing$.

We note that $T(A) \subset A$ is compact for all $T \in G^*$ so that it is sufficient[1] to show that for all $T_1, \ldots, T_\ell \in G^*$ we have $\bigcap_{k=1}^{\ell} T_k(A) \neq \varnothing$.

We have that for all $1 \leqslant k \leqslant \ell$,

$$T_k(A) \supset T_k(T_1 \cdots T_{k-1} T_{k+1} T_\ell)(A)$$

and since all $T_j$'s commute we obtain

$$T_k(T_1 \cdots T_{k-1} T_{k+1} T_\ell) = T_1 \cdots T_\ell$$

and hence

$$\bigcap_{k=1}^{\ell} T_k(A) \supset T_1 \cdots T_\ell(A) \neq \varnothing$$

which proves the claim.

Let now $y \in \bigcap_{T \in G^*} T(A)$. Then it follows that for all $n \geqslant 1$ and $g \in G$ there exists some $x_{n,g} \in A$ with $M_{n,g}(x_{n,g}) = y$, i.e.

$$y = \frac{1}{n} \sum_{k=0}^{n-1} \pi(g^k)(x_{n,g}).$$

It follows that

$$\pi(g)(y) - y = \frac{1}{n}(\pi(g^n)(x_{n,g}) - x_{n,g}).$$

Now, let $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ be the sufficient family of seminorms defining the topology on $V$. Then for all $a \in A$,

$$\|\pi(g)(y) - y\|_\alpha \leqslant \frac{1}{n}(\|\pi(g^n)(x_{n,g})\|_\alpha + \|x_{n,g}\|_\alpha).$$

Let $B_\alpha := \sup_{v \in A} \|v\|_\alpha < +\infty$, since $A$ is compact. We conclude that

$$\|\pi(g)(y) - y\|_\alpha \leqslant \frac{2B_\alpha}{n}$$

for all $n \geqslant 1$, hence $\|\pi(g)(y) - y\|_\alpha = 0$ for all $\alpha \in A$. This yields $\pi(g)(y) = y$ for all $g \in G$. ■

In the context of Example 6.18 we obtain the following corollaries, which we state in terms of group actions. Recall that a group action of a group $G$ on a set $X$ is a map

$$G \times X \to X, \quad (g, x) \mapsto g_* x$$

satisfying the following axioms:

(1) $e_* x = x$ for all $x \in X$.
(2) $(g_1 g_2)_* x = (g_1)_* (g_2)_* (x)$ for all $g_1, g_2 \in G$ and all $x \in X$.

[1] Cf. Proposition 8.2, this is a characterisation of compactness.

If $X$ is a topological space, the group action is by homeomorphisms if for all $g \in G$ the map

$$\psi_g \colon X \to X, \quad x \mapsto g_* x$$

is a homeomorphism.

**Corollary 6.21.** Let $G \times X \to X$ be an action by homeomorphism of an abelian group $G$ on a compact Hausdorff space $X$. Then there exists an invariant probability measure, that is, there exists $\mu \in M^1(X)$ s.t. $(\psi_g)_* \mu = \mu$ for all $g \in G$.

**Proof.** We apply Theorem 6.20 to the space $V = M(X)$ of signed regular Borel measures with weak* topology and the homomorphism $\pi \colon G \to \operatorname{Aut}(V)$ obtained by composing

$$\begin{aligned} G &\longrightarrow \operatorname{Homeo}(X) \xrightarrow{\lambda^*} \operatorname{Aut}(V) \\ g &\longmapsto \qquad \psi_g \qquad \longmapsto \lambda^*(\psi_g). \end{aligned}$$

Then the weak* compact convex subset $M^1(X)$ is invariant under $\pi(g)$ for all $g \in G$ and the corollary follows by invoking the just proven theorem. ■

**Corollary 6.22.** Let $X$ be a compact Hausdorff space and $\psi \in \operatorname{Homeo}(X)$. Then there exists $\mu \in M^1(X)$ with $\psi_*(\mu) = \mu$.

**Proof.** Apply the preceding corollary to the group action

$$\mathbf{Z} \times X \to X, \quad (n, x) \mapsto \psi^n(x).$$

■

In the context of Example 6.19 we obtain the following corollary.

**Corollary 6.23.** Let $G$ be an abelian group. Then there exists a mean $m \in m(G)$ that is invariant under $\lambda^*(g)$ for all $g \in G$. In particular there exist a set function $\mu \colon 2^G \to [0, 1]$ with the properties

(1) $\mu(G) = 1$
(2) $\mu$ is finitely additive
(3) $\mu(gE) = \mu(E)$ for all $g \in G$ and $E \subset G$.

This Corollary is the starting point of the theory of amenable groups: a group $G$ is amenable if there is a mean $m \in m(G)$ that is invariant under "left translations", that is $\lambda^*(g)m = m$ for all $g \in G$. Not all groups are amenable; for instance, the free groups $G = \mathbf{F}(a, b)$ on two generators is not, and this is intimately connected to the paradoxical decomposition mentioned back, in Theorem [tbd., Chapter 2.2] (Banach-Tarski paradox).

**Remark 6.24.** The countable group $G$ has property (F) if there exists a sequence $F_n \subset G$ of finite subsets s.t. for all $g \in G$

$$\frac{|gF_n \triangle F_n|}{|F_n|} \xrightarrow{n \to \infty} 0.$$

Then one can shows hat the conclusion of Theorem 6.20 holds for any countable group $G$ satisfying (F). In particular such a $G$ is amenable; it is a Theorem of Föllner that the converse holds.

Assume next that $X$ is a compact Hausdorff space and $\psi\colon X \to X$ is a homeomorphism. We know now that there exist $\psi$-invariant probability measures on $X$. The case where there is a unique such is particularly interesting.

**Theorem 6.25.** Assume that there is a unique $\psi$-invariant probability measure $\mu$ on $X$. Then for all $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\psi^k(x)) = \int_X f \, d\mu$$

the convergence being uniform in $x \in X$.

**Proof.** Let $(n_k)_{k\geqslant 1}$ be a strictly increasing sequence in $\mathbf{N}$ and $(x_k)_{k\geqslant 1}$ a sequence in $X$. Define the sequence of probability measures,

$$\mu_k := \frac{1}{n_k} \sum_{j=0}^{n_k-1} \delta_{\psi^j(x_k)} \in M^1(X).$$

Let $\nu$ be any accumulation point of this sequence (w.r.t. weak* topology), that is:

$$\nu \in \bigcap_{N\geqslant 1} \overline{\{\mu_k \colon k \geqslant N\}}.$$

Then for all $f \in C(X)$,

$$\mu_k(f\circ\psi) - \mu_k(f) = \frac{1}{n_k}(f(\psi^{n_k}(x_k)) - f(x_k))$$

and hence

$$|\mu_k(f\circ\psi) - \mu_k(f)| \leqslant \frac{2\|f\|_b}{n_k}$$

which implies that $\nu$ is $\psi$-invariant.

(2) If the convergence in the theorem is not uniform, there is $f \in C(X)$ and $\varepsilon > 0$ s.t.

$$\limsup_{n\to\infty} \sup_{x\in X} \left| \frac{1}{n} \sum_{k=0}^{n-1} f(\psi^k(x)) - \int_X f \, d\mu \right| > \varepsilon.$$

Thus there is a strictly increasing sequence $(n_k)_{k\geqslant 1}$ and a sequence $(x_k)_{k\geqslant 1}$ in $X$ with

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} f(\psi^j(x_k)) - \int_X f \, d\mu \right| > \varepsilon$$

which by (1) would lead to a $\psi$-invariant probability measure $\nu \neq \mu$. ∎

In specific situations it is relatively easy to establish uniqueness of the $\psi$-invariant measure, as the following example shows.

**Example 6.26.** In the notations of Chapter 3 "the problem of measure", we defined the probability measure $\lambda \in M^1(\mathbf{R}/\mathbf{Z})$ which on continuous functions $f \in C(\mathbf{R}/\mathbf{Z})$ is given by

$$\int_{\mathbf{R}/\mathbf{Z}} f\, d\lambda = \int_0^1 f(\pi(x))\, d\lambda(x)$$

where $\pi\colon \mathbf{R} \to \mathbf{R}/\mathbf{Z}$ is the canonical projection and $\lambda$ the Lebesgue measure on $\mathbf{R}$ normalised s.t. $\lambda([0,1]) = 1$. For $\alpha \in \mathbf{R}/\mathbf{Z}$ the map

$$T_\alpha\colon \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}, \quad x \mapsto x + \alpha$$

is a homeomorphism. For $\alpha \in \mathbf{Q}/\mathbf{Z}$, let $\alpha = p/q + \mathbf{Z}$ where $p, q \in \mathbf{N}$ are coprime. Then $(T_\alpha)^q = \mathrm{id}_{\mathbf{R}/\mathbf{Z}}$ and for all $x \in \mathbf{R}/\mathbf{Z}$,

$$\mu_x := \frac{1}{q} \sum_{k=0}^{q-1} \delta_{T_\alpha^k(x)} \in M^1(\mathbf{R}/\mathbf{Z})$$

is an $T_\alpha$-invariant probability measure; of course, $\lambda$ itself is $T_\alpha$-invariant for all $\alpha \in \mathbf{R}/\mathbf{Z}$. The point is then that for $\alpha \notin \mathbf{Q}/\mathbf{Z}$, $\lambda$ is the unique $T_\alpha$-invariant probability measure. As a result we get from Theorem 6.25

$$\frac{1}{n} \sum_{k=0}^{n-1} f(x + k\alpha) \to \int_{\mathbf{R}/\mathbf{Z}} f\, d\lambda$$

uniformly for all $f \in C(\mathbf{R}/\mathbf{Z})$.

## 6.3. Extreme points and the Krein-Milman Theorem

Let $V$ be an $\mathbf{R}$-vector space. For $x, y \in V$ we define

$$[x,y] = \{(1-t)x + ty\colon t \in [0,1]\}$$
$$(x,y) = \{(1-t)x + ty\colon t \in (0,1)\}$$

so for example $(x,x) = \{x\}$.

**Definition 6.27.** Let $A \subset V$ be a convex subset.

(1) $x \in A$ is an extreme point of $A$ if $x \in (y,z)$ with $y, z \in A$ implies $x = y = z$.
(2) A convex subset $B \subset A$ is extreme in $A$ if $(y,z) \cap B \neq \varnothing$ with $y, z \in A$ implies $[y,z] \subset B$.

For example, the extreme points of a triangle are its corners and its extreme sets are the entire Triangle, the edges and corners.

**Theorem 6.28.** Let $V$ be a TVS defined by a sufficient family of seminorms and $A \subset V$ convex compact. Then

$$A = \overline{\mathrm{conv}(\mathrm{ex}(A))}.$$

Before entering the proof, which uses the second geometric form of Hahn-Banach, we make the following remark.

**Remark 6.29.** Let $A \subset B \subset C$ be convex subsets. Then, if $A$ is extreme in $B$ and $B$ is extreme in $C$, $A$ is extreme in $C$. Indeed, suppose that $(y,z) \cap A \neq \varnothing$ for $y, z \in C$. Then, since $A \subset B$, $(y,z) \cap B \neq \varnothing$ so since $B$ is extreme in $C$ we have $[y,z] \subset B$ and hence $[y,z] \subset A$ since $A$ is extreme in $B$.

Now lets get to the proof of the theorem.

**Proof.** (1) We first show that every closed convex nonempty extreme subset $B \subset A$ contains an extreme point of $A$. To this end, consider

$$\mathcal{E}(B) := \{c \colon C \subset B,\, C \text{ is nonempty, closed, convex and extreme in } A\}$$

with the ordering $C_1 \leqslant C_2$ if $C_2 \subset C_1$. This is a partially ordered set and we now show that every totally ordered subset has an upper bound. Let $\mathcal{C} \subset \mathcal{E}(B)$ be totally ordered and define $M := \bigcap_{c \in \mathcal{C}} C$. Since $\mathcal{C}$ is totally ordered, given any finite subset $\{C_1, \ldots, C_n\}$ of $\mathcal{C}$ we may assume wlog that $C_1 \subset \cdots \subset C_n$ and hence $\bigcap_{k=1}^n C_k = C_1 \neq \varnothing$. By compactness of $A$ we deduce $M \neq \varnothing$. Clearly, $M$ is closed and convex; in addition, if $(y,z) \cap M \neq \varnothing$ with $y, z \in A$ then for all $C \in \mathcal{C}$, since $C$ is extreme in $A$, we have $[y,z] \subset C$ and hence $[y,z] \subset M$. Thus $M \in \mathcal{E}(B)$ and it is an upper bound of $\mathcal{C}$. By Zorn's lemma there exists a maximal element $Z \in \mathcal{E}(B)$; we claim that $Z$ is a single point.

For the contrary, assume that there exist $x, y \in Z$ with $x \neq y$. By Hahn-Banach there is $F \in V^*$ with $F(x) < F(y)$. Now consider $m := \max\{F(z) \colon z \in Z\}$ which exists since $Z$ is compact and $D = \{z \in Z \colon F(z) = m\}$. Then $D$ is closed and convex; in addition, if $v, w \in Z$ and $(v,w) \cap D \neq \varnothing$ then for some $t \in (0,1)$,

$$m = F((1-t)v + tw) \geqslant F(v) \geqslant F(w).$$

From

$$(1-t)F(v) + tF(w) = F((1-t)v + tw) \geqslant F(v)$$

and $t > 0$ we get $F(w) \geqslant F(v)$ and from

$$(1-t)F(v) + tF(w) \geqslant F(w)$$

with $1 - t > 0$ we get $F(v) \geqslant F(w)$. Thus, $F(v) = F(w) = m$. This shows that $[v,w] \subset D$ and hence $D$ is extreme in $Z$ so extreme in $A$ by Remark 6.29. On the other hand, $F(x) < F(y)$ so that $x \notin D$ which contradicts the maximality of $Z$.

(2) Above we have shown that $\mathrm{ex}(A) \neq \varnothing$. Clearly, $A \supset \overline{\mathrm{conv}(\mathrm{ex}(A))}$; if there now were $x \in A$ and $x \notin \overline{\mathrm{conv}(\mathrm{ex}(A))}$ then by the second geometric form of Hahn-Banach there is $\alpha \in \mathbf{R}$ and $F \in V^*$ such that

$$F(x) > \alpha > F(y)$$

for all $y \in \overline{\mathrm{conv}(\mathrm{ex}(A))}$. Consider now, as above,

$$m := \max\{F(z) \colon z \in A\}$$

$$D := \{z \in A \colon F(z) = m\}.$$

As above, $D$ is closed, convex and extreme in $A$ hence by (1) contains an extreme point $e$ of $A$. But then

$$F(e) \geqslant F(x) > \alpha > F(y)$$

for all $y \in \mathrm{ex}(A)$, a contradiction. ■

Let $G \times X \to X$ be a countable group acting by homeomorphisms on a compact Hausdorff space. An example is $G = \mathbf{Z}$ and the action is

$$\mathbf{Z} \times X \to X$$
$$(m, x) \mapsto \psi^m(x)$$

where $\psi \in \mathrm{Homeo}(X)$. There is a measure theoretic notion of transitivity which which plays a central role in dynamics.

**Definition 6.30.** A $G$-invariant probability measure $\mu \in M^1(X)$ is called *ergodic* if whenever $S \subset X$ is a $G$-invariant measurable set, we have either $\mu(S) = 0$ or $\mu(X \setminus S) = 0$.

If now $M^1(X)^G$ denotes the subsets of $M^1(X)$ of $G$-invariant probability measures then $M^1(X)^G$ is convex and weakly* closed, hence compact.

**Lemma 6.31.** $\mu \in M^1(X)^G$ is ergodic iff $\mu$ is an extreme point of $M^1(X)^G$.

**Proof.** We only prove the ( $\Longleftarrow$ ) direction: if $\mu$ is not ergodic there is $S \subset X$ measurable $G$-invariant with $0 < \mu(S) < 1$. Define then

$$\mu_1 := \frac{1}{\mu(S)}\, \mu|_S\,, \quad \mu_2 := \frac{1}{\mu(X \setminus S)}\, \mu|_{X \setminus S}\,.$$

Then $\mu_1, \mu_2 \in M^1(X)^G$ and

$$\mu = \mu(S)\mu_1 + \mu(X \setminus S)\mu_2.$$

Since $\mu \neq \mu_1$ and $\mu \neq \mu_2$, $\mu$ is not an extreme point. ■

The Kreinn-Milman Theorem then implies the following corollary.

**Corollary 6.32.** If there exists a $G$-invariant probability measure on $X$ then there is an ergodic one. In fact, every $G$-invariant probability measure is a weak* limit of convex combinations of ergodic ones.

## Chapter 7. Fourier analysis and Sobolev embedding theorem

Recall that in Example **??** we defined a family of function spaces on $\mathbf{R}^n$ called Sobolev spaces, denoted $W^{s,p}(\mathbf{R}^d)$ where $p \geqslant 1$ and $p \in \mathbf{N}$. Loosely speaking, $W^{s,p}(\mathbf{R}^d)$ consists of all functions admitting weak derivatives up to order $s$ that are in $L^p(\mathbf{R}^d)$. In this chapter we shall concentrate on $W^{s,2}(\mathbf{R}^d)$ and show that if $s > r + \frac{n}{2}$ this space consists of bounded $C^r$-functions. The means to achieve this is Fourier Analysis and the Plancherel Theorem to which we turn now.

### 7.1. Basic Fourier Analysis on $\mathbf{R}^d$

For a thorough treatment we refer to Ioacobelli, Analysis 4, Chapter 3. Here we will recall the basic definitions and theorems necessary four our purposes. For $x, \xi \in \mathbf{R}^d$ we write $\xi \cdot x$ for the euclidean inner product, and we set

$$\langle f, g \rangle = \int_{\mathbf{R}^d} f \overline{g} \, dx$$

whenever $|fg| \in L^1(\mathbf{R}^d)$.

**Definition 7.1.** For $f \in L^1(\mathbf{R}^d)$ define the Fourier transform[1] of $f$ by

$$\hat{f}(\xi) = \int_{\mathbf{R}^d} f(x) e^{-2\pi i \langle \xi, x \rangle} \, dx$$

Recall that

$$C_0(\mathbf{R}^d) = \Big\{ f \colon \mathbf{R}^d \to \mathbf{C} \colon f \text{ is continuous and } \lim_{|\xi| \to \infty} f(\xi) = 0 \Big\}$$

which together with the usual sup norm $\| \bullet \|_\infty$ is a Banach space.

**Proposition 7.2.** If $f \in L^1(\mathbf{R}^d)$ then $\hat{f} \in C_0(\mathbf{R}^d)$ and the operator

$$\mathcal{F} \colon L^1(\mathbf{R}^d) \to C_0(\mathbf{R}^d), \quad f \mapsto \hat{f}$$

has operator norm $\|\mathcal{F}\| \leqslant 1$.

**Proof.** First let us show that $\hat{f}$ is continuous. This follows from the dominated convergence theorem since $|f(x) e^{-2\pi i \langle \xi + h, x \rangle}| = |f(x)|$ and thus

$$\begin{aligned} \lim_{h \to 0} \hat{f}(\xi + h) &= \lim_{h \to 0} \int_{\mathbf{R}^d} f(x) e^{-2\pi i \langle \xi + h, x \rangle} \, dx \\ &= \int_{\mathbf{R}^d} \lim_{h \to 0} f(x) e^{-2\pi i \langle \xi + h, x \rangle} \, dx = \hat{f}(\xi). \end{aligned}$$

For the second part we give an argument that generalises well to the case of LCA groups (in our case of $\mathbf{R}^d$ one could also proceed with integration

---

[1] There exist different conventions for the definition of the Fourier transform, each making certain properties more concise to state. Another common convention is

$$\hat{f}(\xi) = (2\pi)^{-d/2} \int_{\mathbf{R}^d} f(x) e^{-i \langle \xi, x \rangle} \, dx$$

by parts). First we show it for $f \in C_c(\mathbf{R}^d)$. Comparing $f$ to $\tau_h f$, where $\tau_h f(x) = f(x-h)$, we note that for any $h, \xi \in \mathbf{R}^d$

$$\int_{\mathbf{R}^d} |f(x) - \tau_h f(x)|\, dx \geqslant \left| \int_{\mathbf{R}^d} (f(x) - \tau_h f(x)) e^{-2\pi i \langle \xi, x\rangle}\, dx \right|$$
$$= |\hat{f}(\xi) - e^{-2\pi i \langle \xi, h\rangle} \hat{f}(\xi)| = |\hat{f}(\xi)| \cdot |1 - e^{-2\pi i \langle \xi, h\rangle}|$$

using (1) of Lemma 7.6. By continuity and dominated convergence we know that the left term goes to zero as $h \to 0$. Now for any $(\xi_n)_{n\geqslant 1}$ s.t. $|\xi_n| \to \infty$ set $(h_n)_{n\geqslant 1}$ s.t. $|h_n| = \frac{1}{|\xi_n|}$ and $|1 - e^{-2\pi i \langle \xi, h_n\rangle}| \geqslant 1$ so that the above calculation gives us

$$\lim_{n\to\infty} |\hat{f}(\xi_n)| \leqslant \lim_{n\to\infty} \|f - \tau_{h_n} f\|_{L^1(\mathbf{R}^d)} = 0$$

as desired. Employing density of $C_c(\mathbf{R}^d)$ in $L^1(\mathbf{R}^d)$, for $f \in L^1(\mathbf{R}^d)$ and $\varepsilon > 0$ we pick $g \in C_c(\mathbf{R}^d)$ with $\|f - g\|_{L^1(\mathbf{R}^d)} < \varepsilon$ and find

$$\limsup_{|\xi|\to\infty} |\hat{f}(\xi)| = \limsup_{|\xi|\to\infty} \left| \int_{\mathbf{R}^d} f(x) e^{-2\pi i \langle \xi, x\rangle}\, dx \right| \leqslant \limsup_{|\xi|\to\infty} |\hat{g}(\xi)| + \varepsilon = \varepsilon$$

The last assertion follows from $\left|\hat{f}(\xi)\right| \leqslant \|f\|_{L^1(\mathbf{R}^d)}$. ■

**Lemma 7.3.** Let $1 \leqslant p < +\infty$ and $f \in L^p(\mathbf{R}^d)$. Then

$$\mathbf{R}^d \to L^p(\mathbf{R}^d), \quad h \mapsto \tau_h f$$

is continuous.

**Remark 7.4.** Observe that for $f \in L^\infty(\mathbf{R}^d)$, $h \mapsto \tau_h f$ is continuous iff $f$ coincides almost everywhere with a uniformly continuous function.

**Proof.** It suffices to prove continuity at zero since for $y, h \in \mathbf{R}^d$ we have $\|\tau_{a+h} f - \tau_a f\|_{L^p(\mathbf{R}^d)} = \|\tau_h f - f\|_{L^p(\mathbf{R}^d)}$. Again by density of $C_c(\mathbf{R}^d)$ in $L^1(\mathbf{R}^d)$ we may assume $f \in C_c(\mathbf{R}^d)$ yielding

$$\lim_{h\to 0} \|\tau_h f - f\|^p_{L^p(\mathbf{R}^d)} = \int_{\mathbf{R}^d} \lim_{h\to 0} |f(x-h) - f(x)|\, dx = 0$$

by dominated convergence. ■

One of the major difficulties with the Fourier transform is that for $f \in \|\|_{L^1(\mathbf{R}^d)}$, $\hat{f}$ does not satisfy any global integrability conditions on $\mathbf{R}^d$. The next proposition specifies a class of function in $\|\|_{L^1(\mathbf{R}^d)}$ whose Fourier transform is in $\|\|_{L^p(\mathbf{R}^d)}$ for all $p \geqslant 1$.

Recall that $C_c^k(\mathbf{R}^d)$ is the space of compactly supported functions which $k$ times continuously differentiable. Recall also some multi-index notation: Given $(\alpha_1, \ldots, \alpha_d) \in \mathbf{N}^d$ we define

$$\partial^\alpha f = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} f$$

and for $\xi \in \mathbf{R}^d$ set $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$.

**Proposition 7.5.**

(1) If $f \in C_c^1(\mathbf{R}^d)$, then $\widehat{\partial_{x_j} f}(\xi) = 2\pi i \xi_j \hat{f}(\xi)$
(2) If $f \in C_c^k(\mathbf{R}^d)$ and $|\alpha| \leqslant k$, then $\widehat{\partial^\alpha f}(\xi) = (2\pi i)^{|\alpha|} \xi^\alpha \hat{f}(\xi)$
(3) If $f \in C_c^\infty(\mathbf{R}^d)$, then $\hat{f} \in C_0(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$ for all $p \geqslant 1$.

**Proof.**
(1) Via integration by parts we find (since $f$ has compact support)

$$\widehat{\partial_{x_j} f}(\xi) = \int_{\mathbf{R}^d} (\partial_{x_j} f(x)) e^{-2\pi i \langle \xi, x \rangle} \, dx$$
$$= -\int_{\mathbf{R}^d} f(x) \partial_{x_j} e^{-2\pi i \langle \xi, x \rangle} \, dx = 2\pi i \xi_j \hat{f}(\xi)$$

(2) Follows by induction from (1).
(3) Using (2) we get that $\xi^\alpha \hat{f}(\xi)$ is bounded for all $\alpha$ and so is $\prod_{j=1}^d (1 + \xi_j^2) \hat{f}(\xi)$. ■

Assertion (2) of the above proposition is of considerable interest since it shows that $\mathcal{F}$ converts the operator $\partial^\alpha$ into a simple multiplication.

For $a \in \mathbf{R} \backslash \{0\}$, let us define $\sigma_a f(x) = f(x/a)$. We then have the following properties:

**Lemma 7.6.** For $f \in L^1(\mathbf{R}^d)$ it holds
(1) $\widehat{\tau_h f}(\xi) = e^{-2\pi i \langle \xi, h \rangle} \hat{f}(\xi)$
(2) $\tau_h \hat{f}(\xi) = \hat{f}(\xi - h) = \widehat{e^{2\pi i \langle h, \bullet \rangle} f}(\xi)$
(3) $\widehat{\sigma_a f}(\xi) = a^d \sigma_{1/a} \hat{f}(\xi) = a^d \hat{f}(a\xi)$
(4) $\sigma_a \hat{f}(\xi) = \hat{f}(\xi/a) = a^d \widehat{\sigma_{1/a} f}(\xi)$

**Proof.** These are straightforward calculations. ■

**Example 7.7.** The function $\varphi(x) = e^{-\pi |x|^2}$ satisfies $\hat{\varphi} = \varphi$.

**Proof.** We compute it for $d = 1$, the general case then follows.

$$\hat{\varphi}(\xi) = \int_{\mathbf{R}} e^{-\pi x^2} e^{-2\pi i \xi x} \, dx$$
$$= e^{-\pi \xi^2} \int_{\mathbf{R}} e^{-\pi (x - i\xi)^2} \, dx$$
$$= e^{-\pi \xi^2} \int_{\mathbf{R}} e^{-\pi x^2} \, dx = e^{-\pi \xi^2}.$$

■

Let us now define the inverse Fourier transform:

**Definition 7.8.** For $g \in L^1(\mathbf{R}^d)$ define

$$\check{g}(\xi) = \int_{\mathbf{R}^d} g(\xi) e^{2\pi i \langle \xi, x \rangle} \, d\xi.$$

Note that $\check{g} = \overline{\mathcal{F}\overline{g}}$ so we may easily transfer the already established properties to $\check{g}$. We may also use the notation

$$\mathcal{F}^* g = \check{g}$$

which is justified by

**Lemma 7.9.** If $f, g \in L^1(\mathbf{R}^d)$ then

$$\langle \mathcal{F} f, g \rangle = \langle f, \mathcal{F}^* g \rangle.$$

Now we are in the position to show a version of the Fourier inversion formula.

**Theorem 7.10.** For $f \in C_c^\infty(\mathbf{R}^d)$ we have $\mathcal{F}^*\mathcal{F} f = f$.

**Proof.** It suffices to show that $\mathcal{F}^*\mathcal{F} f(0) = f(0)$ for all $f \in C_c^\infty(\mathbf{R}^d)$. Indeed, assuming this we have

$$f(x) = \tau_{-x} f(0) = \mathcal{F}^*\mathcal{F}\tau_{-x} f(0) = \mathcal{F}^*(e^{-2\pi i \langle x, \bullet \rangle} \mathcal{F} f)(0) = \mathcal{F}^*\mathcal{F} f(x).$$

We now want to show that

$$f(0) = \int_{\mathbf{R}^d} \hat{f}(\xi)\, d\xi$$

for $f \in C_c^\infty(\mathbf{R}^d)$, or in other words that $f(0) = \langle \hat{f}, \mathbf{1} \rangle$. Consider $\varphi(x) = e^{-\pi |x|^2}$ and $\varphi_a(x) = \varphi(x/a)$ and notice that by Example 7.7 $\varphi$ satisfies the inversion formula, and hence also $\varphi_a$ since

$$\mathcal{F}^*\mathcal{F}\varphi_a = \mathcal{F}^*\mathcal{F}\sigma_a\varphi = \mathcal{F}^* a^d \sigma_{1/a} \hat{f} = f.$$

Now $\varphi_a \to \mathbf{1}$ pointwise as $a \to +\infty$ (in fact uniformly on compact sets) and $|f(x)\varphi_a(x)| \leqslant |f(x)|$ meaning dominated convergence gives us $\langle \hat{f}, \varphi_a \rangle \to \langle \hat{f}, \mathbf{1} \rangle$ as $a \to \infty$. On the other hand we have

$$\begin{aligned}
\langle \hat{f}, \varphi_a \rangle &= \langle f, \mathcal{F}^* \varphi_a \rangle \\
&= \langle f, a^n \sigma_{1/a} \mathcal{F}^* \varphi \rangle \\
&= \int_{\mathbf{R}^d} f(x) a^n \overline{\mathcal{F}^* \varphi(ax)}\, dx = \int_{\mathbf{R}^d} f(x/a) \overline{\mathcal{F}^* \varphi(x)}\, dx.
\end{aligned}$$

Again, $x \mapsto f(x/a)$ converges pointwise to $f(0)$ as $a \to \infty$ and since $\left| f(x/a) \overline{\mathcal{F}^* \varphi(x)} \right| \leqslant \|f\|_{L^\infty(\mathbf{R}^d)} |\varphi(x)|$ we can apply dominated convergence to conclude

$$\lim_{a \to \infty} \langle \hat{f}, \varphi_a \rangle = f(0) \int_{\mathbf{R}^d} \overline{\mathcal{F}^* \varphi(x)}\, dx = f(0) \int_{\mathbf{R}^d} \varphi(x)\, dx = f(0).$$

■

**Corollary 7.11.** For all $f, g \in C_c^\infty(\mathbf{R}^d)$ we have $\langle \mathcal{F} f, \mathcal{F} g \rangle = \langle f, g \rangle$. In particular $\|\hat{f}\|_{L^2(\mathbf{R}^d)} = \|f\|_{L^2(\mathbf{R}^d)}$.

**Theorem 7.12** (Plancherel)**.** The map

$$\mathcal{F}\colon L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d) \to C_0(\mathbf{R}^d)$$

extends uniquely to a unitary operator $\mathcal{F}\colon L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$.

We will need the following

**Lemma 7.13.** Let $X, Y$ be Banach spaces with respective norms $\|\bullet\|_X, \|\bullet\|_Y$, $E \subset X$ a vector subspace and $T\colon E \to Y$ a bounded linear operator. Then $T$ extends uniquely to a bounded linear operator $\widetilde{T}\colon \overline{E} \to Y$. If in addition $\|Tx\|_Y = \|x\|_X$ for all $x \in E$ the same will hold for $\widetilde{T}$ and $x \in \overline{E}$.

**Proof.** For $x \in \overline{E}$ and $(x_n)_{n\geqslant 1} \subset E$ with $x_n \to x$ we define

$$\widetilde{T}x = \lim_{n\to\infty} Tx_n.$$

This limit exists since

$$\|Tx_n - Tx_m\|_Y \leqslant \|T\|\|x_n - x_m\|_X$$

and is also independent of the chosen sequence since if $y_n \to x$ then

$$\|Ty_n - Tx_n\|_Y \leqslant \|T\|\|y_n - x_n\|_X \to 0.$$

Lastly, if the assumption in the second assertion holds, for $x \in \overline{E}$ and any $x_n \to x$ we have

$$\|\widetilde{T}x\|_Y = \lim_{n\to\infty} \|Tx_n\|_Y = \lim_{n\to\infty} \|x_n\|_X = \|x\|_X.$$

■

**Proof.** (Theorem 7.12)

Since $\|\mathcal{F}\varphi\|_{L^2(\mathbf{R}^d)} = \|\varphi\|_{L^2(\mathbf{R}^d)}$ for all $\varphi \in C_c^\infty(\mathbf{R}^d)$ and $C_c^\infty(\mathbf{R}^d)$ is dense in $L^2(\mathbf{R}^d)$ (c.f. ), the above lemma tells us that $\mathcal{F}$ extends uniquely to an isometry $\widetilde{\mathcal{F}}\colon L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$. We claim that for $f \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, $\widetilde{\mathcal{F}}f = \hat{f}$. Indeed, if $(\varphi_n)_{n\geqslant 1} \subset C_c^\infty(\mathbf{R}^d)$ converges to $f$ in $L^2(\mathbf{R}^d)$, then by definition $\widetilde{\mathcal{F}}f = \lim_{n\to\infty} \mathcal{F}\varphi_n$. By Corollary 7.11 we have

$$\|\varphi_n - f\|_{L^2(\mathbf{R}^d)} = \|\hat{\varphi}_n - \hat{f}\|_{L^2(\mathbf{R}^d)}$$

so $\hat{\varphi}_n \to f$ in $L^2(\mathbf{R}^d)$ giving us $\widetilde{\mathcal{F}}f = f$.

For surjectivity we note that since $\mathcal{F}^*f = \overline{\mathcal{F}\overline{f}}$, $\mathcal{F}^*$ is also norm-preserving and by density extends to all of $L^2(\mathbf{R}^d)$. Since $\mathcal{F}^*\mathcal{F} = \mathcal{F}\mathcal{F}^* = \mathrm{id}$ on the dense subspace $C_c^\infty(\mathbf{R}^d)$, it holds on all of $L^2(\mathbf{R}^d)$. ■

## 7.2. Convolution

**Definition 7.14.** Given measurable $f, g\colon \mathbf{R}^d \to \mathbf{C}$ and $x \in \mathbf{R}^d$ s.t. $y \mapsto f(x-y)g(y)$ is in $L^1(\mathbf{R}^d)$ we define

$$f * g(x) = \int_{\mathbf{R}^d} f(x-y)g(y)\,dy.$$

We recall

**Proposition 7.15** (Young's inequality)**.** Let $1 \leqslant p, q, r \leqslant +\infty$ s.t. $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Then if $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$, $f * g$ is well defined and it holds

$$\|f * g\|_{L^r(\mathbf{R}^d)} \leqslant \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$

In the case $p = 1$ this gives $\|f * g\|_{L^q(\mathbf{R}^d)} \leqslant \|f\|_{L^1(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}$.

One of the main points of the convolution product is that the Fourier transform turns pointwise products into convolutions, namely for $f, g \in L^1(\mathbf{R}^d)$ we have

$$\begin{aligned} \widehat{f * g}(\xi) &= \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} f(x-y) g(y) e^{-2\pi i \langle \xi, x \rangle} \, dy \, dx \\ &= \int_{\mathbf{R}^d} g(y) e^{-2\pi i \langle \xi, y \rangle} \, dy \int_{\mathbf{R}^d} f(x-y) e^{-2\pi i \langle \xi, x-y \rangle} \, dx = \hat{f}(\xi) \hat{g}(\xi) \end{aligned}$$

whereby we can apply Fubini's theorem due to

$$\int_{\mathbf{R}^d} \int_{\mathbf{R}^d} |f(x-y) g(y)| \, dy \, dx \leqslant \|f\|_{L^1(\mathbf{R}^d)} \|g\|_{L^1(\mathbf{R}^d)}$$

by Young's inequality.

Our next goal will be to construct a sequence $\eta_\varepsilon \in C_c^\infty(\mathbf{R}^d)$ s.t. for $1 \leqslant p < +\infty$ and every $f \in L^p(\mathbf{R}^d)$, $\eta_\varepsilon * f \to f$ in $L^p(\mathbf{R}^d)$ as $\varepsilon \downarrow 0$ and $\eta_\varepsilon * f \in C_c^\infty(\mathbf{R}^d)$: such a sequence $\eta_\varepsilon$ is called "mollifier" or "approximate identity" and is a very useful tool.

**Proposition 7.16.** If $f \in C_c^\infty(\mathbf{R}^d)$ and $g \in L^p(\mathbf{R}^d)$ then $f * g \in C_c^\infty(\mathbf{R}^d)$ and $\partial_{x_i}(f * g) = (\partial_{x_i} f) * g$ for all $1 \leqslant i \leqslant n$.

For the proof, recall differentiation under the integral, namely

**Lemma 7.17.** Let $(X, \mathcal{A}, \mu)$ be a $\sigma$-finite measure space, $I \subset \mathbf{R}$ open and $f \colon I \times X \to \mathbf{C}$ a measurable function s.t.

(1) $f(t, \bullet) \in L^1(X)$ for all $t \in I$
(2) $f(\bullet, x)$ is differentiable for a.e. $x \in X$
(3) there exists $g \in L^1(X)$ s.t. $|\partial_t f(t, x)| \leqslant g(x)$ for all $t \in I$ and a.e. $x \in X$.

Then for every $t_0 \in I$ we have

$$\left( \partial_t \int_X f(t, x) \, dx \right) \Big|_{t = t_0} = \int_X \partial_t f(t_0, x) \, dx$$

**Proof.** (Proposition 7.16)

For the sake of a simple notation we will show differentiability at $x = 0$ for $i = 1$. Now if $B := B_{\leqslant r}(0)$ is s.t. $\operatorname{supp}(f) \subset B$, then

$$\left( \partial_{x_1} \int_{\mathbf{R}^d} f(x-y) g(y) \, dy \right) \Big|_{x=0} = \left( \partial_t \int_{\mathbf{R}^d} f(te_1 - y) g(y) \mathbf{1}_B(te_1 - y) \, dy \right) \Big|_{t=0}$$

and we can apply Lemma 7.17 to

$$h \colon I \times \mathbf{R}^d, \quad h(t, y) = f(te_1 - y) g(y) \mathbf{1}_B(te_1 - y).$$

where $I = (-\varepsilon, \varepsilon)$ is some open interval around zero. Indeed,

$$\|h(t, \bullet)\|_{L^1(\mathbf{R}^d)} \leqslant \|f\|_{L^\infty(\mathbf{R}^d)} \|g\|_{L^p(\mathbf{R}^d)} \mathcal{L}^d(B)^{1/q} < +\infty$$

for all $t \in I$ as well as

$$|\partial_t h(t, y)| \leqslant \|\partial_{x_1} f\|_{L^\infty(\mathbf{R}^d)} |g(y) \mathbf{1}_{B + (-\varepsilon, \varepsilon) e_1}(y)|$$

where $g \mathbf{1}_{B + (-\varepsilon, \varepsilon) e_1} \in L^1(\mathbf{R}^d)$. ■

Now we turn to the construction of approximate identity: fix any $\eta \in C_c^\infty(\mathbf{R}^d)$ with $\int_{\mathbf{R}^d} \eta(x)\, dx = 1$ and for $\varepsilon > 0$ let

$$\eta_\varepsilon(x) = \frac{1}{\varepsilon^d} \eta\Big(\frac{x}{\varepsilon}\Big).$$

Then $\eta_\varepsilon \in C_c^\infty(\mathbf{R}^d)$, $\int_{\mathbf{R}^d} \eta_\varepsilon(x)\, dx = 1$ and if $\operatorname{supp}(\eta) \subset B_{\leqslant r}(0)$ then $\operatorname{supp}(\eta_\varepsilon) \subset B_{\leqslant \varepsilon r}(0)$.

**Proposition 7.18.**

(1) If $f \in C(\mathbf{R}^d)$ then $\eta_\varepsilon * f \to f$ uniformly on compact sets as $\varepsilon \downarrow 0$
(2) If $1 \leqslant p < +\infty$ and $f \in L^p(\mathbf{R}^d)$, then $\eta_\varepsilon * f \to f$ in $L^p(\mathbf{R}^d)$ as $\varepsilon \downarrow 0$.

**Proof.**

(1) We compute

$$\begin{aligned} |\eta_\varepsilon * f(x) - f(x)| &= \left| \int_{\mathbf{R}^d} \eta_\varepsilon (f(x - y) - f(y))\, dy \right| \\ &\leqslant \sup_{y \in B_{\leqslant \varepsilon r}(0)} |f(x - y) - f(x)| \end{aligned}$$

which by continuity of $f$ shows (1).

(2) By density, for $0 < \varepsilon \leqslant 1$ we may choose $\varphi \in C_c(\mathbf{R}^d)$ s.t. $\|\varphi - f\|_{L^p(\mathbf{R}^d)} < \varepsilon$. Then

$$\|\eta_\varepsilon * f - f\|_{L^p(\mathbf{R}^d)} \leqslant \|\eta_\varepsilon * f - \eta_\varepsilon * \varphi\|_{L^p(\mathbf{R}^d)} + \|\eta_\varepsilon \varphi - \varphi\|_{L^p(\mathbf{R}^d)} + \|\varphi - f\|_{L^p(\mathbf{R}^d)}.$$

Now

$$\begin{aligned} \|\eta_\varepsilon * f - \eta_\varepsilon * \varphi\|_{L^p(\mathbf{R}^d)} &= \|\eta_\varepsilon * (f - \varphi)\|_{L^p(\mathbf{R}^d)} \\ &\leqslant \|\eta_\varepsilon\|_{L^1(\mathbf{R}^d)} \|f - \varphi\|_{L^p(\mathbf{R}^d)} = \|f - \varphi\|_{L^p(\mathbf{R}^d)} < \varepsilon. \end{aligned}$$

Lastly, $\operatorname{supp}(\eta_\varepsilon * \varphi) \subset B_{\leqslant r}(0) + \operatorname{supp}(\varphi) := K$ so

$$\begin{aligned} \|\eta_\varepsilon * \varphi - \varphi\|_{L^p(\mathbf{R}^d)}^p &= \int_K |\eta_\varepsilon * \varphi(y) - \varphi(y)|^p\, dy \\ &\leqslant \sup_{y \in K} |\eta_\varepsilon * \varphi(y) - \varphi(y)|^p \mathcal{L}^d(K) \end{aligned}$$

which by (1) vanishes as $\varepsilon \downarrow 0$. ■

This construction immediately gives us the following important corollary:

**Corollary 7.19.** Let $\Omega \subset \mathbf{R}^d$ be an open subset and $1 \leqslant p < +\infty$. Then $C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$.

**Proof.** Since $\Omega$ is locally compact and the Lebesgue measure is Radon, $C_c(\Omega)$ is dense in $L^p(\Omega)$. Now for any $f \in C_c(\Omega)$ we have that $\eta_\varepsilon * f \in C_c^\infty(\mathbf{R}^d)$ for all $\varepsilon > 0$ small enough s.t. $\operatorname{supp}(\eta_\varepsilon * f) = B_{\leqslant \varepsilon r}(0) + \operatorname{supp}(f) \subset \Omega$ and $\eta_\varepsilon * f \to f$ in $L^p(\Omega)$ as $\varepsilon \downarrow 0$. ■

Here is another applicaiton that enters in the proof of Plancherel:

**Lemma 7.20.** Let $1 \leqslant p, q < +\infty$ and $f \in L^p(\mathbf{R}^d) \cap L^q(\mathbf{R}^d)$. Then there is a sequence $(\varphi_n)_{n \geqslant 1} \subset C_c^\infty(\mathbf{R}^d)$ s.t. $\varphi_n \to f$ in $L^p(\mathbf{R}^d)$ and $L^q(\mathbf{R}^d)$.

## 7.3. Weak Derivatives

Let $\Omega \subset \mathbf{R}^d$ be open. Recall that for $f \in C^\infty(\Omega)$ and $\varphi \in C_c^\infty(\Omega)$ integration by parts gives

$$\int_\Omega f \partial^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega (\partial^\alpha f) \varphi.$$

We use this to define weak derivatives:

**Definition 7.21.** Let $f, g \in L^1_{\text{loc}}(\Omega)$, then $g$ is the weak $\alpha$-th partial derivative of $f$ on $\Omega$ if

$$\int_\Omega f \partial^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega h \varphi$$

for all $\varphi \in C_c^\infty(\Omega)$.

Observe that since $\varphi$ and $\partial^\alpha \varphi$ are compactly supported, these integrals make sense. Our first task is to that if such a weak derivative exists, it is unique. This will follow from

**Lemma 7.22.** Let $g \in L^1_{\text{loc}}(\Omega)$. If $\int_\Omega g\varphi = 0$ for all $\varphi \in C_c^\infty(\Omega)$ then $g = 0$ a.e.

**Proof.** Suppose that $g$ does not vanish a.e., then w.l.o.g. $S := \{g > 0\}$ has non-zero measure. Let $\delta := \mathcal{L}^d(S)$ and choose a compact set $K \subset S$ s.t. $\mathcal{L}^d(K) \geqslant \mathcal{L}^d(S) - \delta/2 = \delta/2$. Now choose an approximate identity $\eta_\varepsilon$ s.t. $\operatorname{supp}(\eta_\varepsilon) \subset B_{\leqslant \varepsilon}(0)$ and $\varepsilon_0 > 0$ small enough s.t. $\operatorname{supp}(\eta_{\varepsilon_0} * \mathbf{1}_K) \subset B_{\leqslant \varepsilon_0}(0) + K \subset \Omega$. Then for $0 < \varepsilon \leqslant \varepsilon_0$ we have $\eta_\varepsilon * \mathbf{1}_K \in C_c^\infty(\Omega)$ and

$$\int_\Omega g(\eta_\varepsilon * \mathbf{1}_K) = \int_\Omega g \mathbf{1}_K + \int_\Omega g(\eta_\varepsilon * \mathbf{1}_K - \mathbf{1}_K).$$

The first term is by assumption positive and the second term can be made arbitrarily small: Pass to a subsequence s.t. $\eta_{\varepsilon_k} * \mathbf{1}_K \to \mathbf{1}_K$ pointwise a.e. and apply dominated convergence, using that $\operatorname{supp}(\eta_\varepsilon * \mathbf{1}_K - \mathbf{1}_K) \subset B_{\leqslant \varepsilon_0(0)} + K$ and

$$\begin{aligned} |g(\eta_{\varepsilon_k} * \mathbf{1}_K - \mathbf{1}_K)| &\leqslant \sup_{k \geqslant 1} \|\eta_{\varepsilon_k} * \mathbf{1}_K - \mathbf{1}_K\|_{L^\infty(\Omega)} |g \mathbf{1}_{B_{\leqslant \varepsilon_0}(0) + K}| \\ &\leqslant |g \mathbf{1}_{B_{\leqslant \varepsilon_0}(0) + K}| \end{aligned}$$

We conclude that $\int_\Omega g\varphi > 0$ for $\varphi = \eta_{\varepsilon_k} * \mathbf{1}_K$ and $k$ large enough. ■

If $g$ is the weak $\alpha$-partial derivative of $f$, we write $g = \partial_w^\alpha f$. In particular it follows from Lemma 7.22 that if $f \in C^\infty(\Omega)$ then $\partial_w^\alpha f = \partial^\alpha f$.

**Example 7.24.** Let $\alpha > 0$ and $f(t) = |t|^\alpha$. Then $f \in L^1_{\text{loc}}(\mathbf{R})$ and the weak derivative $\partial_t^w f$ exists and is equal to $\alpha \operatorname{sgn}(t)|t|^{\alpha-1}$.

Now we can define Sobolev spaces $W^{k,p}(\Omega)$:

**Definition 7.25.**

$$W^{k,p}(\Omega) = \{f\colon \Omega \to \mathbf{C}\colon \partial_\alpha^w f \text{ exists for all } |\alpha| \leqslant k \text{ and } \partial_\alpha^w f \in L^p(\Omega)\}.$$

We define the norm on $W^{k,p}(\Omega)$ by:

$$\|f\|_{W^{k,p}(\Omega)} := \sum_{|\alpha|\leqslant k} \|\partial_w^\alpha f\|_{L^p(\Omega)}$$

**Proposition 7.26.** $W^{k,p}(\Omega)$ is a Banach space.

**Proof.** Let $(f_k)_{k\geqslant 1}$ be a Cauchy sequence in $W^{k,p}(\Omega)$. Then for all $\alpha$ with $|\alpha| \leqslant k$, $(\partial_w^\alpha f)_{k\geqslant 1}$ is a Cauchy sequence in $L^p(\Omega)$ and hence has a limit $f^\alpha \in L^p(\Omega)$; let $f = f^\alpha$ for $\alpha = (0, \dots, 0)$ and observe that $f^\alpha \in L^p(\Omega) \subset L^p_{\text{loc}}(\Omega) \subset L^1_{\text{loc}}(\Omega)$. By definition we have for all $|\alpha| \leqslant k$ and $\varphi \in C_c^\infty(\Omega)$:

$$\int_\Omega f_k \partial^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega (\partial_w^\alpha f_k)\varphi$$

but as $\partial_w^\alpha f_k \to f^\alpha$ in $L^p(\Omega)$ and $\varphi \in L^1(\Omega)$ we get

$$\int_\Omega (\partial_w^\alpha f_k)\varphi \to \int_\Omega f^\alpha \varphi.$$

Since $f_k \to f$ in $L^p(\Omega)$ we get

$$\int_\Omega f \partial^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega f^\alpha \varphi$$

and hence $\partial_w^\alpha f = f^\alpha$. ∎

**Remark 7.27.** $W^{k,2}(\Omega)$ is a Hilbert space; in fact for $f, g \in W^{k,2}(\Omega)$

$$\langle f, g\rangle = \sum_{|\alpha|\leqslant k} \langle \partial_w^\alpha f, \partial_w^\alpha g\rangle$$

leads to the norm

$$\|f\| = \Big(\sum_{|\alpha|\leqslant k} \|\partial_w^\alpha f\|^2_{L^2(\Omega)}\Big)^{1/2}$$

which is equivalent to $\|f\|_{W^{2,k}(\Omega)}$.

Let

$$C^\infty_{k,p}(\Omega) := \{f \in C^\infty(\Omega)\colon \|\partial^\alpha f\|_{L^p(\Omega)} < +\infty \text{ for all } |\alpha| \leqslant k\}.$$

Then $C^\infty_{k,p}(\Omega) \subset W^{k,p}(\Omega)$ and it is a fact that the former is dense in the latter for any open $\Omega \subset \mathbf{R}^d$. The proof of this is rather delicate and here we will show it for $\Omega = \mathbf{R}^d$. To this end we collect some simple facts about weak derivatives which will be also useful later on in the proof of the Sobolev embedding theorem.

**Lemma 7.28.**

(1) If $f \in W^{k,p}(\Omega)$ and $|\alpha| + |\beta| \leqslant k$ then $\partial_w^\alpha \partial_w^\alpha f = \partial_w^{\alpha+\beta} f$
(2) If $f \in W^{k,p}(\Omega)$ and $\varphi \in C_c^\infty(\Omega)$ then $\varphi f \in W^{k,p}(\Omega)$
(3) If $\varphi \in C_c^\infty(\mathbf{R}^d)$ and $f \in W^{k,p}(\mathbf{R}^d)$ then $\varphi * f \in C_{k,p}^\infty(\mathbf{R}^d)$ and $\partial^\alpha(\varphi * f) = \varphi * \partial_w^\alpha f$ for all $|\alpha| \leqslant k$.

**Proof.**

(1) For all $\varphi \in C_c^\infty(\Omega)$ we have

$$\begin{aligned}\int_\Omega (\partial_w^\alpha \partial_w^\beta f)\varphi &= (-1)^{|\alpha|} \int_\Omega (\partial_w^\beta f)\partial^\alpha \varphi \\ &= (-1)^{|\alpha|+|\beta|} \int_\Omega f \partial^{\alpha+\beta}\varphi = \int_\Omega (\partial_w^{\alpha+\beta} f)\varphi.\end{aligned}$$

By Lemma 7.22 we get $\partial_w^\alpha \partial_w^\beta f = \partial_w^{\alpha+\beta} f$.

(2) Let $\psi \in C_c^\infty(\mathbf{R}^d)$; we may assume $k \geqslant 1$.

$$\begin{aligned}\int_\Omega \varphi f \partial_j \psi &= \int_\Omega f(\partial_j \varphi \psi - \psi \partial_j \varphi) \\ &= \int_\Omega f \partial_j \varphi \psi - \int_\Omega f \psi \partial_j \varphi \\ &= -\int_\Omega (\partial_j^w f)\varphi\psi - \int_\Omega f\psi\partial_j\varphi \\ &= -\int_\Omega ((\partial_j^w f)\varphi + f\partial_j\varphi)\psi.\end{aligned}$$

Since $\partial_j^w f \in L^p(\Omega)$ and $f \in L^p(\Omega)$ so is $(\partial_j^w f)\varphi + f\partial_j\varphi$ and hence $\partial_j^w(\varphi f)$ exists and is in $L^p(\Omega)$ and $\partial_j^w(\varphi f) = \varphi \partial_j^w f + f \partial_j \varphi$ for all $1 \leqslant j \leqslant d$. One completes the proof by recurrence on $|\alpha|$ using the formula

$$\partial^\alpha(\varphi\psi) = \sum_{0 \leqslant \beta \leqslant \alpha} \binom{\alpha}{\beta} \partial^\beta \varphi \partial^{\alpha-\beta}\psi$$

and integrating by parts.

(3) We know that $\varphi * f \in C^\infty(\mathbf{R}^d)$ by Proposition 7.16, and also $\varphi * \partial_w^\alpha f \in L^p(\mathbf{R}^d)$ for all $|\alpha| \leqslant k$ by Proposition 7.15. Thus it suffices to show that $\partial^\alpha(\varphi * f) = \varphi * \partial_w^\alpha f$ for all $|\alpha| \leqslant k$. We have for all $\psi \in C_c^\infty(\mathbf{R}^d)$

$$\int_{\mathbf{R}^d} (\varphi * \partial_w^\alpha f)\psi = \int_{\mathbf{R}^d} (\partial_w^\alpha f)(\widetilde{\varphi} * \psi)$$

where we set $\widetilde{\varphi}(x) = \varphi(-x)$. The latter equals

$$(-1)^{|\alpha|} \int_{\mathbf{R}^d} f \partial^\alpha(\widetilde{\varphi} * \psi) = (-1)^{|\alpha|} \int_{\mathbf{R}^d} (\varphi * f)\partial^\alpha \psi = \int_{\mathbf{R}^d} (\partial^\alpha(\psi * f))\psi$$

and shows $\partial^\alpha(\varphi * f) = \varphi * \partial_w^\alpha f$. ■

**Proposition 7.29.** For $1 \leqslant p < +\infty$, $C^\infty_{k,p}(\mathbf{R}^d)$ is dense in $W^{k,p}(\mathbf{R}^d)$.

**Proof.** Let $\eta_\varepsilon$ be an approximate unity and $f \in W^{k,p}(\mathbf{R}^d)$. By Lemma 7.28, $\eta_\varepsilon * f \in C^\infty_{k,p}(\mathbf{R}^d)$ and $\partial^\alpha(\eta_\varepsilon * f) = \eta_\varepsilon * \partial^\alpha_w f$ for all $|\alpha| \leqslant k$. By Proposition 7.18 (2) we have for $|\alpha| \leqslant k$, $\eta_\varepsilon * \partial^\alpha_w f \to \partial^\alpha_w f$ in $L^p(\mathbf{R}^d)$ and hence $\partial^\alpha(\eta_\varepsilon * f) \to \partial^\alpha_w f$ in $L^p(\mathbf{R}^d)$. ■

## 7.4. Sobolev embedding theorems

The aim of this section is to prove

**Theorem 7.30** (Sobolev)**.** If $f \in W^{k,2}(\mathbf{R}^d)$ and $k > r + \frac{d}{2}$ then $f \in C^r_b(\mathbf{R}^d)$. Moreover, the inclusion $W^{k,2}(\mathbf{R}^d) \to C^r_b(\mathbf{R}^d)$ is a bounded operator.

**Remark 7.31.** More precisely, if $f \in W^{k,2}(\mathbf{R}^d)$, $f$ coincides a.e. with a $C^r$-function that is bounded. Interestingly, while in the statement of the theorem there is no Fourier transform, the latter is a crucial tool in the proof.

We proceed with three lemmas.

**Lemma 7.32.** If $f \in W^{k,2}(\mathbf{R}^d)$ then for all $|\alpha| \leqslant k$
$$\widehat{\partial^\alpha_w f}(\xi) = (2\pi i)^{|\alpha|}\xi^\alpha \hat{f}(\xi)$$
and in particular $\xi^\alpha \hat{f} \in L^2(\mathbf{R}^d)$ for all $|\alpha| \leqslant k$.

**Proof.** By induction it reduces to the case $k = 1$. Let $\varphi \in C^\infty_c(\mathbf{R}^d)$; since $\partial^w_j f \in L^2(\mathbf{R}^d)$ we have by Plancherel
$$\langle \widehat{\partial^w_j f}, \hat{\varphi}\rangle = \langle \partial^w_j f, \varphi\rangle = -\langle f, \partial_j \varphi\rangle = -\langle \hat{f}, \widehat{\partial_j \varphi}\rangle.$$
By Proposition 7.5 (1) we have
$$\widehat{\partial_j \varphi}(\xi) = 2\pi i \xi_j \hat{\varphi}(\xi)$$
and thus
$$-\langle \hat{f}, \widehat{\partial_j\varphi}\rangle = -\int_{\mathbf{R}^d} \hat{f}(\xi)\overline{\widehat{\partial_j\varphi}(\xi)}\, d\xi = \int_{\mathbf{R}^d} \hat{f}(\xi) 2\pi i \xi_j \overline{\hat{\varphi}(\xi)}\, d\xi = \langle 2\pi i \xi_j \hat{f}, \hat{\varphi}\rangle.$$
Finally, since $\{\hat{\varphi} \colon \varphi \in C^\infty_c(\mathbf{R}^d)\}$ is dense in $L^2(\mathbf{R}^d)$ we obtain the desired result. ■

For the sake of the applications we have in mind we formulate the next lemma in terms of the inverse Fourier transform, which we recall is given by
$$\check{h}(x) = \int_{\mathbf{R}^d} h(\xi) e^{2\pi i \langle \xi, x\rangle}\, d\xi.$$

**Lemma 7.33.** Let $r \in \mathbf{N}$, assume $h \in L^1(\mathbf{R}^d)$ and $\xi^\alpha h \in L^1(\mathbf{R}^d)$ for all $|\alpha| \leqslant r$. Then $\check{h} \in C^r_b(\mathbf{R}^d)$ and $\partial^\alpha \check{h}(x) = (2\pi i)^{|\alpha|}(\xi^\alpha h)^\vee(x)$.

**Proof.** As usual, by induction this reduces to the case $r = 1$ where a differentiation under the integral sign establishes the claim. ■

**Lemma 7.34.** Let $r \geqslant 0$; assume $f \in L^2(\mathbf{R}^d)$ and $\xi^\alpha \hat{f} \in L^1(\mathbf{R}^d)$ for all $|\alpha| \leqslant r$. Then $f \in C_b^r(\mathbf{R}^d)$ and $\partial^{\alpha f = (2\pi i)^{|\alpha|}\left(\xi^\alpha \hat{f}\right)^\vee}$.

**Proof.** Apply the preceding lemma to $h = \hat{f}$: then we get that $\check{h} \in C_b^r(\mathbf{R}^d)$ and $\partial^\alpha \check{h}(x) = (2\pi i)^{|\alpha|}(\xi^\alpha h)^\vee(x)$. Now use the hypothesis $f \in L^2(\mathbf{R}^d)$ to conclude $\check{h} = \check{\hat{f}} = f$. ■

Now we turn to the proof of Theorem 7.30.

**Proof.** By Lemma 7.32 we have $\xi^\alpha \hat{f} \in L^2(\mathbf{R}^d)$ for all $|\alpha| \leqslant k$. In order to show that $f \in C_b^r(\mathbf{R}^d)$ it suffices to show that $\xi^\alpha \hat{f} \in L^1(\mathbf{R}^d)$ for all $|\alpha| \leqslant r$. To this end, write $\xi^\alpha \hat{f} = g h_\alpha$ where

$$g = (1 + |\xi|^k)\hat{f}$$
$$h_\alpha = \frac{\xi^\alpha}{1 + |\xi|^k}.$$

Let's show that $h, g \in L^2(\mathbf{R}^d)$ . This will imply, for $|\alpha| \leqslant r$,

$$\|\xi^\alpha \hat{f}\|_{L^1(\mathbf{R}^d)} \leqslant \|g\|_{L^2(\mathbf{R}^d)} \|h_\alpha\|_{L^2(\mathbf{R}^d)}.$$

Lemma 7.34 will then imply that $f \in C_b^r(\mathbf{R}^d)$ and $\partial^{\alpha f} = (2\pi i)^{|\alpha|}\left(\xi^{\alpha f} \hat{f}\right)^\vee$ and thus $\|\partial^\alpha f\|_{L^\infty(\mathbf{R}^d)} \leqslant \left\|\xi^\alpha \hat{f}\right\|_{L^1(\mathbf{R}^d)}$. On the other hand we will relate $\|g\|_{L^2(\mathbf{R}^d)}$ to the Sobolev norm of $f$, concluding hte proof.

To estimate $\|g\|_{L^2(\mathbf{R}^d)}$, use that $|\xi| \leqslant \sum_{j=1}^d |\xi_j|$ and hence

$$|\xi| \leqslant \sum_{j=1}^d |\xi_j| \leqslant n^{1-1/k} \left( \sum_{j=1}^d |\xi_j|^k \right)^{1/k}$$

i.e. $|\xi|^k \leqslant n^{k-1} \sum_{j=1}^d |\xi_j|^k$. Thus

$$|g| \leqslant n^{k-1} \left( |\hat{f}| + \sum_{j=1}^d |\xi_j^k \hat{f}| \right)$$

which implies, using Lemma 7.32 and Plancherel,

$$\|h\|_{L^2(\mathbf{R}^d)} \leqslant n^{k-1} \left( \|f\|_{L^2(\mathbf{R}^d)} + \sum_{j=1}^d \|\partial_j^k f\|_{L^2(\mathbf{R}^d)} \right) \leqslant n^{k-1} \|f\|_{W^{k,2}(\mathbf{R}^d))}.$$

Next,

$$|h_\alpha(\xi)| \leqslant \frac{|\xi|^\alpha}{1 + |\xi|^k} \leqslant \frac{|\xi|^r}{1 + |\xi|^k}$$

and in polar coordinates

$$\int_{\mathbf{R}^d} \frac{|\xi|^{2r}}{(1 + |\xi|^k)^2} \, d\xi = c_d \int_0^\infty r^{d-1} \frac{r^{2r}}{(1 + r^k)^2} \, dr$$

which converges iff $k > r + \frac{d}{2}$. ■

**Chapter** 8. **Miscellaneous**

## 8.1. Topology

**Proposition 8.1.** A metric space $(X, d)$ is compact if and only if it is complete and totally bounded. $X$ is totally bounded if $\forall \varepsilon > 0$ there exists some finite subset $A \subset X$ s.t.

$$X = \bigcup_{a \in A} B_{\leqslant \varepsilon}(a)$$

i.e. $X$ is the union of finitely many balls of radius $\varepsilon$.

**Proof.** We show that it is equivalent to sequential compactness. First, to see that it implies sequential compactness, for any $(a_n)_{n \geqslant 1} \subset X$ we construct a convergent subsequence $(b_n)_{n \geqslant 1}$ as follows: By total boundedness of $X$ there exist $x_1, \ldots, x_m$ s.t. $X = \bigcup_{k=1}^m B_{\leqslant 1}(x_k)$ so there exists a $x^{(1)} \in \{x_1, \ldots, x_m\}$ s.t. infinitely many sequence members lie in $B_{\leqslant 1}(x_k)$, set $b_1$ to any of these sequence members. Then $b_n$ is defined as one of the infinitely many sequence members lying in the set $\bigcap_{k=1}^n B_{\leqslant 1/k}(x^{(k)})$ with $x^{(k)}$ iteratively chosen as described above. Then $(b_n)_{n \geqslant 1}$ is a Cauchy sequence and since $M$ is complete it converges.

For the other direction, if $(a_n)_{n \geqslant 1}$ is a Cauchy sequence then there exists a convergent subsequence $(a_{n_k})_{k \geqslant 1}$ since $M$ is sequentially compact. Writing $l$ for the limit of this subsequence we see that the entire sequence converges to $l$ since

$$d(a_n, l) \leqslant d(a_n, a_{n_k}) + d(a_{n_k}, l)$$

Next, suppose $M$ was not totally bounded so that there exists an $\varepsilon > 0$ s.t. $M$ can not be covered by finitely many balls of radius $\varepsilon$. Now define the sequence $(x_n)_{n \geqslant 1}$ recursively by $x_1 \in X$ and $x_n \in X \setminus \left(\bigcup_{k=1}^{n-1} B_\varepsilon(x_k)\right)$ so that $d(x_n, x_m) \geqslant \varepsilon$ for all $n, m \geqslant 1$ and hence $(x_n)_{n \geqslant 1}$ can not contain a convergent subsequence, a contradiction. ■

**Proposition 8.2.** Let $X$ be a topological space. Then the following are equivalent:

(1) Every open cover of $X$ admits a finite open subcover.
(2) If $\{F_\alpha\}_{\alpha \in \mathcal{A}}$ is family of closed subsets of $X$ s.t. for every finite subset $J \subset \mathcal{A}$, $\bigcap_{j \in J} F_j \neq \varnothing$, then also $\bigcap_{\alpha \in \mathcal{A}} F_\alpha \neq \varnothing$.

**Proof.** (1) $\Longrightarrow$ (2) Let $\{F_\alpha\}_{\alpha \in A}$ be a family of closed sets such that $\bigcap_{j \in J} F_j \neq \varnothing$ for all finite $J \subset \mathcal{A}$ but $\bigcap_{\alpha \in \mathcal{A}} F_\alpha = \varnothing$. Then $\{X \setminus F_\alpha\}_{\alpha \in \mathcal{A}}$ is an open cover of $X$ and hence, by assumption, admits a finite open subcover $\{U_1, \ldots, U_m\}$. However, this means that $\bigcup_{k=1}^m U_k = X$ so that $\bigcap_{k=1}^m X \setminus U_k = \varnothing$. But $X \setminus U_k \in \{F_\alpha\}_{\alpha \in A}$ which contradicts that finite intersections of sets belonging to this family are non-empty.

(2) $\Longrightarrow$ (1) Let $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ be an open cover of $X$ that does not admit a finite open subcover. Then for every finite $J \subset \mathcal{A}$ we have $\bigcup_{j \in J} U_j \neq X$ meaning $\{X \setminus U_\alpha\}_{\alpha \in \mathcal{A}}$ is a family of closed sets s.t. the intersection of every

finite subfamily is non-empty. By assumption this means $\bigcap_{\alpha \in A}(X \setminus U_\alpha) \neq \varnothing$ contradicting that $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ is an open cover of $X$. ■

## 8.2. Measure Theory

**Theorem 8.3** (Change of variables)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space and let $\phi\colon X \to Y$ be a measurable function from $(X, \mathcal{B})$ to another measurable space $(Y, \mathcal{C})$. Then for all measurable functions $f\colon Y \to [0, +\infty]$ we have

$$\int_X (f \circ \phi)\, d\mu = \int_Y f\, d\phi_* \mu$$

whereby $\phi_*\mu\colon \mathcal{C} \to [0, +\infty]$ is the *pushforward* of $\mu$ by $\phi$ given by $\phi_*\mu(E) := \mu(\phi^{-1}(E))$.

**Proof.** First, if $f = \mathbf{1}_E$ is an indicator function, for $E \in \mathcal{C}$ we have

$$\int_X (f \circ \phi)\, d\mu = \int_X \mathbf{1}_E(\phi(x))\, d\mu(x) = \int_X \mathbf{1}_{\phi^{-1}(E)}(x)\, d\mu(x) = \mu(\phi^{-1}(E))$$

and using the definition of the pushforward we find

$$\mu(\phi^{-1}(E)) = \phi_*\mu(E) = \int_Y \mathbf{1}_E(x)\, d\phi_*\mu(x) = \int_Y f\, d\phi_*\mu.$$

Using linearity one sees that this equality holds for all simple functions so an application of the monotone convergence theorem concludes the proof. ■

## References

[SS11] Elias M Stein and Rami Shakarchi. *Functional analysis: introduction to further topics in analysis.* Vol. 4. Princeton University Press, 2011.

[Tao10] Terence Tao. *An Epsilon of Room, II: pages from year three of a mathematical blog.* American Mathematical Society Providence, RI, 2010.

[Whi66] Robert Whitley. "Projecting m onto c0". In: *The American Mathematical Monthly* 73.3 (1966), pp. 285–286.