Applied Analysis

Lecture Notes
WinterTerm 2011/2012

by Markus Kunze

Contents

Chapter 1

Metric Spaces and Normed Spaces

No fear. No distractions. The ability to let that which does not matter truly slide.
from the movie "Fightclub"

Mathematicians are like Frenchmen: whatever you say to them they translate into their own language and forthwith it is something entirely different.
Johann Wolfgang von Goethe

1.1 Metric spaces

In this section, we introduce the concept of a metric space and give several examples. We start with the definition.

Definition 1.1.1. A *metric space* is a pair (M, d), where M is a set and $d : M \times M \to [0, \infty)$ satisfies

(M1) $d(x, y) = 0$ if and only if $x = y$. (definiteness)

(M2) $d(x, y) = d(y, x)$. (symmetry)

(M3) $d(x, y) \leq d(x, z) + d(z, y)$. (triangle inequality)

for all $x, y, z \in M$.

If $d : M \times M \to [0, \infty)$ satisfies (M1) – (M3), we say that d is a *metric* on M.

Remark 1.1.2. The letter d in the definition above is reminiscent of the word "*d*istance". One should think of $d(x, y)$ as the distance from x to y. With this interpretation, (M1) states that any x different from y has strictly positive distance to y, (M2) asserts that the distance from x to y is the same as that from y to x, whereas (M3) states that the distance from x to y becomes at most larger if we insist on visiting z on the way.

Remark 1.1.3. In Definition 1.1.1 we have indeed let that which does not matter truly slide and defined a metric through certain *axioms* which we believe to form the essence of the notion "distance". From now on, "metric space" and "distance" mean exactly what is stated in Definition 1.1.1, no more, no less. Some things which are true in certain situation might not hold in general and we cannot quote such things in any proofs. We may only appeal to the axioms (M1) – (M3) and results derived from these.

As we shall see this generality has great advantages, since we can identify certain objects as metric spaces (even though from a naive point of view this might not be obvious) and thus apply our general theory in this situation.

Example 1.1.4. The discrete metric.

Let M be any set, then $d : M \times M \to [0, \infty)$, defined by $d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$, is a metric on M. Indeed, (M1) and (M2) are obvious and for (M3) it is enough to note that the right-hand side in the equation has always value 1, unless $x = y = z$, in which case also the left-hand side has value 0.

Example 1.1.5. The real line in the euclidean metric.

On the set $M = \mathbb{R}$, a metric is defined by setting $d_2(x, y) := |x - y|$, where $|z|$ denotes the absolute value of the number z, i.e. $|x| = x$ if $x \geq 0$ and $|x| = -x$ if $x < 0$.

Indeed, (M1) is obvious in this situation and (M2) follows since $(x - y) = -(y - x)$ for all $x, y \in \mathbb{R}$ and, unless $x = y$, exactly one of these numbers is positive and the other negative. As for (M3), let x, y and z be given. We may assume without loss of generality that $y \leq x$ (otherwise, we switch the roles of x and y which we may, by (M2)).

We consider several cases.

If $y \leq z \leq x$, then

$$d(x, y) = x - y = x - z + z - y = d(x, z) + d(z, y)\,.$$

If $z \leq y \leq x$, then

$$d(x, y) = x - z + z - y \leq x - z + y - z = d(x, z) + d(z, y)$$

since $z - y \leq 0 \leq y - z$. The case where $y \leq x \leq z$ is similar.

Example 1.1.6. The complex plane in the euclidean metric.

On the set $M = \mathbb{C}$, a metric is defined by setting

$$d_2(z, w) := |z - w| = \left[(z - w)\overline{(z - w)}\right]^{\frac{1}{2}} = \left([\mathrm{Re}\,(z - w)]^2 + [\mathrm{Im}\,(z - w)]^2\right)^{\frac{1}{2}}.$$

This follows from a more general result in the next section.

Remark 1.1.7. In what follows, the sets $\mathbb{R}$ and $\mathbb{C}$ will always be endowed with the euclidean metric, unless something else is stated explicitly. We will write $\mathbb{K}$ to mean either $\mathbb{R}$ or $\mathbb{C}$. We note that the euclidean metric on $\mathbb{K}$ has the additional property that $|xy| = |x| \cdot |y|$. We will use this fact frequently in what follows.

Example 1.1.8. A sequence space.

Let ℓ^0 denote the space of all (real or complex) sequences. Thus an element $\mathbf{x}$ of ℓ^0 is a sequence: $\mathbf{x} = (x_k)_{k\in\mathbb{N}}$. Then $d_0 : \ell^0 \times \ell^0 \to [0, \infty)$, defined by

$$d_0(\mathbf{x}, \mathbf{y}) := \sum_{k=1}^{\infty} \frac{\min\{|x_k - y_k|, 1\}}{2^k}\,,$$

is a metric on ℓ^0. Note that the series converges since $2^{-k} \min\{|x_k - y_k|, 1\} \leq 2^{-k}$ and $\sum_{k=1}^{\infty} 2^{-k} < \infty$.

Proof. We check (M1) – (M3):

(M1) Since d_0 is defined through a series of nonnegative numbers, $d_0(\mathbf{x},\mathbf{y}) = 0$ implies that $2^{-k}\min\{|x_k - y_k|, 1\} = 0$ for all k which, in turn, implies that $x_k = y_k$ for all k, i.e. $\mathbf{x} = \mathbf{y}$.

(M2) Is immediate since $2^{-k}\min\{|x_k - y_k|, 1\} = 2^{-k}\min\{|y_k - x_k|, 1\}$ for all $k \in \mathbb{N}$.

(M3) Given $\mathbf{x},\mathbf{y},\mathbf{z} \in \ell^0$, we note that for all $k \in \mathbb{N}$ we have $|x_k - y_k| \leq |x_k - z_k| + |z_k - y_k|$; this is (M3) for the euclidean metric on $\mathbb{R}$ resp. $\mathbb{C}$. It follows that

$$\min\{|x_k - y_k|, 1\} \leq \min\{|x_k - z_k| + |z_k - y_k|, 1\}$$

for all $k \in \mathbb{N}$ which, in turn, yields that

$$\min\{|x_k - y_k|, 1\} \leq \min\{|x_k - z_k|, 1\} + \min\{|z_k - y_k|, 1\},$$

since for all nonnegative numbers a, b we have $\min\{a+b, 1\} \leq \min\{a, 1\} + \min\{b, 1\}$. Indeed, if $a + b \leq 1$, then also (since $a, b \geq 0$!) $a \leq 1$ and $b \leq 1$ and the inequality holds. If, on the other hand, $a + b \geq 1$, then $\min\{a + b, 1\} = 1$. For the right-hand side, we distinguish two cases. If both a and b are less then 1, then the right-hand side equals $a + b$ which is greater than 1 by assumption. If, on the other hand, a or b (or both!) are greater than 1, then the right-hand side is at least 1 and the inequality is trivially fulfilled.

Summing over $k \in \mathbb{N}$, the proof of (M3) is complete. □

We next give some examples how new metric spaces can be constructed from known ones.

Proposition 1.1.9. *(Induced metric) Let (M, d) be a metric space and $M_0 \subset M$. Then $d_{M_0} : M_0 \times M_0 \to [0,\infty)$, defined by $d_{M_0}(x, y) := d(x, y)$ defines a metric on M_0. The metric d_{M_0} is called the* induced metric *on M_0.*

Proof. (M1) – (M3) for d_{M_0} follow directly from (M1) – (M3) for d. □

If no confusion is likely, we will not notationally distinguish between d and its restriction to $M_0 \times M_0$ and merely speak of the metric space (M_0, d).

Proposition 1.1.10. *(Product metric)*

Let (M_i, d_i) be metric spaces for $i = 1, \ldots, n$. Then

$$d((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := \max\{d_i(x_i, y_i) \,:\, i = 1, \ldots, n\}$$

defines a metric on $M := M_1 \times \cdots \times M_n$.

The space (M, d) is called the product *of the metric spaces $(M_1, d_1), \ldots, (M_n, d_n)$.*

Proof. Obviously, d is nonnegative. Now let $x = (x_1, \ldots x_n), y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$ be elements of M.

If $d(x, y) = 0$, then, by definition, $d_i(x_i, y_i) = 0$ for all $i \in \{1, \ldots, n\}$. By (M1) for d_i, it follows that $x_i = y_i$ for all i and hence $x = y$. This proves (M1) for d.

Using (M2) for d_i, we have $d_i(x_i, y_i) = d_i(y_i, x_i)$ for all i. Taking the maximum over i, (M2) for d follows.

Finally, fix $i \in \{1, \ldots, d\}$. Then $d(x, y) \leq d_i(x_i, y_i) \leq d_i(x_i, z_i) + d_i(z_i, y_i)$ by (M3) for d_i. This being true for all i, we may take the maximum value on the right-hand side. This yields (M3) for d. □

Example 1.1.11. The n-fold product of $\mathbb{K}$ with itself is $\mathbb{K}^n$ endowed with the metric d_∞, defined by $d_\infty(x, y) = \max\{|x_i - y_i| \,:\, i = 1, \ldots, n\}$.

1.2 Normed Spaces

An important example of metric spaces is given by *normed spaces*. In the case of normed spaces, the set M is assumed to be a *vector space*.

In what follows, $\mathbb{K}$ denotes either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$. We recall that a vector space E over $\mathbb{K}$ is a set E together with two maps $+ : E \times E \to E$ (addition) and $\cdot : \mathbb{K} \times E \to E$ (scalar multiplication) such that $(E, +)$ is an Abelian group, i.e.

(A1) There exists an element $\mathbf{0}$, such that $x + \mathbf{0} = x$ for all $x \in E$.

(A2) For all $x \in E$, there exists an element $-x \in E$ such that $x + (-x) = \mathbf{0}$.

(A3) For all $x, y, z \in E$, we have $(x + y) + z = x + (y + z)$.

(A4) For all $x, y \in E$, we have $x + y = y + x$.

and such that the scalar multiplication satisfies

(S1) For all $x \in E$ and $\lambda, \mu \in \mathbb{K}$, we have $(\lambda + \mu) \cdot x = (\lambda \cdot x) + (\mu \cdot x)$ and $(\lambda\mu) \cdot x = \lambda \cdot (\mu x)$

(S2) For all $\lambda \in \mathbb{K}$ and $x, y \in E$, we have $\lambda \cdot (x + y) = (\lambda \cdot x) + (\lambda \cdot y)$.

This set of axioms has several consequences, most notably, the neutral element $\mathbf{0}$ from (A1) is unique and for every $x \in E$ its inverse element $-x$ from (A2) is unique and equal to $(-1) \cdot x$. For the proof and further properties, we refer to standard literature on linear algebra.

In what follows, we will denote scalar multiplication by mere concatenation and write λx rather than $\lambda \cdot x$. As is customary, we will insist that scalar multiplications are carried out before additions, thus $\lambda x + y$ should be interpreted as $(\lambda x) + y$ rather than $\lambda(x + y)$. Finally, we will write $x - y := x + (-y)$.

Important examples of vector spaces are the spaces $\mathbb{K}^d$, endowed with component-wise addition and scalar multiplication, i.e.

$$(x_1, \ldots, x_d) + (y_1, \ldots, y_d) = (x_1 + y_1, \ldots, x_d + y_d) \quad \text{and} \quad \lambda(x_1, \ldots, x_d) = (\lambda x_1, \ldots, \lambda x_d)$$

for all $x, y \in \mathbb{K}^d$ and $\lambda \in \mathbb{K}$.

Another example is the set ℓ^0 from Example 1.1.8, also endowed with component-wise addition and scalar multiplication.

Definition 1.2.1. Let E be a vector space over $\mathbb{K}$. A *norm* on E is a map $\|\cdot\| : E \to [0, \infty)$ such that

(N1) $\|x\| = 0$ implies $x = \mathbf{0}$. (definiteness)

(N2) $\|\lambda x\| = |\lambda| \cdot \|x\|$ for all $x \in E$ and $\lambda \in \mathbb{K}$. (homogenity)

(N3) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in E$. (triangle inequality)

A *normed space* is a pair $(E, \|\cdot\|)$, where E is a vector space and $\|\cdot\|$ is a norm on E.

Proposition 1.2.2. *Let $(E, \|\cdot\|)$ be a normed space and define $d : E \times E \to [0, \infty)$ by $d(x, y) := \|x - y\|$. Then (E, d) is a metric space.*

Proof. If $d(x,y) = \|x - y\| = 0$, then $x - y = \mathbf{0}$ by (N1) and thus $x = y$. This proves (M1).

Now let $x, y \in E$. Then

$$d(x,y) = \|x - y\| = \| - (y - x)\| \overset{(N2)}{=} | - 1| \cdot \|y - x\| = \|y - x\| = d(y,x)\,,$$

proving (M2).

Finally, for $x, y, z \in E$, we have

$$d(x,y) = \|x - y\| = \|(x - z) + (z - y)\| \overset{(N3)}{\leq} \|x - z\| + \|z - y\| = d(x,z) + d(z,y)\,,$$

proving (M3). □

Definition 1.2.3. The metric d is called the *metric induced by* $\| \cdot \|$.

Remark 1.2.4. In what follows, we will define more terms for metric spaces. If we later apply these to normed spaces, it is tacitly understood that we consider the space as a metric space with respect to the induced metric.

Example 1.2.5. The metric d_∞ from Example 1.1.11 is induced by the norm $\|x\|_\infty := \max\{|x_j| \,:\, j = 1, \ldots, d\}$. We leave it to the reader to verify that $\| \cdot \|_\infty$ is indeed a norm.

Example 1.2.6. Let $1 \leq p < \infty$. Then $\| \cdot \|_p$, defined by

$$\|x\|_p := \Big(\sum_{j=1}^{d} |x_j|^p \Big)^{\frac{1}{p}}$$

is a norm on $\mathbb{K}^d$.

Proof. If $\|x\| = 0$, then $\|x\|_p^p = \sum_{j=1}^{d} |x_j|^p = 0$. Since the summands are nonnegative, it follows that $|x_j|^p = 0$ and hence, since $t \mapsto t^{\frac{1}{p}}$ is an injective map from $[0, \infty)$ to itself, that $|x_j| = 0$ for all $j = 1, \ldots, d$. This proves (N1).

Now let $x \in \mathbb{K}^d$ and $\lambda \in \mathbb{K}$. Then

$$\|\lambda x\|_p = \Big(\sum_{j=1}^{d} |\lambda x_j|^p \Big)^{\frac{1}{p}} = \Big(|\lambda|^p \sum_{j=1}^{d} |x_j|^p \Big)^{\frac{1}{p}} = |\lambda| \Big(\sum_{j=1}^{d} |x_j|^p \Big)^{\frac{1}{p}} = |\lambda| \cdot \|x\|_p\,,$$

proving (N2).

If $p = 1$, (N3) follows directly from the triangle inequality in $\mathbb{K}$. In the case where $p \in (1, \infty)$ the triangle inequality (N3) is called *Minkowski's inequality.* It's proof is given in the theorem below. □

Theorem 1.2.7. *Let $x, y \in \mathbb{K}^d$. Then*

(a) (Hölder's inequality)

For $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, where $\frac{1}{\infty} := 0$, we have

$$\sum_{j=1}^{d} |x_j y_j| \leq \|x\|_p \|y\|_q\,.$$

(b) *(Minkowski's inequality)*

For $p \in [1, \infty]$ *we have* $\|x + y\|_p \le \|x\|_p + \|y\|_p$.

Proof. We only give proofs in the case where $p, q \in (1, \infty)$. The proofs of the (easier) cases where p, q might be 1 or ∞ are left to the reader.

We start with a preliminary result:

If $a, b \in [0, \infty)$, then $a^{\frac{1}{p}} b^{\frac{1}{q}} \le \frac{a}{p} + \frac{b}{q}$.

To see this, consider the function $f : [0, \infty) \to [0, \infty)$, given by $f(t) = t^{\frac{1}{p}} b^{\frac{1}{q}} \big(\frac{t}{p} + \frac{b}{q}\big)^{-1}$. We note that $f(0) = 0$ and $\lim_{t\to\infty} f(t) = 0$. It follows that f is bounded and thus has a global maximum point. If $f(t^*) = \max\{f(t) : t \ge 0\}$, then, by calculus, $f'(t^*) = 0$. A direct computation shows that $f'(t) = 0$ if and only if $t = b$. But $f(b) = 1$, hence $f(a) \le 1$ which is equivalent with the statement above.

(a) To prove Hölder's inequality, first observe that if $x = 0$ or $y = 0$, then both sides in Hölder's inequality are 0, so there is nothing to prove.

We may thus assume that $x \neq 0$ and $y \neq 0$, so that $\|x\|_p \neq 0$ and $\|y\|_p \neq 0$. By the preliminary result,

$$\frac{|x_j|}{\|x\|_p} \frac{|y_j|}{\|y\|_q} \le \frac{1}{p} \frac{|x_j|^p}{\|x\|_p^p} + \frac{1}{q} \frac{|y_j|^q}{\|y\|_q^q}$$

for all $j \in \{1, \dots, d\}$. Summing up, it follows that

$$\frac{1}{\|x\|^{\frac{1}{p}} \|y\|^{\frac{1}{q}}} \sum_{j=1}^{d} |x_j y_j| \le \frac{1}{p\|x\|_p^p} \sum_{j=1}^{d} |x_j|^p + \frac{1}{q\|x\|_q^q} \sum_{j=1}^{d} |y_j|^q = \frac{1}{p} + \frac{1}{q} = 1\,.$$

Multiplying with $\|x\|^{\frac{1}{p}} \|y\|^{\frac{1}{q}}$, Hölder's inequality follows.

(b) To prove Minkowski's inequality, we assume that $\|x + y\|_p \neq 0$, since otherwise there is nothing to prove.

We note that $\frac{1}{p} + \frac{1}{q} = 1$ exactly for $q = (1 - \frac{1}{p})^{-1} = \frac{p}{p-1}$. Hence, by Hölder's inequality,

$$\sum_{j=1}^{d} |x_j| \cdot |x_j + y_j|^{p-1} \le \Big(\sum_{j=1}^{d} |x_j|^p\Big)^{\frac{1}{p}} \Big(\sum_{j=1}^{d} |x_j + y_j|^{(p-1)\cdot\frac{p}{p-1}}\Big)^{\frac{p-1}{p}}$$

and the same result holds with x and y interchanged. Consequently,

$$\sum_{j=1}^{d} |x_j + y_j|^p \le \sum_{j=1}^{d} |x_j| \cdot |x_j + y_j|^{p-1} + \sum_{j=1}^{d} |y_j| \cdot |x_j + y_j|^{p-1} \le (\|x|_p + \|y\|_p) \|x + y\|_p^{p-1}\,.$$

Dividing by $\|x + y\|_p^{p-1}$, Minkowski's inequality follows.

□

Exercise 1.2.8. Sketch the sets $\{x \in \mathbb{R}^2 : \|x\|_j = 1\}$ for $j = 1, 2, \infty$. These sets are called the *unit spheres* of the norms $\|\cdot\|_1, \|\cdot\|_2$ and $\|\cdot\|_\infty$ respectively. Also try to figure out how the unit spheres of the norms $\|\cdot\|_p$ for general $1 < p < \infty$ look like.

Definition 1.2.9. Let M be a set and d_1, d_2 be metrics on M. We say that d_1 *dominates* d_2 if there exists $a > 0$ such that $d_2(x, y) \leq a \cdot d_1(x, y)$ for all $x, y \in M$. We say that d_1 and d_2 are *equivalent*, if d_1 dominates d_2 and d_2 dominates d_1.

Now let E be a vector space over $\mathbb{K}$ and $\|\cdot\|_1, \|\cdot\|_2$ be norms on E. We say that $\|\cdot\|_1$ *dominates* $\|\cdot\|_2$ (or that $\|\cdot\|_1$ *is stronger than* $\|\cdot\|_2$) if there exists a constant $a > 0$ such that $\|x\|_2 \leq a\|x\|_1$. We say that $\|\cdot\|_1$ and $\|\cdot\|_2$ *are equivalent* if $\|\cdot\|_1$ dominates $\|\cdot\|_2$ and $\|\cdot\|_2$ dominates $\|\cdot\|_1$.

Remark 1.2.10. Let E be a vector space over $\mathbb{K}$,$\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on E and d_1 resp. d_2 be the induced metrics. It is easy to see that $\|\cdot\|_1$ dominates (is equivalent with) $\|\cdot\|_2$ if and only if d_1 dominates (is equivalent with) d_2.

Exercise 1.2.11. Let $1 \leq p, q \leq \infty$. Prove that $\|\cdot\|_p$ and $\|\cdot\|_q$ are equivalent norms on $\mathbb{R}^d$.

We now give further examples:

Example 1.2.12. Consider the metric space (ℓ^0, d_0) from example 1.1.8. Then d_0 is not induced by a norm. Indeed, if this was the case, then we must have $d_0(\mathbf{0}, \lambda\mathbf{x}) = |\lambda| d_0(\mathbf{0}, \mathbf{x})$ for all $\lambda \in \mathbb{K}$ and $\mathbf{x} \in \ell^0$. However, if we put $\mathbf{x} = (1, 1, 1, \ldots) \in \ell^0$ and $\lambda = 5$, then $d_0(\mathbf{0}, \lambda\mathbf{x}) = 1 \neq 5d_0(\mathbf{0}, \mathbf{x}) = 5 \cdot 1 = 5$.

Example 1.2.13. The spaces ℓ^p.

Given $1 \leq p < \infty$, we define ℓ^p as

$$\ell^p := \{\mathbf{x} \in \ell^0 \,:\, \sum_{k=1}^{\infty} |x_k|^p < \infty\}.$$

Then ℓ^p is a vector space and $\|\mathbf{x}\|_p := \big(\sum_{k=1}^{\infty} |x_j|^p\big)^{\frac{1}{p}}$ defines a norm on ℓ^p.

Similarly,

$$\ell^\infty := \{\mathbf{x} \in \ell^0 \,:\, \sup_{k \in \mathbb{N}} |x_k| < \infty\}.$$

is a vector space and $\|\mathbf{x}\|_\infty := \sup_{k \in \mathbb{N}} |x_j|^\infty$ defines a norm on ℓ^∞.

Proof. We leave the cases $p = 1$ and $p = \infty$ to the reader.

To begin with, we note that $\sum_{k=1}^{\infty} |x_k|^p = \sup_{n \in \mathbb{N}} \sum_{k=1}^{n} |x_k|^p$. In particular, $\mathbf{x} \in \ell^0$ belongs to ℓ^p if and only if the sequence $\Big(\sum_{k=1}^{n} |x_k|^p\Big)_{n \in \mathbb{N}}$ is bounded.

By linear algebra, ℓ^p is a vector space if and only if it is a subspace of ℓ^0, i.e. for all $\mathbf{x}, \mathbf{y} \in \ell^p$ and $\alpha, \beta \in \mathbb{K}$ also the vector $\alpha\mathbf{x} + \beta\mathbf{y} \in \ell^p$.

Thus, to prove that ℓ^p is a vector space, let $\mathbf{x}, \mathbf{y} \in \ell^p$ and $\alpha, \beta \in \mathbb{K}$ be given. By Theorem 1.2.7 (b), for any $n \in \mathbb{N}$ we have

$$\begin{aligned}\Big(\sum_{k=1}^{n} |\alpha x_k + \beta y_k|^p\Big)^{\frac{1}{p}} &\leq \Big(\sum_{k=1}^{n} |\alpha x_k|^p\Big)^{\frac{1}{p}} + \Big(\sum_{k=1}^{n} |\beta y_k|^p\Big)^{\frac{1}{p}} \\ &= |\alpha|\Big(\sum_{k=1}^{n} |x_k|^p\Big)^{\frac{1}{p}} + |\beta|\Big(\sum_{k=1}^{n} |y_k|^p\Big)^{\frac{1}{p}} \\ &\leq |\alpha| \cdot \|\mathbf{x}\|_p + |\beta| \cdot \|\mathbf{y}\|_p < \infty.\end{aligned}$$

By the observation above, $\alpha\mathbf{x} + \beta\mathbf{y} \in \ell^p$. Moreover, letting $n \to \infty$, (N3) follows by setting $\alpha = \beta = 1$ and (N2) follows by setting $y = \mathbf{0}$, observing that in this case the first inequality is in fact an equality.

(N1) follows from the fact that if $\|\mathbf{x}\|_p = 0$, then $\left(\sum_{k=1}^n |x_k|^p\right)^{\frac{1}{p}} = 0$ for all $n \in \mathbb{N}$. By (N1) for the norm $\|\cdot\|_p$ on $\mathbb{R}^n$, this yields $x_1 = \cdots = x_n = 0$. Since n was arbitrary, $\mathbf{x} = 0$. □

The spaces ℓ^p and ℓ^q are related as follows:

Proposition 1.2.14. *For $1 \le p < q \le \infty$ we have $\ell^p \subset \ell^q$ and $\|\cdot\|_p$ dominates $\|\cdot\|_q$ on ℓ^p, more precisely, $\|\mathbf{x}\|_p \le \|\mathbf{x}\|_\infty$ for all $\mathbf{x} \in \ell^p$.*

Proof. We first consider the case where $q = \infty$. Clearly, $|x_k| = (|x_k|^p)^{\frac{1}{p}} \le \|\mathbf{x}\|_p$ for all k. Hence also $\sup_{k\in\mathbb{N}} |x_k| \le \|\mathbf{x}\|_p$. This proves that $\mathbf{x} \in \ell^\infty$ and $\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_p$.

Now, let $q < \infty$. We first assume additionally that $\|\mathbf{x}\|_p = 1$. By the above, $|x_k| \le \|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_p = 1$ for all $k \in \mathbb{N}$. Since $t^q \le t^q$ for $p < q$ and $0 \le t \le 1$, it follows that $|x_k|^q \le |x_k|^p$ for all $k \in \mathbb{N}$. Hence

$$\sum_{k=1}^\infty |x_k|^q \le \sum_{k=1}^\infty |x_k|^p = \|\mathbf{x}\|_p^p = 1\,.$$

Taking powers $\frac{1}{q}$ on both sides, $\|\mathbf{x}\|_q \le 1$ follows.

Now let $\mathbf{x} \neq 0$ (for $\mathbf{x} = 0$, there is nothing to prove). Put $\mathbf{y} = \|x\|_p^{-1}\mathbf{x}$. Then $\|\mathbf{y}\|_p = 1$ and thus, by the above, $\|\mathbf{x}\|_p^{-1} \cdot \|\mathbf{x}\|_q = \|\mathbf{y}\|_q \le 1$ which is equivalent with the statement. □

Example 1.2.15. Consider on ℓ^∞ the metrics d_0 and d_∞. Then d_∞ dominates d_0 but d_∞ and d_0 are not equivalent.

Proof. Let $\mathbf{x}, \mathbf{y} \in \ell^\infty$. Since $d_\infty(\mathbf{x}, \mathbf{y}) = \sup\{|x_j - y_j| \,:\, j \in \mathbb{N}\}$, we have

$$d_0(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^\infty \frac{\min\{|x_k - y_k|, 1\}}{2^{-j}} \le \sum_{k=1}^\infty \frac{\min\{d_\infty(\mathbf{x}, \mathbf{y}), 1\}}{2^{-j}} \le d_\infty(\mathbf{x}, \mathbf{y})\,,$$

proving that d_∞ dominates d_0. On the other hand, picking $\mathbf{x} = (0, 0, \ldots)$ and $\mathbf{y}_n := (0, \ldots, 0, n, 0, 0, \ldots)$ where the n is at n-th position, we have $d_0(\mathbf{x}, \mathbf{y}_n) = 2^{-n} \le 1$ for all n whereas $d_\infty(\mathbf{x}, \mathbf{y}_n) = n$ for all $n \in \mathbb{N}$. Therefore, there cannot be any constant $a > 0$ such that $n = d_\infty(\mathbf{x}, \mathbf{y}_n) \le a d_0(\mathbf{x}, \mathbf{y}_n) = 2^{-n}$ for all $n \in \mathbb{N}$. □

1.3 Convergent Sequences and Continuous Functions

We now generalize the concept of *convergence*, which lies at the heart of analysis, to the setting of metric spaces.

Definition 1.3.1. Let (M, d) be a metric space. A sequence $(x_n)_{n\in\mathbb{N}}$ in M is said to *converge to $x \in M$*, if for all $\varepsilon > 0$ there exists a number $n_0 \in \mathbb{N}$ such that $d(x_n, x) \le \varepsilon$ for all $n \ge n_0$.

If there exists an element $x \in M$ so that $(x_n)_{n\in\mathbb{N}}$ converges to x then the sequence $(x_n)_{n\in\mathbb{N}}$ is called *convergent*. In this case, x is called its *limit*. Notation: $\lim_{n\to\infty} x_n = x$ or $x_n \to x$.

The following lemma shows that a sequence has at most one limit. Therefore, if (x_n) is a convergent sequence, we are justified to speak of *the* limit of the sequence (x_n).

Lemma 1.3.2. *Let (M, d) be a metric space and (x_n) be a sequence which converges to x and to y. Then $x = y$.*

Proof. Let $\varepsilon > 0$. By assumption, there exist $n_0, n_1 \in \mathbb{N}$ such that $d(x_n, x) \le \varepsilon$ for all $n \ge n_0$ and $d(x_n, y) \le \varepsilon$ for all $n \ge n_1$. Pick $n \ge \max\{n_0, n_1\}$. Then

$$d(x,y) \overset{(N3)}{\le} d(x,x_n) + d(x_n,y) \overset{(N2)}{=} d(x_n,x) + d(x_n,y) \le \varepsilon + \varepsilon = 2\varepsilon$$

by the choice of n. Since $\varepsilon > 0$ was arbitrary, $d(x,y) = 0$. Now (N1) implies that $x = y$. □

Remark 1.3.3. Intuitively speaking, a sequence x_n converges to x if and only if it eventually comes arbitrary close to x. Here, closeness is measured in terms of the distance d.

To be more precise, this idea is formalized in the definition as follows. Given $\varepsilon > 0$ (this is a given threshold of how "close" we want to get to x), we find an $n_0 \in \mathbb{N}$ so that starting from n_0 (i.e. *eventually*) all members of the sequence have distance at most ε from x. Since $\varepsilon > 0$ was arbitrary, the sequence gets arbitrary close to x.

Remark 1.3.4. Let (M, d) be a metric space, (x_n) be a sequence in M and $x \in M$. Then $x_n \to x$ in (M, d) if and only if $d(x_n, x) \to 0$ in $\mathbb{R}$.

Example 1.3.5. Consider the space $(\mathbb{K}^d, \|\cdot\|_p)$ for some $1 \le p \le \infty$. Then a sequence $x_n = (x_1^{(n)}, \ldots, x_d^{(n)})$ converges to $x = (x_1, \ldots, x_d)$ if and only if $x_j^{(n)}$ converges to x_j in $\mathbb{K}$ for all $j = 1, \ldots, d$.

Proof. It follows from Exercise 1.2.11 that x_n converges to x in $(\mathbb{K}^d, \|\cdot\|_p)$ if and only if x_n converges to x in $(\mathbb{K}^d, \|\cdot\|_\infty)$. Indeed, by that exercise there exist $\alpha, \beta > 0$ such that

$$\alpha\|y\|_\infty \le \|y\|_p \le \beta\|y\|_\infty \quad \forall\, y \in \mathbb{R}^d.$$

Now assume that $x_n \to x$ in $(\mathbb{K}^d, \|\cdot\|)$. Then, given $\varepsilon > 0$, there is $n_0 \in \mathbb{N}$ such that $\|x_n - x\|_p \le \alpha\varepsilon$ for all $n \ge n_0$. But then $\|x_n - x\|_\infty \le \alpha^{-1}\|x_n - x\|_p \le \varepsilon$ for all $n \ge n_0$, proving that x_n converges to x in $(\mathbb{R}^d, d_\infty)$.

Conversely, if $x_n \to x$ in $(\mathbb{K}^d, \|\cdot\|_\infty)$, given $\varepsilon > 0$ we find $n_0 \in \mathbb{N}$ such that $\|x_n - x\|_\infty \le \beta^{-1}\varepsilon$ for all $n \ge n_0$. But then $\|x_n - x\|_p \le \varepsilon$ for all $n \ge n_0$, proving that x_n converges to x in $(\mathbb{K}^d, \|\cdot\|_p)$.

By this preliminary observation, it suffices to prove that $x_n \to x$ in $(\mathbb{K}^d, \|\cdot\|_\infty)$ if and only if $x_j^{(n)}$ converges to x_j for all $j = 1, \ldots, d$.

To see this, first assume that $x_n \to x$ in $(\mathbb{K}^d, \|\cdot\|_\infty)$ and fix $j \in \{1, \ldots, d\}$. Then $|x_j^{(n)} - x_j| \le \|x_n - x\|_\infty$, as the latter is the maximum of $|x_j^{(n)} - x_j|$ over $j = 1, \ldots, d$.

By assumption, given $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that $\|x_n - x\|_\infty \le \varepsilon$ for all $n \ge n_0$. But then the above inequality ensures that for such n also $|x_j^{(n)} - x_j| \le \varepsilon$, proving that $x_j^{(n)} \to x_j$ in $\mathbb{K}$.

Conversely assume that $x_j^{(d)}$ converges to x_j in $\mathbb{K}$ for all $j = 1, \ldots, d$. Given $\varepsilon > 0$ we find $n_j \in \mathbb{N}$ such that $|x_j^{(n)} - x_j| \le \varepsilon$ for all $n \ge n_j$. Putting $n_0 := \max\{n_1, \ldots, n_d\}$ we have that for $n \ge n_0$ we have $|x_j^{(n)} - x_j| \le \varepsilon$ for *all* $j = 1, \ldots, d$. Thus we also have $\|x_n - x\|_\infty = \max\{d_2(x_j^{(n)}, x_j) \,:\, j = 1, \ldots, d\} \le \varepsilon$ for such n. This shows that $x_n \to x$ in $(\mathbb{K}^d, \|\cdot\|_\infty)$. □

Exercise 1.3.6. Prove that $\mathbf{x}_n \to \mathbf{x}$ in (ℓ^0, d_0) if and only if $x_k^{(n)} \to x_k$ in $\mathbb{K}$ for all $k \in \mathbb{N}$.

Similarly as in Example 1.3.5, one proves the following

Lemma 1.3.7. *Let M be a set and d_1, d_2 be metrics on M such that d_1 dominates d_2. Then $x_n \to x$ in (M, d_1) implies that $x_n \to x$ in (M, d_2).*

In particular, if d_1 and d_2 are equivalent, then any sequence x_n converges to x in (M, d_1) if and only if it converges to x in (M, d_2).

Example 1.3.8. Let $1 \leq p \leq \infty$. If $\mathbf{x}_n \to \mathbf{x}$ in $(\ell^p, \|\cdot\|_p)$, then $x_j^{(n)} \to x_j$ in $\mathbb{K}$ for all $j \in \mathbb{N}$. Indeed, this follows from the inequality

$$|x_j^{(n)} - x_j| = \big(|x_j^{(n)} - x_j|^p\big)^{\frac{1}{p}} \leq \big(\sum_{j=1}^{\infty} |x_j^{(n)} - x_j|^p\big)^{\frac{1}{p}} = \|\mathbf{x}_n - \mathbf{x}\|_p \,.$$

However, putting $\mathbf{e}_n := (0, \ldots, 0, 1, 0, \ldots)$, where the 1 is at position n, we have that $e_j^{(n)} \to 0$ for all $j \in \mathbb{N}$ (for fixed j, given $\varepsilon > 0$ we have $|e_j^{(n)} - 0| = 0 \leq \varepsilon$ for all $n \geq j =: n_0$). Nevertheless $\|\mathbf{e}_n - \mathbf{0}\|_p = \|\mathbf{e}_n\|_p = 1$ for all $n \in \mathbb{N}$, proving that $\mathbf{e}_n$ does not converge to $\mathbf{0}$ in $(\ell^p, \|\cdot\|_p)$. This shows that on ℓ^p the metric d_0 is not equivalent to the one induced by $\|\cdot\|_p$.

Definition 1.3.9. Let (M_i, d_i) be metric spaces for $i = 1, 2$ and $f : M_1 \to M_2$ be a function. For $x_0 \in M_i$, f is called *continuous at* x_0 if for all sequences (x_n) in M_1 which converge to x_0, the sequence $(f(x_n))$ converges to $f(x_0)$ in M_2.

f is called *continuous* if f is continuous at every point $x_0 \in M_1$.

Example 1.3.10. As a typical example of a function which is *not* continuous, consider the price of sending a letter within Germany as a function of the weight of the letter. Measuring "weight" in grams and "price" in Euro, this function is given by $f : (0, 1000] \to \mathbb{R}$ (the restriction of the domain is due to the fact that Deutsche Post considers anything heavier than 1000 grams not to be a letter)

$$f(x) = \begin{cases} 0.55, & \text{for } 0 < x \leq 20 \\ 0.90, & \text{for } 20 < x \leq 50 \\ 1.45, & \text{for } 50 < x \leq 500 \\ 2.20, & \text{for } 500 < x \leq 1000 \end{cases}$$

and letting $n \to \infty$. Here, of course, $(0, 1000]$ is endowed with the metric induced by the euclidean metric on $\mathbb{R}$. To see that this function is not continuous, observe that $x_n = 50 + \frac{1}{n}$ converges to $x = 50$, whereas $f(x_n) \equiv 1.45$ does not converge to $0.90 = f(x)$.

Exercise 1.3.11. Let (M_j, d_j) for $j \in \{1, \ldots, n\}$ be metric spaces. Prove that $x_n \to x$ in $M_1 \times \cdots \times M_n$, endowed with the product metric, if and only if $x_j^{(n)} \to x_j$ for all $j \in \{1, \ldots, n\}$.

Conclude that if (M, d) is a metric space, then $f = (f_1, \ldots, f_n) : M \to M_1 \times \cdots \times M_n$ is continuous if and only if $f_j : M \to M_j$ is continuous for all $j \in \{1, \ldots, n\}$.

We give some examples of continuous functions:

Proposition 1.3.12. *Let (M, d) be a metric space. Then d is continuous as a function from $M \times M$ (with respect to the product metric) to $\mathbb{R}$.*

Proof. By Exercise 1.3.11, $(x_n, y_n) \to (x, y)$ in $M \times M$ if and only if $x_n \to x$ and $y_n \to y$ in M.

Now let sequences $x_n \to x$ and $y_n \to y$ be given. We have to prove that $d(x_n, y_n) \to d(x, y)$. By (M3), we have

$$d(x_n, y_n) \leq d(x_n, x) + d(x, y) + d(y, y_n)$$

and hence

$$d(x_n, y_n) - d(x, y) \leq d(x_n, x) + d(y, y_n)\,.$$

Similarly, one sees that

$$d(x, y) - d(x_n, y_n) \leq d(x_n, x) + d(y, y_n)\,,$$

so that, combined, it follows that

$$|d(x, y) - d(x_n, y_n)| \leq d(x_n, x) + d(y, y_n)\,.$$

Since the right-hand side converges to 0 as $n \to \infty$, so does the left-hand side, proving that $d(x_n, y_n) \to d(x, y)$ as claimed. □

Proposition 1.3.13. *Let $(E, \|\cdot\|)$ be a normed space. Then the maps*

$$\mathsf{add} : E \times E \to E \quad (x, y) \mapsto x + y$$

and

$$\mathsf{mult} : \mathbb{K} \times E \to E \quad (\alpha, x) \mapsto \alpha x$$

are continuous. Here, $\mathbb{K}$ is endowed with the euclidean metric and product spaces are endowed with the product topology.

Proof. Let $x_n \to x$ and $y_n \to y$. Then

$$\|\mathsf{add}(x_n, y_n) - \mathsf{add}(x, y)\| = \|x_n + y_n - (x + y)\| \overset{(N3)}{\leq} \|x_n - x\| + \|y_n - y\| \to 0$$

as $n \to \infty$.

Now let $x_n \to x$ and $\alpha_n \to \alpha$. Then $|\alpha_n|$ is bounded, say by C. Thus

$$\begin{aligned}\|\mathsf{mult}(\alpha_n, x_n) - \mathsf{mult}(\alpha, x)\| &= \|\alpha_n x_n - \alpha x\| \leq \|\alpha_n x_n - \alpha_n x\| + \|\alpha_n x - \alpha_x\| \\ &\overset{(N2)}{\leq} C\|x_n - x\| + |\alpha_n - \alpha|\|x\| \to 0\end{aligned}$$

as $n \to \infty$. □

Proposition 1.3.14. *Let (M_i, d_i) be a metric space for $i = 1, 2, 3$ and $f : M_1 \to M_2$ and $g : M_2 \to M_3$ be continuous. Then the composition $g \circ f : M_1 \to M_3$, defined by $g \circ f(x) = g(f(x))$ is continuous.*

Proof. Let $x_n \to x$. Then $f(x_n) \to f(x)$ since f is continuous. But then also $g(f(x_n)) \to g(f(x))$ since g is continuous. □

Theorem 1.3.15. *Let (M, d) be a metric space, $(E, \|\cdot\|)$ be a normed space. Then the set $C(M; E)$ of continuous functions from M to E is a vector space with respect to pointwise addition and scalar multiplication.*

If $(E, \|\cdot\|) = (\mathbb{K}, d_2)$, then $C(M) := C(M, \mathbb{K})$ is even an algebra, i.e. in addition to the vector-space structure it also has a multiplicative structure in that with f and g also the function $f \cdot g : x \mapsto f(x)g(x)$ is continuous.

Proof. Let $f, g \in C(M; E)$. By Exercise 1.3.11, the map $\Phi : M \to E \times E$, given by $\Phi(w) = (f(w), g(w))$ is continuous. Since $\mathsf{add} : E \times E \to E$ is continuous by Proposition 1.3.13, it follows from Proposition 1.3.14 that $f + g = \mathsf{add} \circ \Phi$ is continuous and hence an element of $C(M; E)$.

Using mult instead of add, one sees that scalar multiples of continuous functions and, in the case where $E = \mathbb{K}$, products of continuous functions, are continuous. □

Example 1.3.16. As a corollary of Theorem 1.3.15, we see that polynomials are continuous functions. Indeed, the function $id : \mathbb{R} \to \mathbb{R}$, defined by $id(x) = x$ is obviously continuous. Since the continuous functions are an algebra, it follows by induction that the monomials $p_n : \mathbb{R} \to \mathbb{R}$, defined by $p_n(x) = x^n$ are continuous. Finally, linear combinations of these functions, and those are exactly the polynomials, are continuous.

1.4 Open and Closed Sets

We next introduce the concepts of open and closed sets. These are so-called "topological" concepts and appear naturally when extending concepts such as continuity beyond the setting of metric spaces. Nevertheless, they are also important concepts for metric spaces.

Definition 1.4.1. Let (M, d) be a metric space.

1. For $x \in M$ and $\varepsilon > 0$, $B(x, \varepsilon)$ denotes the set $\{y \in M : d(x, y) < \varepsilon\}$. $B(x, \varepsilon)$ is called the *open ball of radius ε, centered at x.*

2. For $x \in M$ and $\varepsilon > 0$, $\bar{B}(x, \varepsilon)$ denotes the set $\{y \in M : d(x, y) \le \varepsilon\}$. $B(x, \varepsilon)$ is called the *closed ball of radius ε, centered at x.*

3. A set $O \subset M$ is called *open*, if for every $x \in O$ there exists an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset O$.

4. A set $C \subset M$ is called *closed*, $C^c := M \setminus C$ is open.

Let us show that the names "open ball" and "closed ball" are justified.

Example 1.4.2. For $x \in M$ and $\varepsilon > 0$, the set $B(x, \varepsilon)$ is open and the set $\bar{B}(x, \varepsilon)$ is closed.

Proof. Let $y \in B(x, \varepsilon)$. If $x = y$, then $B(y, \varepsilon) \subset B(x, \varepsilon)$. Now assume that $x \neq y$ so that $d(x, y) > 0$ by (M1).

Put $\delta := \varepsilon - d(x, y) > 0$. We claim that $B(y, \delta) \subset B(x, \varepsilon)$. To see this, let $z \in B(y, \delta)$. Then $d(x, z) \le d(x, y) + d(y, z) < d(x, y) + \delta = \varepsilon$, proving that $y \in B(x, \varepsilon)$. This proves that $B(x, \varepsilon)$ is open.

To prove that $\bar{B}(x, \varepsilon)$ is closed, let $y \in \bar{B}(x, \varepsilon)^c$. Then $d(x, y) > \varepsilon$. Put $\delta := \frac{1}{2}(d(x, y) - \varepsilon) > 0$. Then $B(y, \delta) \subset \bar{B}(x, \varepsilon)^c$. Indeed, let $z \in B(y, \delta)$. Aiming for a contradiction, suppose that $z \in \bar{B}(x, \varepsilon)$ so that $d(x, z) \le \varepsilon$. Then $d(x, y) \le d(x, z) + d(z, y) \le \varepsilon + \frac{1}{2}(d(x, y) - \varepsilon)$, which yields that $d(x, y) \le \varepsilon$ and hence $y \in \bar{B}(x, \varepsilon)$ — a contradiction. Hence $B(y, \delta) \subset \bar{B}(x, \varepsilon)^c$, proving that $\bar{B}(x, \varepsilon)^c$ is open. □

Proposition 1.4.3. *Let (M, d) be a metric space and let $\mathcal{O}$ denote the collection of all open subsets of M. Then*

(O1) $\emptyset \in \mathcal{O}$ *and* $M \in \mathcal{O}$.

(O2) If $U_1, \dots, U_n \in \mathcal{O}$, then also $\bigcap_{j=1}^n U_j \in \mathcal{O}$.

(O3) If A is an index set and $U_\alpha \in \mathcal{O}$, then also $\bigcup_{\alpha\in A} U_\alpha \in \mathcal{O}$.

Proof. (O1) is obvious. For (O2), let $x \in \bigcap_{j=1}^n U_j$. Since U_j is open, there exist $\varepsilon_1, \dots, \varepsilon_n$ such that $B(x, \varepsilon_j) \subset U_j$. Put $\varepsilon := \min\{\varepsilon_1, \dots, \varepsilon_n\} > 0$. Then $B(x,\varepsilon) \subset \bigcap_{j=1}^n B(x,\varepsilon_j) \subset \bigcap_{j=1}^n U_j$, proving that the latter set contains an open ball centered at x. Since $x \in \bigcap_{j=1}^n U_j$ was arbitrary, (O2) follows.

Finally, let $x \in \bigcup_{\alpha\in A} U_\alpha$. Then there exists α_0 such that $x \in U_{\alpha_0}$. Since U_{α_0} is open, there exists an $\varepsilon > 0$ such that $B(x,\varepsilon) \subset U_{\alpha_0} \subset \bigcup_{\alpha\in A} U_\alpha$. This proves that $\bigcup_{\alpha\in A} U_\alpha$ is open. □

Corollary 1.4.4. *Let (M, d) be a metric space and let $\mathcal{C}$ denote the collection of all closed subsets of M. Then*

(C1) $\emptyset \in \mathcal{C}$ and $M \in \mathcal{C}$.

(C2) If $F_1, \dots, F_n \in \mathcal{C}$, then also $\bigcap_{j=1}^n F_j \in \mathcal{C}$.

(C3) If A is an index set and $F_\alpha \in \mathcal{C}$, then also $\bigcap_{\alpha\in A} F_\alpha \in \mathcal{C}$.

Proof. Since closed sets are exactly the complements of open sets, (C1) follows from (O1) and the fact that $\emptyset^c = M$ and $M^c = \emptyset$, whereas (C2) and (C3) follow from (O2) and (O3) and DeMorgan's law

$$\bigcup_{\beta\in B} S_\beta^c = \Big(\bigcap_{\beta\in B} S_\beta\Big)^c \qquad \bigcap_{\beta\in B} S_\beta^c = \Big(\bigcap_{\beta\in B} S_\beta\Big)^c$$

for all index sets B and sets $S_\beta \subset M$. □

Exercise 1.4.5. In $\mathbb{R}$, find a sequence of open sets whose intersection is not open and a sequence of closed sets whose union is not closed.

Exercise 1.4.6. Show that if d_1 and d_2 are equivalent metrics on the set M, then U is open in (M, d_1) if and only if U is open in (M, d_2).

Show that $d(x, y) := |\arctan x - \arctan y|$ defines a metric on $\mathbb{R}$ which produces the same open sets as the euclidean metric d_2 and show that d and d_2 are not equivalent.

Hint: You may use the fact that the functions arctan and tan are continuous for the euclidean metric.

Exercise 1.4.7. In ℓ^∞, consider the set

$$B = \{\mathbf{x} \in \ell^\infty \,:\, |x_j| < 1 \;\forall\, j \in \mathbb{N}\}\,.$$

Decide whether B is open if ℓ^∞ is endowed with (i) the metric d_0 and (ii) the metric induced by the norm $\|\cdot\|_\infty$.

Definition 1.4.8. Let (M, d) be a metric space and $A \subset M$.

1. The *interior of* A, denoted by A°, is the union of all open sets contained in A.
2. The *closure of* A, denoted by $\bar{A}$, is the intersection of all closed sets which contain A.
3. The *boundary of* A, denoted by ∂A, is the set $\bar{A} \cap \overline{A^c}$.
4. The set A is called *dense* in M if $\bar{A} = M$.

5. (M, d) is called *separable* if there exists a countable set A which is dense in M.

We note that by (O3) A° is an open set and by (C3) $\bar{A}$ is a closed set for any $A \subset M$. We next give a convenient criterion to determine A° and $\bar{A}$.

Lemma 1.4.9. *Let (M, d) be a metric space and $A \subset M$. Then*

1. *$x \in A^\circ$ if and only if there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset A$.*

2. *$x \in \overline{A}$ if and only if there exists a sequence $(x_n) \subset A$ which converges to x.*

Proof. (1) If $x \in A^\circ$, then there exists an open set U which contains x and is itself contained in A. Since U is open, there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U \subset A$.

Conversely, if $x \in A$ is such that $B(x, \varepsilon) \subset A$ for some $\varepsilon > 0$, then $B(x, \varepsilon)$ is, by Example 1.4.2, an open set containing x. Consequently, x is contained in the union of all open sets contained in A, i.e. A°.

(2) Let $x \in \bar{A}$. We claim that for every $n \in \mathbb{N}$ the ball $B(x, n^{-1})$ intersects A. Indeed, if this was false, then $B(x, n^{-1})^c$ would be a closed set containing A but not x. But then $\bar{A}$ would not contain x — a contradiction.

Thus for every $n \in \mathbb{N}$, there exists an $x_n \in A$ with $d(x_n, x) < n^{-1}$. It follows that the sequence $(x_n) \subset A$ converges to x. Indeed, given $\varepsilon > 0$, there exists an $n_0 \in \mathbb{N}$ such that $n_0^{-1} < \varepsilon$. It follows that for $n \geq n_0$ we have $d(x_n, x) < n^{-1} \leq n_0^{-1} < \varepsilon$.

Conversely, suppose that (x_n) is a sequence in A converging to $x \in M$ and let C be a closed set containing A. Suppose that $x \in C^c$. Since C^c is open, there exists an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset C^c$. But then $x_n \notin B(x, \varepsilon)$ for all $n \in \mathbb{N}$ — a contradiction. It follows that every closed set C which contains A also contains x, hence $\overline{A}$ contains x. ☐

Definition 1.4.10. Let A be a subset of the metric space (M, d). A point $x \in M$ is called *accumulation point of* A if there exists a sequence $(x_n) \subset A \setminus \{x\}$ which converges to x.

Corollary 1.4.11. *Let (M, d) be a metric space. A set A is closed if and only if A contains all its accumulation points.*

Proof. Obviously, $A \subset \bar{A}$, and A is closed if and only if $A = \bar{A}$ and hence if and only if $\bar{A} \setminus A \subset A$. But, by Lemma 1.4.9, the former set consists of accumulation points of A. ☐

Let us now give some examples of interior and closure of sets

Example 1.4.12. Let $(E, \|\cdot\|)$ be a normed space and consider E with the induced metric. Then $\overline{B(x, \varepsilon)} = \bar{B}(x, \varepsilon)$, $\bar{B}(x, \varepsilon)^\circ = B(x, \varepsilon)$.

The corresponding equalities are *not* true in a general metric space.

Proof. Clearly $\overline{B(x, \varepsilon)} \subset \bar{B}(x, \varepsilon)$ as the latter is a closed set (Example 1.4.2) containing $B(x, \varepsilon)$. Now, if $y \in \bar{B}(x, \varepsilon)$, then $x_n := x + (1 - 1/n)(y - x) \in B(x, \varepsilon)$ (since $d(x, x_n) = (1-1/n)\|y-x\| \leq (1-1/n)\varepsilon < \varepsilon$) and $x_n \to y$ (since $d(y, x_n) = n^{-1}\|y-x\| \to 0$ as $n \to \infty$). Hence $y \in \overline{B(x, \varepsilon)}$ by Lemma 1.4.9

On the other hand, $B(x, \varepsilon)$ is an open set (also by Example 1.4.2) contained in $\bar{B}(x, \varepsilon)$ and hence in the interior of that set. Moreover, if $y \in \bar{B}(x, \varepsilon) \setminus B(x, \varepsilon)$ then every $B(y, \delta)$ contains an element of $B(x, \varepsilon)^c$, for example $x + (1 + \delta/2\varepsilon)(y - x)$. Thus no open set contained in $\bar{B}(x, \varepsilon)$ can contain y. This proves that $y \notin \bar{B}(x, \varepsilon)^\circ$.

To see that the equalities are not true in a general metric space, consider $M = \{1, 2\}$ with the discrete metric. Then $B(1, 1) = \{1\} = \overline{\{1\}} \neq \bar{B}(1, 1) = \{1, 2\}$ and $\bar{B}(1, 1) = \{1, 2\} = \{1, 2\}^\circ \neq B(1, 1) = \{1\}$. ☐

Exercise 1.4.13. Let $(E, \|\cdot\|)$ be a normed space and $W \subset E$ be a subspace of E. Show that also $\overline{W}$ is a subspace of E. Is also W° a subspace of E?

We now continue with examples of separable and not separable spaces.

Example 1.4.14. Since every real number is the limit of a sequence of rational numbers (this follows from the fact that every real number has a decimal expansion), and since the rational numbers are countable, $\mathbb{R}$ is separable.

Example 1.4.15. For $1 \le p < \infty$, the normed space $(\ell^p, \|\cdot\|_p)$ is separable, the normed space $(\ell^\infty, \|\cdot\|_\infty)$ is not separable.

Proof. We only consider the case $\mathbb{K} = \mathbb{R}$.

First let $1 \le p < \infty$ and put

$$A := \{\mathbf{x} \in \ell^\infty \,:\, \exists\, n \in \mathbb{N}\,,\, q_1, \ldots, q_n \in \mathbb{Q} \text{ such that } \mathbf{x} = (q_1, \ldots, q_n, 0, 0, \ldots)\}\,.$$

Then A is a countable set.

For every $x \in \mathbb{R}$, there exists a sequence $(\alpha_n) \subset \mathbb{Q}$ which converges to x. We claim that this implies that

$$\{\mathbf{x} \in \ell^p \,:\, \exists\, k \in \mathbb{N} \text{ such that } x_j = 0\ \forall\, j \ge k\} \subset \bar{A}\,.$$

Indeed, for $\mathbf{x} = (x_1, \ldots, x_k, 0 \ldots)$, we can find sequences $(\alpha_j^{(n)}) \subset \mathbb{Q}$ converging to x_j for $j = 1, \ldots, k$. Then $\mathbf{x}_n = (\alpha_1^{(n)}, \ldots, \alpha_k^{(n)}, 0, \ldots)$ is an element of A and $\|\mathbf{x} - \mathbf{x}_n\|_p = \big(\sum_{j=1}^k |x_j - \alpha_j^{(n)}|^p\big)^{\frac{1}{p}} \to 0$ by Example 1.3.5. It follows from Lemma 1.4.9 that $x \in \bar{A}$.

Now let $\mathbf{x} \in \ell^p$ and put $\mathbf{x}_n := (x_1, x_2, x_n, 0, 0, \ldots)$. Then

$$\|x - x_n\|_p^p \le \sum_{k=n}^\infty |x_k|^p \to 0$$

as $n \to \infty$. Thus, using Lemma 1.4.9 again, it follows that $x \in \bar{A}$. Since $x \in \ell^p$ was arbitrary, $\bar{A} = \ell^p$.

Next consider $(\ell^\infty, \|\cdot\|_\infty)$. Given a set $N \subset \mathbb{N}$, we define $\mathbf{x}_N := (x_j^N)_{j\in\mathbb{N}}$ where $x_j^N = 1$ if $j \in N$ and $x_j^N = 0$ if $j \ne N$. Then $\|\mathbf{x}_N - \mathbf{x}_M\|_\infty = 1$ whenever $N \ne M$. Noting that there are uncountably many subsets of $\mathbb{N}$, it follows that every dense subset of $(\ell^\infty, \|\cdot\|_\infty)$ must be uncountable as well. Indeed, if A is a dense subset of $(\ell^\infty, \|\cdot\|_\infty)$, given $N \subset \mathbb{N}$ we choose an element $\mathbf{a}_N \in A \cap B(\mathbf{x}_N, 2^{-1})$. It follows that $\mathbf{a}_N \ne \mathbf{a}_M$ for $N \ne M$, whence $\Phi : \mathcal{P}(\mathbb{N}) \to A$, given by $\Phi(N) := \mathbf{a}_N$ is an injection, proving that A is uncountable. □

The proof of separability in Example 1.4.15 can actually be applied in many more situations. The general result is presented below. In the statement, span$S$ refers to the linear span of the set S, i.e. the set (actually, vector space)

$$\Big\{\sum_{j=1}^n \lambda_j x_j \,:\, n \in \mathbb{N} \,:\, \lambda_j \in \mathbb{K}\,,\, x_j \in S \text{ for } 1 \le j \le n\Big\}$$

and $\overline{\text{span}}S$ denotes the closure of that set.

Proposition 1.4.16. *Let $(E, \|\cdot\|)$ be a normed space. Then E is separable if and only if there exists a countable set S such that $\overline{\text{span}}S = E$.*

Proof. If E_0 countable and dense, then we may take $S = E_0$ as $E = \overline{E_0} \subset \overline{\text{span}} E_0 = E$.

Conversely, let S be countable such that span$S$ is dense. For $\mathbb{K}_{\mathbb{Q}} := \mathbb{Q}$ if $\mathbb{K} = \mathbb{R}$ and $\mathbb{K}_{\mathbb{Q}} := \{q + ir \,:\, q, r \in \mathbb{Q}\} = \mathbb{Q} + i\mathbb{Q}$, define

$$E_0 := \Big\{ \sum_{j=1}^{n} \alpha_j x_j \,:\, n \in \mathbb{N}\,,\, \alpha_j \in \mathbb{K}_{\mathbb{Q}}\,,\, x_j \in S_0 \Big\}\,.$$

Then E_0 is countable. Moreover, by the continuity of add and mult, it follows that span$S$, and hence E, is contained in the closure of E_0. Thus E_0 is dense. □

We can now give an equivalent description of continuity in terms of open sets. We note that in the following proposition the function $f : M_1 \to M_2$ is not assumed to be bijective; by $f^{-1}(S)$ we denote the *pre-image* of the set S, i.e. the set of all $x \in M_1$ with $f(x) \in S \subset M_2$.

Proposition 1.4.17. *For $i = 1, 2$, let (M_i, d_i) be metric spaces and let $\mathcal{O}_i$ resp. $\mathcal{C}_i$ denote the collection of open resp. closed sets in (M_i, d_i). For a function $f : M_1 \to M_2$, the following are equivalent:*

(a) f is continuous.

(b) $f^{-1}(F) \in \mathcal{C}_1$ for all $F \in \mathcal{C}_2$.

(c) $f^{-1}(U) \in \mathcal{O}_1$ for all $U \in \mathcal{O}_2$.

(d) For every $x \in M_1$ and all $\varepsilon > 0$ there exists $\delta > 0$ such that $d_1(x, y) < \delta$ implies that $d_2(f(x), f(y)) < \varepsilon$.

Moreover, f is continuous at x_0 if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $d_1(x_0, y) < \delta$ implies that $d_2(f(x), f(y)) < \varepsilon$.

Proof. (a) ⇒ (b) Let f be continuous and F be a closed subset of M_2. Suppose that $f^{-1}(F)$ is not closed. As a consequence of Corollary 1.4.11 there exists a sequence (x_n) in $f^{-1}(F)$ converging to a point $x \notin f^{-1}(F)$. It follows that $f(x_n) \in F$ for all $n \in \mathbb{N}$ whereas $f(x) \notin F$. However, since f is continuous, $f(x_n) \to f(x)$. This shows that $f(x)$ is an accumulation point of F not contained in F, contradicting the closedness of F. Thus $f^{-1}(F)$ must be closed.

(b) ⇒ (c) If $U \in \mathcal{O}_2$, then $U^c \in \mathcal{C}_2$. By (b), $f^{-1}(U)^c = f^{-1}(U^c) \in \mathcal{C}_1$, that is, $f^{-1}(U)^c$ is closed, hence $f^{-1}(U)$ is open.

(c) ⇒ (d) Let $x \in M_1$ and $\varepsilon > 0$. Then $B(f(x), \varepsilon) \in \mathcal{O}_2$ by Example 1.4.2, hence $f^{-1}(B(f(x), \varepsilon))$ is open by (c). Thus there is a $\delta > 0$ such that $B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon))$. This shows that the statement of (d) is true for this δ.

(d) ⇒ (a) Let $x_n \to x$ and $\varepsilon > 0$ be given. Pick a $\delta > 0$ such that (d) holds for the given x and ε. Since $x_n \to x$, there exists an n_0 such that $d_1(x_n, x) < \delta$ for all $n \geq n_0$. But then $d_2(f(x_n), f(x)) < \varepsilon$ for all $n \geq n_0$. Since $\varepsilon > 0$ was arbitrary, this proves that $f(x_n) \to f(x)$.

Let us now prove the addendum. Assume that f is continuous at x_0 and, aiming for a contradiction, that there exists some $\varepsilon > 0$ such that for every $\delta > 0$ there exists a y_δ with $d_1(x_0, y_\delta) < \delta$ but $d_2(f(x_0), f(y_\delta)) \geq \varepsilon$. Then the sequence $x_n := y_{n^{-1}}$ converges to x_0, since $d_1(x_0, x_n) < n^{-1}$, but $d_2(f(x_0), f(x_n)) \geq \varepsilon$, hence $f(x_n) \not\to f(x_0)$, a contradiction to the continuity at x_0.

The converse implication follows as in the proof of (d) ⇒ (a). □

Corollary 1.4.18. *If d_1, d_2 are equivalent metrics on M and ρ_1, ρ_2 are equivalent metrics on N, then a function $f : M \to N$ is continuous as a function from (M, d_1) to (N, ρ_1) if and only if it is continuous as a function from (M, d_2) to (N, ρ_2).*

Proof. Combine Proposition 1.4.17 with Exercise 1.4.6. □

1.5 Complete Metric Spaces

The definition of convergence has a single, yet crucial, shortcoming:

In order to show that $x_n \to x$, we have to know the limit x in advance. However, frequently, one wants to *define* an object as the limit of some sequence. Prominent examples are irrational numbers, such as $\sqrt{2}, \pi, e$ which one wants to define as the limit of a sequence of rational numbers.

Definition 1.5.1. Let (M, d) be a metric space. A sequence $(x_n) \subset M$ is called a *Cauchy sequence* if for all $\varepsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq n_0$.

The space (M, d) is called *complete* if every Cauchy sequence converges.

A complete normed space is called a *Banach space*

In the definition of Cauchy sequence, we require that the elements of a sequence come arbitrarily close to each other, rather than to a limit. Thus one can verify that a sequence is a Cauchy sequence *without* having to know any limit in advance. If the metric space in question is complete, we can then conclude that the sequence does converge.

Exercise 1.5.2. Let $(E, \|\cdot\|)$ be a normed space and x_k be a sequence in E. Then the partial sums $s_n := \sum_{k=1}^{n} x_k$ is well-defined. If the sequence s_n is convergent, then we say that the series $\sum_{k=1}^{\infty} x_k$ is convergent. If the (real-valued) series $\sum_{k=1}^{\infty} \|x_k\|$ converges, we say that the series $\sum_{k=1}^{\infty} x_k$ is absolutely convergent.

Show that $(E, \|\cdot\|)$ is a Banach space if and only if every absolutely convergent series is convergent.

We next collect some easy properties of Cauchy sequences. We recall that a *subsequence* of a sequence $(x_n)_{n\in\mathbb{N}}$ is a sequence $(x_{\varphi(k)})_{k\in\mathbb{N}}$, where $\varphi : \mathbb{N} \to \mathbb{N}$ is strictly increasing. We often write (x_{n_k}) instead of $(x_{\varphi(k)})$.

We also recall that a subset A of a metric space (M, d) is called *bounded*, if there is an $x \in M$ and an $r > 0$ such that $A \subset B(x, r)$. We note that if $(E, \|\cdot\|)$ is a normed space, then a subset A is bounded if and only if $A \subset B(\mathbf{0}, r)$ for some $r > 0$, if and only if $\sup_{x\in A} \|x\| < \infty$.

Lemma 1.5.3. *Let (M, d) be a metric space and (x_n) be a sequence in M.*

(a) If (x_n) converges, then (x_n) is a Cauchy sequence. Moreover, every subsequence of (x_n) converges to the same limit as (x_n).

(b) If (x_n) is a Cauchy sequence, then (x_n) converges if and only if (x_n) has a convergent subsequence.

(c) If (x_n) is a Cauchy sequence, then $\{x_n : n \in \mathbb{N}\}$ is bounded. In particular, if M is a vector space and d is induced by a norm $\|\cdot\|$, then $\|x_n\|$ is bounded.

Proof. (a) Let $x_n \to x$. Then, by (M3), $d(x_n, x_m) \le d(x_n, x) + d(x, x_m) \le \varepsilon$ for all $n, m \ge n_0$, where we have picked n_0 such that $d(x_n, x) \le \varepsilon/2$ for all $n \ge n_0$. Thus (x_n) is a Cauchy sequence.

To see the second part, let (x_{n_k}) be a subsequence of (x_n). Given $\varepsilon > 0$, there exists n_0 such that $d(x_n, x) \le \varepsilon$ for all $n \ge n_0$. Since $k \mapsto n_k$ is strictly increasing, there exists an index k_0 such that $n_k \ge n_0$ for all $k \ge k_0$. But then $d(x_{n_k}, x) \le \varepsilon$ for all $k \ge k_0$, proving that $x_{n_k} \to x$.

(b) Let (x_n) be a Cauchy sequence and assume that $x_{n_k} \to x$. We prove that $x_n \to x$. to that end, let $\varepsilon > 0$ be given. We find $n_0 \in \mathbb{N}$ such that $d(x_n, x_m) \le \varepsilon/2$ for all $n, m \ge n_0$. Moreover, we find k_0 such that $d(x, x_{n_k}) \le \varepsilon/2$ for all $k \ge k_0$. Possibly taking a larger k_0, we may assume that $n_k \ge n_0$ for all $k \ge k_0$. Then for $n \ge 0$ we find by inserting an x_{n_k} for $k \ge k_0$ that $d(x_n, x) \le d(x_n, x_{n_k}) + d(x_{n_k}, x) \le \varepsilon$, i.e. $x_n \to x$ as claimed.

The converse is trivial.

(c) Pick n_0 such that $d(x_n, x_m) \le 1$ for all $n, m \ge n_0$. Put

$$r := 1 + \max\{1, d(x_j, x_{n_0}) \,:\, j = 1, \ldots, n_0 - 1\}\,.$$

Then $\{x_n \,:\, n \in \mathbb{N}\} \subset B(x_{n_0}, r)$, as is easy to see. □

We also note the following equivalent description of convergent sequences.

Lemma 1.5.4. *Let (M, d) be a metric space. Then $x_n \to x$ if and only if every subsequence of x_n has a subsequence converging to x.*

Proof. If x_n converges to x then, by Lemma 1.5.3, every subsequence converges also to x.

Conversely, assume that every subsequence of x_n has a subsequence which converges to x. Suppose that x_n does not converge to x. Then there exists $\varepsilon > 0$ and a subsequence x_{n_k} such that $d(x_{n_k}, x) \ge \varepsilon$ for all $k \in \mathbb{N}$. By assumption, for a suitable subsequence $x_{n_{k_l}}$, we have $x_{n_{k_l}} \to x$ and hence $d(x_{n_{k_l}}, x) \to 0$ — a contradiction. □

We give some examples of complete metric spaces.

Example 1.5.5. The following metric spaces are complete: $\mathbb{K}, (\mathbb{K}^d, \|\cdot\|_p), (\ell^p, \|\cdot\|_p)$ for $1 \le p \le \infty$ and (ℓ^0, d_0).

Proof. The completeness of $\mathbb{R}$ is part of the definition of the real numbers. This completeness is exploited in the proof of completeness of the other spaces.

Let us first consider $(\mathbb{R}^d, \|\cdot\|_p)$. If (x_n) is a Cauchy sequence in $(\mathbb{R}, \|\cdot\|_p)$ then, as a consequence of Exercise 1.2.11, (x_n) is a Cauchy sequence in $(\mathbb{R}^d, \|\cdot\|_\infty)$. It thus follows that $x_j^{(n)}$ is a Cauchy sequence in $\mathbb{R}$ for all $j \in \{1, \ldots, d\}$. Consequently, there exist $x_1, \ldots, x_d \in \mathbb{R}$ such that $x_j^{(n)} \to x_j$ for $j = 1, \ldots, d$. Using Exercise 1.2.11 again it follows that $x_n \to x = (x_1, \ldots, x_d)$ in $(\mathbb{R}^d, \|\cdot\|_p)$.

Since $(\mathbb{C}, |\cdot|)$ is "essentially" the same as $(\mathbb{R}^d, \|\cdot\|_2)$, this also shows that $(\mathbb{C}, |\cdot|)$ is complete. Now repeat the above to show that $(\mathbb{C}^d, \|\cdot\|_p)$ is complete.

In view of Exercise 1.3.6, the above proof can also be adopted to prove the completeness of (ℓ^0, d_0). We leave the details to the reader.

As for $(\ell^p, \|\cdot\|_p)$, the estimate $|x_j - y_j| \le \|\mathbf{x} - \mathbf{y}\|_p$ shows that if $(\mathbf{x}_n)$ is a Cauchy sequence in $(\ell^p, \|\cdot\|_p)$, then $(x_j^{(n)})$ is a Cauchy sequence in $\mathbb{K}$ for all $j \in \mathbb{N}$. We claim that $\mathbf{x} := (x_j)_{j\in\mathbb{N}} \in \ell^p$ and that $\mathbf{x}_n \to \mathbf{x}$ in $(\ell^p, \|\cdot\|_p)$.

To see this, first consider the case where $p = \infty$. Observe that $\|\mathbf{x}_n\|_\infty$ is bounded, say by C. But then it follows that $|x_j^{(n)}| \le C$ for all $j, n \in \mathbb{N}$ and consequently $\mathbf{x} \in \ell^\infty$ with $\|\mathbf{x}\|_\infty \le C$. Now let $\varepsilon > 0$ be given. Fix $j \in \mathbb{N}$ and pick $n_0 \in \mathbb{N}$ such that $\|\mathbf{x}_n - \mathbf{x}_m\| \le \varepsilon$ for all $n, m \ge n_0$. It follows that $|x_j^{(n)} - x_j^{(m)}| \le \|\mathbf{x}_n - \mathbf{x}_m\|_\infty \le \varepsilon$ for all $n, m \ge n_0$. Upon $m \to \infty$ it follows that $|x_j^{(n)} - x_j| \le \varepsilon$ for all $n \ge n_0$. Since $j \in \mathbb{N}$ was arbitrary, it follows that $\|\mathbf{x}_n - \mathbf{x}\|_\infty \le \varepsilon$ for all $n \ge n_0$. This proves that $\mathbf{x}_n \to \mathbf{x}$ in $(\ell^\infty, \|\cdot\|_\infty)$.

Now let $1 \le p < \infty$. By Minkowski's inequality, for every $N \in \mathbb{N}$, we have

$$\Big(\sum_{j=1}^N |x_j|^p\Big)^{\frac{1}{p}} \le \Big(\sum_{j=1}^N |x_j - x_j^{(n)}|^p\Big)^{\frac{1}{p}} + \|\mathbf{x}_n\|_p$$

Since $\mathbf{x}_n$ is a Cauchy sequence, $\|\mathbf{x}_n\|_p$ is bounded, say by C. Since $x_j^{(n)} \to x_j$ for all j, we have $\big(\sum_{j=1}^N |x_j - x_j^{(n)}|^p\big)^{\frac{1}{p}} \le 1$ if n is chosen large enough. Hence the above shows that $\Big(\sum_{j=1}^N |x_j|^p\Big)^{\frac{1}{p}} \le C + 1$ for all $N \in \mathbb{N}$. But this implies that $\mathbf{x} \in \ell^p$. To see that $\mathbf{x}_n \to \mathbf{x}$ in $(\ell^p, \|\cdot\|_p)$, let $\varepsilon > 0$ be given and pick $n_0 \in \mathbb{N}$ such that $\|\mathbf{x}_n - \mathbf{x}_m\|_p \le \varepsilon$ for all $n, m \ge \varepsilon$. Fix $N \in \mathbb{N}$. Then

$$\Big(\sum_{j=1}^N |x_j^{(n)} - x_j^{(m)}|^p\Big)^{\frac{1}{p}} \le \|\mathbf{x}_n - \mathbf{x}_m\|_p \le \varepsilon \quad \forall\, n, m \ge n_0\,.$$

Upon $m \to \infty$, it follows that

$$\Big(\sum_{j=1}^N |x_j^{(n)} - x_j|^p\Big)^{\frac{1}{p}} \le \varepsilon \quad \forall\, n \ge n_0\,.$$

Since N was arbitrary, $N \to \infty$ yields $\|\mathbf{x}_n - \mathbf{x}\|_p \le \varepsilon$ for all $n \ge n_0$. □

The following lemma describes when a subset of a complete metric space is itself complete in the induced metric.

Lemma 1.5.6. *Let (M, d) be a complete metric space. Then (M_0, d_{M_0}) is complete if and only if M_0 is closed in M.*

Proof. First assume that (M_0, d_{M_0}) is complete. Let x be an accumulation point of M_0 in M. Then there exists a sequence (x_n) in M_0, convergent to x. It follows that (x_n) is a Cauchy sequence in (M, d) and hence, since the sequence lies in M_0, in (M_0, d_{M_0}). By completeness, x_n converges to some y in the space (M_0, d). But then it also converges to y in (M, d). By Lemma 1.3.2, $x = y \in M_0$. Hence M_0 contains all its accumulation points and is thus closed by Corollary 1.4.11

Conversely, assume that M_0 is closed in M and let (x_n) be a Cauchy sequence in (M_0, d_{M_0}). Then it is a Cauchy sequence in (M, d), hence convergent to some $x \in M$. But since M_0 is closed, $x \in M_0$ and it follows that x_n converges to x in (M_0, d_{M_0}). This proves that (M_0, d_{M_0}) is complete. □

Exercise 1.5.7. Let $c := \{\mathbf{x} \in \ell^\infty : \lim_{j\to\infty} x_j \text{ exists}\}$ and $c_0 := \{\mathbf{x} \in \ell^\infty : \lim_{j\to\infty} x_j = 0\}$. Show that $(c, \|\cdot\|_\infty)$ and $(c_0, \|\cdot\|_\infty)$ are complete metric spaces.

Also show that (ℓ^∞, d_0) is not complete.

1.6 Compact Sets

We now introduce a concept which produces convergent sequences through a different mechanism. We actually introduce different notions which, as we shall see in Theorem 1.6.8, are all related to the same concept.

Definition 1.6.1. Let (M, d) be a metric space. A subset K of M is called

1. *sequentially compact* if every sequence in K has a subsequence converging to an element of K.
2. *compact* if every open cover of K has a finite subcover, i.e. if $(U_i)_{i\in I}$ is a family of open sets with $K \subset \bigcup_{i\in I} U_i$, then there exist $i_1, \ldots, i_n$ with $K \subset \bigcup_{k=1}^n U_{i_k}$.
3. *totally bounded*, if, given $\varepsilon > 0$ there exist $x_1, \ldots, x_n \in K$ such that $K \subset \bigcup_{j=1}^n B(x_j, \varepsilon)$.

Let us illustrate these notions in some examples:

Example 1.6.2. Let $x_n \to x$. Then the set $K := \{x_n : n \in \mathbb{N}\} \cup \{x\}$ is compact. Indeed, if $(U_i)_{i\in I}$ is an open covering of K, then there is some i_0 such that $x \in U_{i_0}$. Since U_{i_0} is open, there is an $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U_{i_0}$. Since $x_n \to x$, there exists an n_0 such that $x_n \in B(x, \varepsilon) \subset U_{i_0}$ for all $n > n_0$. Picking $i_1, \ldots, i_{n_0}$ such that $x_j \in U_{i_j}$ for $j = 1, \ldots, n_0$, we see that $U_{i_0}, \ldots, U_{i_{n_0}}$ is a finite subcovering of K.

Example 1.6.3. The set $(0, 1]$ is not compact in $\mathbb{R}$. Indeed if $U_n := (n^{-1}, 2)$, then $(U_n)_{n\in\mathbb{N}}$ is an open covering of $(0, 1]$ which has no finite subcovering.

Example 1.6.4. If K is a closed and bounded subset of $\mathbb{R}$ then it is sequentially compact.

Indeed, since K is bounded, there exist $-\infty < a < b < \infty$ such that $K \subset [a, b]$. Now let a sequence x_n in K be given. Then either $I_{1,1} := [a, a + \frac{1}{2}(b - a)]$ or $I_{1,2} := [a + \frac{1}{2}(b - a), b]$ contains infinitely many of the x_n. We call that interval I_1 and define the subsequence $x_{n,1}$ to consist of the elements of x_n which lie in I_1

We now iterate this procedure. Suppose we have already constructed intervals I_k, for $1 \le k \le N$ of length $(b - a)2^{-k}$ with $I_{k+1} \subset I_k$ for all $1 \le k \le N - 1$ and subsequences $(x_{n,k})_{n\in\mathbb{N}}$ such that $(x_{n,k})_{n\in\mathbb{N}} \subset I_k$ for $1 \le k \le N$ and $(x_{n,k+1})$ is a subsequence of $(x_{n,k})$ for all $1 \le k \le N - 1$.

Suppose $I_N = [a', b']$. We then put $I_{N+1,1} := [a', a' + \frac{1}{2}(b' - a')]$ and $I_{N+1,2} := [a' + \frac{1}{2}(b' - a'), b']$. Then one of those intervals contains infinitely many of the $(x_{n,k})_{n\in\mathbb{N}}$. We call that interval I_{N+1} and define the subsequence $(x_{n,k+1})_{n\in\mathbb{N}}$ to consist of those members of $(x_{n,k})_{n\in\mathbb{N}}$ which lie in I_{N+1}.

We now define the subsequence $(x_{n_k})_{k\in\mathbb{N}}$ by $x_{n_k} := x_{k,k}$. Then $x_{n_k} \in I_N$ for all $k \ge N$. In particular, (x_{n_k}) is a Cauchy sequence, since $|x_{n_k} - x_{n_l}| \le 2^{-N}(b - a)$ for all $k, l \ge N$. Since $\mathbb{R}$ is complete, x_{n_k} converges to some element x of $\mathbb{R}$. Since K is closed, we must have $x \in K$.

Example 1.6.5. Let K be a bounded and closed subset of $(\mathbb{R}^d, \|\cdot\|_p)$ (for $1 \le p \le \infty$). Then K is sequentially compact.

Indeed, if (x_n) is a sequence in K then (x_n) is bounded. But then, by equivalence of norms, (x_n) is also bounded with respect to $\|\cdot\|_\infty$ and it follows that all components $(x_j^{(n)})$ are bounded. By Example 1.6.4, that bounded sequence on the real line have convergent subsequences. Thus for a subsequence n_{k_1}, the first component of x_n converges. Passing to a further subsequence n_{k_2}, we can assume that the first and the second component converges. Passing to further and further subsequences, we arrive at a subsequence n_{k_d} such that all d components of x_n converge. We denote the vector of the respective limits by x. By Example 1.2.11, $x_{n_{k_d}}$ converges to x in $(\mathbb{R}^d, \|\cdot\|_p)$. Since K is closed, it follows that $x \in K$.

Example 1.6.6. The set $K := \{\mathbf{x} \in \ell^p : \|\mathbf{x}\|_p \leq 1\}$ is bounded and closed in $(\ell^p, \|\cdot\|_p)$ but not sequentially compact. Indeed $\mathbf{x}_n := (0, \ldots, 0, 1, 0, \ldots)$ is a sequence in K which, since $d_p(x_n, x_m) \geq 1$ for all $n \neq m$, has no convergent subsequence.

Example 1.6.7. The set $\mathbb{N}$ is not totally bounded in $(\mathbb{R}, d_2)$.

We now characterize compact subsets of metric spaces.

Theorem 1.6.8. *Let (M, d) be a metric space and $K \subset M$. The following are equivalent.*

1. *K is compact.*
2. *K is sequentially compact.*
3. *K is complete (i.e. (K, d) is complete) and totally bounded.*

Proof. (1) $\Rightarrow$ (2). Suppose that K is not sequentially compact. Then there exists a sequence (x_n) in K which has no convergent subsequence. Then $A := \{x_n : n \in \mathbb{N}\}$ is an infinite set with no accumulation points (in K). Therefore, A is closed in K. Moreover, every $y \in A$ has an open neighborhood U_y which does not intersect $A \setminus \{y\}$, since otherwise y would be an accumulation point of A. But then $K \setminus A$ together with U_y is an open cover of K which has no finite subcover (since this would imply that A is finite). This is a contradiction to (1), hence K must be sequentially compact.

(2) $\Rightarrow$ (3) By sequential compactness, every Cauchy sequence in K must have a convergent subsequence and thus, by Lemma 1.5.3 be convergent. This proves that K is complete.

Now let $\varepsilon > 0$ be given. We pick $x_1 \in K$. If $K \subset B(x_1, \varepsilon)$, we are done. Otherwise, we proceed inductively. Let $x_1, \ldots, x_m$ already be chosen. If $K \subset \bigcup_{j=1}^m B(x_j, \varepsilon)$, we are done. Otherwise, pick x_{m+1} arbitrary in $K \setminus \bigcup_{j=1}^m B(x_j, \varepsilon)$.

If this procedure does not stop, we obtain a sequence (x_m) in K such that $d(x_i, x_j) \geq \varepsilon$ for all $i \neq j$. But then this sequence cannot have a convergent subsequence, contradicting sequential compactness. Therefore at some point this procedure stops, proving that K is totally bounded.

(3) $\Rightarrow$ (1) As an intermediate step, we prove that K is sequentially compact and then prove that (2) and (3) together imply (1).

To prove sequential compactness, set a sequence (x_n) in K be given. Since K is totally bounded, we may cover K with a finite number of balls of radius 1. Then at least one ball, say B_1, contains infinitely many of the x_n. We put $J_1 := \{n \in \mathbb{N} : x_n \in B_1\}$. Using again that K is totally bounded, we may cover K with a finite number of balls of radius 2^{-1}. Then at least one ball, say B_2, contains infinitely many of the x_n for $n \in J_1$. We put $J_2 := \{n \in J_1 : x_n \in B_2\}$.

Proceeding inductively, we obtain infinite sets J_k with $J_{k+1} \subset J_k$ and Balls B_k of radius k^{-1} such that $x_n \in B_k$ for all $n \in J_k$. Now pick $n_1 \in J_1$ and then, having picked n_k, pick $n_{k+1} \in J_{k+1}$ such that $n_{k+1} > n_k$ (this is possible since J_{k+1} is infinite). It follows that $n_k \in J_{k_0}$ for all $k \geq k_0$ and thus $d(x_{n_k}, x_{n_l}) \leq k_0^{-1}$ for all $k, l \geq k_0$. Thus (x_{k_l}) is a Cauchy sequence and hence, by completeness of K, convergent.

We next prove that every open covering $(U_i)_{i\in I}$ has a *Lebesgue number* $\delta > 0$, i.e. given any set of diameter less than δ, there exists an element U_i containing it.

Suppose to the contrary that for every $n \in \mathbb{N}$ there exists a set C_n of diameter less than n^{-1} so that no U_i contains it. Pick $x_n \in C_n$. By sequential compactness, x_n has a convergent subsequence (x_{n_k}), denote the limit of this subsequence by x. Then x is contained in some U_{i_0}. Since U_{i_0} is open, it contains $B(x, \varepsilon)$ for some $\varepsilon > 0$. For large enough k, $x_{n_k} \in B(x, \varepsilon/2)$. If furthermore k is so large that $n_k^{-1} < \varepsilon/2$, then C_k, having diameter less than $\varepsilon/2$ and containing x_{n_k}, is contained in $B(x,\varepsilon)$ and hence in U_{i_0} — a contradiction.

We are now ready to finish the proof.

Let $(U_i)_{i\in I}$ be an open covering of K and let δ be its Lebesgue number. Since K is totally bounded, $K \subset \bigcup_{j=1}^n B(x_j, \delta/2)$ for certain $x_1, \ldots, x_n$. For every $j \in \{1, \ldots, n\}$, the set $B(x_j, \delta)$ having diameter less than δ, is contained in some U_{i_j}. But then $(U_{i_j})_{j=1}^n$ is a finite subcovering of K. □

Let us note some further consequences

Corollary 1.6.9. *Let (M, d) be a compact metric space, i.e. M is a compact subset of the metric space (M, d). Then*

1. *M is complete.*
2. *M is separable.*

Proof. (1) is immediate from Theorem 1.6.8(3). For (2) observe that since M is totally bounded, for every $n \in \mathbb{N}$ there exist $x_{1,n}, \ldots, x_{k_n,n}$ such that $M \subset \bigcup_{j=1}^{k_n} B(x_{j,n}, n^{-1})$. Then the set $\{x_{j,n} : n \in \mathbb{N}, 1 \le j \le k_n\}$ is countable and dense in M. □

We end this section with two results about continuous functions on compact sets:

Theorem 1.6.10. *Let K be a compact subset of the metric space (M, d) and $f : M \to \mathbb{R}$ be a continuous function, where we endow $\mathbb{R}$ with the metric d_2. Then there exist x_* and x^* in K such that $f(x_*) \le f(x) \le f(x^*)$ for all $x \in K$.*

In other words, continuous functions on a compact set attain their maximum and minimum. In particular, they are bounded.

Proof. Let $\alpha := \sup\{f(x) : x \in K\}$. Then there exist $x_n \in K$ such that $f(x_n) \uparrow \alpha$. Since K is compact, x_n has a convergent subsequence (x_{n_k}). Call its limit x^*. Since f is continuous, $f(x^*) = \lim_{k\to\infty} f(x_{n_k}) = \lim_{n\to\infty} f(x_n) = \alpha$. Therefore $f(x^*) \ge f(x)$ for all $x \in K$.

The proof of the other assertion is similar. □

Exercise 1.6.11. Prove the following more general result. If (M_i, d_i), for $i = 1, 2$ are metric spaces and $f : M_1 \to M_2$ is continuous, then, for every compact set $K \subset M_1$, we have that $f(K) \subset M_2$ is compact.

Theorem 1.6.12. *Let K be a compact subset of the metric space (M, d) and $f : K \to \mathbb{R}$ be a continuous function.*

Then f is equicontinuous on K, i.e. given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $x, y \in K$ are such that $d(x, y) < \delta$.

Proof. Suppose that f is not equicontinuous. Then there exists some $\varepsilon_0 \geq 0$ such that for every $n \in \mathbb{N}$ there exist $x_n, y_n \in M$ with $d(x_n, y_n) \leq n^{-1}$ and $|f(x_n) - f(y_n)| \geq \varepsilon_0$ for all $n \in \mathbb{N}$.

By compactness, a subsequence x_{n_k} converges, say to x. But then also y_{n_k} converges to x, since $d(x_n, y_n) \leq n^{-1} \to 0$. Since f is continuous, $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$. Since d is continuous by Proposition 1.3.12, $|f(x_{n_k}) - f(y_{n_k})| \to |f(x) - f(x)| = 0$, in contradiction to $|f(x_{n_k}) - f(y_{n_k})| \geq \varepsilon_0$ for all k. □

1.7 Spaces of Continuous Functions

Let (M, d) be a metric space. By Theorem 1.3.15 the space $C(M) := C(M, \mathbb{K})$ is a vector space. We denote by $C_b(M)$ the set of all *bounded* continuous functions from $M \to \mathbb{K}$. This is also a vector space. Note that if M is compact, then, by Theorem 1.6.10, $C(M) = C_b(M)$.

Now, for $f \in C_b(M)$, we may define $\|f\|_\infty := \sup_{x \in M} |f(x)| < \infty$, since f is bounded.

Theorem 1.7.1. $\|\cdot\|_\infty$ *is a norm on* $C_b(M)$. *With the induced metric,* $C_b(M)$ *is complete.*

Proof. It is straightforward to verify that $\|\cdot\|_\infty$ defines a norm on $C_b(M)$ and even on $\mathcal{F}_b(M)$, the space of all (not necessarily continuous) bounded Functions from M to $\mathbb{K}$. Similarly as proving that $(\ell^\infty, \|\cdot\|_\infty)$ is complete in Example 1.5.5, one proofs that $\mathcal{F}_b(M)$ is complete in the metric induced by $\|\cdot\|_\infty$.

By Lemma 1.5.6, to prove that $C_b(M)$ is complete, it suffices to prove that it is closed in $\mathcal{F}_b(M)$.

To that end, let a sequence $f_n \in C_b(M)$ be given such that f_n converges with respect to $\|\cdot\|_\infty$ to some $f \in \mathcal{F}_b(M)$. We have to prove that f is continuous. So let $x_m \to x$. Then

$$\begin{aligned} |f(x) - f(x_m)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_m)| + |f_n(x_n) + f(x_m)| \\ &\leq 2\|f - f_n\|_\infty + |f_n(x) - f_n(x_m)| \, . \end{aligned}$$

Given $\varepsilon > 0$ we may first pick $n \in \mathbb{N}$ such that $2\|f - f_n\|_\infty \leq \varepsilon/2$. Since f_n is continuous, we may then pick $m_0 \in \mathbb{N}$ such that $|f_n(x) - f_n(x_m)| \leq \varepsilon/2$ for all $m \geq m_0$. Now the above estimate yields $|f(x) - f(x_m)| \leq \varepsilon$ for all $m \geq m_0$. But this implies that f is continuous and hence an element of $C_b(M)$ as claimed. □

Remark 1.7.2. Recall that a sequence $f_n : M \to \mathbb{K}$ of functions is said to converge *uniformly* to $f : M \to \mathbb{K}$ on a set $S \subset M$, if for all $\varepsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that $|f_n(x) - f(x)| \leq \varepsilon$ for all $n \geq n_0$ *and all* $x \in S$.

Obviously, $f_n \to f$ with respect to $\|\cdot\|_\infty$ if and only if $f_n \to f$ uniformly on M.

Exercise 1.7.3. On the space $C(\mathbb{R})$ of all (not necessarily bounded) continuous functions from $\mathbb{R}$ to $\mathbb{R}$, define

$$d(f, g) := \sum_{k=1}^{\infty} \frac{\min\{1, \sup_{t \in [-k,k]} |f(t) - g(t)|\}}{2^k} \, .$$

Show that d is a norm on $C(\mathbb{R})$ and that $(C(\mathbb{R}), d)$ is complete. Furthermore, show that $f_n \to f$ with respect to d if and only if $f_n \to f$ uniformly on all compact subsets of $\mathbb{R}$.

Example 1.7.4. Let us show how completeness of $C_b(M)$ can be used to construct continuous functions.

Let (M,d) be a metric space and f_k be a sequence in $C_b(M)$ such that $\|f_k\|_\infty \le a_k$ for some numbers a_k with $\sum_{k=1}^\infty a_k < \infty$. Then the series $\sum_{k=1}^\infty f_k(x)$ converges for all $x \in M$ and $f : x \mapsto \sum_{k=1}^\infty f_k(x)$ is a bounded, continuous function.

Indeed, let $S_n := \sum_{k=1}^n f_k$. Then $S_n \in C_b(M)$ by the vector space structure. Moreover, for $m \le n$, we have

$$\|S_n - S_m\|_\infty = \Big\| \sum_{k=m+1}^n f_k \Big\|_\infty \le \sum_{k=m+1}^n \|f_k\|_\infty \le \sum_{k=m+1}^n a_k \to 0$$

as $n, m \to \infty$. This proves that S_n is a Cauchy sequence in $C_b(M)$ and thus convergent.

We know already that $C(M)$ is a vector space. Let us give some more structural properties of continuous functions. Given two functions f, g, we define their product fg pointwise. The *maximum* of f and g is denoted by $f \vee g$ and defined by $(f \vee g)(x) = \max\{f(x), g(x)\}$. The minimum is $f \wedge g$ and defined by $(f \wedge g)(x) = \min\{f(x), g(x)\}$. Finally $f^+ := f \vee \mathbf{0}$ is the *positive part* of f, $f^- := (-f) \vee \mathbf{0}$ is the negative part of f. Finally $|f| = f^+ + f^-$.

Proposition 1.7.5. *Let $\mathbb{K} = \mathbb{R}$, (M,d) be a metric space and $f, g \in C(M)$ $(C_b(M))$. Then also $fg, f \wedge g, f^+, f^-$ belong to $C(M)$ $(C_b(M))$. If $g(x) \neq 0$ for all $x \in M$, then also $\frac{f}{g}$ belongs to $C(M)$.*

Proof. The continuity of fg was already proved in Theorem 1.3.15.

f^+ is continuous as composition of f with the continuous function $\varphi : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = 0$ for $x \le 0$ and $f(x) = x$ for $x \ge 0$. Continuity of f^- and $|f|$ is immediate from $f^- = (-f)^+$ and $|f| = f^+ + f^-$.

Noting that $2(f \vee g) = (g - f)^+ + f$ and $f \wedge g = -((-g) \vee (-f))$, the continuity of these functions follow.

Finally, if $g(x) \neq 0$ for all $x \in M$, then g is continuous as a function from M to $\mathbb{K} \setminus \{0\}$. Since $\varphi : t \mapsto t^{-1}$ is a continuous function from $\mathbb{K} \setminus \{0\}$ to $\mathbb{K} \setminus \{0\}$ the composition $\varphi \circ g : x \mapsto g(x)^{-1}$ is continuous and thus also $\frac{f}{g} = f\varphi \circ g$. □

We now show that the space $C_b(M)$ is in some sense "rich", i.e. it contains many functions.

Theorem 1.7.6. *(Urysohn)*

Let (M,d) be a metric space and A, B be disjoint closed subsets of M. Then there exists a continuous function $f : M \to [0,1]$ such that $f(x) = 1$ for all $x \in A$ and $f(x) = 0$ for all $x \in B$.

Proof. We first prove a preliminary result.

For a closed set $F \subset M$, we put $d(x,F) := \inf\{d(x,y) : y \in F\}$. We claim that $x \mapsto d(x,F)$ is a continuous function with $d(x,F) = 0$ for all $x \in F$ and $d(x,F) > 0$ for all $x \in F^c$.

Let us first prove the continuity of $d(\cdot,F)$. For $x, y \in M$ and $z \in F$, we have, by the triangle inequality,

$$d(x,F) \le d(x,z) \le d(x,y) + d(y,z) \le d(x,y) + d(y,F)$$

and hence $d(x,F) - d(y,F) \le d(x,y)$. Exchanging the role of x and y, it follows that $d(y,F) - d(x,F) \le d(x,y)$ and hence $|d(x,F) - d(y,F)| \le d(x,y)$. This clearly implies that $d(\cdot,F)$ is continuous.

Obviously, $d(x, F) = 0$ if $x \in F$. Now assume that $d(x, F) = 0$. Then, by definition, there exists a sequence $x_n \in F$ with $d(x_n, x) \to d(x, F) = 0$. Hence, $x_n \to x$ and thus, since F is closed, $x \in F$. This proves that $d(x, F) > 0$ for $x \in F^c$.

The proof of the claim is now complete.

Now let disjoint, closed sets A and B be given. We define $f : M \to [0, 1]$ by

$$f(x) := \frac{d(x, A)}{d(x, A) + d(x, B)} .$$

This is well-defined since A and B are disjoint and hence, by the above, either $d(x, A)$ or $d(x, B)$ is strictly positive. Moreover, it is a continuous function. Noting that $f(x) = d(x, A)\big[d(x, A) + 0\big] = 1$ for $x \in A$ and $f(x) = 0\big[0 + d(x, B)\big]^{-1} = 0$ for $x \in B$, the proof of the theorem is complete. □

We next prove the *Stone-Weierstrass theorem* which will allow us to prove that if K is a compact metric space, then $C(K)$ is separable. Let us first show that if K is not compact, then $C(K)$ need not be separable.

Example 1.7.7. Consider $M = \mathbb{R}$. The function $f_n : x \mapsto (1 - 2|x - n|)^+$ is continuous and $f_n(x) = 0$ whenever $x \notin (n - \frac{1}{2}, n + \frac{1}{2})$. For a subset $N \subset \mathbb{N}$, we define $g_N := \sum_{j \in N} f_j$. Then $\|g_N - g_M\|_\infty = 2$ whenever $N \neq M$. As in Example 1.4.15, it follows that $C(\mathbb{R})$ is not separable.

We need some preparation

Definition 1.7.8. Let (M, d) be a metric space and $\mathcal{A}$ be a subset of $C(M)$.

(a) $\mathcal{A}$ is called unital algebra if (i) $\mathcal{A}$ is a sub vector space of $C(M)$, i.e. $f, g \in \mathcal{A}$ and $\alpha, \beta \in \mathbb{K}$ implies that $\alpha f + \beta g \in \mathcal{A}$, (ii) $\mathcal{A}$ is closed under multiplication, i.e. $f, g \in \mathcal{A}$ implies $fg \in \mathcal{A}$ and (iii) the function $\mathbb{1} : M \to \mathbb{K}$, defined by $\mathbb{1}(x) = 1$ for all $x \in M$, belongs to $\mathcal{A}$.

(b) $\mathcal{A}$ said to separate points (in M), if for all $x, y \in M$ with $x \neq y$, there exists $f \in \mathcal{A}$ with $f(x) \neq f(y)$.

(c) $\mathcal{A}$ is said to be closed under conjugation, if $f \in \mathcal{A}$ implies $\bar{f} \in A$.

(c) $\mathcal{A}$ is called a *lattice* if $f, g \in \mathcal{A}$ implies that $f \wedge g, g \vee g \in \mathcal{A}$.

Example 1.7.9. Let $K = [a, b]$ be a compact interval in $\mathbb{R}$. Then the polynomials $\mathbb{R}[t] := \{t \mapsto \sum_{k=0}^n a_k t^k \,:\, n \in \mathbb{N}\,, a_1, \ldots, a_n \in \mathbb{K}\}$ are an unital algebra which is closed under conjugation and separates points in $[a, b]$.

To see the latter, let $x \neq y$, then $f : t \mapsto t - x$ satisfies $f(x) = 0$ and $f(y) = y - x \neq 0$.

Example 1.7.10. Consider $[0, 2\pi]$. A *trigonometric polynomial* is a function $p : [0, 2\pi] \to \mathbb{C}$ of the form $p(t) = \sum_{k=-n}^n a_k e^{ikt}$ where $n \in \mathbb{N}$ and $a_1, \ldots, a_n \in \mathbb{C}$. Using that $e^{ikt}e^{ilt} = e^{i(k+l)t}$, it is easy to see that the trigonometric polynomials form a unital algebra. Since $\overline{e^{ikt}} = e^{-ikt}$, the trigonometric polynomials are closed under conjugation.

However, the trigonometric polynomials do not separate points in $[0, \pi]$, since $p(0) = p(2\pi)$ for all trigonometric polynomials. Nevertheless, the trigonometric polynomials separate points in $(0, 2\pi)$. Indeed, if $e^{it} = e^{is}$, then it follows that $t - s = 2\pi n$ for some $n \in \mathbb{Z}$. Hence, if $t, s \in (0, 2\pi)$ and $t \neq s$, then $e^{it} \neq e^{is}$.

Theorem 1.7.11. *Let (K, d) be a compact metric space and $\mathcal{A} \subset C(M)$ be an unital algebra which separates points and (in case $\mathbb{K} = \mathbb{C}$) is closed under conjugation. Then $\mathcal{A}$ is dense in $(C(K), \|\cdot\|_\infty)$.*

We prepare the proof of Theorem 1.7.11 with some lemmas. The first one is a result from calculus.

Lemma 1.7.12. *The series $1 - \sum_{k=1}^\infty (-1)^{n-1}\binom{\frac{1}{2}}{n} t^n$ converges uniformly on $[0, 1]$, its limit being $\sqrt{1-t}$.*

Sketch of proof. One verifies that the series is the Taylor series of the function $\sqrt{1-t}$. Its radius of convergence is 1, hence it converges uniformly on $[0, s]$ for every $0 < s < 1$. Having identified its limit in t as $\sqrt{1-t}$, one then shows that the series also converges for $t = 1$ and invokes Abel's theorem to infer the uniform convergence on $[0, 1]$. □

Lemma 1.7.13. *Let (K, d) be a compact metric space and $\mathcal{A} \subset C(K)$ be an unital algebra. Then $\bar{\mathcal{A}}$ is also a unital algebra. Moreover, $\bar{\mathcal{A}}$ is a lattice.*

Proof. That also $\bar{A}$ is an unital algebra follows directly from the continuity of the maps $(f, g) \mapsto f + g$ and $(f, g) \mapsto fg$.

To simplify notation, we now assume that $\mathcal{A}$ is already closed and proof that then $\mathcal{A}$ is a lattice.

Let $f \in \mathcal{A}$ with $f \geq 0$ be given. Replacing f with αf for a suitable constant α, we may assume that $0 \leq f \leq \mathbb{1}$. Put $g := \mathbb{1} - f \in \mathcal{A}$. Since $\mathcal{A}$ is an algebra,

$$h_n := \mathbb{1} - \sum_{k=1}^n (-1)^{n-1}\binom{\frac{1}{2}}{n} g^n \in \mathcal{A}\,.$$

It is a consequence of Lemma 1.7.12 that h_n converges uniformly on K to $\sqrt{\mathbb{1} - g} = \sqrt{f}$. Thus, since $\mathcal{A}$ is closed, $\sqrt{f} \in \mathcal{A}$.

It now follows that $|f| = \sqrt{f^2} \in \mathcal{A}$ and hence $f \wedge g = \frac{1}{2}(f + g - |f - g|)$ and $f \vee g = \frac{1}{2}(f + g + |f - g|)$ belong to $\mathcal{A}$, proving that $\mathcal{A}$ is a lattice. □

Proof of Theorem 1.7.11. Let us first consider the case $\mathbb{K} = \mathbb{R}$.

By Lemma 1.7.13, $\bar{\mathcal{A}}$ is an unital algebra which is also a lattice. Moreover, it separates points since $\mathcal{A}$ does.

Let $f \in C(K)$ be given. Since $\bar{\mathcal{A}}$ separates points, given $x \neq y$ we find a function $h \in \bar{\mathcal{A}}$ such that $h(x) \neq h(y)$. Define $f_{x,y} : K \to \mathbb{R}$ by

$$f_{x,y}(v) := f(x)\mathbb{1} + (f(y) - f(x))\frac{h(v) - h(x)}{h(y) - h(x)}\,.$$

Then $f_{x,y} \in \mathcal{A}$ and $f_{x,y}(x) = f(x)$ and $f_{x,y}(y) = y$.

Now let $\varepsilon > 0$ be given. For fixed x, let

$$U_x := \{v \in K \,:\, f_{x,y}(v) < f(v) + \varepsilon\}\,.$$

Then U_x is an open set since $U_x = [f_{x,y} - f]^{-1}(-\infty, \varepsilon)$, i.e. U_x is the continuous pre-image of an open set. Moreover, $x \in U_x$, since $f_{x,y}(x) = f(x) < f(x) + \varepsilon$.

It follows that $(U_x)_{x \in K}$ is an open cover of K. By compactness, there exist finitely many $x_1, \ldots, x_n$ so that $K \subset \bigcup_{k=1}^n U_{x_k}$. Define $g_y := \min\{f_{x_k,y} \,:\, 1 \leq k \leq n\}$. Then $g_y \in \bar{\mathcal{A}}$ since

$\bar{\mathcal{A}}$ is a lattice. Moreover, $g_y(y) = f(y)$ and $g_y(v) < f(v) + \varepsilon$ for all $y, v \in K$. Indeed, given $v \in K$ there exists $k \in \{1, \dots, n\}$ such that $v \in U_{x_k}$. But then $g_y(v) \leq f_{x_k,y}(v) < f(v) + \varepsilon$ by the definition of g_y and U_{x_k}.

Now put $V_y := \{u \in K \,:\, g_y(u) > f(u) - \varepsilon\}$. As above, one sees that V_y is open, contains y and $K \subset \bigcup_{y \in K} V_y$. Using compactness again, we find $y_1, \dots, y_m$ such that $K \subset \bigcup_{j=1}^m V_{y_j}$. Define $g := \max\{g_{y_j} \,:\, 1 \leq j \leq m\}$. Since $\bar{\mathcal{A}}$ is a lattice, $g \in \bar{\mathcal{A}}$. Moreover, for every $u \in K$, we have

$$f(u) - \varepsilon < g(u) < f(u) + \varepsilon\,.$$

Indeed, for $u \in K$ we find k with $u \in U_{x_k}$ and j with $u \in V_{y_j}$. Hence, picking j_* such that $g(u) = g_{y_{j_*}}(u)$, we have

$$f(u) - \varepsilon \leq g_{y_j}(u) \leq g(u) \leq g_{y_{j_*}}(u) \leq f_{x_k, y_{j_*}}(u) < f(u) + \varepsilon$$

It follows that $\|f - g\|_\infty < \varepsilon$. Hence, since $f \in C(K)$ and $\varepsilon > 0$ were arbitrary, $\bar{\mathcal{A}}$ is dense in $C(K)$. Since it is also closed, $\bar{\mathcal{A}} = C(K)$ follows.

The case $\mathbb{K} = \mathbb{C}$ can be reduced to the case $\mathbb{K} = \mathbb{R}$. Indeed, since $\mathcal{A}$ is closed under conjugation, for $f \in \mathcal{A}$ also $\operatorname{Re} f = \frac{1}{2}(f + \bar{f})$ and $\operatorname{Im} f = \frac{1}{2i}(f - \bar{f})$ belong to $\mathcal{A}$.

We denote by $\mathcal{A}_\mathbb{R}$ the unital algebra which contains all real-valued functions. Then $\mathcal{A}_\mathbb{R}$ separates points because $\mathcal{A}$ does. Hence, by the real case, $\mathcal{A}_\mathbb{R}$ is dense in $C(K, \mathbb{R})$. But since $\mathcal{A} = \{f + ig \,:\, f, g \in \mathcal{A}_\mathbb{R}\}$, it follows that $\mathcal{A}$ is dense in $C(K, \mathbb{K})$. □

Corollary 1.7.14. *If K is compact, then $C(K)$ is separable.*

Proof. By Corollary 1.6.9, K is separable, hence there exists a dense sequence $\{x_n \,:\, n \in \mathbb{N}\}$. For all $n, m \in \mathbb{N}$, the sets $\bar{B}(x_n, (2m)^{-1})$ and $B(x_n, m^{-1})^c$ are closed and disjoint. Hence, by Theorem 1.7.6, there exist continuous (real-valued) function $f_{n,m} : K \to [0, 1]$ with $f_{n,m}(x) = 0$ for all $x \in \bar{B}(x_n, (2m)^{-1})$ and $f_{n,m} = 1$ for all $x \in B(x_n, m^{-1})^c$.

We define P as the set of all finite products of functions $f_{n,m}$. This is a countable set. We then define

$$\mathcal{A} := \Big\{ \sum_{k=1}^K a_k g_k \,:\, K \in \mathbb{N}\,,\, a_k \in \mathbb{K}\,,\, g_k \in P \Big\}\,.$$

Then $\mathcal{A}$ is unital algebra which is closed under conjugation. Moreover, it separates points. Indeed if $x \neq y$, then $d(x, y) =: d > 0$. Pick m such that $(2m)^{-1} < d/4$ and then n such that $d(x, x_n) \leq (2m)^{-1}$. Then $f_{n,m}(x) = 0$ and, since $d(y, x_n) \geq d(y, x) - d(x, x_n) > 4d - (2m)^{-1} > 3/(2m) > m^{-1}$, $f_{n,m}(y) = 1$.

By Theorem 1.7.11, $\mathcal{A}$ is dense in $C(M)$. Since $\mathcal{A}$ is the linear span of P, it has a countable basis. Thus the space $C(K)$ is separable by Proposition 1.4.16. □

The Stone Weierstrass theorem allows us also to deduce approximation results:

Corollary 1.7.15. *For every $f \in C([a, b])$, there exists a sequence of polynomials p_n which converges uniformly on $[a, b]$ to f.*

Proof. This is immediate from Theorem 1.7.11 and Example 1.7.9. □

Unfortunately, the trigonometric polynomials do not separate $[0, 2\pi]$, hence we cannot invoke the Stone-Weierstrass theorem directly to deduce that we can approximate continuous functions on $[0, 2\pi]$ with trigonometric polynomials. However, we have

Corollary 1.7.16. *For every $f \in C([0,2\pi])$ with $f(0) = f(2\pi)$ there exists a trigonometric polynomial which converges uniformly to f.*

Proof. If we put $\mathcal{B} := \{f \in C([0,2\pi]) : f(0) = f(2\pi)\}$, then $\mathcal{B}$ is a closed unital algebra which is also closed under conjugation. Let $\mathcal{A}$ denote the closure of the the trigonometric polynomials. Since the trigonometric polynomials separate the points in $(0,2\pi)$, we see as in the proof of Theorem 1.7.11 that for any $x, y \in (0,2\pi)$ we find $p_{x,y} \in \mathcal{A}$ with $p_{x,y}(x) = f(x)$ and $p_{x,y}(y) = f(y)$. We complement this by putting

$$p_{0,y}(t) = p_{2\pi,y}(t) := f(0) + (f(y) - f(0))\frac{e^{it} - 1}{e^{iy} - 1}$$

and

$$p_{x,0}(t) = p_{x,2\pi}(t) := f(0) + (f(x) - f(0)\frac{e^{it} - 1}{e^{ix} - 1}$$

for $x, y \in (0,2\pi)$. Repeating the proof of 1.7.11, we see that any $f \in \mathcal{B}$ lies in the $\mathcal{A}$. □

1.8 Banach's Fixed Point Theorem

In this section, we present a result which is of great importance in applications.

Definition 1.8.1. Let M be a set and $\varphi : M \to M$ be a map. A *fixed point* of φ is an element $x^* \in M$ with $\varphi(x^*) = x^*$.

Given a map φ from some set M to itself, we define its *iterates* inductively by $\varphi^1 = \varphi$ and $\varphi^{n+1} = \varphi \circ \varphi^n$ for $n \geq 1$.

Theorem 1.8.2. *(Banach's fixed point theorem)*
Let (M,d) be a complete metric space and $\varphi : M \to M$ be a map such that there exists a sequence $q_n \geq 0$ with $\sum_{n=1}^\infty q_n < \infty$ with

$$d(\varphi^n(x), \varphi^n(y)) \leq q_n d(x,y) \quad \forall\, x, y \in M\,,\ n \geq 1\,.$$

Then φ has a unique fixed point x^*.

Proof. Let us first prove uniqueness. To that end, assume that x^* and y^* are fixed points of φ, i.e. $\varphi(x^*) = x^*$ and $\varphi(y^*) = y^*$. Then, by induction, $\varphi^n(x^*) = x^*$ and $\varphi^n(y^*) = y^*$. Thus,

$$d(x^*, y^*) = d(\varphi^n(x^*), \varphi^n(y^*)) \leq q_n d(x^*, y^*) \to 0$$

as $n \to \infty$, since $\sum_{j=1}^\infty q_n < \infty$. Thus $d(x^*, y^*) = 0$ and hence $x^* = y^*$.

We now prove existence of a fixed point. To that end, let $x_0 \in M$ be arbitrary and define x_n inductively by $x_n = \varphi(x_{n-1})$ for $n \geq 1$. Then

$$d(x_n, x_{n-1}) = d(\varphi^{n-1}(x_1), \varphi^{n-1}(x_0)) \leq q_{n-1} d(x_1, x_0)\,.$$

Thus, for $n \geq m$, we have

$$d(x_m, x_n) \leq \sum_{j=0}^{n-m-1} d(x_{m+j}, x_{m+j+1}) \leq \sum_{j=0}^{n-m-1} q_{m+j} d(x_1, x_0) = d(x_1, x_0) \sum_{j=m}^{n-1} q_j\,,$$

proving that x_n is a Cauchy sequence. By completeness, x_n converges to some element x^* of x.

We now prove that x^* is a fixed point of φ. To that end, first observe that φ is continuous. Indeed, if $z_n \to z$, then $d(\varphi(z_n), \varphi(z)) \leq q_1 d(z_n, z) \to 0$ as $n \to \infty$. Hence

$$\varphi(x^*) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x^* .$$

□

Definition 1.8.3. Let (M, d) be a metric space. A map $\phi : M \to M$ is called *Lipschitz continuous (with Lipschitz constant L)* if $d(\phi(x), \phi(y)) \leq d(x, y)$. It is called a *strict contraction*, if it is Lipschitz continuous with Lipschitz constant $L \in [0, 1)$.

Example 1.8.4. Let $f : \mathbb{R} \to \mathbb{R}$ be continuously differentiable with $\sup_{t\in\mathbb{R}} |f'(t)| \leq L$. Then f is Lipschitz continuous with Lipschitz constant L. Indeed, by the mean-value theorem, given $x, y \in \mathbb{R}$, there exists ξ between x and y such that $f(x) - f(y) = f'(\xi)(x - y)$. Thus

$$|f(x) - f(y)| = |f'(\xi)(x - y)| = |f'(\xi)| \cdot |x - y| \leq L|x - y| .$$

Corollary 1.8.5. *(Classical Banach fixed point theorem)*

Let (M, d) be a complete metric space and $\varphi : M \to M$ be a strict contraction. Then φ has a unique fixed-point.

Proof. The assumptions of Theorem 1.8.2 are satisfied with $q_n = L^n$ which is summable since $L \in [0, 1)$. □

Exercise 1.8.6. Consider $M := [1, \infty)$ as a metric space with the metric induced by $| \cdot |$. Show that $\varphi : M \to M$, given by $\varphi(x) = x + \frac{1}{x}$, satisfies $|\varphi(x) - \varphi(y)| < |x - y|$ but does not have a fixed point.

The usefulness of Banach's fixed point theorem arises from the fact that many problems can be reformulated as fixed point problems. As an illustration, we discuss ordinary differential equations. A similar approach also works for stochastic differential equations.

Definition 1.8.7. Let I be a compact interval, $t_0 \in I, u_0 \in \mathbb{R}$ and $f : I \times \mathbb{R} \to \mathbb{R}$ be a continuous function. A *solution* to the ordinary differential equation ODE_{f,t_0,u_0}, given by

$$\mathsf{ODE}_{f,t_0,u_0} \begin{cases} u'(t) & = f(t, u(t)) \quad t \in [0, T] \\ u(t_0) & = u_0 , \end{cases}$$

is a continuously differentiable function u^* such that ODE_{f,t_0,u_0} holds for $u = u^*$.

We now reformulate ODE_{f,t_0,u_0} as a fixed point problem.

Lemma 1.8.8. *Given a compact interval $I, t_0 \in I, u_0 \in \mathbb{R}$ and $f : I \times \mathbb{R} \to \mathbb{R}$ continuous, define $\varphi_{f,t_0,u_0} : C(I) \to C(I)$ by*

$$\big[\varphi_{f,t_0,u_0}(u)\big](t) = u_0 + \int_{t_0}^{t} f(s, u(s))\, ds .$$

Then u^* solves ODE_{f,t_0,u_0} if and only if $\varphi_{f,t_0,u_0}(u^*) = u^*$.

Proof. If u^* solves ODE_{f,t_0,u_0}, then by the fundamental theorem of calculus, for all $t \in I$, we have

$$u^*(t) - u_0 = u^*(t) - u^*(t_0) = \int_{t_0}^t (u^*)'(s)\,ds = \int_{t_0}^t f(s, u^*(s))\,ds\,.$$

Thus $\varphi_{f,t_0,u_0}(u^*) = u^*$.

Conversely, if $\varphi_{f,t_0,u_0}(u^*) = u^*$, then, by the fundamental theorem of calculus, u^* is continuously differentiable and

$$(u^*)'(t) = \frac{d}{dt}\int_{t_0}^t f(s, u^*(s))\,ds = f(t, u^*(t))$$

for all $t \in I$. Since $u^*(t_0) = u_0$, it follows that u^* solves ODE_{f,t_0,u_0}. □

Theorem 1.8.9. *(Picard-Lindelöf)*

Let I be a compact interval and $f : I \times \mathbb{R} \to \mathbb{R}$ be a continuous function such that there exists $L \geq 0$ with

$$|f(t,x) - f(t,y)| \leq L|x-y| \quad \forall x, y \in \mathbb{R}\,,\, t \in I\,.$$

Then, for every $t_0 \in I$ and $u_0 \in \mathbb{R}$ there exists a unique solution $u^* : [0,T] \to \mathbb{R}$ of the differential equation ODE_{f,t_0,u_0}.

Proof. Let $\varphi := \varphi_{f,t_0,u_0}$ be as on Lemma 1.8.8. By that Lemma it suffices to show that φ has a unique fixed point. Since $(C(I), \|\cdot\|_\infty)$ is complete, it suffices to show that the hypothesis of Theorem 1.8.2 is satisfied.

To make notation simpler, we will assume that $I = [-T, T]$ and $t_0 = 0$. We claim that $|[\varphi^n(u)](t) - [\varphi^n(v)](t)| \leq \frac{L^n|t|^n}{n!}\|u - v\|_\infty$ for all $u, v \in C(I)$ and $t \in I$. It follows that $\|\varphi^n(u) - \varphi^n(v)\|_\infty \leq \frac{L^nT^n}{n!}\|u - v\|_\infty$ for all $u, v \in C([0,T])$. Note that the latter is summable and hence the hypothesis of Theorem 1.8.2 is satisfied. Thus, proving the claim finishes the proof.

We proceed by induction, first considering $t \geq 0$. For $n = 1$, we have

$$\big|[\varphi(u)](t) - [\varphi(v)](t)\big| \leq \int_0^t |f(s,u(s)) - f(s,v(s))|\,ds \leq \int_0^t L|u(s) - v(s)|\,ds \leq Lt\|u - v\|_\infty\,.$$

Now assume that $|[\varphi(u)](t) - [\varphi(v)](t)| \leq \frac{L^nt^n}{n!}\|u - v\|_\infty$. Then

$$\begin{aligned} |[\varphi^{n+1}(u)](t) - [\varphi^{n+1}(v)](t)| &\leq \int_0^t |f(s, [\varphi^n(u)](s)) - f(s, \varphi^n(v)](s))|\,ds \\ &\leq \int_0^t L|[\varphi^n(u)](s) - [\varphi^n(v)](s)|\,ds \\ &\leq \int_0^t L\frac{L^ns^n}{n!}\|u - v\|_\infty\,ds = \frac{L^{n+1}t^{n+1}}{(n+1)!}\|u - v\|_\infty\,. \end{aligned}$$

This finishes the proof in the case where $t \geq 0$. The case where $t \leq 0$ is similar. □

Exercise 1.8.10. The proof of Theorem 1.8.2 shows that for any initial value x_0 the sequence $\varphi^n(x_0)$ converges to the unique fixed-point of φ. This gives a possibility to *construct* the fixed point. In particular, we can *construct* solutions to (ODE). This is called the *Picard-Iteration*

For the ordinary differential equation $u'(t) = tu(t)$, use the Picard-Iteration to construct the solution of this equation with $u(0) = 1$.

1.9 Bounded Linear Operators

In this section, we study the continuity of *linear* maps between normed vector spaces. The following gives the main characterization of such maps.

Proposition 1.9.1. *Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be normed vector spaces and $T : E \to F$ be a linear map. The following are equivalent:*

(a) *T is continuous.*

(b) *T is continuous at 0.*

(c) *There exists a constant C such that $\|Tx\|_F \leq C\|x\|_E$ for all $x \in E$.*

(d) $\sup\{\|Tx\|_F : \|x\|_E \leq 1\} < \infty$.

Proof. (a) $\Rightarrow$ (b) is trivial. For (b) $\Rightarrow$ (c), note that by the addendum in Proposition 1.4.17, there exists $\delta > 0$ such that $\|Tx - T0\|_F = \|Tx\|_F \leq \delta$ for all $\|x\|_E \leq 1$. Now let $x \neq 0$. Then $\|\|x\|_E^{-1}x\|_E = 1$ and thus, by linearity of T, $\|x\|_E^{-1}\|Tx\|_F = \|T(\|x\|_E^{-1}x)\|_E \leq \delta$, that is (c) holds with $C = \delta$.

(c) $\Rightarrow$ (d): If (c) holds, then $\sup\{\|Tx\|_F : \|x\|_E \leq 1\} \leq C < \infty$.

(d) $\Rightarrow$ (a): Let $x_n \to x$. We may assume without loss, that $x_n \neq x$ for all $n \in \mathbb{N}$. Then $\|x_n - x\|_E^{-1}(x_n - x)$ has norm 1 and hence, by (d), $\|T(\|x_n - x\|_E(x_n - x))\|_F \leq C$, where $C := \sup\{\|Tx\|_F : \|x\|_E \leq 1\} < \infty$. Thus, as in (b) $\Rightarrow$ (c), $\|Tx_n - Tx\|_F = \|T(x_n - x)\|_F \leq C\|x_n - x\|_E \to 0$ as $n \to \infty$, proving that T is continuous. □

Definition 1.9.2. Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be normed vector spaces. A *bounded operator* from E to F is a continuous linear map $T : E \to F$. We write $\mathscr{L}(E, F)$ for the set of all bounded operators from E to F and $\|T\|_{\mathscr{L}(E,F)} := \sup\{\|Tx\|_F : \|x\|_E \leq 1\}$. If it is clear which norms we use on E and F, we also write $\|T\|$ instead of $\|T\|_{\mathscr{L}(E,F)}$. If $E = F$, we write $\mathscr{L}(E)$ instead of $\mathscr{L}(E, E)$, if $F = \mathbb{K}$, we write E^* instead of $\mathscr{L}(E, \mathbb{K})$. We call E^* the *dual space* of E.

Below, we prove that $\mathscr{L}(E, F)$ is a vector space and that $\|\cdot\|_{\mathscr{L}(E,F)}$ defines a norm on that space. In the proof we will use the following fact:

Lemma 1.9.3. *Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be normed vector spaces and $T \in \mathscr{L}(E, F)$. Then $\|Tx\|_F \leq \|T\|_{\mathscr{L}(E,F)}\|x\|_E$ for all $x \in E$. Moreover,*

$$\|T\|_{\mathscr{L}(E,F)} = \inf\{C > 0 : \|Tx\|_F \leq C\|x\|_E \;\forall\, x \in E\}\,.$$

Proof. Let $x \in E \setminus \{0\}$. Then $y := \|x\|_E^{-1}x$ satisfies $\|y\|_E = 1$, hence $\|Ty\|_F \leq \|T\|_{\mathscr{L}(E,F)}$ by definition. On the other hand, by linearity of T and homogenity of the norm, $\|Ty\|_F = \|x\|_E^{-1}\|Tx\|_F$ and hence $\|Tx\|_F \leq \|T\|_{\mathscr{L}(E,F)}\|x\|_E$ follows. This also proves that

$$\|T\|_{\mathscr{L}(E,F)} \geq \inf\{C > 0 : \|Tx\|_F \leq C\|x\|_E \;\forall\, x \in E\}$$

since $C = \|T\|_{\mathscr{L}(E,F)}$ belongs to the set on the right-hand side. To see the reversed inequality, assume that $\|T\|_{\mathscr{L}(E,F)} < \inf\{C > 0 : \|Tx\|_F \leq C\|x\|_E \;\forall\, x \in E\}$. Then there exists $C < \|T\|_{\mathscr{L}(E,F)}$ with $\|Tx\|_F \leq C\|x\|_E$ for all $x \in E$. But this implies $\|T\|_{\mathscr{L}(E,F)} = \sup\{\|Tx\|_F \; \|x\|_E \leq 1\} \leq C < \|T\|_{\mathscr{L}(E,F)}$ which is absurd. □

Theorem 1.9.4. *Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be normed vector spaces. Then $\mathscr{L}(E,F)$ is a vector space (with respect to pointwise addition and scalar multiplication) and $\|\cdot\|_{\mathscr{L}(E,F)}$ defines a norm on $\mathscr{L}(E,F)$. If F is complete, then also $(\mathscr{L}(E,F), \|\cdot\|_{\mathscr{L}(E,F)})$ is complete.*

Proof. It is well-known that linear maps form a vector space with respect to pointwise addition and scalar multiplication. Moreover, as a consequence of Proposition 1.3.13, sums and scalar multiples of continuous maps are continuous (however, the arguments below will also give an alternative proof of this fact). Hence $\mathscr{L}(E,F)$ is a vector space. It thus only remains to show that $\|\cdot\|_{\mathscr{L}(E,F)}$ is a norm on $\mathscr{L}(E,F)$.

If $T \neq \mathbf{0}$, then there exists $x \in E$ with $Tx \neq 0$. Replacing x with $\|x\|_E^{-1}x$, we see that there exist $x \in E$ with $\|x\|_E = 1$ such that $Tx \neq 0$. This shows that $\|T\|_{\mathscr{L}(E,F)} \neq 0$ if $T \neq \mathbf{0}$. This proves (N1).

If $T \in \mathscr{L}(E,F)$ and $\lambda \in \mathbb{K}$, then $\|(\lambda T)x\|_F = \|T(\lambda x)\|_F \leq \|T\|_{\mathscr{L}(E,F)}\|\lambda x\|_E = |\lambda|\|T\|_{\mathscr{L}(E,F)}\|x\|_E$ for all $x \in E$. This shows that λT is bounded and that $\|\lambda T\|_{\mathscr{L}(E,F)} \leq |\lambda|\|T\|_{\mathscr{L}(E,F)}$. It also shows that $\|T\|_{\mathscr{L}(E,F)} = \|\lambda^{-1}\lambda T\|_{\mathscr{L}(E,F)} \leq |\lambda^{-1}|\|\lambda T\|_{\mathscr{L}(E,F)}$, which gives the reversed inequality. This proves (N2).

As for the triangle inequality, let $T, S \in \mathscr{L}(E,F)$. Then, for $x \in E$ with $\|x\|_E \leq 1$,

$$\|(T+S)x\|_F = \|Tx + Sx\|_F \leq \|Tx\|_F + \|Sx\|_F \leq \|T\|_{\mathscr{L}(E,F)} + \|S\|_{\mathscr{L}(E,F)}\,.$$

Taking the supremum over $x \in E$ with $\|x\| \leq 1$, the triangle inequality follows.

Now assume that F is complete and let T_n be a Cauchy sequence in $\mathscr{L}(E,F)$. Then, for $x \in E$, we have

$$\|T_n x - T_m x\|_F \leq \|T_n - T_m\|_{\mathscr{L}(E,F)}\|x\|_E\,,$$

proving that $T_n x$ is a Cauchy sequence in F, hence, by completeness, convergent to some vector $y =: Tx$.

To see that T is linear let $x_1, x_2 \in E$ and $\lambda \in \mathbb{K}$. Then

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

Here, we have used the linearity of the T_n in the second and Proposition 1.3.13 in the second equality.

We next prove that $T \in \mathscr{L}(E,F)$ and $\|T_n - T\|_{\mathscr{L}(E,F)} \to 0$. Given $\varepsilon > 0$, pick n_0 such that $\|T_n - T_m\|_{\mathscr{L}(E,F)} \leq \varepsilon$ for all $n, m \geq n_0$. Now let $x \in E$ with $\|x\|_E \leq 1$. Then $\|T_n x - T_m x\|_F \leq \varepsilon$ for all $n, m \geq n_0$. With $m \to \infty$, it follows that $\|T_n x - Tx\|_F \leq \varepsilon$ for all $n \geq n_0$. This proves

(i) $T \in \mathscr{L}(E,F)$, since

$$\|Tx\|_F \leq \|T_{n_0}x\|_F + \|Tx - T_{n_0}x\|_F \leq \|T_{n_0}\|_{\mathscr{L}(E,F)} + 2\varepsilon$$

for all $\|x\|_E \leq 1$ and

(ii) $\|T_n - T\|_{\mathscr{L}(E,F)} \leq \varepsilon$ for all $n \geq n_0$, which follows by taking the supremum over $x \in E$ with $\|x\|_E \leq 1$. ☐

Exercise 1.9.5. Let $(E_i, \|\cdot\|_i)$ be normed vector spaces for $i = 1, 2, 3$. Show that if $T \in \mathscr{L}(E_1, E_2)$ and $S \in \mathscr{L}(E_2, E_3)$, then $ST \in \mathscr{L}(E_1, E_3)$ and

$$\|ST\|_{\mathscr{L}(E_1,E_3)} \leq \|S\|_{\mathscr{L}(E_2,E_3)}\|T\|_{\mathscr{L}(E_1,E_2)}\,.$$

Example 1.9.6. Consider the normed space $(\ell^p, \|\cdot\|_p)$ for $1 \leq p \leq \infty$. For $\mathbf{m} \in \ell^\infty$, we define $T_\mathbf{m} : \ell^p \to \ell^p$ by

$$T_\mathbf{m}\mathbf{x} = (m_1x_1, m_2x_2, m_3x_x, \ldots) .$$

Then $T_\mathbf{m} \in \mathscr{L}(\ell^p)$ and $\|T_\mathbf{m}\|_{\mathscr{L}(\ell^p)} = \|\mathbf{m}\|_{\ell^\infty}$.

Proof. We give the proof for $1 \leq p < \infty$ and leave the case $p = \infty$ to the reader.

For $\mathbf{x} \in \ell^p$, we have

$$\|T_\mathbf{m}\mathbf{x}\|_p^p = \sum_{k=1}^\infty |m_kx_k|^p = \sum_{k=1}^\infty |m_k|^p|x_k|^p \leq \sum_{k=1}^\infty \|\mathbf{m}\|_\infty |x_k|^p = \|\mathbf{m}\|_\infty^p \|\mathbf{x}\|_p^p .$$

This proves that $T_\mathbf{m} \in \mathscr{L}(\ell^p)$ and that $\|T_\mathbf{m}\|_{\mathscr{L}(\ell^p)} \leq \|\mathbf{m}\|_\infty$. To see equality, consider $\mathbf{e}_j \in \ell^p$. Then $\|\mathbf{e}_j\|_p = 1$ and $\|T_\mathbf{m}e_j\|_p = \|m_j\mathbf{e}_j\|_p = |m_j|$. Thus $\|T_\mathbf{m}\|_{\mathscr{L}(\ell^p)} \geq |m_j|$ for all $j \in \mathbb{N}$ and hence $\|T_\mathbf{m}\|_{\mathscr{L}(\ell^p)} \geq \|\mathbf{m}\|_\infty$. ☐

Exercise 1.9.7. Let $1 \leq p, q \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Given $\mathbf{a} \in \ell^q$, define

$$\varphi_\mathbf{a}(\mathbf{x}) = \sum_{k=1}^\infty a_kx_k .$$

Note that this is well-defined by Hölder's inequality. Show that $\varphi_\mathbf{a}$ is a continuous linear map from ℓ^p to $\mathbb{K}$, hence $\varphi_\mathbf{a} \in (\ell^p)^*$. Also show that $\|\varphi_\mathbf{a}\|_{(\ell^p)^*} = \|\mathbf{a}\|_q$.

Example 1.9.8. Let E be a finite-dimensional vector space and $\|\cdot\|$ be a norm on E. If $\mathfrak{v} = (v_1, \ldots, v_d)$ is a basis of E, we define $T_\mathfrak{v} : E \to \mathbb{K}^n$ by

$$T_\mathfrak{v} \sum_{j=1}^d \alpha_jv_j = (\alpha_1, \ldots, \alpha_d) .$$

Since $\mathfrak{v}$ is a basis of E, the value $T_\mathfrak{v}x$ is well-defined for all $x \in E$. The map $T_\mathfrak{v}$ is called the *coordinate map* associated with the basis $\mathfrak{v}$. It is in fact bijective and its inverse is given by

$$T_\mathfrak{v}^{-1}(x_1, \ldots, x_n) = \sum_{j=1}^d x_jv_j .$$

If we endow $\mathbb{K}^d$ with any of the norms $\|\cdot\|_p$, for $1 \leq p \leq \infty$, then both $T_\mathfrak{v}$ and $T_\mathfrak{v}^{-1}$ are continuous.

Proof. Let us first prove that $T_\mathfrak{v}^{-1}$ is continuous. Assume that $x_n := (x_1^{(n)}, \ldots, x_d^{(n)}) \to x = (x_1, \ldots, x_d)$ in $\mathbb{K}^d$ with respect to $\|\cdot\|_p$. By Example 1.3.5 this is equivalent with $x_j^{(n)} \to x_j$ for all $1 \leq j \leq d$. Thus,

$$\|T_\mathfrak{v}^{-1}x_n - T_\mathfrak{v}^{-1}x\|_E = \Big\| \sum_{j=1}^d (x_j^{(n)} - x_j)v_j \Big\|_E \leq \sum_{j=1}^d |x_j^{(n)} - x_j| \cdot \|v_j\|_E \to 0$$

as $n \to \infty$, proving that $T_\mathfrak{v}^{-1}$ is continuous.

It was seen in Example 1.6.5, that $K := \{x \in \mathbb{K}^d : \|x\|_p = 1\}$, being closed and bounded, is compact. Thus, by Theorem 1.6.10, the continuous function $x \mapsto \|T_\mathfrak{v}^{-1}x\|_E$

attains a minimum m on K. Note that $m > 0$, since otherwise, $\|T_{\mathfrak{v}}^{-1}x\|_E = 0$ for some $x \neq 0$, i.e. $T_{\mathfrak{v}}^{-1}x \neq 0$.

Thus $m \leq \|T_{\mathfrak{v}}^{-1}x\|_E$ for all $\|x\|_p = 1$. Considering $\|x\|_p^{-1}x$ for an arbitrary $x \neq 0$, it follows that $m\|x\|_p \leq \|T_{\mathfrak{v}}^{-1}x\|_E$ for all $x \in \mathbb{K}^d$ or, putting $x = T_{\mathfrak{v}}y$, $\|T_{\mathfrak{v}}y\|_p \leq m^{-1}\|y\|_E$. This proves the boundedness of $T_{\mathfrak{v}}$. □

In the last example, we have actually proved

Theorem 1.9.9. *Let E be a finite dimensional vector space. Then any two norms on E are equivalent.*

Proof. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on E. Fix a basis $\mathfrak{v}$ of E. On $\mathbb{K}^d$, where d is the dimension of E, we use the $\|\cdot\|_\infty$-norm. Then, by Example 1.9.8, $T_{\mathfrak{v}}$ and $T_{\mathfrak{v}}^{-1}$ are bounded from E to $(\mathbb{K}^d, \|\cdot\|_\infty)$ resp. from $(\mathbb{K}^d, \|\cdot\|_\infty)$ to E when E is endowed with either the $\|\cdot\|_1$ or the $\|\cdot\|_2$-norm.

Hence

$$\begin{aligned}\|x\|_1 = \|T_{\mathfrak{v}}^{-1}T_{\mathfrak{v}}x\|_1 &\leq \|T_{\mathfrak{v}}^{-1}\|_{\mathscr{L}((E,\|\cdot\|_1),\mathbb{K}^d)}\|T_{\mathfrak{v}}x\|_\infty \\ &\leq \|T_{\mathfrak{v}}^{-1}\|_{\mathscr{L}((E,\|\cdot\|_1),\mathbb{K}^d)}\|T_{\mathfrak{v}}\|_{\mathscr{L}(\mathbb{K}^d,(E,\|\cdot\|_2))}\|x\|_2 \,.\end{aligned}$$

This shows that $\|x\|_1 \leq C\|x\|_2$ for a suitable constant C. Interchanging the roles of $\|\cdot\|_1$ and $\|\cdot\|_2$, it follows that also $\|x\|_2 \leq c\|x\|_1$ for a suitable constant c. □

We end this section with the following useful result:

Proposition 1.9.10. *Let $(E, \|\cdot\|_E)$ be a normed vector space, $(F, \|\cdot\|_F)$ be a complete normed vector space and E_0 be a dense subspace of E. Given $T \in \mathscr{L}(E_0, F)$ there exists a unique operator $\tilde{T} \in \mathscr{L}(E, F)$ with $Tx = \tilde{T}x$ for all $x \in E_0$. Moreover, $\|T\|_{\mathscr{L}(E_0,F)} = \|\tilde{T}\|_{\mathscr{L}(E,F)}$.*

Proof. Let $x \in E$. By density, there exists a sequence $(x_n) \subset E_0$ with $x_n \to x$. Since $\|Tx_n - Tx_m\|_F \leq \|T\|_{\mathscr{L}(E_0,F)}\|x_n - x_m\| \to 0$ as $n, m \to \infty$, it follows that (Tx_n) is a Cauchy sequence in F. By completeness, Tx_n converges to some $\tilde{T}x \in F$. Note that $\tilde{T}x$ does not depend on the approximating sequence (x_n). Indeed, if (y_n) was another sequence in E_0 converging to x and $Ty_n \to z$, then

$$\|z - \tilde{T}x\|_F \leq \|z - Ty_n\| + \|T\| \cdot \|y_n - x_n\| + \|Tx_n - \tilde{T}x\| \to 0$$

as $n \to \infty$ since $\|y_n - x_n\| \to 0$. Hence $z = \tilde{T}x$.

Using this uniqueness, it is easy to see that $\tilde{T}$ is linear, cf. the proof of Theorem 1.9.4. Moreover, $\tilde{T}x = Tx$ for all $x \in E_0$. To see that $\tilde{T} \in \mathscr{L}(E, F)$, let $x \in E$ and x_n be a sequence in E_0 converging to x. Then

$$\|\tilde{T}x\|_F = \|\lim_{n\to\infty} Tx_n\|_F = \lim_{n\to\infty}\|Tx_n\|_F \leq \lim_{n\to\infty}\|T\|_{\mathscr{L}(E_0,F)}\|x_n\|_E = \|T\|_{\mathscr{L}(E_0,F)}\|x\|_E$$

where we have used the continuity of the norm. This proves that $\tilde{T} \in \mathscr{L}(E, F)$ and $\|\tilde{T}\|_{\mathscr{L}(E,F)} \leq \|T\|_{\mathscr{L}(E_0,F)}$. The other inequality is trivial. □

Chapter 2

Measure and Integration

Man is the measure of all things.
Pythagoras

Lebesgue is the measure of almost all things.
Anonymous

The objective of measure theory is to assign to subsets A of a given "ground set" Ω a nonnegative number $\mu(A)$, called the measure of A, which, in one way or another, measures its size.

As an instructive example, let us look at determining the area of a two dimensional object. Our ground set Ω is $\mathbb{R}^2$ and, for $A \subset \mathbb{R}^2$, the measure $\mu(A)$ should be its area. If R is a rectangle with side lengths a and b, then $\mu(R)$ should be $a \cdot b$. With this information alone and some geometric considerations, one can already determine the area of more complicated objects. For example, the area of a right triangle T, where the two shorter sides have length a resp. b should be $\frac{1}{2}a \cdot b$, since two of these triangles can be assembled into a rectangle of side lengths a and b.

Proceeding, we can determine the area of any object which can be decomposed into finitely many right triangles, in particular for polygons. We could then use polygons to approximate more complex object, e.g. a circle. Already Archimedes (287 BC – 212 BC) approximated the circle with regular polygons to obtain an approximation for the area of a circle.

Analyzing the operations above, the "measure" μ should have the following properties:

1. If $A_1, \ldots, A_n$ are disjoint, then $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$.
2. If B is obtained by translating and rotating A, then $\mu(A) = \mu(B)$.
3. If A_n is an increasing sequence with $\bigcup_{k=1}^{\infty} A_k = B$, then $\mu(B) = \sup_{n \in \mathbb{N}} \mu(A_n)$.

It is not clear at all, whether a measure μ with the above properties can be defined at all on all subsets of $\mathbb{R}^2$. In three dimensions, *Banach* and *Tarski* proved, making use of the axiom of choice, that one can cut a 3-dimensional ball into finitely many (actually, five will do) pieces and then, rotating and translating these pieces, reassemble them into *two* copies of the ball (thus doubling its volume?).

As you can imagine, one cannot imagine how these pieces look like, but they must be rather irregular.

To get around such problems, one defines measures μ merely on a collection Σ of "well-behaved" subsets. How these "well-behaved" subsets look like might well depend on the set Ω and the measure μ itself.

Let us look at yet another example, this one from probability theory.

In an elementary probability course, you may have encountered *Laplace experiments*, i.e. random experiments with only finitely many possible outcomes. Typical examples are the tossing of a coin or the rolling of a die. Here the ground set Ω consists of all possible outcomes, in the examples above

$$\Omega_1 = \{\mathsf{H}, \mathsf{T}\} \quad \text{for coin tossing} \qquad \Omega_2 = \{1, 2, 3, 4, 5, 6\} \quad \text{for rolling a die}.$$

A subset A of Ω is then called an *event*: $A_1 = \{\mathsf{H}\}$ corresponds in the first experiment to the event "the coin came up heads", in the second experiment, $A_2 = \{1, 3, 5\}$ corresponds to the event "an odd number was rolled".

Such an event is now assigned a "probability" $\mathbb{P}(A)$ by putting

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}$$

where $\#S$ denotes the number of elements in a set S.

It is readily checked that $\mu = \mathbb{P}$ satisfies conditions 1. and 3. above. On the other hand, condition 2. depends on the euclidean structure of the ground set Ω and does not make much sense in this case.

2.1 σ-Algebras and Their Generators

We begin with the concept of a σ-algebra, which has the right structure for the collection of "well-behaved" sets on which a measure can be defined.

Definition 2.1.1. Let Ω be a nonempty set. A subset Σ of the power set $\mathscr{P}(\Omega)$ is called *σ-algebra* (on Ω) if

(S1) $\Omega \in \Sigma$.

(S2) $A \in \Sigma$ implies $A^c \in \Sigma$.

(S3) $A_k \in \Sigma$ for all $k \in \mathbb{N}$ implies $\bigcup_{k\in\mathbb{N}} A_k \in \Sigma$.

A *measurable space* is a pair (Ω, Σ), where Ω is a nonempty set and Σ is a σ-algebra on Ω. The elements A of Σ are called *measurable sets.*

Example 2.1.2. Let Ω be a nonempty set. Then $\{\emptyset, \Omega\}$ is a σ-algebra on Ω; it is the smallest σ-algebra on Ω. Moreover, $\mathscr{P}(\Omega)$ is a σ-algebra on Ω; it is the largest σ-algebra on Ω. Finally, if $A \subset \Omega$, then $\{\emptyset, A, A^c, \Omega\}$ is a σ-algebra on Ω.

Proposition 2.1.3. *Let (Ω, Σ) be a measurable space. Then*

(a) $\emptyset \in \Omega$.

(b) If $A_1, \ldots, A_n \in \Sigma$, *then* $A_1 \cup \cdots \cup A_n \in \Omega$ *and* $A_1 \cap \cdots \cap A_n \in \Omega$.

(c) If $A_k \in \Omega$ for all $k \in \mathbb{N}$ then $\bigcap_{k\in\mathbb{N}} A_k \in \Omega$.

Proof. (a) follows directly from (S1) and (S2).

For (b), put $B_k := A_k$ for $k = 1, \dots, n$ and $B_k = \emptyset$ for $k \geq n+1$. Then $B_k \in \Sigma$ for all $k \in \mathbb{N}$ and thus $\bigcup_{k\in\mathbb{N}} B_k = A_1 \cup \dots \cup A_n \in \Sigma$ by (S3). Moreover, $A_1 \cap \dots \cap A_n = (A_1^c \cup \dots \cup A_n^c)^c \in \Sigma$ by (S2) and what was just proved.

(c) By DeMorgan's law, $\bigcap_{k\in\mathbb{N}} A_k = \Big(\bigcup_{k\in\mathbb{N}} A_k^c\Big)^c \in \Sigma$ by (S2) and (S3). □

Lemma 2.1.4. *Let Ω be a nonempty set, I an nonempty (index-) set and Σ_i, for $i \in I$, be a σ-algebra on Ω for all $i \in I$. Then $\bigcap_{i\in I} \Sigma_i$ is a σ-algebra on Ω.*

Proof. By (S1), $\Omega \in \Sigma_i$ for all $i \in I$ and thus $\Omega \in \bigcap_{i\in I} \Sigma_i$, proving (S1).

If $A \in \bigcap_{i\in I} \Sigma_i$, then $A \in \Sigma_i$ for all $i \in I$. Hence, by (S2), $A^c \in \Sigma_i$ for all $i \in I$ and thus $A^c \in \bigcap_{i\in I} \Sigma_i$ proving (S2).

Finally, If $A_k \in \bigcap_{i\in I} \Sigma_i$ for all $k \in \mathbb{N}$, then $A_k \in \Sigma_i$ for all $k \in \mathbb{N}, i \in I$ and thus, by (S3), $\bigcup_{k\in\mathbb{N}} A_k \in \Sigma_i$ for all $i \in I$ proving $\bigcup_{k\in\mathbb{N}} A_k \in \bigcap_{i\in I} \Sigma_i$ and thus (S3). □

Definition 2.1.5. Let Ω be a nonempty set, $\mathcal{A} \subset \mathscr{P}(\Omega)$. Then

$$\sigma(\mathcal{A}) := \bigcap_{\mathcal{A}\subset\Sigma\subset\mathscr{P}(\Omega)\ \sigma\text{- algebra}} \Sigma$$

is a σ-algebra by Lemma 2.1.4. It is called the *σ-algebra generated by $\mathcal{A}$*. If Σ is a σ-algebra on Ω then any $\mathcal{A} \subset \mathscr{P}(\Omega)$ with $\sigma(\mathcal{A}) = \Sigma$ is called a *generator of* Σ.

Example 2.1.6. If $A \subset \Omega$, then $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$. Moreover, $\sigma(\emptyset) = \{\emptyset, \Omega\}$.

Definition 2.1.7. Let (M, d) be a metric space, $\mathcal{O}$ be the collection of all open subsets of M. Then $\sigma(\mathcal{O})$ is called the *Borel σ-algebra* of (M, d) and denoted by $\mathscr{B}(M, d)$ or, if d is understood, by $\mathscr{B}(M)$.

Example 2.1.8. Let M be a set and d be the discrete metric on M. Then $\mathscr{B}(M, d) = \mathscr{P}(M)$.

Proposition 2.1.9. *Let (M, d) be a metric space and denote by $\mathcal{C}$ the collection closed subsets of M, by $\mathcal{K}$ the collection of compact subsets of M.*

(a) $\mathscr{B}(M, d) = \sigma(\mathcal{C})$.

(b) If M is the countable union of compact sets, then $\mathscr{B}(M, d) = \sigma(\mathcal{K})$.

(c) If M is separable, then $\mathscr{B}(M, d)$ is generated by all open balls $B(x, \varepsilon)$ for $x \in M$ and $\varepsilon > 0$. In fact, it is generated by a countable collection of such balls.

Proof. (a) Let Σ be a σ-algebra containing all closed sets. By (S2), Σ contains all open sets and thus $\mathscr{B}(M, d)$. Hence $\mathscr{B}(M, d) \subset \sigma(\mathcal{C})$. Conversely, if Σ be a σ-algebra containing all open sets then, by (S2), Σ contains all closed sets and thus $\sigma(\mathcal{C})$. Hence $\sigma(\mathcal{C}) \subset \mathscr{B}(M, d)$.

(b) Since $\mathcal{K} \subset \mathcal{C}$ we clearly have $\sigma(\mathcal{K}) \subset \sigma(\mathcal{C}) \subset \mathscr{B}(M, d)$. Now assume that $M = \bigcup_{k\in\mathbb{N}} K_n$ where K_n is compact. If F is closed then $F \cap K_n$ is compact and hence an element of $\sigma(\mathcal{K})$. Since $\bigcup_{n\in\mathbb{N}} K_n = \Omega$, we have $\bigcup_{n\in\mathbb{N}} F \cap K_n = F$. Hence, by (S3), $F \in \sigma(\mathcal{K})$. Thus $\sigma(\mathcal{K})$ contains all closed sets and thus the σ-algebra generated by them, i.e. $\mathscr{B}(M, d)$.

(c) We only have to show that there is a countable union of open balls such that any open set is contained in the σ-algebra generated by these balls. To that end, let $\{x_n : n \in \mathbb{N}\}$ be a countable dense subset of M. Consider the open balls $B(x_n, k^{-1})$ for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$. This is a countable collection of open balls and every open subset U of M is a countable union of such balls. Indeed, if U is empty, there is nothing to prove. On the other hand, if $U \neq \emptyset$, then U is the union of all these balls contained in U (note that this is a countable union). Indeed, if $x \in U$ then $B(x, 2k^{-1}) \subset U$ for some k. Since $\{x_n : n \in \mathbb{N}\}$ is dense, there is some n with $x_n \in B(x, k^{-1})$ hence $x \in B(x_n, k^{-1})$. Moreover, $B(x_n, k^{-1}) \subset B(x, 2k^{-1}) \subset U$.

It now follows by (S3) that U belongs to the σ-algebra generated by these open balls. □

Exercise 2.1.10. Endow $\mathbb{R}$ with the discrete metric d and let $\mathcal{K}$ denote the collection of all compact subsets of $(\mathbb{R}, d)$. Show that

$$\sigma(\mathcal{K}) = \big\{A \subset \mathbb{R} : A \text{ is countable or } A^c \text{ is countable}\big\}.$$

Show that this is also the σ-algebra generated by the open balls in $(\mathbb{R}, d)$. Thus (b) and (c) in Proposition 2.1.9 do not hold without further assumptions.

Next, we introduce the concept of a *measurable map*.

Definition 2.1.11. Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be measurable spaces. A map $f : \Omega_1 \to \Omega_2$ is called *measurable*, more precisely Σ_1/Σ_2-measurable, if $f^{-1}(A) \in \Sigma_1$ for all $A \in \Sigma_2$.

Exercise 2.1.12. Show that the composition of measurable maps is measurable.

Exercise 2.1.13. On $\mathbb{R}$, consider the σ-algebra Σ from Example 2.1.10. Determine all functions $f : \mathbb{R} \to \mathbb{R}$ that are $\Sigma/\mathscr{B}(\mathbb{R}, |\cdot|)$-measurable and all continuous functions that are $\mathscr{B}(\mathbb{R}, |\cdot|)/\Sigma$-measurable.

Lemma 2.1.14. *Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be a measurable spaces and $f : \Omega_1 \to \Omega_2$ be a map. Moreover, let $\mathcal{A} \subset \mathscr{P}(\Omega_2)$ be a generator of Σ_2, i.e. $\sigma(\mathcal{A}) = \Sigma_2$. Then f is Σ_1/Σ_2-measurable if and only if $f^{-1}(A) \in \Sigma_1$ for all $A \in \mathcal{A}$.*

Proof. We use the *principle of good sets*:

Let $\mathscr{G} := \{A \in \Sigma_2 : f^{-1}(A) \in \Sigma_1\}$. Then $\mathscr{G}$ is a σ-algebra. Indeed, $f^{-1}(\Omega_2) = \Omega_1 \in \Sigma_1$. Moreover, it follows from the properties of the preimage that

$$f^{-1}(A^c) = (f^{-1}(A))^c \quad \text{and} \quad f^{-1}(\bigcup_{k\in\mathbb{N}} A_k) = \bigcup_{k\in\mathbb{N}} f^{-1}(A_k).$$

These equalities and the axioms of a σ-algebra for Σ_1 imply that $\mathscr{G}$ contains A^c if it contains A and it contains $\bigcup_{k\in\mathbb{N}} A_k$ provided it contains A_k for every $k \in \mathbb{N}$.

Thus, if $\mathscr{G}$ contains $\mathcal{A}$, then it contains $\sigma(\mathcal{A}) = \Sigma_2$, i.e. if $f^{-1}(A) \in \Sigma_1$ for all $A \in \mathcal{A}$, then f is Σ_1/Σ_2-measurable. The converse is obvious. □

Example 2.1.15. Let (M_1, d_1) and (M_2, d_2) be metric spaces. Then every continuous function $f : M_1 \to M_2$ is $\mathscr{B}(M_1, d_1)/\mathscr{B}(M_2, d_2)$-measurable. Indeed, by Proposition 1.4.17, $f^{-1}(U)$ is open in (M_1, d_1) for all open sets $U \subset M_2$- In particular, $f^{-1}(U) \in \mathscr{B}(M_1, d_1)$. Since the open sets in M_2 generate the Borel σ-algebra, it follows from Lemma 2.1.14 that f is measurable.

We next prove that, in a certain sense, the Borel σ-algebra is the smalles σ-algebra such that all continuous functions are measureable. To that end, we first introduce the following concept.

Definition 2.1.16. Let Ω be a nonempty set and $(\tilde{\Omega}, \tilde{\Sigma})$ be a measurable space. Moreover, let $\mathscr{F}$ be a collection of functions $f : \Omega \to \tilde{\Omega}$.

Then $\sigma(\mathscr{F}) := \sigma(\{f^{-1}(A) : f \in \mathscr{F}, A \in \tilde{\Sigma}\})$ is called the σ-algebra generated by $\mathscr{F}$.

Remark 2.1.17. In the situation of Definition 2.1.16, $\sigma(\mathscr{F})$ is the smallest σ-algebra Σ such that every $f \in \mathscr{F}$ is $\Sigma/\tilde{\Sigma}$-measurable. Using the principle of good sets, one sees that if $\mathcal{A}$ is a generator of $\tilde{\Sigma}$, then $\sigma(\mathscr{F})$ is generated by $\{f^{-1}(A) : f \in \mathscr{F}, A \in \mathcal{A}\}$.

Exercise 2.1.18. Let $\Omega := \{0, 1\}^2$ and $f : \Omega \to \mathbb{R}$ be given by $f(x_1, x_2) = x_1 + x_2$. Endow $\mathbb{R}$ with its Borel σ-algebra and determine $\sigma(\{f\})$. Also decide whether the map $g : \Omega \to \mathbb{R}$ given by $g(x_1, x_2) = x_1$ is $\sigma(\{f\})$-measurable.

We now prove that we can actually *characterize* the Borel σ-algebra as the one generated by the continuous functions.

Proposition 2.1.19. *Let (M, d) be a metric space and consider $\mathbb{K}$ as a measurable space endowed with the Borel σ-algebra induced by $|\cdot|$. Then $\mathscr{B}(M, d) = \sigma(C(M; \mathbb{K}))$.*

Proof. Since $f^{-1}(U)$ is open in (M, d) for all open subsets U of $\mathbb{K}$ and $f \in C(M, \mathbb{R})$, we have $\sigma(\{f^{-1}(U) : f \in C(M, \mathbb{R}), U \subset \mathbb{R} \text{ open}\}) \subset \mathscr{B}(M, d)$. By Lemma 2.1.14, $\sigma(C(M, d)) \subset \mathscr{B}(M, d)$.

Conversely, let F be a closed subset of (M, d). Then $F_n := \{x \in M : d(x, F) \geq n^{-1}\}$ is also closed and disjoint from F. By Theorem 1.7.6 there exist continuous functions $f_n : M \to [0, 1]$ such that $f_n(x) = 1$ for $x \in F$ and $f_n(x) = 0$ for $x \in F_n$. But then $F = \bigcap_{k \in \mathbb{N}} f_n^{-1}(\{1\})$. It follows that $\sigma(C(M, \mathbb{R}))$ contains all closed sets and hence $\mathscr{B}(M, d)$. □

Example 2.1.20. Consider the normed space $(C([0, 1]), \|\cdot\|_\infty)$. A set $A \subset C([0, 1])$ is called *cylinder set* if there exists $n \in \mathbb{N}$, $t_1, \ldots, t_n \in [0, 1]$ and $A_1, \ldots, A_n \in \mathscr{B}(\mathbb{K})$ such that

$$A = \{f \in C([0, 1]) : f(t_j) \in A_j \ \forall\, j = 1, \ldots, n\}.$$

Then the Borel σ-algebra $\mathscr{B}(C([0, 1]))$ is generated by the cylinder sets.

Proof. Define the evaluation maps $\pi_t : C([0, 1]) \to \mathbb{K}$ by $\pi_t(f) = f(t)$. Then π_t is a continuous function from $C([0, 1])$ to $\mathbb{K}$. Thus, by Proposition 2.1.19, $A = \bigcap_{j=1}^n \pi_{t_j}^{-1}(A_j) \in \mathscr{B}(C([0, 1])$. Denoting the σ-algebra generated by the cylinder sets by Σ, it follows that $\Sigma \subset \mathscr{B}(C([0, 1]))$.

To prove the converse inclusion, fix $g \in C([0, 1])$ and $\varepsilon > 0$ and consider $\overline{B}(g, \varepsilon) = \{f : \|f - g\|_\infty \leq \varepsilon\}$. Let $\{r_j : j \in \mathbb{N}\} = [0, 1] \cap \mathbb{Q}$. Clearly, $F := \bigcap_{j \in \mathbb{N}} \pi_{r_j}^{-1}([g(r_j) - \varepsilon, g(r_j) + \varepsilon]) \in \Sigma$.

We claim that $\overline{B}(g, \varepsilon) = F$. Indeed, the inclusion "$\subset$" is clear. To see the converse, let $f \in F$. Then $|f(r) - g(r)| \leq \varepsilon$ for all $r \in [0, 1] \cap \mathbb{Q}$. Now let $t \in [0, 1]$. By density, there exists a sequence $r_k \in [0, 1] \cap \mathbb{Q}$ converging to t. Since f and g are continuous, $|f(t) - g(t)| = \lim_{k \to \infty} |f(r_k) - g(r_k)| \leq \varepsilon$, proving that $f \in \overline{B}(g, \varepsilon)$.

Since g and ε were arbitrary, it follows that Σ contains any closed ball. But then also $B(g, \varepsilon) = \bigcup_{n \in \mathbb{N}} \bar{B}(g, (1 - n^{-1})\varepsilon) \in \Sigma$. Since $C([0, 1])$ is separable by Corollary 1.7.14, Proposition 2.1.9 yields $\mathscr{B}(C([0, 1])) \subset \Sigma$.

□

Corollary 2.1.21. $\mathscr{B}(C([0, 1])) = \sigma(\pi_t : t \in [0, 1])$.

Exercise 2.1.22. Consider the metric space (ℓ^0, d_0). Show that the Borel σ-algebra $\mathscr{B}(\ell^0)$ is generated by the coordinate maps $\pi_n : \ell^0 \to \mathbb{K}$, defined by $\pi_n(\mathbf{x}) = x_n$.

Exercise 2.1.23. We define $\mathscr{B}_t := \sigma(\pi_s : s \in [0,t])$ and $\mathscr{B}_{t+} := \bigcap_{s>t} \mathscr{B}_s$. Prove that $\mathscr{B}_t \neq \mathscr{B}_{t+}$ for all $t \in [0,1)$.

Hint: Consider the set $A_{\max}(t) = \{f \in C([0,1]) : f \text{ has a local maximum at } t\}$.

Prove that $A_{\max}(t) \in \mathscr{B}_{t+}$. Show then that if $B \in \mathscr{B}_t$, then $f \in B$ and $f(s) = g(s)$ for all $s \in [0,t]$ implies that $g \in B$. Use this to prove that $A_{\max}(t) \notin \mathscr{B}_t$.

Definition 2.1.24. Let (Ω_k, Σ_k) be a measurable space for $k = 1, \ldots, n$. The *product* of the spaces (Ω_k, Σ_k) is the measurable space $(\prod_{k=1}^n \Omega_k, \bigotimes_{k=1}^n \Sigma_k)$, where $\prod_{k=1}^n \Omega_i$ is the cartesian product of the sets Ω_k, i.e. the set of all tuples $(x_1, \ldots, x_n)$ where $x_k \in \Omega_k$ and $\bigotimes_{k=1}^n \Sigma_k$ is generated by the cuboids $A_1 \times \cdots \times A_n$ where $A_k \in \Sigma_k$.

Exercise 2.1.25. Let (M_1, d_1) and (M_2, d_2) be separable metric spaces. Show that the product σ-algebra $\mathscr{B}(M_1, d_1) \otimes \mathscr{B}(M_2, d_2)$ is the Borel σ-algebra of the product space $M_1 \times M_2$.

Exercise 2.1.26. Let (Ω, Σ) and, for $k = 1, \ldots, n$, also (Ω_k, Σ_k) be measure spaces. Let $f_k : \Omega \to \Omega_k$ be a function and define $f : \Omega \to \prod_{k=1}^n \Omega_k$ by $f(x) = (f_1(x), \ldots, f_n(x))$. Show that f is $\Sigma / \bigotimes_{k=1}^n \Sigma_k$-measurable if and only if f_k is Σ/Σ_k-measurable for all $1 \leq k \leq n$.

2.2 Measures

Definition 2.2.1. Let (Ω, Σ) be a measurable space. A *measure* on (Ω, Σ) is a map $\mu : \Sigma \to [0, \infty]$ such that

(M1) $\mu(\emptyset) = 0$.

(M2) If A_k is a sequence of pairwise disjoint sets in Σ, then $\mu(\bigcup_{k\in\mathbb{N}} A_k) = \sum_{k\in\mathbb{N}} \mu(A_k)$.

A *measure space* is a triple (Ω, Σ, μ) where (Ω, Σ) is a measurable space and μ is a measure on (Ω, Σ).

A measure space (Ω, Σ, μ) (or sometimes the measure μ) is called *finite* if $\mu(\Omega) < \infty$. If $\mu(\Omega) = 1$, then (Ω, Σ, μ) is called probability space and μ is called *probability measure*. The measure space is called *σ-finite* if there exists a sequence Ω_n with $\mu(\Omega_n) < \infty$ and $\bigcup_{n\in\mathbb{N}} \Omega_n = \Omega$.

Remark 2.2.2. The requirement (M2) is called *σ-additivity*. This assumption replaces the items 1. and 3. from the introduction. As we will see below in Proposition 2.2.4, it implies these. In fact, it is equivalent to them.

Example 2.2.3. We give some preliminary examples:

(a) Let (Ω, Σ) be a measurable space. Then $\mathbf{0} : \Sigma \to [0, \infty)$ given by $\mathbf{0}(A) = 0$ is a measure on (Ω, Σ).

(b) Let (Ω, Σ) be a measurable space and $x \in \Omega$. Then $\delta_x : \Sigma \to \{0, 1\}$ defined by $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$ defines a probability measure on (Ω, Σ), the so-called *Dirac measure* in x.

(c) Consider $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. Then

$$\zeta(A) := \begin{cases} \infty, & \text{if } A \text{ is infinite} \\ \#A, & \text{if } A \text{ is finite} \end{cases}$$

where $\#A$ is the number of elements in A, defines a σ-finite measure on $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$, the so-called *counting measure* on $\mathbb{N}$.

(d) On $\mathbb{R}$, consider $\Sigma = \{A \subset \mathbb{R} : A \text{ is countable or } A^c \text{ is countable}\}$. Define

$$\zeta(A) := \begin{cases} \infty, & \text{if } A \text{ is infinite} \\ \#A, & \text{if } A \text{ is finite} \end{cases}$$

Then ζ defines a measure on $(\mathbb{R}, \Sigma)$, which is not σ-finite.

We next collect basic properties of measures.

Proposition 2.2.4. *Let (Ω, Σ, μ) be a measure space.*

(a) If $A, B \in \Sigma$ with $A \subset B$, then $\mu(A) \le \mu(B)$. (Monotonicity)

(b) If $A, B \in \Sigma$ with $A \subset B$ and $\mu(B) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

(c) If A_k is a sequence in Σ (not necessarily disjoint), then $\mu(\bigcup_{k\in\mathbb{N}} A_k) \le \sum_{k=1}^\infty \mu(A_k)$. ($\sigma$-subadditivity).

(d) If A_k is an increasing sequence, i.e. $A_k \subset A_{k+1}$ for all $k \in \mathbb{N}$, and $A = \bigcup_{k\in\mathbb{N}} A_k$ (we write $A_k \uparrow A$), then $\mu(A_k) \uparrow \mu(A)$. (continuity from below)

(e) If $A_k \downarrow A$, i.e. $A_{k+1} \subset A_k$ for all $k \in \mathbb{N}$ and $A = \bigcap_{k\in\mathbb{N}} A_k$, and $\mu(A_1) < \infty$, then $\mu(A_k) \downarrow \mu(A)$. (continuity from above)

Proof. (a) B is the disjoint union of B and $B \setminus A$. Hence, by (M2), $\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A)$, since $\mu(A \setminus B) \ge 0$. If $\mu(B) < \infty$ then also $\mu(A) < \infty$ and subtracting $\mu(A)$, also (b) follows.

(c) Define $B_1 := A_1, B_2 := A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2), \dots B_n := A_n \setminus (A_1 \cup \dots \cup A_{n-1})$. By the properties of a σ-algebra, $B_k \in \Sigma$ for all $n \in \mathbb{N}$. Moreover, $B_k \subset A_k$ for all $k \in \mathbb{N}$, the sets B_k are disjoint and $\bigcup_{k\in\mathbb{N}} A_k = \bigcup_{k\in\mathbb{N}} B_k$. Hence, by (M2)

$$\mu\Big(\bigcup_{k=1}^\infty A_k\Big) = \mu\Big(\bigcup_{k=1}^\infty B_k\Big) = \sum_{k=1}^\infty \mu(B_k) \le \sum_{k=1}^\infty \mu(A_k)\,,$$

where we have used (a) in the last estimate.

(d) By (a) $\mu(A_k)$ increases in k. Now put $A_0 = \emptyset$ and then $B_k := A_k \setminus A_{k-1}$. Then the B_k are pairwise disjoint, belong to Σ and $\bigcup_{k\in\mathbb{N}} B_k = A$. Hence, by (M2)

$$\mu(A) = \sum_{k=1}^\infty \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k) = \lim_{n\to\infty} \mu(\bigcup_{k=1}^n B_k) = \lim_{n\to\infty} \mu(A_n)\,.$$

(e) is immediate from (b) and (d), since $A_1 \setminus A_n \uparrow A_1 \setminus A$ hence

$$\mu(A_1) - \mu(A) = \mu(A_1 \setminus A) = \lim \mu(A_1 \setminus A_n) = \mu(A_1) - \lim \mu(A_n)\,.$$

$\square$

A convenient way to obtain new measures from known ones is through measurable maps:

Lemma 2.2.5. *Let* (Ω, Σ, μ) *be a measure space and* $(\tilde{\Omega}, \tilde{\Sigma})$ *be a measurable space. If* $f : \Omega \to \tilde{\Omega})$ *is measurable, then* $\mu_f : \tilde{\Sigma} \to [0, \infty]$*, defined by* $\mu_f(A) := \mu(f^{-1}(A))$ *is a measure. It is called* the push-forward of μ under f.

Proof. This follows from the fact that $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(\bigcup_{k\in\mathbb{N}} A_k) = \bigcup_{k\in\mathbb{N}} f^{-1}(A_k)$ where the latter sets are disjoint if the A_k are. □

We now study the question, whether a measure is already uniquely determined by its values on a smaller set than Σ. Of particular importance is the question whether it is uniquely determined by its values on a generator of Σ.

Let us first look at an example.

Example 2.2.6. Let $\Omega = \{0, 1, 2, 3\}$ and $\Sigma = \sigma(\{0, 1\}, \{1, 2\})$. Then $\mu = \delta_0 + \delta_1 + \delta_2$ and $\nu = 2\delta_1 + \delta_3$ satisfy $\mu(\{0, 1\}) = \nu(\{0, 1\}) = 2$ and $\mu(\{1, 2\}) = \nu(\{1, 2\}) = 2$ but $\mu \neq \nu$, since $\mu(\{1\}) = 1 \neq 2 = \nu(\{1\})$.

The key in studying uniqueness lies in considering so-called *Dynkin systems.*

Definition 2.2.7. Let Ω be a nonempty set. A *Dynkin system* is a collection $\mathscr{D} \subset \mathscr{P}(\Omega)$ such that

(D1) $\Omega \in \mathscr{D}$.

(D2) If $A \in \mathscr{D}$ then $A^c \in \mathscr{D}$.

(D3) If A_k is a sequence of *pairwise disjoint* subsets of $\mathscr{D}$, then $\bigcup_{k\in\mathbb{N}} A_k \in \mathscr{D}$.

Obviously, every σ-algebra is a Dynkin system. In fact, the only difference between a Dynkin system and a σ-algebra that in (D3), different from (S3), the sequence A_k is required to consist of disjoint subsets. Similarly to $\sigma(\mathcal{A})$ there is a smallest Dynkin system containing a given $\mathcal{A}$, denoted by $d(\mathcal{A})$, the Dynkin system generated by $\mathcal{A}$.

Lemma 2.2.8. *A Dynkin system* $\mathscr{D}$ *is a* σ*-algebra if and only if whenever* A *and* B *belong to* $\mathscr{D}$ *then also* $A \cap B \in \mathscr{D}$*. For the latter we say that* $\mathscr{D}$ *is* stable under intersections.

Proof. Since every σ-algebra is stable under intersections (see Proposition 2.1.3), we only need to prove that if a Dynkin system is stable under intersection, then it satisfies (S3) (since (S1) and (S2) clearly hold).

So let a sequence A_k of (not necessarily disjoint) sets in $\mathscr{D}$ be given. We put $B_1 := A_1$ and then inductively $B_{k+1} := A_n \cap (B_1 \cup \cdots \cup B_n)^c$. One then proofs that the sequence B_k consists of disjoint sets, which then in turn proves that $B_k \in \mathscr{D}$, since the latter was assumed to be stable under intersections.

Moreover, $\bigcup_{k\in\mathbb{N}} A_k = \bigcup_{k\in\mathbb{N}} B_k$ and the latter belongs to $\mathscr{D}$ by the pairwise disjointness and (D3). This proves (S3). □

Lemma 2.2.9. *(Dynkin's* π*-*λ *theorem)*

Let Ω *be a nonempty set and* $\mathcal{A} \subset \mathscr{P}(\Omega)$ *be stable under intersections. Then* $d(\mathcal{A}) = \sigma(\mathcal{A})$.

Proof. Clearly, $d(\mathcal{A}) \subset \sigma(\mathcal{A})$ since $\sigma(\mathcal{A})$ is a Dynkin system containing $\mathcal{A}$. To prove the converse inclusion, by Lemma 2.2.8 it suffices to prove that $d(\mathcal{A})$ is stable under intersections.

To that end, fix $B \in \mathcal{A}$ and define $\mathscr{G}_B := \{A \in d(\mathcal{A}) : A \cap B \in d(\mathcal{A})\}$. Then $\mathscr{G}_B$ is a Dynkin system. Indeed, $\Omega \cap B = B \in d(\mathcal{A})$ proving (D1), also (D3) easily follows since if the sets A_k are disjoint, then $\bigcup_{k\in\mathbb{N}} A_k \cap B = \bigcup_{k\in\mathbb{N}} A_k \cap B$ and the latter union is also disjoint. For (D2), let $A \in \mathscr{G}_B$ and observe that $A^c \cap B = B \cap (A \cap B)^c = (B^c \cup (A \cap B))^c$. By assumption, $A \cap B \in \mathscr{G}_B$. Since $B^c \cap (A \cap B) = \emptyset$, it follows that $A^c \cap B \in \mathscr{G}_B$.

By assumption $\mathcal{A} \subset \mathscr{G}_B$ for every $B \in \mathcal{A}$ and hence $d(\mathcal{A}) \subset \mathscr{G}_B$ for all $B \in \mathcal{A}$. Now put $\mathscr{G} := \{B \in d(\mathcal{A}) : A \cap B \in d(\mathcal{A}) \ \forall A \in d(\mathcal{A})\}$. By what was done so far, $\mathcal{A} \subset \mathscr{G}$. Similarly as above, one sees that $\mathscr{G}$ is a Dynkin system. It now follows that $\mathscr{G} = d(\mathcal{A}) = \sigma(\mathcal{A}) = \Sigma$, as a consequence of Lemma 2.2.8. □

We can now prove the following result on uniqueness of measures.

Theorem 2.2.10. *Let (Ω, Σ) be a measurable space and $\mathcal{A}$ be a generator of Σ which is stable under intersections. If μ and ν are two finite measures on (Ω, Σ) with $\mu(\Omega) = \nu(\Omega)$ such that $\mu(A) = \nu(A)$ for all $A \in \mathcal{A}$, then $\mu = \nu$.*

Proof. Let $\mathscr{G} = \{A \in \Sigma : \mu(A) = \nu(A)\}$.

Then $\Omega \in \mathscr{G}$ and if $A \in \mathscr{G}$ then $A^c \in \mathscr{G}$ since $\mu(A^c) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^c)$. Moreover, if A_k is a sequence of disjoint sets in $\mathscr{G}$, then

$$\mu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k) = \sum_{k=1}^{\infty} \nu(A_k) = \nu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big)$$

since μ an ν are measures. Thus $\mathscr{G}$ is a Dynkin system. Since $\mathcal{A} \subset \mathscr{G}$ by assumption, it follows from Lemma 2.2.9 that $\Sigma = \sigma(\mathcal{A}) = d(\mathcal{A}) \subset \mathscr{G}$, proving that $\mu(A) = \nu(A)$ for all $A \in \Sigma$.

□

Corollary 2.2.11. *Let (Ω, Σ) be a measurable space and $\mathcal{A}$ be a generator of Σ with is stable under intersections such that there exists an increasing sequence $\Omega_n \in \mathcal{A}$ with $\bigcup \Omega_n = \Omega$. If μ and ν are two σ-finite measures on (Ω, Σ) such that $\mu(A) = \nu(A) < \infty$ for all $A \in \mathcal{A}$, then $\mu = \nu$.*

Proof. For fixed n, consider $\Sigma_n := \sigma(\{A \cap \Omega_n : A \in \mathcal{A}\})$. By Theorem 2.2.10, $\mu(A) = \nu(A)$ for all $A \in \Sigma_n$. Now if $A \in \Sigma$, then $\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A)$, where we have continuity from below for μ and ν and the fact that $A \cap \Omega_n \in \Sigma_n$ for all $A \in \Sigma$, which follows from the fact that $\mathcal{A}$ generates Σ. □

2.3 Construction of Measures

In order to construct measures, one first defines it on a system of sets much smaller than a σ-algebra.

Definition 2.3.1. Let Ω be a nonempty set. A *ring* on Ω is a subset $\mathscr{R}$ of $\mathscr{P}(\Omega)$ such that

(R1) $\emptyset \in \mathscr{R}$.

(R2) if $A, B \in \mathscr{R}$ then $A \cup B \in \mathscr{R}$ and $B \setminus A \in \mathscr{R}$.

A *pre-measure* is a function $\mu : \mathscr{R} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ and if A_k is a sequence of disjoint sets in $\mathscr{R}$ with $\bigcup_{k\in\mathbb{N}} A_k \in \mathscr{R}$, then $\mu(\bigcup_{k\in\mathbb{N}} A_k) = \sum_{k\in\mathbb{N}} \mu(A_k)$.

Remark 2.3.2. The properties in Proposition 2.2.4 remain valid for pre-measures, if one additionally requires that the countable unions/intersections appearing belong to $\mathscr{R}$. In particular, pre-measures are σ-subadditive and monotone.

Example 2.3.3. Let $\Omega = \mathbb{R}$ and $\mathscr{R}$ be the collection of all finite unions of bounded (possibly empty), right-open intervals. A typical element of $\mathscr{R}$ is of the form

$$[a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_n, b_n)$$

with $-\infty < a_1 \leq b_1 < a_2 \leq b_2 < \cdots < a_n \leq b_n < \infty$.

Then $\mathscr{R}$ is a ring. The map $\lambda : \mathscr{R} \to [0,\infty)$, given by

$$\lambda([a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_n, b_n)) = \sum_{j=1}^{n} (b_j - a_j)\,,$$

defines a pre-measure on $\mathscr{R}$, called *Lebesgue pre-measure.*

We should note that in the above example, $\sigma(\mathscr{R}) = \mathscr{B}(\mathbb{R})$, as is easy to see. The question arises, whether λ can be extended to a measure on $\mathscr{B}(\mathbb{R})$. It follows from the following theorem that this is the case.

Theorem 2.3.4. *(Carathéodory's extension theorem)*

Let $\mathscr{R}$ be a ring on the nonempty set Ω and μ be a pre-measure on $\mathscr{R}$. Then μ extends to a measure on $\sigma(\mathscr{R})$.

Proof. For any $B \subset E$, define

$$\mu^*(B) := \inf \sum_{n\in\mathbb{N}} \mu(A_n)$$

where the infimum is taken over all sequences $(A_k) \subset \mathbb{R}$ such that $B \subset \bigcup_{k\in\mathbb{N}} A_k$. If no such sequence exists, we put $\mu^*(B) = \infty$. We now proceed in several steps:

Step1: We prove that μ^* is σ-subadditive, i.e. If B_n is a sequence of subsets of Ω, then $\mu^*(\bigcup_{n\in\mathbb{N}} B_n) \leq \sum_{n\in\mathbb{N}} \mu^*(B_n)$.

If $\mu^*(B_n) = \infty$ for some n, there is nothing to prove, so let us assume that $\mu^*(B_n) < \infty$ for all $n \in \mathbb{N}$. By definition, for all $\varepsilon > 0$ there exist sequences $(A^{\varepsilon}_{n,m})_{m\in\mathbb{N}} \subset \mathscr{R}$ such that

(i) $B_n \subset \bigcup_{m\in\mathbb{N}} A^{\varepsilon}_{n,m}$

(ii) $\mu^*(B_n) \geq \sum_{m\in\mathbb{N}} \mu(A^{\varepsilon}_{n,m}) - \varepsilon 2^{-n}$.

In this situation, $B := \bigcup_{n\in\mathbb{N}} B_n \subset \bigcup_{n\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} A^{\varepsilon}_{n,m}$ and hence, by the definition of μ^*,

$$\mu^*(B) \leq \sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} \mu(A^{\varepsilon}_{n,m}) \leq \sum_{n\in\mathbb{N}} \big(\mu^*(B_n) + \varepsilon 2^{-n}\big) \leq \varepsilon + \sum_{n\in\mathbb{N}} \mu^*(B_n)\,.$$

Since $\varepsilon > 0$ was arbitrary, the σ-subadditivity follows.

Step 2: We prove that $\mu^*(A) = \mu(A)$ for all $A \in \mathscr{R}$.

Clearly, $\mu^*(A) \leq \mu(A)$ for $A \in \mathscr{R}$, since $A \subset A \cup \emptyset \cup \emptyset \ldots$. On the other hand, if $A \in \mathscr{R}$ and $A \subset \bigcup_{n\in\mathbb{N}} A_n$ for some sequence A_n in $\mathscr{R}$, then

$$\mu(A) \leq \sum_{k=1}^{\infty} \mu(A \cap A_k) \leq \sum_{k=1}^{\infty} \mu(A_k)$$

since μ is countably subadditive and monotone.

We now define $\mathscr{M} := \{A \subset \Omega \,:\, \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A) \;\; \forall\, B \subset \Omega\}$.

Step 3: We prove that $\mathscr{M}$ is a σ-algebra and μ^* is a measure on $(\Omega, \mathscr{M})$.

By Lemma 2.2.8, it suffices to prove that $\mathscr{M}$ is a Dynkin system and closed under intersections. Clearly, (D1) and (D2) hold. Now let A_n be a sequence of disjoint sets in $\mathscr{M}$. For $B \subset \Omega$ we have

$$\begin{aligned}
\mu(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\
&= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c)\,,
\end{aligned}$$

since A_1 and A_2 are disjoint, hence $A_1 \cap A_2 = \emptyset, A_1 \cap A_2^c = A_1$ and $A_1^c \cap A_2 = A_2$.

Proceeding in a similar way, we obtain for all $n \in \mathbb{N}$ that

$$\mu^*(B) = \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c) + \sum_{k=1}^{n} \mu^*(B \cap A_k) \geq \mu^*(B \cap (\bigcup_{k=1}^{\infty} A_k)^c) \sum_{k=1}^{n} \mu^*(B \cap A_k)\,.$$

Hence, upon $n \to \infty$, we find

$$\mu^*(B) \geq \sum_{k=1}^{\infty} \mu^*(B \cap A_k) + \mu^*(B \cap (\bigcup_{k\in\mathbb{N}} A_k)^c) \geq \mu^*(B \cap \bigcup_{k\in\mathbb{N}} A_k) + \mu^*(B \cap (\bigcup_{k\in\mathbb{N}} A_k)^c)$$

where we have used Step 1 in the last estimate. Since the reverse inequality holds by subadditivity, $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{M}$, i.e. (D3) holds. Moreover, taking $B = \Omega$, we have proved the σ-additivity of μ^* on $\mathscr{M}$.

It thus remains to prove that $\mathscr{M}$ is closed under intersections. To that end, let $A_1, A_2 \in \mathscr{M}$. Then, for $B \subset \Omega$, we have

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \quad \text{since}\, A_1 \in \mathscr{M} \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \quad \text{since}\, A_2 \in \mathscr{M} \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\
&\qquad \text{since}\, A_1^c \subset (A_1 \cap A_2)^c \\
&= \mu^*(B \cap A_2 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\
&\qquad \text{since}\, B \cap (A_1 \cap A_2)^c \cap A_1 = B \cap A_1 \cap A_1^c \cup B \cap A_2 \cap A_1^c = B \cap A_2 \cap A_1^c \\
&= \mu^*(B \cap (A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c) \quad \text{since}\, A_1 \in \mathscr{M}\,.
\end{aligned}$$

This proves that $A_1 \cap A_2 \in \mathscr{M}$.

Step 4: We finish the proof.

It remains to prove that $\mathscr{R} \subset \mathscr{M}$, for this implies $\sigma(\mathscr{R}) \subset \mathscr{M}$ and thus, since μ^* is a measure on $\mathscr{M}$ (by Step 3), it is a measure on $\sigma(\mathscr{R})$ which, by Step 2, extends μ.

Thus, let $A \in \mathscr{R}$ be given. We need to prove that $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A)$ for all $B \subset \Omega$. In fact, by subadditivity, it suffices to prove that $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \setminus A)$ for all $B \subset \Omega$.

If $\mu^*(B) = \infty$, there is nothing to prove, so assume that $\mu^*(B) < \infty$: Given $\varepsilon > 0$, we find a sequence $(A_k)_{k\in\mathbb{N}} \subset \mathscr{R}$ such that $B \subset \bigcup_{k\in\mathbb{N}} A_k$ and $\mu^*(B) \geq \sum_{k\in\mathbb{N}} \mu(A_k) - \varepsilon$. Noting that $B \cap A \subset \bigcup_{k\in\mathbb{N}} A_k \cap A$ and $B \setminus A \subset \bigcup A_k \setminus A$ and that $A_k \cap A$ and $A_k \setminus A$ belong to $\mathscr{R}$ for all $k \in \mathbb{N}$, it follows that

$$\mu^*(B \cap A) + \mu^*(B \setminus A) \leq \sum_{k\in\mathbb{N}} \mu(A_n \cap A) + \sum_{k\in\mathbb{N}} \mu(A_n \setminus A) = \sum_{k\in\mathbb{N}} \mu(A_k) \leq \mu^*(B) + \varepsilon\,.$$

Since $\varepsilon > 0$ was arbitrary, we are done.

□

Corollary 2.3.5. *There exists a unique measure λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with $\lambda([a, b)) = b - a$ for all $b > a$.*

Proof. Apply Theorem 2.3.4 in Example 2.3.3 to obtain existence of such a measure. Uniqueness follows from Corollary 2.2.11. □

Definition 2.3.6. The measure λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with $\lambda([a, b)) = b - a$ for all $b > a$ is called *Lebesgue measure.*

It is also possible to define d-dimensional versions of Lebesgue measure:

Example 2.3.7. Similar as in the one-dimensional situation, one can prove that there exists a unique measure λ_d on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that

$$\lambda_d([a_1, b_1) \times \cdots \times [a_d, b_d)) = (b_1 - a_1) \cdot \ldots \cdot (b_d - a_d)\,.$$

λ_d is called *d-dimensional Lebesgue measure.*

2.4 Measurable Functions

We have already defined the notion of measurable map from a measurable space (Ω_1, Σ_1) to a second measurable space (Ω_2, Σ_2). Of particular importance is the situation where $(\Omega_2, \Sigma_2) = (\mathbb{K}, \mathscr{B}(\mathbb{K}))$. Here, $\mathbb{K}$ is as before either $\mathbb{R}$ or $\mathbb{C}$ and $\mathscr{B}(\mathbb{K})$ is the Borel σ-algebra generated by $|\cdot|$.

Definition 2.4.1. Let (Ω, Σ) be a measurable space. A *measurable function* is a measurable map from (Ω, Σ) to $(\mathbb{K}, \mathscr{B}(\mathbb{K}))$.

Example 2.4.2. Let Ω be a set and $A \subset \Omega$. The *indicator function of A* is the function $\mathbb{1}_A : \Omega \to \mathbb{R}$ defined by $\mathbb{1}_A(x) = 1$ iff $x \in A$ and $\mathbb{1}_A(x) = 0$ iff $x \notin A$.

If (Ω, Σ) is a measurable space, then $\mathbb{1}_A$ is a measurable function if and only if $A \in \Sigma$. Indeed, If $S \in \mathscr{B}(\mathbb{R})$, then

$$\mathbb{1}_A^{-1}(S) = \begin{cases} \emptyset, & \text{if } S \cap \{0, 1\} = \emptyset \\ A, & \text{if } 1 \in S \text{ and } 0 \notin S \\ A^c, & \text{if } 1 \notin S \text{ and } 1 \in S \\ \Omega, & \text{if } 1 \in S \text{ and } 0 \in S\,. \end{cases}$$

Proposition 2.4.3. *Let (Ω, Σ) be a measurable space, $f, g : \Omega \to \mathbb{K}$ be measurable and $\lambda \in \mathbb{K}$. Then $\lambda f, f \cdot g, f + g$ and, if $\mathbb{K} = \mathbb{R}$, $f \vee g$ and $f \wedge g$ are measurable. Moreover, if f_n is a sequence of measurable functions from Ω to $\mathbb{K}$ which converges pointwise to the measurable function f, i.e. $f_n(x) \to f(x)$ for all $x \in \Omega$, then f is measurable.*

Proof. We define $\Phi : \Omega \to \mathbb{K}^2$ by $\Phi(x) = (f(x), g(x))$. Then Φ is measurable by Exercise 2.1.26. Moreover, the maps $\mathsf{add}, \mathsf{mult}, \vee, \wedge : \mathbb{K}^2 \to \mathbb{K}$ are continuous and hence measurable. Thus the first assertion follows since the composition of measurable maps is measurable.

For the second assertion, let C be a closed subset of $\mathbb{K}$. We put $C^\varepsilon := \{x \in \mathbb{K} : d(x, C) \leq \varepsilon\}$. Since $f_C : x \mapsto d(x, C)$ is continuous $C^\varepsilon = f_C^{-1}([0, \varepsilon])$ is a closed set as a consequence of Proposition 1.4.17. We claim that

$$f^{-1}(C) = \bigcap_{k \in \mathbb{N}} \bigcup_{n_0 \in \mathbb{N}} \bigcap_{n \geq n_0} f_n^{-1}(C^{\frac{1}{k}}).$$

Note that this finishes the proof since the set on the right-hand side belongs to Σ.

It remains to prove the claim. First assume $x \in f^{-1}(C)$. Then $f(x) \in C$. Since $f_n(x) \to f(x)$ given $k \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that $|f_n(x) - f(x)| \leq k^{-1}$ for all $n \geq n_0$. Thus, $f_n(x) \in C^{\frac{1}{k}}$ for all $n \geq n_0$, proving that x belongs to the set on the right-hand side.

Conversely assume that x belongs to the set on the right-hand side. This means that for all $k \in \mathbb{N}$ there exists an $n_0 \in \mathbb{N}$ such that $f_n(x) \in C^{\frac{1}{k}}$ for all $n \geq n_0$. Since $C^{\frac{1}{k}}$ is closed and $f_n(x) \to f(x)$, it follows that $f(x) \in C^{\frac{1}{k}}$ for all $k \in \mathbb{N}$. But then $f(x) \in \bigcap_{k \in \mathbb{N}} C^{\frac{1}{k}} = C$, since C is closed. □

Definition 2.4.4. Let (Ω, Σ) be a measurable space. A *simple function* is a measurable function $f : \Omega \to \mathbb{K}$ taking finitely many values.

The following Lemma gives a description of simple functions.

Lemma 2.4.5. *Let (Ω, Σ) be a measurable space and $f : \Omega \to \mathbb{K}$ be a simple function. Then*

$$f(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$$

where $f(\Omega) = \{a_1, \ldots, a_n\}$ and $A_k = f^{-1}(\{a_k\}) \in \Sigma$. Note that in this case, the sets A_k are disjoint and the a_k are all distinct. We call this the standard representation *of f.*

Proposition 2.4.6. *Let (Ω, Σ) be a measurable space and $f : \Omega \to \mathbb{K}$ be a measurable function. Then there exists a sequence of simple functions $f_n : \Omega \to \mathbb{K}$ with $|f_n(x)| \leq 2|f(x)|$ for all $n \in \mathbb{N}$ which converges pointwise to f. Moreover, if $\mathbb{K} = \mathbb{R}$, then f_n can be chosen as real functions. If $f \geq 0$, then the sequence can be chosen to consist of positive functions and to be increasing, i.e. $f_n(x) \leq f_{n+1}(x) \uparrow f(x)$.*

Proof. Let us first consider the case where $\mathbb{K} = \mathbb{R}$. For $n \in \mathbb{N}, k \in \mathbb{N}_0$, define

$$A_{n,k} := [k2^{-n}, (k+1)2^{-n}) \quad \text{and} \quad B_{n,k} := [-(k+1)2^{-n}, -k2^{-n}).$$

Note that $\mathbb{R} = \bigcup_{k \in \mathbb{N}_0} A_{n,k} \cup B_{n,k}$ for all $n \in \mathbb{N}$.

Now put

$$f_n := \sum_{k=0}^{2^{2^n}} \frac{k}{2^n} \mathbb{1}_{f^{-1}(A_{n,k})} - \sum_{k=0}^{2^{2^n}} \frac{k}{2^n} \mathbb{1}_{f^{-1}(B_{n,k})}$$

Then f_n is a simple function and positive whenever f is positive (the latter follows from the fact that in this case $f^{-1}(B_{n,k}) = \emptyset$ for all k, n).

Moreover, $f_n(x) \to f(x)$. Indeed, if $x \in \Omega$ then there exists an n_0 such that $|f(x)| \leq 2^{n_0}$. Then $|f_n(x) - f(x)| \leq 2^{-n}$ for all $n \geq n_0$.

If $f \geq 0$ then $f_n \leq f_{n+1}$. Indeed, if $f_n(x) = k2^{-n}$ then $f(x) \in [k2^{-n}, (k+1)2^{-n})$. But then $f(x) \in [(2k)2^{-(n+1)}, (2k+1)2^{-(n+1)})$, in which case $f_{n+1}(x) = (2k)2^{-(n+1)} = f(x)$, or $f(x) \in [(2k+1)2^{-(n+1)}, (2k+2)2^{-(n+1)})$, in which case $f_{n+1}(x) = (2k+1)2^{-(n+1)} > f(x)$.

In the case where $\mathbb{K} = \mathbb{C}$, we find simple functions g_k and h_k which converge to $\operatorname{Re} f$ and $\operatorname{Im} f$ respectively and put $f_n := g_k + ih_k$. □

We will sometimes need a somewhat more general result.

Remark 2.4.7. (Extended real line)

We put $\bar{\mathbb{R}} := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ which we endow with the σ-algebra $\mathscr{B}(\bar{\mathbb{R}})$, defined as $\sigma(\mathscr{B}(\mathbb{R}) \cup \{\{-\infty\}, \{\infty\}\})$. Then a function $f : (\Omega, \Sigma) \to \bar{\mathbb{R}}$ is measurable if and only if $f^{-1}(\{\infty\}), f^{-1}(\{-\infty\}) \in \Sigma$ and $f^{-1}(A) \setminus \{\pm\infty\} \in \Sigma$ for all $A \in \mathscr{B}(\mathbb{R})$.

Similarly, $\mathscr{B}([0,\infty])$ is defined as $\sigma(\mathscr{B}([0,\infty)) \cup \{\infty\})$. We note that there exist metrics on $\bar{\mathbb{R}}$ resp. $[0,\infty]$ such that $\mathscr{B}(\bar{\mathbb{R}})$ resp. $\mathscr{B}([0,\infty])$ is the Borel σ-algebra for this metric. We also remark that Proposition 2.4.6 generalizes to this situation.

2.5 The Lebesgue Integral

Given a measure space (Ω, Σ, μ) we now introduce the Lebesgue integral $\int_\Omega f\, d\mu$ for complex-valued, measurable functions which are "integrable". We proceed in several steps and first define the integral for (real-valued) measurable functions f taking values in $[0,\infty]$.

Definition 2.5.1. Let (Ω, Σ, μ) be a measure space. If $f : \Omega \to [0,\infty]$ is a simple function, with standard representation $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$, we put

$$\int_\Omega f\, d\mu := \sum_{k=1}^n a_k \mu(A_k)$$

where we set, by convention, $0 \cdot \infty := 0$ and $\infty + \infty = \infty$.

Lemma 2.5.2. *If (Ω, Σ, μ) is a measure space, $f, g : \Omega \to [0,\infty]$ are simple functions and $\lambda \in (0,\infty)$, then $\int_\Omega \lambda f + g\, d\mu = \lambda \int_\Omega f\, d\mu + \int_\Omega g\, d\mu$. Moreover, if $f \leq g$, then $\int_\Omega f\, \mu \leq \int_\Omega g\, d\mu$.*

Proof. If $\sum_{k=1}^n a_k \mathbb{1}_{A_k}$ is the standard representation of f, then $\sum_{k=1}^n \lambda a_k \mathbb{1}_{A_k}$ is the standard representation of λf. Now $\lambda \int_\Omega f\, d\mu = \int_\Omega \lambda f\, d\mu$ follows from the definition of the integral. Now let $\lambda = 1$ and $\sum_{l=1}^m b_l \mathbb{1}_{B_k}$ be the standard representation of g. If $\sum_{j=1}^w c_j \mathbb{1}_{C_j}$ is the standard representation of $f + g$, then $C_j = \bigcup_{a_k + b_l = c_j} A_k \cap B_l$. Thus

$$\begin{aligned}\int_\Omega f\, d\mu + \int_\Omega g\, d\mu &= \sum_{k=1}^n a_k \mu(A_k) + \sum_{l=1}^n b_l \mu(B_l) = \sum_{k=1}^n \sum_{l=1}^m (a_k + b_l)\mu(A_k \cap B_l) \\ &= \sum_{j=1}^w c_k \mu(C_j) = \int_\Omega f + g\, d\mu\,,\end{aligned}$$

where we have used the finite additivity of μ and the fact that the sets $A_k \cap B_l$ are pairwise disjoint.

Now assume that $f \leq g$. Consider the function $g - f$, with the convention that $\infty - \infty = \infty - c = \infty$ for all $c \in [0, \infty)$. Then $f - g$ is a nonnegative, simple function. Hence,

$$\int_\Omega g\, d\mu = \int_\Omega g - f + f\, d\mu = \int_\Omega g - f\, d\mu + \int_\Omega f\, d\mu \geq \int_\Omega f\, d\mu\,,$$

since the integral of a nonnegative, simple function is clearly nonnegative. □

We can now define the Lebesgue integral for an arbitrary measurable function $f : \Omega \to [0, \infty]$.

Definition 2.5.3. Let (Ω, Σ, μ) be a measure space and $f : \Omega \to [0, \infty]$ is measurable. We put

$$\int_\Omega f\, d\mu = \sup \Big\{ \int_\Omega g\, d\mu \, : \, 0 \leq g \leq f\,,\, g\,\text{simple} \Big\}\,.$$

We say that f is *integrable*, if $\int_\Omega f\, d\mu < \infty$.

Example 2.5.4. If δ_x is Dirac measure on (Ω, Σ) where $x \in \Omega$, then for all measurable $f : \Omega \to [0, \infty]$ we have

$$\int_\Omega f\, d\delta_x = f(x)$$

and f is integrable iff $f(x) < \infty$.

Proof. First, let f be a simple function. Then $\delta_x\big[f^{-1}(\{t\})\big] = 1$ if $x \in f^{-1}(\{t\})$, i.e. $f(x) = t$ and $\delta_x\big[f^{-1}(\{t\})\big] = 0$ else. Thus in this case $\int_\Omega f\, d\delta_x = f(x)$.

Now let $f : \Omega \to [0, \infty]$ be a measurable function. If g is a simple function with $g \leq f$, then

$$\int_\Omega g\, d\delta_x = g(x) \leq f(x)\,.$$

Taking the supremum over such g, it follows that $\int_\Omega f\, d\mu \leq f(x)$.

For the reversed inequality, let $g(y) := f(x)\mathbb{1}_{f^{-1}(\{f(x)\})}(y)$. Since since f is measurable and $\{f(x)\} \in \mathscr{B}(\mathbb{R})$, $f^{-1}(\{f(x)\}) \in \Sigma$ so that g is a simple function. Moreover, $g \leq f$. Hence, $\int_\Omega f\, d\delta_x \geq \int_\Omega g\, d\delta_x = f(x)$. □

Exercise 2.5.5. Consider the measure space $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \zeta)$. A measurable function $a : \mathbb{N} \to [0, \infty]$ is a sequence $(a_n)_{n\in\mathbb{N}} = (a(n))_{n\in\mathbb{N}}$ in $[0, \infty]$. Show that a is integrable iff $a \in \ell^1$ and in this case,

$$\int_\mathbb{N} a_n\, d\zeta = \sum_{n\in\mathbb{N}} a_n\,.$$

Theorem 2.5.6. *(Monotone convergence theorem)*

Let (Ω, Σ, μ) be a measure space and $f_n : \Omega \to [0, \infty]$ be an increasing sequence of measurable functions. Then $f(x) := \sup_{n\in\mathbb{N}} f_n(x)$ is measurable and $\int_\Omega f\, d\mu = \sup_{n\in\mathbb{N}} \int_\Omega f_n\, d\mu$.

Proof. As a consequence of Proposition 2.4.3, f is measurable. Now fix $k \in \mathbb{N}$. If g is a simple function with $g \leq f_k$, then $g \leq f_n$ for all $n \geq k$ and thus $g \leq f$. By definition $\int_\Omega g\, d\mu \leq \int_\Omega f\, d\mu$. Taking the supremum over simple functions $g \leq f_k$, it follows that $\int_\Omega f_k\, d\mu \leq \int_\Omega f\, d\mu$. This being true for all f, it follows that $\sup_{k\in\mathbb{N}} \int_\Omega f_k\, d\mu \leq \int_\Omega f\, d\mu$.

To prove the converse inequality let g be a simple function with $g \leq f$. For $\varepsilon > 0$ let

$$B_n^\varepsilon := \{x \, : \, (1 + \varepsilon)f_n(x) \geq g(x)\}\,.$$

Note that B_n^ε is measurable, since the function $\varphi := (1+\varepsilon)f_n - g$ is measurable by Proposition 2.4.3 and $B_n^\varepsilon = \varphi^{-1}([0,\infty))$.

Since the sequence f_n is increasing, we have we have $B_n^\varepsilon \subset B_{n+1}^\varepsilon$ for all $n \in \mathbb{N}$ and $\varepsilon > 0$. Moreover, $\varepsilon > 0$ implies $\bigcup_{n\in\mathbb{N}} B_n^\varepsilon = \Omega$. Indeed, if $x \in \Omega \setminus \bigcup_{n\in\mathbb{N}} B_n^\varepsilon$, then $(1+\varepsilon)f_n(x) < g(x)$ for all $n \in \mathbb{N}$ and consequently, since $f_n(x) \to f(x)$ as $n \to \infty$, it follows that $(1+\varepsilon)f(x) \leq g(x)$ and thus $f(x) < g(x)$ — a contradiction.

We now have

$$\int_\Omega g\mathbb{1}_{B_n^\varepsilon}\, d\mu \leq (1+\varepsilon)\int_\Omega f_n \mathbb{1}_{B_n^\varepsilon}\, d\mu \leq (1+\varepsilon)\int_\Omega f_n\, d\mu \leq (1+\varepsilon)\sup_{k\in\mathbb{N}}\int_\Omega f_k\, d\mu\,.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\int_\Omega g\mathbb{1}_{B_n}\, d\mu \leq \sup_{k\in\mathbb{N}}\int_\Omega f_k\, d\mu$. We claim that

$$\int_\Omega g\, d\mu = \sup_{n\in\mathbb{N}}\int_\Omega g\mathbb{1}_{B_n}\, d\mu\,.$$

To see this, let $\sum_{k=1}^K a_k\mathbb{1}_{A_k}$ be the standard representation of g. Then the standard representation of $g\mathbb{1}_{B_n}$ is $\sum_{k=1}^K a_k\mathbb{1}_{A_k\cap B_n}$. Since $A_k \cap B_n \uparrow A_k$ as $n \to \infty$, the claim follows at once from the definition of the integral for simple functions and Proposition 2.2.4 (d).

Consequently, $\int_\Omega g\, d\mu \leq \sup_{n\in\mathbb{N}} f_n\, d\mu$. Since $g \leq f$ was arbitrary, it follows that $\int_\Omega f\, d\mu \leq \sup_{n\in\mathbb{N}} f_n\, d\mu$. □

Corollary 2.5.7. *Let (Ω,Σ,μ) be a measure space, $f,g : \Omega \to [0,\infty]$ and $\lambda > 0$. Then $\int_\Omega \lambda f + g\, d\mu = \lambda\int_\Omega f\, d\mu + \int_\Omega g\, d\mu$. Moreover, if $f \leq g$, then $\int_\Omega f\, d\mu \leq \int_\Omega g\, d\mu$.*

Proof. Let (f_n) and (g_n) be sequences of simple functions such that $f_n \uparrow f$ and $g_n \uparrow g$. Then $\lambda f_n + g_n \uparrow \lambda f + g$. Using monotone convergence and Lemma 2.5.2, we obtain

$$\int_\Omega \lambda f + g\, d\mu = \lim_{n\to\infty}\int_\Omega \lambda f_n + g_n\, d\mu = \lim_{n\to\infty}\lambda\int_\Omega f_n\, d\mu + \int_\Omega g_n\, d\mu = \lambda\int_\Omega f\, d\mu + \int_\Omega g\, d\mu$$

The second assertion was seen in the proof of Theorem 2.5.6 □

Definition 2.5.8. Let (Ω,Σ,μ) be a measure space. A *null set* is a set $N \subset \Omega$ such that there exists a set $M \in \Sigma$ with $N \subset M$ and $\mu(M) = 0$.

Now let $p = p(x)$ be a property which, depending on x, may be true or false. We say that p holds *almost everywhere* or *for almost every* $x \in \Omega$, if μ is a probability measure we also say also *almost surely*, if $\{x \in \Omega : p(x) \text{ is false}\}$ is a null set.

Example 2.5.9. Consider $(\mathbb{R},\mathscr{B}(\mathbb{R}),\delta_0)$. Then almost every $x \in \mathbb{R}$ is equal to 0.

Consider $(\mathbb{R},\mathscr{B}(\mathbb{R}),\lambda)$. Then almost every $x \in \mathbb{R}$ is not equal to 0. In fact, almost every $x \in \mathbb{R}$ is irrational.

Corollary 2.5.10. *Let (Ω,Σ,μ) be a measure space and $f : \Omega \to [0,\infty]$ be measurable. Then $\int_\Omega f\, d\mu = 0$ if and only if $f(x) = 0$ for almost every $x \in \Omega$.*

Proof. If f is a simple function, then the assertion is obvious.

In the general case, let a measurable $f : \Omega \to [0,\infty]$ with $\int_\Omega f\, d\mu = 0$ be given. Let f_n be an increasing sequence of simple functions converging to f. Such a sequence exists by Proposition 2.4.6. By definition, $\int_\Omega f_n\, d\mu = 0$ for all $n \in \mathbb{N}$. Hence, by the case above, $\{x \in \Omega : f_n(x) \neq 0\}$ is a null set, whence there exists $M_n \in \Sigma$ with $\mu(M_n) = 0$ such that

$f_n(x) = 0$ for all $x \notin M_n$. Put $M := \bigcup_{n\in\mathbb{N}} M_n$. Then $M \in \Sigma$ and $\mu(M) \le \sum_{n\in\mathbb{N}} \mu(M_n) = 0$. Moreover, $x \notin M$ implies that $f(x) = 0$. Hence $f = 0$ almost everywhere.

If, conversely, $f = 0$ almost everywhere, then there exists a measurable set M with $f(x) = 0$ for all $x \notin M$. If g is a simple function with $g \le f$ then $g^{-1}(\{x\}) \subset M$ for all $x > 0$. By the definition of the integral for simple functions, $\int_\Omega g\,d\mu = 0$ and hence, since $g \le f$ was arbitrary, $\int_\Omega f\,d\mu = 0$. ☐

Exercise 2.5.11. Let (Ω, Σ, μ) be a measure space and $f : \Omega \to [0, \infty]$ be measurable. Show that $\nu : \Sigma \to [0, \infty]$, defined by

$$\nu(A) := \int_A f\,d\mu := \int_\Omega \mathbb{1}_A f\,d\mu$$

defines a measure on (Ω, Σ).

Theorem 2.5.12. *(Fatou's Lemma)*

Let (Ω, Σ, μ) be a measure space, $f_n : \Omega \to [0, \infty]$ be measurable and put $f(x) := \liminf_{n\to\infty} f_n(x) (:= \sup_{k\ge1} \inf_{n\ge k} f_n(x))$. Then f is measurable and

$$\int_\Omega f\,d\mu \le \liminf_{n\to\infty} \int_\Omega f_n\,d\mu\,.$$

Proof. Let $g_k(x) := \inf_{n\ge k} f_n(x)$. Then g_k is measurable as the limit of the functions $h_{k,w} := f_k \wedge f_{k+1} \wedge \ldots \wedge f_{k+w}$ which are themselves measurable by Proposition 2.4.3.

Then $g_k \uparrow f$ and thus, by monotone convergence, $\sup_{k\in\mathbb{N}} \int_\Omega g_k\,d\mu = \int_\Omega f\,d\mu$. On the other hand,

$$\sup_{k\in\mathbb{N}} \int_\Omega g_k\,d\mu \le \liminf_{n\to\infty} \int_\Omega f_n\,d\mu\,.$$

Indeed, for every $n \ge k$ we have $g_k \le f_n$ and thus $\int_\Omega g_k\,d\mu \le \int_\Omega f_n\,d\mu$. Since this is true for all $n \ge k$, we have $\int_\Omega g_k\,d\mu \le \inf_{n\ge k} \int_\Omega f_n\,d\mu$. Taking the supremum over $k \in \mathbb{N}$ on both sides the above inequality follows. ☐

Definition 2.5.13. Let (Ω, Σ, μ) be a measure space, $f : \Omega \to \mathbb{K}$ be measurable. Then f is called *integrable* if $\int_\Omega |f|\,d\mu < \infty$. We write $f \in \mathscr{L}^1(\Omega, \Sigma, \mu)$.

If $\mathbb{K} = \mathbb{R}$, we put

$$\int_\Omega f\,d\mu = \int_\Omega f^+\,d\mu - \int_\Omega f^-\,d\mu\,.$$

Note that if $|f|$ is integrable then f^+ and f^- are both integrable nonnegative functions. If $\mathbb{K} = \mathbb{C}$, we put

$$\int_\Omega f\,d\mu = \int_\Omega \operatorname{Re} f\,d\mu + i \int_\Omega \operatorname{Im} f\,d\mu\,.$$

Note that if $|f|$ is integrable, then $\operatorname{Re} f$ and $\operatorname{Im} f$ are both integrable real-valued functions.

If f is an integrable (real or complex) measurable function and $A \in \Sigma$, we define

$$\int_A f\,d\mu := \int_\Omega f \mathbb{1}_A\,d\mu$$

Lemma 2.5.14. *Let (Ω, Σ, μ) be a measure space.*

(a) For all integrable f, we have $\left|\int_\Omega f\,d\mu\right| \le \int_\Omega |f|\,d\mu$.

(b) If f is integrable and $\lambda \in \mathbb{K}$, then λf is integrable and $\int_\Omega \lambda f\, d\mu = \lambda \int_\Omega f\, d\mu$.

(c) If f and g are integrable, then $f+g$ is integrable and $\int_\Omega f + g\, d\mu = \int_\Omega f\, d\mu + \int_\Omega g\, d\mu$.

Remark 2.5.15. (b) and (c) can be expressed by saying that the integrable functions form a vector space and the map Int, defined by $\mathsf{Int}(f) := \int_\Omega f\, d\mu$ is a linear map from the integrable functions to $\mathbb{K}$.

Proof. Let us first consider the case where $\mathbb{K} = \mathbb{R}$.

(a) We have

$$\Big|\int_\Omega f\, d\mu\Big| = \Big|\int_\Omega f^+\, d\mu - \int_\Omega f^-\, d\mu\Big| \leq \int_\Omega f^+\, d\mu + \int_\Omega f^-\, d\mu = \int_\Omega |f|\, d\mu$$

where we have used Corollary 2.5.7 in the last step.

(b) First note that by Corollary 2.5.7 $\int_\Omega |\lambda f|\, d\mu = |\lambda| \int_\Omega |f|\, d\mu < \infty$ if f is integrable. This proves that λf is integrable whenever f is.

Now, if $\lambda > 0$, then $(\lambda f)^+ = \lambda f^+$ and $(\lambda f)^- = \lambda f^-$. Thus, using Corollary 2.5.7

$$\int_\Omega \lambda f\, d\mu = \int_\Omega \lambda f^+\, d\mu - \int_\Omega \lambda f^-\, d\mu = \lambda \int_\Omega f^+\, d\mu - \lambda \int_\Omega f^-\, d\mu = \lambda \int_\Omega f\, d\mu\,.$$

If, on the other hand, $\lambda < 0$, then $(\lambda f)^+ = -\lambda f^-$ and $(\lambda f)^- = -\lambda f^+$. Thus, in this case,

$$\int_\Omega \lambda f\, d\mu = \int_\Omega -\lambda f^-\, d\mu - \int_\Omega -\lambda f^+\, d\mu = \lambda \int_\Omega f^+\, d\mu - \lambda \int_\Omega f^-\, d\mu = \lambda \int_\Omega f\, d\mu\,.$$

(c) Since $|f+g| \leq |f| + |g|$, it follows that $\int_\Omega |f+g|\, d\mu \leq \int_\Omega |f| + |g|\, d\mu = \int_\Omega |f|\, d\mu + \int_\Omega |g|\, d\mu < \infty$ if f and g are integrable.

Moreover, by definition and Corollary 2.5.7,

$$\begin{aligned}
\int_\Omega f\, d\mu + \int_\Omega g\, d\mu &= \int_\Omega f^+\, d\mu - \int_\Omega f^-\, d\mu + \int_\Omega g^+\, d\mu - \int_\Omega g^-\, d\mu \\
&= \int_\Omega f^+ + g^+\, d\mu - \int_\Omega (f^- + g^-)\, d\mu \\
&\overset{(*)}{=} \int_\Omega (f+g)^+\, d\mu - \int_\Omega (f+g)^-\, d\mu \\
&= \int_\Omega f + g\, d\mu\,.
\end{aligned}$$

Here, $(*)$ follows from integrating the identity $f^+ + g^+ + (f+g)^- = (f+g)^+ + f^- + g^-$ and using Corollary 2.5.7.

In the case where $\mathbb{K} = \mathbb{C}$, for (b) and (c) one uses that $\int_\Omega f\, d\mu = \int_\Omega \operatorname{Re} f\, d\mu + i \int_\Omega \operatorname{Im} f\, d\mu$. We omit the easy computations. For (a), we use that for every complex number z we have $|z| = \sup_{t\in\mathbb{R}} \operatorname{Re}(e^{it} z)$. Now for $t \in \mathbb{R}$ we have

$$\operatorname{Re}\Big(e^{it}\int_\Omega f\, d\mu\Big) = \operatorname{Re}\int_\Omega e^{it} f\, d\mu = \int_\Omega \operatorname{Re} e^{it} f\, d\mu \leq \int_\Omega |f|\, d\mu\,.$$

Taking the supremum over $t \in \mathbb{R}$, (a) follows. □

Theorem 2.5.16. *(Dominated convergence theorem)*

Let (Ω, Σ, μ) be a measurable space, f_n be a sequence of integrable functions with *(a)* $\tilde{f}(x) := \lim_{n\to\infty} f_n(x)$ *exists for almost every* $x \in \Omega$, *say outside the set* $N \in \Sigma$ *with* $\mu(N) = 0$.

(b) There exists an integrable function g *with* $|f_n(x)| \leq g(x)$ *for almost every* $x \in \Omega$. *Then* $f : \Omega \to \mathbb{K}$, *defined by* $f(x) = \tilde{f}(x)$ *if* $x \notin N$ *and* $f(x) = 0$ *if* $x \in N$, *is integrable and*

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu \, .$$

Proof. Changing f_n and f on a set of measure zero, we may assume that (a) and (b) hold everywhere. By Proposition 2.2.4, f is measurable. Since $|f| \leq g$, it follows that f is integrable.

Now observe that $|f_n - f| \leq 2g$ and hence $2g - |f_n - f| \geq 0$. By Fatou's Lemma 2.5.12,

$$\begin{aligned}\int_\Omega 2g \, d\mu &= \int_\Omega \liminf_{n\to\infty} 2g - |f_n - f| \, d\mu \leq \liminf_{n\to\infty} \int_\Omega 2g - |f_n - f| \, d\mu \\ &= \int_\Omega 2g \, d\mu - \limsup_{n\to\infty} \int_\Omega |f_n - f| \, d\mu \, .\end{aligned}$$

It follows that $\limsup_{n\to\infty} \int_\Omega |f_n - f| \, d\mu = 0$ and thus $\lim_{n\to\infty} \int_\Omega |f_n - f| \, d\mu = 0$. By Lemma 2.5.14, $\big| \int_\Omega f_n \, d\mu - \int_\Omega f \, d\mu \big| \leq \int_\Omega |f_n - f| \, d\mu \to 0$. This proves the claim. □

Example 2.5.17. Let us give some example, that condition (b) in Theorem 2.5.16 is necessary.

Consider $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$. If we put $f_n := n \mathbb{1}_{(0,n^{-1})}$ then f_n is a sequence of simple functions converging to 0 everywhere. However, $\int_\mathbb{R} f_n \, d\lambda \equiv 1 \not\to 0 = \int_\mathbb{R} 0 \, d\lambda$.

We next compare the Lebesgue integral with the Riemann integral. Let us first recall the definition of the Riemann integral on a interval $[a, b]$, where $-\infty < a < b < \infty$.

A *partition* of $[a, b]$ is a finite sequence $\pi := (a : t_0 < t_1 < \cdots < t_n = b)$. The *mesh size* of π is $|\pi| := \max_{1\leq j\leq n} |t_j - t_{j-1}|$. Given a partition π of $[a, b]$, an *associated vector of sample points* (or associated sample points) is a vector $\xi = (\xi_1, \ldots, \xi_n)$ with $\xi_j \in (t_{j-1}, t_j)$. Given a function $f : [a, b] \to \mathbb{R}$, a partition π of $[a, b]$ and a vector of associated sample points ξ, the *Riemann sum* $S(f, \pi, \xi)$ is defined by

$$S(f, \pi, \xi) := \sum_{j=1}^{n} f(\xi_j)(t_j - t_{j-1}) \, .$$

A function f is called *Riemann integrable* if there exists a number $A \in \mathbb{R}$ such that for every sequence of partitions π_n with $|\pi_n| \to 0$ and every sequence of associated sample points ξ_n we have $S(f, \pi_n, \xi_n) \to A$ as $n \to \infty$. In this case, f is called *Riemann integrable* and A is called *The Riemann integral of* f *over* $[a, b]$. Notation: $R\text{-}\int_a^b f(t) \, dt$.

As is well-known, every continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable.

Theorem 2.5.18. *If* $f : [a, b] \to \mathbb{K}$ *is continuous, then* f *is an integrable function on* $([a, b], \mathscr{B}([a, b]), \lambda)$. *Moreover,*

$$\int_{[a,b]} f \, d\lambda = R\text{-}\int_a^b f(t) \, dt \, .$$

Proof. Let a sequence of partitions $\pi_n := (t_0^{(n)}, \ldots, t_{k_n}^{(n)})$ with $|\pi_n| \to 0$ be given and let $\xi_n = (\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)})$ be a sequence of associated sample points. Put

$$f_n := \sum_{j=1}^{k_n} f(\xi_j^{(n)}) \mathbb{1}_{[t_{j-1}^{(n)}, t_j^{(n)})} .$$

then f_n is a simple function and $\int_{[a,b]} f \, d\lambda = S(f, \pi_n, \xi_n)$.

Moreover, (a) $|f_n| \leq \|f\|_\infty$ and the latter is integrable on our measure space and (b) $f_n(t) \to f(t)$ for all $t \in [a,b]$. Indeed, for fixed $t \in [a,b]$, we have $|f_n(t) - f(t)| = |f(\xi_{j_n}^{(n)}) - f(t)|$, where $\xi_{j_n}^{(n)}$ is the sample point in the interval $[t_{j_n}^{(n)}, t_{j_n-1}^{(n)})$ and j_n is chosen such that t lies in this interval. But then $|\xi_{j_n}^{(n)} - t| \leq |t_{j_n}^{(n)} - t_{j_n-1}^{(n)}| \leq |\pi_n| \to 0$ and hence, by the continuity of f, $f|(\xi_{j_n}^{(n)}) - f(t)| \to 0$.

Hence the dominated convergence theorem 2.5.16 can be used and yields that f is integrable and

$$\int_{[a,b]} f \, d\lambda = \lim_{n\to\infty} \int_{[a,b]} f_n \, d\lambda = \lim_{n\to\infty} S(f, \pi_n, \xi_n) .$$

Since, on the other hand, the Riemann sums converge to $R\text{-}\int_a^b f(t)\, dt$, the assertion follows. □

Remark 2.5.19. Actually, the continuity assumption in Theorem 2.5.18 is not needed, if one is willing to enlarge the σ-algebra. It can be proved that if $f : [a,b] \to \mathbb{K}$ is Riemann integrable then it equals almost everywhere a measurable function which is integrable and the Riemann and the Lebesgue integral coincide.

There is also an extension of Theorem 2.5.18 to improper Riemann integrals. We recall that if $-\infty < a < b \leq \infty$ and $f : [a,b) \to \mathbb{R}$ is continuous, then f is called improperly Riemann integrable on $[a,b)$ if the limit $\lim_{r\uparrow} R\text{-}\int_a^r f(t)\, dt$ exists. The limit is then called the *improper Riemann integral of f over* $[a,b)$ and denoted by $R\text{-}\int_a^{b-} f(t)\, dt$.

Theorem 2.5.20. *Let $-\infty < a < b \leq \infty$ and $f : [a,b) \to \mathbb{R}$ be continuous, such that the improper Riemann integral $R\text{-}\int_a^{b-} |f(t)|\, dt$ exists, then f is integrable on $([a,b), \mathscr{B}([a,b)), \lambda)$, the improper Riemann integral $R - \text{-}\int_a^{b-} f(t) dt$ exists and*

$$\int_{[a,b)} f \, d\lambda = R\text{-}\int_a^{b-} f(t)\, dt .$$

Proof. Pick a sequence $(b_n) \subset (a,b)$ with $a_n \uparrow a$. By monotone convergence and Theorem 2.5.18,

$$\int_{[a,b)} |f| \, d\lambda = \lim_{n\to\infty} \int_{[a,b_n]} |f| \, d\lambda = \lim_{n\to\infty} R\text{-}\int_a^{b_n} |f(t)|\, dt = R\text{-}\int_a^{b-} |f(t)|\, dt < \infty .$$

It follows that $|f|$ is integrable on $[a,b)$. Moreover, since $\mathbb{1}_{[a,b_n)} f$ converges to f pointwise and $|\mathbb{1}_{[a,b_n)} f| \leq |f|$, the dominated convergence theorem yields

$$\int_{[a,b)} f \, d\lambda = \lim_{n\to\infty} \int_{[a,b_n)} f \, d\lambda = \lim_{n\to\infty} R\text{-}\int_a^{b_n} f(t)\, dt = R\text{-}\int_a^{b-} f(t)\, dt ,$$

where we have used Theorem 2.5.18 in the end. □

Remark 2.5.21. Similar results as in Theorem 2.5.20 also hold for improper Riemann integrals which are improper at the right-hand side or which are improper on both sides.

Example 2.5.22. In Theorem 2.5.20, the assumption that $R\text{-}\int_{[a,b)} |f(t)|\,dt$ is crucial and cannot be omitted. An example is given by $f : [1,\infty) \to \mathbb{R}$, defined by $f(t) = \frac{\sin t}{t}$ In this case, by partial integration, we obtain

$$R\text{-}\int_1^x \frac{\sin t}{t}\,dt = [\frac{-\cos t}{t}|_0^x - R\text{-}\int_0^x \frac{\cos t}{t^2}\,dt \to \cos 1 - R\text{-}\int_0^\infty \frac{\cos t}{t^2}\,dt$$

as $x \to \infty$. The latter improper Riemann integral exists since $|t^{-2}\cos t| \leq t^{-2}$ and the latter is integrable. It follows that the improper Riemann integral $R\text{-}\int_0^\infty \frac{\cos t}{t^2}\,dt$ exists.

On the other hand, on each interval $[k\pi, (k+1)\pi)$, we have $|f(t)| \geq |\sin(t)|((k+1)\pi)^{-1}$. It thus follows that

$$\int_{[0,\infty)} f\,d\lambda \geq \sum_{k=1}^n \frac{1}{(k+1)\pi}\int_{[k\pi,(k+1)\pi)} |\sin(t)|\,d\lambda(t) = \Big(\sum_{k=1}^n \frac{1}{k+1}\Big)\cdot\frac{1}{\pi}R\text{-}\int_0^\pi |\sin(t)|\,dt\,.$$

Since the harmonic series diverges, it follows that f is not (Lebesgue-) integrable on $(0,\infty)$.

Remark 2.5.23. In what follows, we will also use the differential dt in Lebesgue integrals instead of the (formally correct) $d\lambda$. We will thus write

$$\int_{[a,b]} f\,dt \quad \text{or} \quad \int_a^b f\,dt$$

to denote the Lebesgue integral of f on the interval $[a,b]$.

This is especially helpful when the function f depends on more than one variable. To have this feature also at hand for general measures we will sometimes, e.g. in the next section, write $\int_\Omega f(x)\,d\mu(x)$ instead of $\int_\Omega f\,d\mu$ to emphasize that we are integrating with respect to the variable x.

Exercise 2.5.24. Consider the situation of Exercise 2.5.11, i.e. (Ω,Σ,μ) is a measure space, $f : \Omega \to [0,\infty]$ is measurable and $\nu(A) := \int_A f\,d\mu$.

Show that g is integrable with respect to ν if and only if gf is integrable with respect to μ and in this case,

$$\int_\Omega g\,d\nu = \int_\Omega gf\,d\mu\,.$$

We close this section by considering the integration under a push-forward measure which is of great importance for applications.

Theorem 2.5.25. *Let (Ω,Σ,μ) be a measure space, $(M,\mathscr{F})$ be a measurable space and $\Phi : (\Omega,\Sigma) \to (M,\mathscr{F})$ be measurable. We denote the push-forward of μ under Φ by μ_Φ. Then for a measurable $f : (M,\mathscr{F}) \to (\mathbb{K},\mathscr{B}(\mathbb{K}))$ we have $f\circ\Phi \in \mathscr{L}^1(\Omega,\Sigma,\mu)$ if and only if $f \in \mathscr{L}^1(M,\mathscr{F},\mu_\Phi)$. In this case,*

$$\int_\Omega f\circ\Phi\,d\mu = \int_M f\,d\mu_\Phi\,.$$

Proof. First, let $f = \sum_{j=1}^n a_k \mathbb{1}_{A_k}$ be a simple function with $a_k \in [0, \infty]$.

$$f \circ \Phi = \sum_{j=1}^n a_k \mathbb{1}_{\Phi^{-1}(A_k)}$$

And thus, by the definition of the push-forward measure,

$$\int_\Omega f \circ \Phi \, d\mu = \sum_{j=1}^n a_k \mu(\Phi^{-1}(A)) = \sum_{j=1}^n a_k \mu_\Phi(A) = \int_M f \, d\mu_\Phi \, .$$

It follows that for positive, simple functions the assertion holds true.

Now let $f : M \to [0, \infty]$ be measurable and f_n be a sequence of simple functions with $f_n \uparrow f$ pointwise. Then, by monotone convergence and the above,

$$\int_\Omega f \circ \Phi \, d\mu = \sup_{n \in \mathbb{N}} \int_\Omega f_n \circ \Phi \, d\mu = \sup_{n \in \mathbb{N}} \int_M f_n \, d\mu_\Phi = \int_M f \, d\mu_\Phi \, .$$

This shows that the assertion holds for arbitrary measurable positive f.

Since $|f \circ \Phi| = |f| \circ \Phi$, it follows that $f \in \mathscr{L}^1(M, \mathscr{F}, \mu_\Phi)$ if any only if $f \circ \Phi \in \mathscr{L}^(\Omega, \Sigma, \mu)$. The general formula follows by splitting real valued functions f into the positive functions f^+ and f^- and complex valued functions f into $\operatorname{Re} f$ and $\operatorname{Im} f$. □

2.6 Integrals Depending on a Parameter

Suppose that (Ω, Σ, μ) is a measure space and (M, d) is a metric space. If we are given a map $f : M \times \Omega \to \mathbb{C}$ such that $f(t, \cdot)$ is integrable for all $t \in M$, we may define $F(t) := \int_\Omega f(t, x) \, d\mu(x)$. It is then natural to ask how F depends on the parameter t. In this short section, we use the dominated convergence theorem to prove some results in this direction.

Proposition 2.6.1. *Let (M, d) be a metric space and (Ω, Σ, μ) be a measure space. Furthermore, let $f : M \times \Omega \to \mathbb{K}$ be such that*

(a) $x \mapsto f(t, x) \in \mathscr{L}^1(\Omega, \Sigma, \mu)$ for all $t \in M$.

(b) $t \mapsto f(t, x)$ is continuous for almost all $x \in \Omega$.

(c) There exists $g \in \mathscr{L}^1(\Omega, \Sigma, \mu)$ such that $|f(t, x)| \leq g(x)$ for all $(t, x) \in M \times \Omega$.

Then $F : M \to \mathbb{C}$, defined by $F(t) = \int_\Omega f(t, x) \, d\mu(x)$ is continuous.

Proof. Let $t_n \to t$ in M. Then $f(t_n, x) \to f(t, x)$ for all $x \in \Omega$ by (b). Since $|f(t_n, x)| \leq g(x)$ for all $x \in \Omega$ by assumption and $g \in L^1(\Omega)$, it follows from the Dominated Convergence Theorem 2.5.16 that

$$F(t_n) = \int_\Omega f(t_n, x) \, d\mu(x) \to \int_\Omega f(t, x) \, d\mu(x) = F(t) \, .$$

This proves the continuity of F. □

Proposition 2.6.2. *Let I be an interval in $\mathbb{R}$ and (Ω, Σ, μ) be a measure space. Furthermore, let $f : I \times \Omega \to \mathbb{K}$ be such that*

(a) $x \mapsto f(t, x) \in L^1(\Omega, \Sigma, \mu)$ for all $t \in M$.

(b) $t \mapsto f(t,x)$ *is continuously differentiable for all* $x \in \Omega$.

(c) There exists $g \in L^1(\Omega, \Sigma, \mu)$ *such that* $|\frac{\partial}{\partial t} f(t,x)| \le g(x)$ *for all* $(t,x) \in M \times \Omega$.

Then $F : I \to \mathbb{K}$, *defined by* $F(t) = \int_\Omega f(t,x)\, d\mu(x)$ *is differentiable. Moreover,* $\frac{\partial}{\partial t} f(t,x)$ *is integrable for all* $t \in I$ *and*

$$F'(t) = \frac{d}{dt} \int_\Omega f(t,x)\, d\mu(x) = \int_\Omega \frac{\partial}{\partial t} f(t,x)\, d\mu(x)\,.$$

Proof. Fix $t \in I$ and let t_n be a sequence in I, converging to t. Define $h_n, h : \Omega \to \mathbb{C}$ by $h_n(x) := (t_n - t)^{-1} F(t_n, x) - F(t,x)$ and $h(x) = \frac{\partial}{\partial t} F(t,x)$. Then h_n is integrable for every $n \in \mathbb{N}$ as a linear combination of integrable functions. Moreover, $h_n(x) \to h(x)$ for all $x \in \Omega$ by assumption. By the mean-value theorem, $h_n(x) = \frac{\partial}{\partial t} F(\xi_n, x)$ for some ξ_n between t and t_n. In particular, $|h_n| \le g$. Thus the dominated convergence theorem yields that h is integrable and

$$\frac{F(t_n) - F(t)}{t_n - t} = \int_\Omega h_n(x)\, d\mu(x) \to \int_\Omega h(x)\, d\mu(x) = \int_\Omega \frac{\partial}{\partial t} F(t,x)\, d\mu(x)\,.$$

This finishes the proof. □

2.7 The L^p-Spaces

Definition 2.7.1. Let (Ω, Σ, μ) be a measure space. For $f : \Omega \to \mathbb{K}$ measurable and $1 \le p < \infty$, put

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

We define $\mathscr{L}^p(\Omega, \Sigma, \mu) := \{f : \Omega \to \mathbb{K} \text{ measurable} : \|f\|_p < \infty\}$. If it is clear which Σ and μ we use (or we make a statement over generic measure spaces), we will just write $\mathscr{L}^p(\Omega)$.

We now prove $\mathscr{L}^p(\Omega)$ is a vector space and that $\|\cdot\|_p$ is *nearly* a norm on $\mathscr{L}^p(\Omega)$.

Proposition 2.7.2. *Let* (Ω, Σ, μ) *be a measure space and* $1 \le p < \infty$. *Then the following hold true:*

(a) For all $f \in \mathscr{L}^p(\Omega)$, *we have* $\|f\|_p \ge 0$ *and* $\|f\|_p = 0$ *if and only if* $f = 0$ *almost everywhere.*

(b) For all $f \in \mathscr{L}^p(\Omega)$ *and* $\lambda \in \mathbb{K}$, *we have* $\lambda f \in \mathscr{L}^p(\Omega)$ *and* $\|\lambda f\|_p = |\lambda| \|f\|_p$.

(c) For all $f, g \in \mathscr{L}^p(\Omega)$ *we have* $f + g \in \mathscr{L}^p(\Omega)$ *and* $\|f+g\|_p \le \|f\|_p + \|g\|_p$.

Proof. (a) $\|f\|_p \ge 0$ is obvious and the second assertion follows from Corollary 2.5.10.

(b) By Corollary 2.5.7,

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

In particular, if $\|f\|_p < \infty$ then $\|\lambda f\|_p < \infty$.

(c) Let us first assume that f and g are simple functions, say $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$ and $g = \sum_{k=1}^n b_k \mathbb{1}_{B_k}$. We may assume without loss of generality that $m = n$ and $A_k = B_k$ for all $k = 1, \ldots, m$. We can moreover assume that the A_k are disjoint. Then

$$\begin{aligned}\|f+g\|_p &= \Big\| \sum_{k=1}^n (a_k + b_k)\mathbb{1}_{A_k} \Big\|_p = \Big(\sum_{k=1}^n |a_k + b_k|^p \mu(A_k)\Big)^{\frac{1}{p}} \\ &\leq \Big(\sum_{k=1}^n (|a_k|\mu(A_k)^{\frac{1}{p}} + |b_k|\mu(A_k)^{\frac{1}{p}})^p\Big)^{\frac{1}{p}} \\ &\leq \Big(\sum_{k=1}^n |a_k|^p\mu(A_k)\Big)^{\frac{1}{p}} + \Big(\sum_{k=1}^n |b_k|^p\mu(A_k)\Big)^{\frac{1}{p}} = \|f\|_p + \|g\|_p\,,\end{aligned}$$

where we have used Minkowski's inequality 1.2.7. This proves (c) for simple functions.

Now let $f, g \in \mathscr{L}^p(\Omega)$. By Proposition 2.4.6, there exist sequences f_n, g_n of simple functions converging pointwise to f resp. g with $|f_n| \leq 2|f|$ and $|g_n| \leq 2|g|$. In particular, f_n and g_n belong to $\mathscr{L}^p(\Omega)$ for all $n \in \mathbb{N}$. It also follows that $|f_n + g_n|^p \to |f + g|^p$ pointwise and hence, by Fatou's Lemma 2.5.12,

$$\begin{aligned}\Big(\int_\Omega |f+g|^p\, d\mu\Big)^{\frac{1}{p}} &= \Big(\int_\Omega \liminf_{n\to\infty} |f_n + g_n|^p\, d\mu\Big)^{\frac{1}{p}} = \liminf_{n\to\infty} \Big(\int_\Omega |f_n+g_n|^p\, d\mu\Big)^{\frac{1}{p}} \\ &\leq \liminf_{n\to\infty} \Big(\int_\Omega |f|^p\, d\mu\Big)^{\frac{1}{p}} + \liminf_{n\to\infty} \Big(\int_\Omega |g_n|^p\, d\mu\Big)^{\frac{1}{p}} = \|f\|_p + \|g\|_p\,.\end{aligned}$$

Here we have used the continuity of $t \mapsto |t|^{\frac{1}{p}}$ in the second, the first case above in the third and the dominated convergence theorem in the third step. ☐

By Proposition 2.7.2 $\|\cdot\|_p$ satisfies all properties of a norm on $\mathscr{L}^p(\Omega)$ except for (N1), i.e. we can have that $\|f\|_p = 0$ without f being constantly zero. In order to overcome this difficulty, we identify functions which are equal almost everywhere. More precisely, on $\mathscr{L}^p(\Omega, \Sigma, \mu)$ we introduce the equivalence relation $\sim$ by $f \sim g :\Leftrightarrow f = g$ almost everywhere. We then consider the equivalence classes $[f] := \{g : f \sim g\}$ as our primary objects.

Exercise 2.7.3. Consider the measure space $([0,1], \mathscr{B}([0,1]), \lambda)$. Show that if f, g are continuous functions with $f = g$ almost everywhere, then $f = g$ everywhere.

Now endow $([0,1], \mathscr{B}([0,1]))$ with the Dirac measure δ_1. For a continuous function f, determine its equivalence class $[f]$.

Definition 2.7.4. For a measure space (Ω, Σ, μ), we define

$$L^p(\Omega, \Sigma, \mu) := \{[f] : f \in \mathscr{L}^p(\Omega, \Sigma, \mu)\}\,.$$

We put $\|[f]\|_p := \|f\|_p$ and define $\lambda[f] := [\lambda f]$ and $[f] + [g] := [f + g]$. In this way, $L^p(\Omega, \Sigma, \mu)$ becomes a normed vector space.

Remark 2.7.5. That $L^p(\Omega, \Sigma, \mu)$ is a normed vector space is an immediate consequence of Proposition 2.7.2 and the definition of the norm, scalar multiplication and addition on L^p. However, one needs to check that these are well defined, i.e. do not depend on the choice of the particular representative of $[f]$. We thus have to show, e.g., that if $f \sim g$ then $\|f\|_p = \|g\|_p$ (this follows from Corollary 2.5.10), that $\lambda f = \lambda g$ almost everywhere, etc.

Remark 2.7.6. As is customary, we will not distinguish between f and $[f]$ and treat elements of $L^p(\Omega, \Sigma, \mu)$ as functions, rather than as equivalence classes, and understand equalities, inequalities, etc. only to hold almost everywhere.

Theorem 2.7.7. *Let (Ω, Σ, μ) be a measure space and $1 \leq p < \infty$. Then $(L^p(\Omega), \|\cdot\|_p)$ is complete.*

The proof of Theorem 2.7.7 rests on the following Lemma which is also of independent interest.

Lemma 2.7.8. *Let (Ω, Σ, μ) be a measure space and let $1 \leq p < \infty$. If $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(L^p(\Omega), \|\cdot\|_p)$ (in particular, if (f_n) is a converging sequence), then there exists a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ which converges almost everywhere to a function $f \in L^p(\Omega)$. Moreover, there exists a function $g \in L^p(\Omega)$ such that $|f_{n_k}| \leq g$ for all $k \in \mathbb{N}$.*

Proof. Since (f_n) is a Cauchy sequence, there exists a subsequence (f_{n_k}) with $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$. We now define

$$h_k := f_{n_{k+1}} - f_{n_k} \quad \text{and} \quad h := \sum_{k=1}^{\infty} |h_k| \,.$$

Then, using Proposition 2.7.2, we see that for every $N \in \mathbb{N}$, we have

$$\Big\| \sum_{k=1}^{N} |h_k| \Big\|_p \leq \sum_{k=1}^{N} \|f_{n_{k+1}} - f_{n_k}\|_p \leq \sum_{k=1}^{N} 2^{-k} \leq 1 \,.$$

It follows from monotone convergence, that $h \in L^p(\Omega)$ with $\|h\|_p \leq 1$. In particular, the series $\sum_{k=1}^{\infty} h_k(x)$ converges (absolutely) for almost all $x \in \Omega$. Noting that $\sum_{k=1}^{N} h_k = f_{n_{N+1}} - f_{n_1}$, it follows that f_{n_k} converges almost surely to $f_{n_1} + \sum_{k=1}^{\infty} h_k =: f$. Moreover,

$$|f_{n_k}| = |h_1 + \cdots + h_{k-1} + f_{n_1}| \leq h + |f_{n_1}| =: g \,.$$

Noting that $g \in L^p(\Omega)$, the proof is complete. □

Let us give some examples illustrating Lemma 2.7.8

Example 2.7.9. Consider the measure space $([0,1], \mathscr{B}([0,1]), \lambda)$. Then $f_n(t) := t^n$ converges to $f(t) \equiv 0$ in $L^p([0,1])$ for all $1 \leq p < \infty$. Indeed,

$$\|f_n - f\|_p = \Big(\int_0^1 t^{np} \, dt \Big)^{\frac{1}{p}} = \Big(\frac{1}{np+1} \Big)^{\frac{1}{p}} \to 0 \,.$$

Moreover, $f_n(t)$ converges to $f(t)$ for all $t \in [0,1)$. Since the singleton $\{1\}$ has Lebesgue measure zero, f_n converges to f almost everywhere, but f_n does not converge to f pointwise.

Example 2.7.10. Consider the measure space $((0,1], \mathscr{B}((0,1]), \lambda)$. If $m = 2^n + k$ for $n \in \mathbb{N}$ and $0 \leq k \leq 2^n - 1$, put $f_m = \mathbb{1}_{(k2^{-n},(k+1)2^{-n}]}$. Then $f_m \to 0$ in $L^p((0,1])$, since $\|f_{2^n+k}\|_p = 2^{-\frac{n}{p}} \to 0$. By Lemma 2.7.8, f_m has a subsequence converging to 0 almost everywhere (an example being f_{m_n}, where $m_n = 2^n$, so that $f_{m_n} = \mathbb{1}_{(0,2^{-n}]}$) but the whole sequence f_m does not converge to 0 almost everywhere.

In view of Lemma 1.5.4, this shows that there exists no metric d such that $f_n \to f$ almost everywhere if and only if $f_n \to f$ with respect to d.

Proof of Theorem 2.7.7. Let (f_n) be a Cauchy sequence in $(L^p(\Omega), \|\cdot\|_p)$. By Lemma 2.7.8, there exists a subsequence f_{n_k} which converges almost surely to a function $f \in L^p(\Omega)$ and is dominated by a function $g \in L^p(\Omega)$. By the dominated convergence theorem, $\|f_{n_k} - f\|_p \to 0$, that is, $f_{n_k} \to f$ with respect to $\|\cdot\|_p$. By Lemma 1.5.3, the whole sequence f_n converges to f with respect to $\|\cdot\|_p$. □

We now complement the scale of L^p-spaces by introducing the space $L^\infty(\Omega, \Sigma, \mu)$. The only difficulty is that, as before, we are formally dealing with equivalence classes of functions rather than with functions themselves.

Definition 2.7.11. Let (Ω, Σ, μ) be a measure space. An equivalence class $[f]$ is said to belong to $L^\infty(\Omega, \Sigma, \mu)$ if there exists a constant $c > 0$ such that $|f| \leq c$ almost everywhere. In this case, we put $\|[f]\|_\infty := \inf\{c > 0 : |f| \leq c \text{ } a.e.\}$.

In practice, we again do not distinguish between f and $[f]$. As is to be expected, the normed space $(L^\infty(\Omega), \|\cdot\|_\infty)$ turns out to be complete. We formulate this result and leave the proof to the reader.

Proposition 2.7.12. *Let (Ω, Σ, μ) be a measure space. Then $(L^\infty(\Omega), \|\cdot\|_\infty)$ is a complete normed space.*

Example 2.7.13. Consider the measure space $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \zeta)$. Then $L^p(\mathbb{N}, \mathscr{P}(\mathbb{N}), \zeta) = \ell^p$ for $1 \leq p \leq \infty$.

We now also extend Hölder's inequality to the L^p setting.

Theorem 2.7.14. *Let (Ω, Σ, μ) be a measure space and $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$, then $fg \in L^1(\Omega)$ and $\|fg\|_1 \leq \|f\|_p \|g\|_q$.*

Proof. The proof is similar as that of Minkowski's inequality in Proposition 2.7.2. We give a rough sketch.

If f and g are simple functions, then the claim follows from Hölder's inequality 1.2.7 for finite sums. The general case follows from an approximation argument using Fatou's lemma and dominated convergence. □

Exercise 2.7.15. Work out the details of the proof of Theorem 2.7.14.

Exercise 2.7.16. Let (Ω, Σ, μ) be a measure space and $1 \leq p, q \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Given $g \in L^q(\Omega)$, define $\varphi_g : L^p(\Omega) \to \mathbb{K}$ by

$$\varphi_g(f) = \int_\Omega fg \, d\mu \, .$$

Show that $\varphi_g \in (L^p(\Omega))^*$ and $\|\varphi_g\|_{(L^p(\Omega))^*} = \|g\|_q$.

Corollary 2.7.17. *Let (Ω, Σ, μ) be a* finite *measure space and $1 \leq p \leq q \leq \infty$. Then $L^q(\Omega, \Sigma, \mu) \subset L^p(\Omega, \Sigma, \mu)$. Moreover, if $f_n \to f$ in $(L^q(\Omega, \Sigma, \mu), \|\cdot\|_q)$, then $f_n \to f$ in $(L^p(\Omega, \Sigma, \mu), \|\cdot\|_p)$.*

Proof. Let us first consider the case where $q = \infty$. In this case, if $f \in L^\infty(\Omega)$, then $f \leq \|f\|_\infty \mathbb{1}_\Omega$ almost everywhere. Hence

$$\|f\|_p^p = \int_\Omega |f|^p \, d\mu \leq \int_\Omega \|f\|_\infty^p \, d\mu = \mu(\Omega)\|f\|_\infty^p \, .$$

This proves that $L^\infty(\Omega) \subset L^p(\Omega)$. Now let $f_n \to f$ in $L^\infty(\Omega)$. Then $\|f_n - f\|_p \leq \mu(\Omega)^{\frac{1}{p}} \|f_n - f\|_\infty \to 0$, proving that $f_n \to f$ in $L^p(\Omega)$.

Now let $1 \leq p < q \neq \infty$ and fix $f \in L^q(\Omega)$. Then $r := \frac{q}{p} \in (1, \infty)$. With $s = \frac{q}{q-p}$, we have $\frac{1}{r} + \frac{1}{s} = 1$. By the above, $\mathbb{1}_\Omega \in L^s(\Omega)$. Moreover, $|f|^p \in L^r(\Omega)$. By Hölder's inequality 2.7.14 $|f|^p \mathbb{1}_\Omega \in L^1(\Omega)$ and

$$\int_\Omega |f|^p \, d\mu \leq \|\mathbb{1}_\Omega\|_s \||f|^p\|_r = \mu(\Omega)^{\frac{1}{s}} \Big(\int_\Omega |f|^q \Big)^{\frac{1}{r}} = \mu(\Omega)^{\frac{q-p}{q}} \|f\|_q^p .$$

This proves that $f \in L^p(\Omega)$. Moreover, taking p-th roots on both sides, $\|f\|_p \leq \mu(\Omega)^{q-p} \|f\|_q$ follows. As above, this inequality also shows that if $f_n \to f$ in $(L^q(\Omega, \Sigma, \mu), \| \cdot \|_q)$, then $f_n \to f$ in $(L^p(\Omega, \Sigma, \mu), \| \cdot \|_p)$. □

Theorem 2.7.18. *Let (M, d) be a metric space, μ be a finite measure on $(M, \mathscr{B}(M))$. Then $C_b(M)$ is dense in $L^p(M, \mathscr{B}(M), \mu)$ for all $1 \leq p < \infty$.*

Proof. Let $E := \overline{C_b(M)}^{\|\cdot\|_p}$ be the closure of $C_b(M)$ in $L^p(M, \mathscr{B}(M), \mu)$. By continuity of addition and scalar multiplication, E is a closed subspace of L^p.

Define $\mathscr{G} := \{A \in \mathscr{B}(M) : \mathbb{1}_A \in E\}$. Then $\mathscr{G}$ is a Dynkin system. Indeed, $\mathbb{1}_\Omega$ is continuous and integrable since $\mu(\Omega) < \infty$. Hence $\Omega \in \mathscr{G}$. If $A \in \mathscr{G}$, then $A^c \in \mathscr{G}$ since $\mathbb{1}_{A^c} = \mathbb{1}_\Omega - \mathbb{1}_A$ and E is a vector space. Finally, let (A_k) be a sequence of pairwise disjoint sets in $\mathscr{G}$. Then $\bigcup_{k=1}^n A_k \in \mathscr{G}$ since $\mathbb{1}_{\bigcup_{k=1}^n A_k} = \sum_{k=1}^n \mathbb{1}_{A_k}$ and E is a vector space. Moreover, $\mathbb{1}_{\bigcup_{k=1}^n A_k} \to \mathbb{1}_{\bigcup_{k=1}^\infty A_k}$ pointwise and dominated by the integrable function $\mathbb{1}_\Omega$. By dominated convergence, $\bigcup_{k=1}^\infty A_k \in \mathscr{G}$.

Now let $F \subset M$ be closed. With the help of Urysohn's theorem we find, as in the proof of Proposition 2.1.19, a sequence f_n of continuous functions with $0 \leq f_n \leq 1$ and $f_n \to \mathbb{1}_F$ pointwise. It follows from dominated convergence that $F \in \mathscr{G}$.

Since $\mathscr{C}$, the collection of all closed subsets of M, is a generator of $\mathscr{B}(M)$ which is stable under intersections, Dynkin's π-λ theorem yields $\mathscr{B}(M) = d(\mathscr{C}) \subset \mathscr{G} \subset \mathscr{B}(M)$. Hence $\mathbb{1}_A \in E$ for all $A \in \mathscr{B}(M)$.

By linearity, E contains all simple functions. Now an approximation argument yields that $E = L^p(M)$. □

Corollary 2.7.19. *Let (K, d) be a compact metric space, μ be a finite measure on $(K, \mathscr{B}(K))$. Then $L^p(K, \mathscr{B}(K), \mu)$ is separable for all $1 \leq p < \infty$.*

Proof. By Corollary 1.7.14, there exists a countable set $\mathcal{S} \subset C(K)$ which is dense with respect to the norm $\| \cdot \|_\infty$. Put $E := \overline{\mathcal{S}}^{\|\cdot\|_\infty}$. Then $C(K) \subset E$. Indeed, given $f \in C(K)$, there exists a sequence f_n in $\mathcal{S} \subset E$ such that $f_n \to f$ with respect to $\| \cdot \|_\infty$. By Corollary 2.7.17, $f_n \to f$ in L^p, hence $f \in E$. Now theorem 2.7.18 yields $L^p(K, \mathscr{B}(K), \mu) = \overline{C(K)}^{\|\cdot\|_p} \subset E$. □

Remark 2.7.20. In what follows, $L^p(\Omega, \Sigma, \mu)$ will always be endowed with the norm $\| \cdot \|_p$. We will therefore drop it from our notation and say, e.g., that $f_n \to f$ in $L^p(\Omega, \Sigma, \mu)$ or that f_n is a Cauchy sequence in $L^p(\Omega, \Sigma, \mu)$ with the understanding, that this is to be understood with respect to $\| \cdot \|_p$.

2.8 Convergence in Measure

In the last section, we have established the L^p-spaces as the appropriate generalizations of the sequence spaces ℓ^p to the setting of general measure spaces. What is missing is a generalization of the sequence space ℓ^0. We recall that convergence in (ℓ^0, d_0) is nothing else than pointwise convergence.

In the general setting, since we identify functions which are equal almost everywhere, one would expect that the appropriate generalization is to require convergence almost everywhere. However as we have seen in Example 2.7.10, there is no metric such that convergence almost everywhere is convergence with respect to that metric. In other words: *almost everywhere convergence is not metrizable.*

Hence, we have to take a different approach here. For the sake of simplicity, we confine ourselves with the situation of *finite* measure spaces. Throughout this section, (Ω, Σ, μ) denotes a finite measure space.

Definition 2.8.1. We denote by $L^0(\Omega, \Sigma, \mu)$ the (vector space!) of all (equivalence classes) of measurable functions $f : \Omega \to \mathbb{K}$. We say that a sequence $(f_n) \subset L^0(\Omega, \Sigma, \mu)$ *converges in measure* (if μ is a probability measure, one also says *converges in probability* to $f \in L^0(\Omega, \Sigma, \mu)$, if for all $\varepsilon > 0$, we have

$$\mu(\{x \in \Omega : |f_n(x) - f(x)| > \varepsilon\}) \to 0$$

as $n \to \infty$. To indicate that $f_n \to f$ in measure we will write $f_n \stackrel{m}{\to} f$. If μ is a probability measure, we will write $f_n \stackrel{P}{\to} f$ instead.

In what follows we will briefly write $\{|f_n - f| > \varepsilon\}$ instead of $\{x \in \Omega : |f_n(x) - f(x)| > \varepsilon\}$.

Remark 2.8.2. It should be noted that $\mu(\{x \in \Omega : |f_n(x) - f(x)| > \varepsilon\})$ does not depend on the choice of the representatives. Indeed, if $\tilde{f}_n = f_n$ and $\tilde{f} = f$ almost everywhere, there exists sets N_n and N of measure zero such that $\tilde{f}_n(x) = f_n(x)$ for all $x \notin N_n$ and $\tilde{f}(x) = f(x)$ for all $x \notin N$. If we put $M := N \cup \bigcup_{n \in \mathbb{N}} N_n$, then M has measure zero and $|f_n(x) - f(x)| = |\tilde{f}_n(x) - \tilde{f}(x)|$ for all $n \notin M$. Hence $\{|\tilde{f}_n - \tilde{f}| > \varepsilon\}$ and $\{|f_n - f| > \varepsilon\}$ differ at most by a set of measure zero.

We will first show that convergence in measure is indeed induced by a metric.

Definition 2.8.3. Let (Ω, Σ, μ) be a finite measure space. Define

$$d_0(f, g) := \int_\Omega |f - g| \wedge 1 \, d\mu \, .$$

Proposition 2.8.4. *d_0 defines a metric on $L^0(\Omega)$.*

Proof. Obviously, $d_0(f, g) \geq 0$. Moreover, if $d_0(f, g) = 0$, then, by Corollary 2.5.10, $|f - g| \wedge 1 = 0$ almost everywhere and hence $f = g$ almost everywhere, i.e. $f = g$ in $L^0(\Omega)$. This proves (M1). (M2) is obvious.

To see (M3), observe that $|f - g| \wedge 1 \leq |f - h| \wedge 1 + |h - g| \wedge 1$. Indeed, if $|f - g| \leq 1$ then this is clear if $|f - h| \geq 1$ or $|h - g| \geq 1$. If $|f - g|, |f - h|$ and $|h - g|$ are less then 1, then it follows from the triangle inequality in $\mathbb{R}$.

If, on the other hand $|f - g| \geq 1$ and $|f - h| \wedge 1 + |h - g| \wedge 1$ was strictly less than 1, then we must have $|f - h| + |h - g| < 1$. Thus, using the triangle inequality in $\mathbb{R}$ again, it follows that $|f - g| < 1$ — a contradiction. ☐

Theorem 2.8.5. *Let (Ω, Σ, μ) be a finite measure space, (f_n) be a sequence in $L^0(\Omega)$ and $f \in L^0(\Omega)$. The following are equivalent:*

(a) f_n converges to f in measure.

(b) $f_n \to f$ with respect to d_0.

(c) Every subsequence of f_n has a further subsequence which converges to f pointwise almost surely.

Proof. (a) $\Rightarrow$ (b): Assume that $f_n \to f$ in measure. Given $\varepsilon \in (0,1)$, pick $\delta > 0$ such that $\delta\mu(\Omega) \leq \varepsilon/2$ and then n_0 such that $\mu(\{|f_n - f| > \delta\}) \leq \varepsilon/2$ for all $n \geq n_0$. Then, for $n \geq n_0$, we have

$$\begin{aligned} d(f_n, f) &= \int_\Omega |f_n - f| \wedge 1\, d\mu = \int_{\{|f_n - f| > \delta\}} |f_n - f| \wedge 1\, d\mu + \int_{\{|f_n - f| \leq \delta\}} |f_n - f| \wedge 1\, d\mu \\ &\leq \mu(\{f_n - f| > \delta\}) + \delta\mu(\Omega) \leq \varepsilon\,. \end{aligned}$$

This proves that $f_n \to f$ with respect to d_0.

(b) $\Rightarrow$ (c): Assume that $f_n \to f$ with respect to d_0. Then also every subsequence f_{n_k} converges to f with respect to d_0. Hence $|f_{n_k} - f| \wedge 1 \to 0$ in $L^1(\Omega)$. By Lemma 2.7.8, there is a subsequence $|f_{n_{k_l}} - f| \wedge 1$ which converges to 0 almost everywhere. But then $f_{n_{k_l}} \to f$ almost everywhere. This proves (c).

(c) $\Rightarrow$ (a): Suppose that (a) was false, i.e. there exists an $\varepsilon_0 \in (0,1)$ such that $\mu(\{|f_n - f| > \varepsilon_0\})$ does not converge to 0. Then there exists a δ_0 and a subsequence n_k, such that $\mu(\{|f_{n_k} - f| > \varepsilon_0\}) \geq \delta_0$. But then

$$\varepsilon_0\delta_0 \leq \varepsilon_0\mu(\{|f_{n_k} - f| > \varepsilon_0\}) \leq \int_{\{|f_{n_k} - f| > \varepsilon_0\}} |f_{n_k} - f| \wedge 1\, d\mu \leq \int_\Omega |f_{n_k} - f| \wedge 1\, d\mu\,.$$

However, by (c), there is a subsequence $f_{n_{k_l}}$ which converges to f pointwise almost everywhere. By dominated convergence, this implies that $\int_\Omega |f_{n_{k_l}} - f|\, d\mu \to 0$, a contradiction. ☐

We now show that L^p-convergence implies convergence in measure. This follows almost at once from Chebyshev's inequality:

Theorem 2.8.6. *(Chebyshev)*
Let $1 \leq p < \infty$ andf $\in L^p(\Omega)$. Then

$$\mu(\{|f| > \varepsilon\}) \leq \frac{\|f\|_p^p}{\varepsilon^p}\,.$$

Proof.

$$\|f\|_p^p = \int_\Omega |f|^p\, d\mu \geq \int_{\{|f| > \varepsilon\}} |f|^p\, d\mu \geq \int_{\{|f| > \varepsilon\}} \varepsilon^p\, d\mu \geq \varepsilon^p\mu(\{|f| > \varepsilon\})\,.$$

☐

Corollary 2.8.7. *Let $1 \leq p \leq \infty$ and let $f_n \to f$ in $L^p(\Omega)$. Then $f_n \to f$ in measure.*

Proof. By Corollary 2.7.17, $f_n \to f$ in $L^\infty(\Omega)$ implies that $f_n \to f$ in $L^p(\Omega)$ for all $1 \le p < \infty$. Hence it suffices to consider the case where $1 \le p < \infty$. However, for such p Chebyshev's inequality yields

$$\mu(\{|f_n - f| > \varepsilon\} \le \frac{\|f_n - f\|_p}{\varepsilon^p} \to 0$$

for all $\varepsilon > 0$. □

It is natural to ask, whether for a sequence f_n in L^p which converges in measure to a function $f \in L^p$ we automatically have that $f_n \to f$ in L^p. The following example shows that this is in general wrong.

Example 2.8.8. Consider $((0,1), \mathscr{B}((0,1)), \lambda)$. Define $f_n := n\mathbb{1}_{(0,n^{-1})}$ and $f \equiv 0$. Then, for every $p \in [0,\infty]$, we have $f_n, f \in L^p(0,1)$. Moreover, $f_n \to f$ pointwise and hence, by Theorem 2.8.5, $f_n \to f$ in measure. However, for $1 \le p < \infty$, $\|f_n - f\|_p = n^{p-1} \not\to 0$ and $\|f_n - f\|_\infty = n \not\to 0$, so that $f_n \not\to f$ in $L^p(0,1)$.

Definition 2.8.9. Let $M \subset L^1(\Omega, \Sigma, \mu)$. Then M is called *equi-integrable* (or *uniformly integrable*) if

$$\lim_{r\to\infty} \sup_{f\in M} \int_{\{|f|>r\}} |f|\, d\mu = 0\,.$$

A subset $M \subset L^p(\Omega, \Sigma, \mu)$ is called *p-equi-integrable* if the set $\{|f|^p : f \in M\} \subset L^1(\Omega, \Sigma, \mu)$ is equi-integrable.

Example 2.8.10. If $f \in L^p(\Omega)$, then the set $M := \{g : |g| \le f \text{ a.e.}\}$ is p-equi-integrable. Indeed, for all $g \in M$ we have

$$\int_{\{|g|^p>r\}} |g|^p\, d\mu \le \int_{\{|g|^p>r\}} |f|^p\, d\mu \le \int_{\{|f|^p>r\}} |f|^p\, d\mu$$

since $\{|g|^p > r\} \subset \{|f|^p > r\}$. Since $\int_{\{|f|^p>r\}} |f|^p\, d\mu \to 0$ for $r \to \infty$ by dominated convergence, it follows that M is p-equi-integrable.

Example 2.8.11. If M is bounded in $L^p(\Omega)$, then M is q-equi-integrable for all $q < p$. Indeed, if $\sup\{\|f\|_p : f \in M\} =: C < \infty$, then, by Hölder's inequality, applied to $r = \frac{p}{q}$ and $\frac{1}{s} = 1 - \frac{q}{p}$, and Chebyshev's inequality

$$\begin{aligned}\int_{\{|f|^q>r\}} |f|^q\, d\mu &\le \mu(\{|f|^q > r\})^{\frac{1}{s}} \|f\|_p^{\frac{q}{p}} = \mu(\{|f| > r^{\frac{1}{q}}\})^{\frac{1}{s}} \|f\|_p^{\frac{q}{p}} \\ &\le \|f\|_p^{p\frac{1}{s}} \|f\|_p^{\frac{q}{p}} r^{-\frac{1}{q}sp} = \|f\|_p^{q-p} r^{-(\frac{p}{q}-1)} \le C^{q-p} r^{-(\frac{p}{q}-1)}\end{aligned}$$

for all $f \in M$. Since $\frac{p}{q} - 1 > 0$, $r^{-(\frac{p}{q}-1)} \to 0$ as $r \to \infty$, the q-equi-integrability of M follows.

Let us first give an equivalent description of equi-integrability.

Lemma 2.8.12. *A subset $M \subset L^1(\Omega, \Sigma, \mu)$ is equi-integrable if and only if $\sup_{f\in M} \|f\|_1 < \infty$ and, for every $\varepsilon > 0$, there exists a number $\delta > 0$ such that $\mu(A) \le \delta$ implies that*

$$\sup_{f\in M} \int_A |f|\, d\mu \le \varepsilon\,. \tag{2.1}$$

Proof. First assume that M is equi-integrable. Then

$$\int_\Omega |f|\,d\mu = \int_{\{|f|\le 1\}} |f|\,d\mu + \int_{\{|f|>1\}} |f|\,d\mu \le \mu(\Omega) + \sup_{f\in M}\int_{\{|f|>1\}} |f|\,d\mu < \infty$$

for all $f \in M$, proving that $\sup_{f\in M}\|f\|_1 < 1$. Next, given $\varepsilon > 0$, first pick r such that $\int_{\{|f|>r\}} |f|\,d\mu \le \varepsilon/2$ for all $f \in M$. Now pick δ such that $r\delta \le \varepsilon/2$. Then for every $A \in \Sigma$ with $\mu(A) \le \delta$ we have

$$\int_A |f|\,d\mu = \int_{\{|f|>r\}\cap A} |f|\,d\mu + \int_{\{|f|\le r\}\cap A} |f|\,d\mu \le \varepsilon/2 + r\delta \le \varepsilon\,.$$

Conversely, assume that the second condition holds and put $C := \sup_{f\in M}\|f\|_1$. By Chebyshev's inequality 2.8.6, $\mu(\{|f| > r\}) \le r^{-1}C \to 0$ as $r \to \infty$. Thus, given $\varepsilon > 0$ we may pick δ as in the second condition. But if we then pick r so large that $r^{-1}C \le \delta$, then we may use $A = \{|f| > r\}$ in (2.1). This proves that M is equi-integrable. □

Theorem 2.8.13. *Let $1 \le p < \infty$ and (f_n) be a sequence in $L^p(\Omega)$ which converges in measure to a function $f \in L^p(\Omega)$. The following are equivalent:*

(a) $f_n \to f$ in $L^p(\Omega)$.

(b) $\|f_n\|_p \to \|f\|_p$.

(c) The sequence f_n is p-equi-integrable.

Proof. (a) ⇒ (b) is clear, since norms are continuous functions.

(b) ⇒ (c) If $\|f_n\|_p \to \|f\|_p$, then in particular, $\|f_n\|_p$ is bounded, say by M. By Chebyshev's inequality 2.8.6, $\mu(\{|f_n|^p > r\}) \le r^{-1}M^p$, proving that the sequence f_n is p-equi integrable.

(c) ⇒ (a) We have

$$\begin{aligned}\|f_n - f\|_p &\le \int_{\{|f_n-f|^p>r\}} |f_n - f|^p\,d\mu + \int_{\{|f_n-f|\le r\}} |f_n - f|^p\,d\mu\\ &\le r^{-1}\sup_{n\in\mathbb{N}}\|f_n - f\|_p^p + \int_{\{|f_n-f|\le r\}} |f_n - f|^p\,d\mu\,.\end{aligned}$$

By equi-integrability $\sup_{n\in\mathbb{N}}\|f_n - f\|_p < \infty$. Thus, given $\varepsilon > 0$, we may first pick r large enough so that $r^{-1}\sup_{n\in\mathbb{N}}\|f_n - f\|_p^p \le \varepsilon/2$. Since $f_n \to f$ in measure, for fixed r, the second integral converges to 0 by dominated convergence. To be more precise, since $f_n \to f$ in measure, every subsequence has a subsequence which converges almost surely. But then dominated convergence yields convergence of the integrals along that subsequence. Thus we have to have convergence of the integrals along the whole sequence and hence $f_n \to f$ in $L^p(\Omega)$. □

Theorem 2.8.13 can be seen as a generalization of the dominated convergence theorem. This follows from Example 2.8.10, which shows that domination implies equi-integrability and the fact that almost sure convergence implies convergence in measure. On the other hand, Theorem 2.8.13 can be used to establish convergence of integrals also in absence of a dominating function:

Example 2.8.14. Consider the finite measure space $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \mu)$, where $\mu = \sum_{k=1}^\infty 2^{-k}\delta_k$. We let $f_n := 2^n n^{-1}\mathbb{1}_n$. Then the sequence f_n is equi-integrable. Indeed,

$$\int_{\{f_n>r\}} f_n\, d\mu = \begin{cases} n^{-1} & \text{if } 2^n n^{-1} > r \\ 0 & \text{else} \end{cases} .$$

Hence $\lim_{r\to\infty} \sup_{n\in\mathbb{N}} \int_{\{f_n>r\}} f_n\, d\mu = \lim_{r\to\infty} n(r)^{-1} = 0$, where

$$n(r) = \inf\{n \in \mathbb{N} \,:\, 2^n n^{-1} > r\} \uparrow \infty \quad \text{as} \quad r \uparrow \infty\,.$$

We note that smallest function g with $f_n \le g$ is given by $g(k) = 2^k k^{-1}$. But this function is not integrable since $\int_{\mathbb{N}} g\, d\mu = \sum_{k=1}^\infty k^{-1} = \infty$.

2.9 Product Measures

Let $(\Omega_i, \Sigma_i, \mu_i)$ be measure spaces for $i = 1, 2$. In Section 2.1 we have already introduced the product $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$ of the measurable spaces (Ω_1, Σ_1) and (Ω_2, Σ_2). We now want to define a measure $\mu_1 \otimes \mu_2$ on this product space which is the product of the measures μ_1 and μ_2, in the sense that

$$\mu_1 \otimes \mu_2(A \times B) = \mu_1(A)\mu_2(B)$$

for all $A \in \Sigma_1$ and $B \in \Sigma_2$. Note that by Corollary 2.2.11, if μ_1 and μ_2 are σ-finite, then there exists at most one such measure. Throughout this section, $(\Omega_i, \Sigma_i, \mu_i)$ are *σ-finite* measure spaces for $i = 1, 2$.

For a set $Q \subset \Omega_1 \times \Omega_2$ and $x \in \Omega_1, y \in \Omega_2$, we define the *cuts* Q_x and Q_y by

$$Q_x := \{y \in \Omega_2 \,:\, (x, y) \in Q\} \quad \text{and} \quad Q_y := \{x \in \Omega_1 \,:\, (x, y) \in Q\}\,.$$

Lemma 2.9.1. *For $x \in \Omega_1, y \in \Omega_2$ and $Q \in \Sigma_1 \otimes \Sigma_2$ we have $Q_x \in \Sigma_2$ and $Q_y \in \Sigma_1$.*

Proof. We put $\mathscr{G} := \{Q \in \Sigma_1 \otimes \Sigma_2 \,:\, Q_x \in \Sigma_2\}$. We claim that $\mathscr{G}$ is a σ-algebra on $\Omega_1 \times \Omega_2$. Clearly (S1) holds, since $(\Omega_1 \times \Omega_2)_x = \Omega_2$. (S2) and (S3) follow from the identities

$$(Q^c)_x = (Q_x)^c \quad \text{and} \quad \Big(\bigcup_{n\in\mathbb{N}} Q_n\Big)_x = \bigcup_{n\in\mathbb{N}} (Q_n)_x$$

which hold for every $Q \in \Sigma_1 \otimes \Sigma_2$, resp. every sequence Q_n in $\Sigma_1 \otimes \Sigma_2$.

To finish the proof, it suffices to observe that for $A \in \Sigma_1$ and $B \in \Sigma_2$ we have $(A\times B)_x = B$ or $(A \times B)_x = \emptyset$ depending on whether $x \in A$ or $x \notin A$. Since $B, \emptyset \in \Sigma_2$, it follows that every rectangle $A \times B$ belongs to $\mathscr{G}$. As these rectangles generate the product σ-algebra, $\mathscr{G} = \Sigma_1 \otimes \Sigma_2$.

The proof for the cuts Q_y is completely similar. ∎

By Lemma 2.9.1, $\mu_1(Q_y)$ and $\mu_2(Q_x)$ are well-defined for all $x \in \Omega_1$ and $y \in \Omega_2$.

Lemma 2.9.2. *If μ_1 and μ_2 are σ-finite, then the maps*

$$x \mapsto \mu_2(Q_x) \quad \text{and} \quad y \mapsto \mu_1(Q_y)$$

are well-defined and Σ_1- resp. Σ_2-measurable for all $Q \in \Sigma_1 \otimes \Sigma_2$.

Proof. We put $\varphi_Q(x) := \mu_2(Q_x)$ and prove that φ_Q is Σ_1-measurable for all $Q \in \Sigma_1 \otimes \Sigma_2$. The proof for the second map is entirely similar.

Let us first assume that $\mu_2(\Omega_2) < \infty$. Then

$$\mathscr{D} := \{Q \in \Sigma_1 \otimes \Sigma_2 \, : \, \varphi_Q \text{ is } \Sigma_1\text{-measurable}\}$$

is a Dynkin system. Indeed, $\varphi_{\Omega_1 \times \Omega_2}(x) \equiv \mu_2(\Omega_2)$ is constant, hence measurable whence (D1). For (D2) note that if φ_Q is measurable then $\varphi_{Q^c} = \varphi_{\Omega_1 \times \Omega_2} - \varphi_Q$ is measurable as difference of measurable functions. Finally, if Q_n is a sequence of pairwise disjoint sets in $\mathscr{D}$, then $\varphi_{\bigcup_n Q_n} = \sum_{n\in\mathbb{N}} \varphi_{Q_n}$ is measurable by Proposition 2.4.3.

Next observe that if $A \in \Sigma_1$ and $B \in \Sigma_2$, then $\varphi_{A\times B} = \mu_2(B)\mathbb{1}_A$ is measurable, hence $A \times B \in \mathscr{D}$. Since these rectangles generate $\Sigma_1 \otimes \Sigma_2$ and are stable under intersections, it follows that $\mathscr{D} = \Sigma_1 \otimes \Sigma_2$.

Now assume that μ_2 is merely σ-finite. Then there exists a sequence B_n with $\mu_2(B_n) < \infty$ and $B_n \uparrow \Omega_2$.

In this case, $\mu_{2,n} : B \mapsto \mu_2(B \cap B_n)$ is a finite measure on Σ_2, whence, by the above, $\varphi_{Q,n} : x \mapsto \mu_{2,n}(Q_x)$ is measurable for all $Q \in \Sigma_1 \otimes \Sigma_2$. Since $\mu_{2,n}(Q_x) \uparrow \mu_2(Q_x)$, $\varphi_Q = \sup_n \varphi_{Q,n}$ is measurable as the pointwise limit of measurable functions. □

We can now prove the existence of the product measure.

Theorem 2.9.3. *Let $(\Omega_i, \Sigma_i, \mu_i)$ be σ-finite measure spaces for $i = 1, 2$. Then there exists a unique measure π on $(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2)$ such that $\pi(A \times B) = \mu_1(A)\mu_2(B)$ for all $A \in \Sigma_1$ and $B \in \Sigma_2$. Moreover, for all $Q \in \Sigma_1 \otimes \Sigma_2$ we have*

$$\pi(Q) = \int_{\Omega_1} \mu_2(Q_x)\, d\mu_1(x) = \int_{\Omega_2} \mu_1(Q_y)\, d\mu_2(y)\,. \tag{2.2}$$

Proof. We define $\pi(Q) := \int_{\Omega_1} \mu_2(Q_x)\, d\mu_1(x)$ for all $Q \in \Sigma_1 \otimes \Sigma_2$. This is well-defined by Lemma 2.9.2. Then π is a measure. Indeed, $\pi(\emptyset) = 0$ and, if Q_n is a sequence of disjoint sets in $\Sigma_1 \otimes \Sigma_2$, then $(Q_n)_x$ is a sequence of disjoint sets in Σ_2. Thus

$$\begin{aligned}\pi\Big(\bigcup_{n\in\mathbb{N}} Q_n\Big) &= \int_{\Omega_1} \mu_2\big(\bigcup_{n\in\mathbb{N}} (Q_n)_x\big)\, d\mu_1(x) = \int_{\Omega_1} \sum_{n\in\mathbb{N}} \mu_2((Q_n)_x)\, d\mu_1(x)\\ &= \sum_{n\in\mathbb{N}} \int_{\Omega_1} \mu_2((Q_n)_x)\, d\mu_1(x) = \sum_{n\in\mathbb{N}} \pi(Q_n)\end{aligned}$$

where we have used monotone convergence in the third step. Since

$$\pi(A \times B) = \int_{\Omega_1} \mu_2((A \times B)_x)\, d\mu_1(x) = \int_{\Omega_1} \mu_2(B)\mathbb{1}_A\, d\mu_1 = \mu_1(A)\mu_2(B)$$

there exists a measure with the required properties. Its uniqueness follows from Corollary 2.2.11.

Now define $\tilde\pi(Q) := \int_{Q_2} \mu_1(Q_y)\, d\mu_2(y)$. Repeating the above computations, we see that $\tilde\pi$ is also a measure with with $\tilde\pi(A \times B) = \mu_1(A)\mu_2(B)$. Consequently, by uniqueness, $\pi = \tilde\pi$ and thus (2.2) holds. □

Here is a neat application.

Proposition 2.9.4. *Let (Ω, Σ, μ) be a σ-finite measure space and $f \geq 0$ be measurable. Then*

$$\int_\Omega f\, d\mu = \int_0^\infty \mu(\{f \geq t\})\, dt\,.$$

Proof. Consider the set $G := \{(x,t) \in \Omega \times [0,\infty) \,:\, f(x) \geq t\}$. Then $G \in \Sigma \otimes \mathscr{B}([0,\infty))$. Indeed, the maps $\Phi, \Psi : \Omega \times [0,\infty)$, given by $\Phi(x,t) = f(x)$ and $\Psi(t,x) = t$, are clearly $\Sigma \otimes \mathscr{B}([0,\infty))$-measurable hence so is $\Phi - \Psi$. But then $G = (\Phi - \Psi)^{-1}([0,\infty))$ is measurable.

By Theorem 2.9.3, we have, on the one hand,

$$(\mu \otimes \lambda)(G) = \int_0^\infty \mu(G_t)\, dt = \int_0^\infty \mu(\{f \geq t\})\, dt$$

and, on the other hand,

$$(\mu \otimes \lambda)(G) = \int_\Omega \lambda(G_x)\, d\mu(x) = \int_\Omega f(x)\, d\mu(x)\,.$$

□

We next turn to the question on when a map $f : \Omega_1 \times \Omega_2 \to \mathbb{K}$ is integrable with respect to $\mu_1 \otimes \mu_2$ and how to compute the integral. We start with a lemma about measurability.

Lemma 2.9.5. *Let (Ω_i, Σ_i) for $i = 1,2,3$ be a measurable space. If $f : \Omega_1 \times \Omega_2 \to \Omega_3$ is $\Sigma_1 \otimes \Sigma_2/\Sigma_3$-measurable, then $f(x,\cdot)$ is Σ_2/Σ_3-measurable for all $x \in \Omega_1$ and $f(\cdot,y)$ is Σ_1/Σ_3-measurable for all $y \in \Omega_2$.*

Proof. For $A \in \Sigma_3$, we have

$$f(x,\cdot)^{-1}(A) = \{y \,:\, f(x,y) \in A\} = \big(f^{-1}(A)\big)_x$$

and

$$f(\cdot,y)^{-1}(A) = \{x \,:\, f(x,y) \in A\} = \big(f^{-1}(A)\big)_y\,.$$

Hence the claim follows from Lemma 2.9.1. □

Theorem 2.9.6. *(Tonelli)*

Let $(\Omega_i, \Sigma_i, \mu_i)$ be σ-finite measure spaces for $i = 1,2$. Moreover, let $f : \Omega_1 \times \Omega_2 \to [0,\infty]$ be $\Sigma_1 \otimes \Sigma_2/\mathscr{B}([0,\infty])$-measurable. then the maps

$$y \mapsto \int_{\Omega_1} f(x,y)\, d\mu_1(x) \quad \text{and} \quad x \mapsto \int_{\Omega_2} f(x,y)\, d\mu_2(y)$$

are measurable and

$$\int_{\Omega_1\times\Omega_2} f\, d\mu_1 \otimes \mu_2 = \int_{\Omega_2}\int_{\Omega_1} f(x,y)\, d\mu_1(x)\, d\mu_2(y) = \int_{\Omega_1}\int_{\Omega_2} f(x,y)\, d\mu_2(x)\, d\mu_1(y)\,. \tag{2.3}$$

In particular, if one of the iterated integrals is finite, then f is integrable with respect to $\mu_1 \otimes \mu_2$.

Proof. Put $\Omega := \Omega_1 \times \Omega_2, \Sigma := \Sigma_1 \times \Sigma_2$ and $\pi := \mu_1 \otimes \mu_2$. First, let f be a simple function, say $f = \sum_{j=1}^n \alpha_j \mathbb{1}_{Q_j}$. Then $f(x,\cdot) = \sum_{j=1}^n \alpha_j \mathbb{1}_{Q_j}(x,\cdot) = \sum_{j=1}^n \alpha_j \mathbb{1}_{(Q_j)_x}(\cdot)$. Thus

$$\int_{\Omega_2} f(x,y)\, d\mu_2(y) = \sum_{j=1}^n \alpha_j \mu_2((Q_j)_x)\, .$$

Taking Lemma 2.9.2 into account, it follows in particular that $x \mapsto \int_{\Omega_2} f(x,y)\, d\mu_2(y)$ is measurable. Moreover, by equation (2.2),

$$\int_{\Omega_1}\int_{\Omega_2} f(x,y)\, d\mu_2(y)\, d\mu_1(y) = \sum_{j=1}^n \alpha_j \int_{\Omega_1} \mu_2((Q_j)_x)\, d\mu_1(x) = \sum_{j=1}^n \alpha_j \pi(Q) = \int_\Omega f\, d\pi\, .$$

Now let f be an arbitrary, nonnegative, measurable function and f_n be a sequence of simple functions increasing to f. Such a sequence exists by Proposition 2.4.6. By the above, $f_n(x,\cdot)$ is a sequence of simple functions which, moreover, increases to $f(x,\cdot)$. Thus, by monotone convergence,

$$\varphi(x) := \int_{\Omega_2} f(x,y)\, d\mu_2(y) = \sup_{n\in\mathbb{N}} \int_{\Omega_2} f_n(x,y)\, d\mu_2(y) =: \sup_{n\in\mathbb{N}} \varphi_n(x)$$

for all $x \in \Omega_1$. In particular, φ is measurable as the pointwise limit of simple functions. It now follows that

$$\int_\Omega f\, d\pi = \sup_{n\in\mathbb{N}} \int_\Omega f_n\, d\pi = \sup_{n\in\mathbb{N}} \int_{\Omega_1} \varphi_n\, d\mu_1 = \int_{\Omega_1} \varphi\, d\mu_1\, .$$

Here, we have used monotone convergence on (Ω, Σ, π), then the result above for simple functions and then monotone convergence on $(\Omega_1, \Sigma_1, \mu_1)$. This proves the first equality in (2.3). The rest follows by interchanging the roles of x and y. □

Theorem 2.9.7. *Let $(\Omega_i, \Sigma_i, \mu_i)$ be σ-finite measure spaces for $i = 1, 2$. If $f : \Omega_1 \times \Omega_2 \to \mathbb{K}$ is $\Sigma_1 \otimes \Sigma_2$-measurable and integrable with respect to $\mu_1 \otimes \mu_2$ then $f(x,\cdot)$ is integrable with respect to μ_2 for μ_1-a.e. $x \in \Omega_1$ and $f(\cdot, y)$ is integrable with respect to μ_1 for μ_2-a.e. $y \in \Omega_2$. Moreover,*

$$\int_{\Omega_1\times\Omega_2} f\, d\mu_1 \otimes \mu_2 = \int_{\Omega_2}\int_{\Omega_1} f(x,y)\, d\mu_1(x)\, d\mu_2(y) = \int_{\Omega_1}\int_{\Omega_2} f(x,y)\, d\mu_2(x)\, d\mu_1(y)\, . \quad (2.4)$$

Proof. First let $\mathbb{K} = \mathbb{R}$. By Tonelli,

$$\int_{\Omega_2}\int_{\Omega_1} |f(x,y)|\, d\mu_1(x)\, d\mu_2(y) = \int_{\Omega_1}\int_{\Omega_2} |f(x,y)|\, d\mu_2(x)\, d\mu_1(y) = \int_{\Omega_1\times\Omega_2} |f|\, d\mu_1 \otimes \mu_2 < \infty$$

by assumption. Hence $x \mapsto \int_{\Omega_1} |f(x,y)|\, d\mu_2(y)$ is μ_1 integrable and thus, in particular, μ_1-a.e. finite. This proves that $f(x,\cdot)$ is μ_2-integrable for μ_1-a.e. x.

By the definition of the integral and Tonelli,

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f\, d\mu_1 \otimes \mu_2 &= \int_{\Omega_1\times\Omega_2} f^+\, d\mu_1 \otimes \mu_2 - \int_{\Omega_1\times\Omega_2} f^-\, d\mu_1 \otimes \mu_2\\ &= \int_{\Omega_1}\int_{\Omega_2} f^+\, d\mu_2\, d\mu_2 - \int_{\Omega_1}\int \Omega_2 f^-\, d\mu_1\, d\mu_2\\ &= \int_{\Omega_1}\int_{\Omega_2} f\, d\mu_1\, d\mu_2\, .\end{aligned}$$

This proves one equality in (2.4), the other follows by interchanging the roles of x and y. If $\mathbb{K} = \mathbb{C}$, consider $\operatorname{Re} f$ and $\operatorname{Im} f$ separately. □

Example 2.9.8. Consider $(\Omega_i, \Sigma_i, \mu_i) = ((0,\infty), \mathscr{B}(0,\infty), \lambda)$ for $i = 1, 2$. Then the product space is $((0,\infty)^2, \mathscr{B}((0,\infty)^2), \lambda_2)$. Consider $f : (0,\infty)^2 \to \mathbb{R}$, given by $f(x,y) = ye^{-(1+x^2)y^2}$. Since f is positive, we may evaluate one-dimensional integrals as improper Riemann integrals. We obtain

$$\int_0^\infty |f(x,y)|\,dy = \Big[-\frac{1}{2}\frac{1}{1+x^2}e^{-(1+x^2)y^2}\Big|_0^\infty = \frac{1}{2}\frac{1}{1+x^2}\,.$$

Hence

$$\int_0^\infty \int_0^\infty |f(x,y)|\,dy\,dx = \int_0^\infty \frac{1}{2}\frac{1}{1+x^2}\,dx = [\frac{1}{2}\arctan x|_0^\infty = \frac{\pi}{4}\,.$$

It follows from Tonelli's Theorem that f is integrable with respect to λ_2. Moreover, the integral is given by $\int_{(0,\infty)^2} f\,d\lambda_2 = \frac{\pi}{4}$. On the other hand, interchanging the order of integration (which is possible by Tonelli's Theorem), we see that

$$\frac{\pi}{4} = \int_0^\infty \int_0^\infty ye^{-(1+x^2)y^2}\,dx\,dy = \int_0^\infty e^{-y^2}\int_0^\infty ye^{x^2y^2}\,dx\,dy = \int_0^\infty e^{-y^2}\,dy\int_0^\infty e^{-z^2}\,dz\,,$$

where we have used the substitution $z = xy$ in the last step. It thus follows that

$$\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}\,.$$

By symmetry,

$$\int_{\mathbb{R}} e^{-x^2}\,dx = \sqrt{\pi}\,.$$

By substituting $x = \frac{t}{\sqrt{2}}$, it follows that

$$\int_{\mathbb{R}} e^{-\frac{t^2}{2}}\,dx = \sqrt{2\pi}\,.$$

Remark 2.9.9. It is possible to generalize the results of this section also to finite products of σ-finite measure spaces. The main task is to prove that the product σ-algebra $\Sigma_1\otimes\cdots\otimes\Sigma_n$ is the product of the two σ-algebras Σ_1 and $\Sigma_2\otimes\cdots\otimes\Sigma_n$. One can then proceed by induction. We leave the details to the reader.

2.10 Independent Random Variables

In this section, we first provide some probabilistic interpretation of the measure-theoretic facts introduced so far. This basically is a change in notation and terminology. We then proceed to introduce the notion of independence, which is of great importance in probability theory.

As we have noted already, a probability space is a measure space (Ω, Σ, μ) with $\mu(\Omega) = 1$, i.e. μ is a probability measure. In probability theory, it is more common to denote the measure with $\mathbb{P}$. The elements $A \in \Sigma$ are called *events* rather than measurable sets and for real- or complex valued measurable functions on Ω one uses capital letters X, Y, Z rather than f, g, h and calls them *random variables*. A measurable map from (Ω, Σ) to another measurable space $(M, \mathscr{F})$ is often called a *random element*. If $M = \mathbb{K}^d$ and $\mathscr{F} = \mathscr{B}(\mathbb{K}^d)$, one also calls such a map a *random vector*.

Now let $X : \Omega \to \mathbb{K}$ be a random variable. If $X \in L^1(\Omega, \Sigma, \mathbb{P})$, one says that X has *finite expectation* and writes

$$\mathbb{E}X := \int_\Omega X\,d\mathbb{P}\,.$$

If $X \in L^p(\Omega, \Sigma, \mathbb{P})$, one says that X has *finite moments up to order* p. Note that this means that $\mathbb{E}|X|^p < \infty$. However, as is clear from Corollary 2.7.17, it follows that $\mathbb{E}|X|^q < \infty$ for all $1 \leq q < p$. Of particular importance is the space $L^2(\Omega, \Sigma, \mathbb{P})$ of random variables with finite second moments. For $X \in L^2(\Omega, \Sigma, \mathbb{P})$,

$$\mathsf{Var}(X) := \mathbb{E}(X - \mathbb{E}X)^2$$

is called the *variance* of X. Note that $(X - \mathbb{E}X)^2 = X^2 + 2X\mathbb{E}X + (\mathbb{E}X)^2$ is integrable (has finite expectation) as both X and $\mathbb{E}X$, being a constant, belong to $L^2(\Omega, \Sigma, \mathbb{P})$.

If $X, Y \in L^2(\Omega, \Sigma, \mathbb{P})$, then

$$\mathsf{Cov}(X, Y) := \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

is called the *Covariance* of X and Y.

The question remains of how to compute the expectation of a random variable. To that end, one should note that the probability space $(\Omega, \Sigma, \mathbb{P})$ is often of little interest itself and not known in detail. What *is* known in applications is the push-forward μ_X of $\mathbb{P}$ under a random variable X. In this case, expectation, variances, etc. can be computed with the help of Theorem 2.5.25.

Example 2.10.1. Standard Gaussian measure is the measure γ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given by

$$\gamma(A) := \int_A e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}}$$

for all $A \in \mathscr{B}(\mathbb{R})$. As a consequence of Example 2.9.8, γ is a probability measure.

A random variable X defined on some probability space $(\Omega, \Sigma, \mathbb{P})$ is called a *standard Gaussian random variable*, if the distribution of X is γ.

A standard Gaussian random variable has finite moments of all order, i.e. $X \in L^p(\Omega, \Sigma, \mathbb{P})$ for all $1 \leq p < \infty$. Note, however, that $X \notin L^\infty(\Omega, \Sigma, \mathbb{P})$.

Proof. By Theorem 2.5.25, we have $|X|^p \in L^1(\Omega, \Sigma, \mathbb{P})$ if and only if the measurable function $f : t \mapsto |t|^p$ belongs to $L^1(\mathbb{R}, \mathscr{B}(\mathbb{R}), \gamma)$ and in this case

$$\int_\Omega |X|^p \, d\mathbb{P} = \int_\mathbb{R} t^p \, d\gamma(t) = \int_\mathbb{R} |t|^p e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}}$$

Where the last equality follows from Exercise 2.5.11.

To see that this integral is finite for all p, it suffices to note that for $0 < \varepsilon < \frac{1}{2}$ the map $t \mapsto |t|^p e^{-\frac{t^2}{2}} e^{-\varepsilon t^2}$ is bounded on $\mathbb{R}$. Thus, there exists a constant c_p such that $|t|^p e^{-\frac{t^2}{2}} \leq c_p e^{-\varepsilon t^2}$. Since the latter is integrable with respect to Lebesgue measure, it follows that also $t \mapsto |t|^p e^{-\frac{t^2}{2}}$ is integrable with respect to Lebesgue measure, hence the latter integral is finite. $\square$

Example 2.10.2. Let (X, Y) be a random vector which is equally distributed on the triangle

$$T := \{(x, y) \in \mathbb{R}^2 \,:\, 0 \leq x \leq 2, 0 \leq y \leq \frac{x}{2}\},$$

i.e. the distribution $\mu_{(X,Y)}$ has with respect to the 2-dimensional Lebesgue measure the density $\frac{1}{\lambda_2(T)} \mathbb{1}_T$. We compute $\mathsf{Cov}(X, Y)$.

Noting that $X = \varphi(X,Y)$ for $\varphi(t,s) = t$, we obtain with the rules for integrating with respect to push-forward measures and measures with density

$$\mathbb{E}X = \int_\Omega \varphi(X,Y)\, d\mathbb{P} = \int_{\mathbb{R}^2} \varphi \mathbb{1}_T \, d\lambda_2 = \int_0^2 t \int_0^{t/2} dsdt = \int_0^2 \frac{t^2}{2}\, dt = \frac{4}{3}\,,$$

where we have used Fubini's theorem in the third step. Similarly, with $\psi(t,s) = s$, we have

$$\mathbb{E}Y = \int_{\mathbb{R}^2} \psi \mathbb{1}_T \, d\lambda_2 = \int_0^2 \int_0^{t/2} s\, ds\, dt = \int_0^2 \frac{t^2}{8}\, dt = \frac{1}{3}\,.$$

Hence, with $\rho(t,s) = (t - \frac{4}{3})(s - \frac{1}{3})$ we obtain

$$\begin{aligned}\mathsf{Cov}(X,Y) &= \int_\Omega \rho(X,Y)\, d\mathbb{P} = \int_{\mathbb{R}^2} \rho(X,Y)\, d\mu_{X,Y} = \int_0^2 \big(t - \frac{4}{3}\big) \int_0^{t/2} \big(s - \frac{1}{3}\big)\, ds\, dt \\ &= \int_0^2 \big(t - \frac{4}{3}\big)\Big[\frac{s^2}{2} - \frac{s}{3}\Big|_{s=0}^{t/2} dt = \int_0^2 \frac{t^3}{8} - \frac{t^2}{3} + \frac{2t}{9}\, dt = \frac{1}{18}\,.\end{aligned}$$

Often, a probability space carries more than one random variable which is of interest. In this situation, it is not sufficient to merely know the distribution function of each of them. Rather, one has to understand how the random variables interact with each other. In the example above, we knew explicitly the distribution of the *vector* (X,Y). Another important concept is that of *independence*.

Definition 2.10.3. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. We call a family $(A_i)_{i\in I} \subset \Sigma$ of events *independent* if for every $n \in \mathbb{N}$ and distinct $i_1, \ldots, i_n \in I$ we have

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n}) = \prod_{k=1}^n \mathbb{P}(A_{i_k})\,.$$

If $(\mathscr{F}_i)_{i\in I}$ is a family of sub-σ-algebras of Σ, We say that the σ-algebras $\mathscr{F}_i$ are *independent*, if for all $n \in \mathbb{N}$, distinct $i_1, \ldots, i_n \in I$ and $A_k \in \mathscr{F}_{i_k}$ for $k = 1, \ldots, n$, we have

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \prod_{k=1}^n \mathbb{P}(A_k)\,. \tag{2.5}$$

We say that a family of random elements $(X_i)_{i\in I}$ are *independent*, if the σ-algebras $(\sigma(X_i))_{i\in I}$ are independent. Similarly, we say that a random variable X is independent of a σ-algebra $\mathscr{F}$, if $\mathscr{F}$ and $\sigma(X)$ are independent.

Let us note first that to prove independence, it suffices to check (2.5) on generators stable under intersection.

Lemma 2.10.4. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and, for every i in an index set I, $\mathscr{F}_i \subset \Sigma$ be a σ-algebra and S_i be a generator of $\mathscr{F}_i$ which is stable under intersections. Then the $\mathscr{F}_i$ are independent if and only if for all $n \in \mathbb{N}$, distinct $i_1, \ldots, i_n \in I$ and $C_k \in S_{i_k}$ for $k = 1, \ldots, n$, we have*

$$\mathbb{P}(C_1 \cap \ldots \cap C_n) = \prod_{k=1}^n \mathbb{P}(C_k)\,.$$

Proof. We may assume that $S_i \neq 0$ for all t. By assumption, equation (2.5) holds whenever $A_j \in S_{i_j}$ for all $1 \leq j \leq n$. Fix $A_1, \ldots, A_{n-1}$ and denote by $\mathscr{D}$ the collection of $A \in \mathscr{F}_{i_n}$ such that (2.5) holds for $A_{i_n} = A$. Then $\mathscr{D}$ is a Dynkin system (we leave the easy proof to the reader) and contains S_{i_n}. Thus, by Dynkin's π - λ theorem, $\mathscr{D} = \mathscr{F}_{i_n}$. Thus (2.5) holds for arbitrary $A_n \in \mathscr{F}_{i_n}$ and $A_k \in S_{i_k}$ for $1 \leq k \leq n-1$. Arguing as above, we see that we can allow also arbitrary $A_{n-1} \in \mathscr{F}_{i_{n-1}}$ and thus, proceeding by induction, we can allow that $A_k \in \mathscr{F}_{i_k}$ for all $1 \leq k \leq n$. □

The next result relates independence with product measures. This will also give examples of independent random variables.

Theorem 2.10.5. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $(X_i)_{i \in I}$ be a family of random variables with distribution μ_i. Then the family $(X_i)_{i \in I}$ is independent if and only if for every $n \in \mathbb{N}$ and distinct $i_1, \ldots, i_n$ in I the distribution of the random vector $(X_{i_1}, \ldots, X_{i_n})$ is $\mu_{i_1} \otimes \cdots \otimes \mu_{i_n}$.*

Proof. Assuming independence, fix $n \in \mathbb{N}$ and distinct $i_1, \ldots, i_n$ in I. Let $B = B_1 \times \cdots \times B_n$ be such that $B_k \in \mathscr{B}(\mathbb{K})$ for $k = 1, \ldots, n$. Then

$$\begin{aligned} \mathbb{P}((X_{i_1}, \cdots, X_{i_n})^{-1} B) &= \mathbb{P}(\{X_{i_1} \in B_1\} \cap \cdots \cap \{X_{i_n} \in B_n\}) \\ &= \prod_{k=1}^{n} \mathbb{P}(\{X_{i_k} \in B_k\}) = \prod_{k=1}^{n} \mathbb{P}\mu_{i_k}(B_k) \end{aligned}$$

which shows that on cuboids $B = B_1 \times \cdots \times B_n$ the distribution of $(X_{i_1}, \cdots, X_{i_n})$ agrees with $\mu_{i_1} \otimes \cdots \otimes \mu_{i_n}$. Since the cuboids are a generator of $\mathscr{B}(\mathbb{K}^n)$ which is stable under intersections, the equality of the measures follows from Theorem 2.2.10.

The converse implication is trivial. □

Example 2.10.6. On the probability space $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \bigotimes_{k=1}^n \gamma)$, the random variables $X_k : x \mapsto x_k$ are independent and standard Gaussian.

Theorem 2.10.7. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $(\mathscr{F}_n)_{n \in \mathbb{N}}$ be a sequence of independent sub-σ-algebras of Σ. The* tail σ-algebra *$\mathscr{T}$ is defined as*

$$\mathscr{T} := \bigcap_{n \in \mathbb{N}} \sigma\Big(\bigcup_{k \geq n} \mathscr{F}_k \Big) .$$

Then the tail σ-algebra $\mathscr{T}$ is $\mathbb{P}$-trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathscr{T}$.

Proof. We fix $A \in \mathscr{T}$ and define $\mathscr{G} := \{B \in \Sigma : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$. Then $\mathscr{G}$ is a Dynkin system, as is easy to see.

Now define $\Sigma_n := \sigma(\mathscr{F}_1 \cup \cdots \cup \mathscr{F}_n)$ and $\mathscr{A}_n := \sigma(\bigcup_{k \geq n} \mathscr{F}_k)$. Then for every $n \in \mathbb{N}$ the σ-algebras Σ_n and $\mathscr{A}_n$ are independent. Indeed, given $B_j \in \mathscr{F}_j$ for $j = 1, \ldots, n$ and $C_k \in \mathscr{F}_{n+k}$ for $k = 1, \ldots, m$, put $B = B_1 \cap \ldots B_n$ and $C = C_1 \cap \ldots \cap C_m$. Then, by independence,

$$\mathbb{P}(B \cap C) = \mathbb{P}(B_1 \cap \ldots \cap B_n \cap C_1 \cap \ldots C_m) = \prod_{j=1}^{n} \mathbb{P}(B_j) \prod_{k=1}^{m} \mathbb{P}(C_k) = \mathbb{P}(B)\mathbb{P}(C) .$$

Since Σ_n is generated by sets B of this form and $\mathscr{A}_n$ is generated by sets C of this form and these generators are stable under intersections, the claimed independence follows from Lemma 2.10.4.

Since $A \in \mathscr{T} \subset \mathscr{A}_n$ for all $n \in \mathbb{N}$, it follows that $\mathbb{P}(B \cap A) = \mathbb{P}(B)\mathbb{P}(A)$ for all $B \in \Sigma_n$, i.e. $\Sigma_n \subset \mathscr{G}$ for all $n \in \mathbb{N}$. By Dynkin's π-λ theorem, $\sigma(\bigcup_{n\in\mathbb{N}} \Sigma_n) = d(\bigcup_{n\in\mathbb{N}} \Sigma_n) \subset \mathscr{G}$. Since clearly $\mathscr{T} \subset \sigma(\bigcup_{n\in\mathbb{N}} \Sigma_n)$ it follows that $A \in \mathscr{G}$, hence $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$, i.e. $\mathbb{P}(A) \in \{0, 1\}$. □

Let us give some typical examples of members of the tail σ-algebra:

Example 2.10.8. Consider the random experiment of rolling a die infinitely often, independently of each other. We let X_n be the outcome in the n-th turn. Then the random variables X_n, hence the σ-algebras $\mathscr{F}_n := \sigma(X_n)$ are independent.

Now $A := \{\text{infinitely often a 6 is rolled}\} \in \mathscr{T}$. Indeed,

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} \{X_k = 6\} \in \bigcap_{n\in\mathbb{N}} \sigma\Big(\bigcup_{k\geq n} \mathscr{F}_k\Big)$$

since $\{X_k = 6\} \in \mathscr{F}_k$.

Also the event

$$B := \{\text{eventually, only even numbers are rolled}\} = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} \{X_k \in \{2, 4, 6\}\}$$

belongs to $\mathscr{T}$, since

$$B = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} \bigcap_{l\geq k} \{X_l \in \{2, 4, 6\}\}\,.$$

On the other hand, the event

$$C = \{X_{1000321} = 5\}$$

does not belong to $\mathscr{T}$, since otherwise C should be independent of every $A \in \mathscr{F}_n$ for all $n \in \mathbb{N}$, but $C \in \mathscr{F}_{1000321}$ and $\frac{1}{6} = \mathbb{P}(C) = \mathbb{P}(C \cap C) \neq \mathbb{P}(C)\mathbb{P}(C) = \frac{1}{36}$.

Chapter 3

Hilbert spaces

The proof of the Hilbert basis theorem is not mathematics, it is theology.
Camille Jordan

We now continue to study a special class of Banach spaces, namely Hilbert spaces, in which the presence of a so-called "inner product" allows us to define angles between elements. In particular, we can introduce the concept of orthogonality. This has far-reaching consequences.

3.1 Definition and Examples

Definition 3.1.1. Let H be a vector space over $\mathbb{K}$. An *inner product* (or a *scalar product*) on H is a map $(\,\cdot\,|\,\cdot\,) : H \times H \to \mathbb{K}$ such that

(IP1) For all $x \in H$ we have $(x\,|\,x) \geq 0$ and $(x\,|\,x) = 0$ if and only if $x = 0$.

(IP2) For all $x, y \in H$, we have $(x\,|\,y) = \overline{(y\,|\,x)}$.

(IP3) For all $x, y, z \in H$ and $\lambda \in \mathbb{K}$, we have $(\lambda x + y\,|\,z) = \lambda(x\,|\,z) + (y\,|\,z)$.

The pair $(H, (\,\cdot\,|\,\cdot\,))$ is called *inner product space* or *pre-Hilbert space.*

Example 3.1.2. (a) On $H = \mathbb{R}^d$,

$$(x\,|\,y) := \sum_{j=1}^{d} x_j \bar{y}_j$$

defines an inner product.

(b) On $H = \ell^2$,

$$(\mathbf{x}\,|\,\mathbf{y}) := \sum_{j=1}^{\infty} x_j \bar{y}_j$$

defines an inner product.

(c) On $C([a,b])$,

$$(f \,|\, g) := \int_a^b f(t)\overline{g(t)}\, dt$$

defines an inner product. If $w : [a,b] \to \mathbb{R}$ is such that there exist constants $0 < \varepsilon < M$, such that $\varepsilon \leq w(t) \leq M$ for all $t \in [a,b]$, then also

$$(f \,|\, g)_w := \int_a^b f(t)\overline{g(t)} w(t)\, dt$$

defines an inner product on $C([a,b])$.

(d) If (Ω, Σ, μ) is a measure space, then

$$(f \,|\, g) := \int_\Omega f\bar{g}\, d\mu$$

defines an inner product on $L^2(\Omega, \Sigma, \mu)$.

Note that for examples (b), (c) and (d), it follows from Hölder's inequality that the sum resp. the integral is well-defined.

Remark 3.1.3. Condition (IP1) is called *definiteness* of $(\,\cdot\,|\,\cdot\,)$, (IP2) is called *symmetry*. Note that if $\mathbb{K} = \mathbb{R}$, then (IP2) reduces to $(x \,|\, y) = (y \,|\, x)$. (IP3) states that $(\,\cdot\,|\,\cdot\,)$ is linear in the first component. Note that it follows from (IP2) and (IP3) that

$$(x \,|\, \lambda y + z) = \bar{\lambda}(x \,|\, y) + (x \,|\, z)\,.$$

If $(H, (\,\cdot\,|\,\cdot\,))$ is an inner product space, we put

$$\|x\| := \sqrt{(x \,|\, x)}\,.$$

Lemma 3.1.4. *(Cauchy-Schwarz)*
Let $(H, (\,\cdot\,|\,\cdot\,))$ be an inner product space. Then

$$|(x \,|\, y)| \leq \|x\| \cdot \|y\| \tag{3.1}$$

and equality holds if and only if x and y are linearly dependent.

Proof. Let $\lambda \in \mathbb{K}$. Then

$$\begin{aligned} 0 &\leq (x + \lambda y \,|\, x + \lambda y) = (x \,|\, x) + \lambda(x \,|\, x) + \bar{\lambda}(x \,|\, y) + |\lambda|^2 (y \,|\, y) \\ &= \|x\|^2 + 2\mathrm{Re}\,(\bar{\lambda}(x \,|\, y)) + |\lambda|^2 \|y\|^2\,. \end{aligned}$$

If $y = 0$, then $(x \,|\, y) = 0$ and the claimed inequality holds true. If $y \neq 0$, we may put $\lambda = -\overline{(x \,|\, y)}\|y\|^{-2}$. Then we obtain

$$0 \leq \|x\|^2 - 2\mathrm{Re}\,\frac{|(x \,|\, y)|^2}{\|y\|^2} + \frac{|-\overline{(x \,|\, y)}|^2}{\|y\|^4}\|y\|^2 = \|x\|^2 - \frac{|(x \,|\, y)|^2}{\|y\|^2}$$

that is (3.1). By (IP1), equality holds if and only if $x = \lambda y$. □

Lemma 3.1.5. *Let $(H, (\,\cdot\,|\,\cdot\,))$ be an inner product space. Then $\|\cdot\| := \sqrt{(\,\cdot\,|\,\cdot\,)}$ defines a norm on H.*

Proof. If $\|x\| = 0$, then $(x \,|\, x) = 0$ hence $x = 0$ by (IP1). This proves (N1). For (N2), let $x \in H$ and $\lambda \in \mathbb{K}$ be given. Then $\|\lambda x\|^2 = (\lambda x \,|\, \lambda x) = \lambda\bar{\lambda}(x \,|\, x) = |\lambda|^2\|x\|^2$. (N3) holds, since

$$\|x + y\|^2 = \|x\|^2 + 2\mathrm{Re}\,(x \,|\, y) + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$$

where we used the Cauchy-Schwarz inequality. □

Definition 3.1.6. A Hilbert space is an inner product space $(H, (\,\cdot\,|\,\cdot\,))$ which is complete with respect to the norm $\|\cdot\| := \sqrt{(\,\cdot\,|\,\cdot\,)}$.

Example 3.1.7. The inner product spaces from Example 3.1.2 (a), (b) and (d) are Hilbert spaces. This follows from Example 1.5.5 and Theorem 2.7.7. The inner product space from Example 3.1.2 (c) is not a Hilbert space. To see this, let $[a, b] = [-1, 1]$. Consider

$$f_n : t \mapsto \begin{cases} 0, & t \in [-1, 0] \\ nt, & t \in (0, n^{-1}] \\ 1, & t \in (n^{-1}, 1]\,. \end{cases}$$

Then $f_n \to f := \mathbb{1}_{(0,1)}$ in $L^2(-1, 1)$, in particular, it is a Cauchy-sequence in that space. Since $f_n \in C([-1, 1])$ and the norm is the same as that in $L^2(-1, 1)$, it follows that f_n is a Cauchy sequence in $(C([-1, 1]), (\,\cdot\,|\,\cdot\,))$. However, f does not converge in $(C([-1, 1]), (\,\cdot\,|\,\cdot\,))$. Indeed, if $f_n \to g$ in that $C([-1, 1])$, we must have $f_n \to f$ in $L^2(-1, 1)$ but then we must have $f = g$ almost everywhere. Since there is no continuous function on $[-1, 1]$ which coincides with $\mathbb{1}_{(0,1)}$ almost everywhere, no such function g can exist.

Lemma 3.1.8. *(Parallelogram identity)*
Let H be an inner product space. Then

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

for all $x, y \in H$.

Proof.

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

□

Exercise 3.1.9. Show that in $(\mathbb{R}^d, \|\cdot\|_p), (\ell^p, \|\cdot\|_p)$ and $(L^p([0, 1]), \|\cdot\|_p)$ the parallelogram identity fails. Hence there is no inner product $(\,\cdot\,|\,\cdot\,)_p$ such that $\|x\|_p = \sqrt{(x \,|\, x)_p}$.

Lemma 3.1.10. *Let H be an inner product space. Then $(\,\cdot\,|\,\cdot\,) : H \times H \to \mathbb{K}$ is continuous.*

Proof. Let $x_n \to x$ and $y_n \to y$. Then $M := \sup_{n\in\mathbb{N}} \|x_n\| < \infty$. By the Cauchy-Schwarz inequality,

$$|(x_n \,|\, y_n) - (x \,|\, y)| \leq |(x_n \,|\, y_n - y)| + |(x_n - x \,|\, y)| \leq M\|y_n - y\| + \|y\|\|x_n - x\| \to 0\,.$$

□

3.2 Orthogonal Projection

Definition 3.2.1. Let $(H, (\cdot\,|\,\cdot))$ be an inner product space. We say that two vectors $x, y \in H$ are *orthogonal* (and write $x \perp y$) if $(x\,|\,y) = 0$.

Given a subset $S \subset H$, the *annihilator* $S^\perp$ of S is defined by

$$S^\perp := \{y \in H \,:\, y \perp x \;\forall\, x \in S\}\,.$$

If S is a subspace, then $S^\perp$ is also called the *orthogonal complement* of S.

Lemma 3.2.2. *(Pythagoras)*

Let H be an inner product space. If $x \perp y$, then

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

Proof.

$$\|x + y\|^2 = \|x\|^2 + 2\mathrm{Re}\,(x\,|\,y) + \|y\|^2 = \|x\|^2 + \|y\|^2\,.$$

□

Proposition 3.2.3. *Let H be an inner product space and $S \subset H$.*

(a) $S^\perp$ is a closed, linear subspace of H.

(b) $\overline{\mathrm{span}}S \subset (S^\perp)^\perp$.

(c) $\overline{\mathrm{span}}S \cap S^\perp = \{0\}$.

Proof. (a) If $x, y \in S^\perp$ and $\lambda \in \mathbb{K}$, then, for $z \in S$, we have $(\lambda x + y\,|\,z) = \lambda(x\,|\,z) + (y\,|\,z) = 0$, hence $\lambda x + y \in S^\perp$. This shows that $S^\perp$ is a linear subspace. If x_n is a sequence in $S^\perp$ which converges to x, then, for $z \in S$, we infer from Lemma 3.1.10 that $(x\,|\,z) = \lim\,(x_n\,|\,z) = 0$.

(b) By (a), $(S^\perp)^\perp$ is a closed linear subspace which, moreover, contains S. Thus $\overline{\mathrm{span}}S \subset (S^\perp)^\perp$.

(c) If $x \in \overline{\mathrm{span}}S \cap S^\perp$, then, by (b), $x \in S^\perp \cap (S^\perp)^\perp$ hence $x \perp x$. But this means $(x\,|\,x) = 0$ which, by (IP1), forces $x = 0$. □

We now come to the main result of this section:

Theorem 3.2.4. *Let $(H, (\cdot\,|\,\cdot))$ be a Hilbert space and $K \subset H$ be a closed linear subspace. Then for every $x \in H$, there exists a unique element $P_K x$ of K such that*

$$\|P_K x - x\| = \min\{\|y - x\| \,:\, y \in K\}\,.$$

Proof. Let $d := \inf\{\|y - x\| \,:\, y \in K\}$. By the definition of the infimum, there exists a sequence y_n in K with $\|y_n - y\| \to d$. Applying the parallelogram identity 3.1.8 to the vectors $x - y_n$ and $x - y_m$, we obtain

$$\begin{aligned}2(\|x - y_n\|^2 + \|x - y_m\|^2) &= \|(x - y_n) + (x - y_m)\|^2 + \|x - y_n - (x - y_m)\|^2\\ &= \|2x - y_n - y_m\|^2 + \|y_n - y_m\|^2 = 4\Big\|x - \frac{1}{2}(y_n + y_m)\Big\|^2 + \|y_n - y_m\|^2\,.\end{aligned}$$

Since $z_{nm} := \frac{1}{2}(y_n + y_m) \in K$, we have $\|x - z_{nm}\|^2 \geq d^2$ and thus

$$\|y_n - y_m\|^2 \leq 2(\|x - y_n\|^2 + \|x - y_m\|^2) - 4d^2\,.$$

By the choice of the sequence y_n, the right-hand side of this equation converges to 0 as $n, m \to \infty$, proving that y_n is a Cauchy sequence. Since H is complete, y_n converges to some vector $P_K x$. Since K is closed, $P_K x \in K$. We have thus proved existence.

As for uniqueness, if $\|z - x\| = \min\{\|y - x\| : y \in K\}$, then, by the parallelogram identity,

$$2\|x - P_K x\|^2 + 2\|x - z\|^2 = \|z - P_K x\|^2 + 4\|x - \frac{1}{2}(P_K x + z)\|^2$$

and thus

$$4d^1 = 4\|x - \frac{1}{2}(P_K x + z)\|^2 + \|P_K x - z\|^2 \geq 4d^2 + \|P_K x - z\|^2$$

proving that $\|P_K x - z\| = 0$, hence $P_K x = z$.

□

Definition 3.2.5. The map $P_K : H \to H$ from Theorem 3.2.4 is called the *orthogonal projection onto K.*

We now collect some properties of P_K.

Proposition 3.2.6. *Let H be a Hilbert space, K be a closed subspace of H and P_K be the orthogonal projection onto K.*

(a) For all $x, y \in H$, we have $P_K x = y$ if and only if $y \in K$ and $x - y \in K^\perp$.

(b) P_K is a bounded linear operator on H.

(c) $P_K^2 = P_k$ and $(P_K x \,|\, y) = (x \,|\, P_K y)$ for all $x, y \in H$.

Proof. (a) If $y \in K$ and $x - y \in K^\perp$, then for every $z \in K$ we have $z - y \in K$ and thus $x - y \perp y - z$. By Pythagoras,

$$\|x - y\|^2 = \|x - y\|^2 + \|y - z\|^2 \geq \|x - y\|^2 .$$

Thus $\|x - y\| = \min\{\|x - z\| : z \in K\}$, proving that $P_K x = y$.

Conversely, if $P_K x = y$, then clearly $y \in K$. Assume that $x - P_K x \notin K^\perp$. Then there exists $0 \neq z \in K$ with $(x - y \,|\, z) \neq 0$. We may assume that $(x - y \,|\, z) = 1$ (otherwise, we replace z with $\overline{(x - y \,|\, z)}^{-1} z$).

Then, for $\lambda \in \mathbb{R}$,

$$\|x - y - \lambda z\|^2 = \|x - y\|^2 - 2\mathrm{Re}\, \lambda (x - y \,|\, z) + \lambda^2 \|z\|^2 = \|x - y\|^2 - 2\lambda + \lambda^2 \|z\|^2 .$$

The latter is strictly less than $\|x - y\|^2$ if $\lambda^2 \|z\|^2 < 2\lambda$, i.e. $\lambda < 2\|z\|^{-2}$. Hence we find an element in K (for example $y + \|z\|^{-2} z$) which is closer to x than y. But then $y \neq P_K x$.

(b) Let $x, y \in H$ and $\lambda \in \mathbb{K}$. By (a), $x - P_K x, y - P_K y \in K^\perp$. Since $K^\perp$ is a subspace by Proposition 3.2.3, $\lambda x - \lambda P_K x + y - P_K y = (\lambda x + y) - (\lambda P_K x + P_K y) \in K^\perp$. Since $\lambda P_K x + P_K y \in K$, it follows from (a) that $P_K(\lambda x + y) = \lambda P_K x + P_K y$, i.e. P_K is linear. As for the boundedness, observe that $x = P_K x + (x - P_K x)$ where $P_K x \perp x - P_K x$ by (a). Thus, by Pythagoras,

$$\|x\|^2 = \|P_K x\|^2 + \|x - P_K x\|^2 \geq \|P_k\|^2 ,$$

proving the boundedness of P_K.

(c) $P_K x \in K$ and $0 = P_K x - P_K x \in K^\perp$. Hence, by (a), $P_K P_K x = P_K x$.
For the second part, observe that

$$(P_K x \,|\, y) = (P_k x \,|\, P_K y) + (P_K x \,|\, y - P_K y) = (P_K x \,|\, P_K y)$$

since $y - P_K y \in K^\perp$ and $P_K x \in K$ by (a). Similarly, one sees that $(x \,|\, P_K y) = (P_K x \,|\, P_K y)$. □

Example 3.2.7. Let $H = \ell^2$ and, for $J \subset \mathbb{N}$, $K_J := \{\mathbf{x} \in \ell^2 \,:\, x_j = 0 \,\forall j \in J\}$. Using that $\mathbf{x}_n \to x$ implies that $x_i^{(n)} \to x_i$ for all $i \in \mathbb{N}$, it is easy to see that K_j is closed. Then $K_J^\perp = K_{J^c}$ and $P_{K_J}\mathbf{x} = \mathbf{y}$, where $y_j = 0$ for $j \in J$ and $y_j = x_j$ for $j \notin J$. Indeed, $\mathbf{x} - \mathbf{y} = \mathbf{z}$, where $z_j = x_j$ for $j \in J$ and $z_j = 0$ for $j \notin J$, thus $\mathbf{z} \in K_J^\perp$.

We can now refine Proposition 3.2.3

Corollary 3.2.8. *If H is a Hilbert space and K is a linear subspace of H, then $\overline{K} = (K^\perp)^\perp$.*

Proof. We have seen already that $\overline{K} \subset (K^\perp)^\perp$. Now let $y \in (K^\perp)^\perp$. Then $y = P_{\bar{K}} y + (1 - P_{\bar{K}})y =: y_1 + y_2$. Thus $\|y_2\|^2 = (y_2 \,|\, y_2) = (y_2 \,|\, y) - (y_2 \,|\, y_1) = 0$, since $y_2 \in \bar{K}^\perp = K^\perp$ and $y \in (K^\perp)^\perp$ and $y_1 \in \bar{K}$. It follows that $y_2 = 0$, hence $y = y_1 \in \bar{K}$. one sees that □

An important consequence of Theorem 3.2.4 is

Theorem 3.2.9. *(Fréchet-Riesz)*
Let H be a Hilbert space. Then $\varphi \in H^*$ if and only if there exists a $y \in H$ such that $\varphi(x) = (x \,|\, y)$ for all $x \in H$.

Proof. If $\varphi(x) = (x \,|\, y)$, then φ is continuous as a consequence of Lemma 3.1.10.

Conversely, let $\varphi \in H^*$ be given. Then $K := \ker \varphi$ is a closed subspace of H: If $K = H$, pick $y = 0$. If $K \neq H$, there exists $x_0 \in H$ with $\varphi(x_0) \neq 0$. Put $z = x_0 - P_K x_0$. Since $x_0 \notin K$, we have $z \neq 0$ and may thus define $w = \|z\|^{-1} z$. Then $\|w\| = 1$ and $w \in K^\perp$. In particular, $\varphi(w) \neq 0$.

Now for $x \in H$, we have $\varphi(x) = \frac{\varphi(x)}{\varphi(w)}\varphi(w) =: \lambda\varphi(w)$ and hence, by linearity, $\varphi(x - \lambda w) = 0$ and thus $x - \lambda w \in K$. Put $y := \overline{\varphi(w)}w$. Then

$$(x \,|\, y) = \varphi(w)(x \,|\, w) = \varphi(w)\big[(x - \lambda w \,|\, w) + (\lambda w \,|\, w)\big] = \varphi(w)\lambda\|w\|^2 = \varphi(x)\,.$$

□

Exercise 3.2.10. Let H be a Hilbert space and K be a closed, linear subspace of H with $K \neq H$. Given $x_0 \in H \setminus K$, show that there exists $\varphi \in H^*$ such that $\varphi(x_0) = 1$ and $\varphi(x) = 0$ for all $x \in K$.

3.3 The Radon-Nikodym Theorem

In this section, we use the Theorem of Fréchet-Riesz to give a short prove of the famous Radon-Nikodym theorem, due to von Neumann. This illustrates the power of Hilbert space techniques.

Let (Ω, Σ) be a measurable space. We recall that if μ and ν are two measures on (Ω, Σ), then we say that μ has density h with respect to ν if

$$\mu(A) = \int_A h \, d\nu\,.$$

We now ask the question when a measure μ has such a density. Certainly, if this is the case and $\nu(A) = 0$, then also $\mu(A) = 0$. If this is the case, we say that μ is *absolutely continuous with respect to* ν and write $\mu \ll \nu$.

Theorem 3.3.1. *Let μ, ν be finite measures on the measurable space (Ω, Σ) with $\mu \ll \nu$. Then there exists a positive $h \in L^1(\Omega, \Sigma, \nu)$ such that*

$$\mu(A) = \int_A h \, d\nu \quad \forall \, A \in \Sigma \, .$$

Proof. For the measure $\rho := \mu + \nu$, consider the real Hilbert space $L^2(\Omega, \Sigma, \rho)$. Then $\varphi : L^2(\Omega, \Sigma, \rho) \to \mathbb{R}$, defined by

$$\varphi(f) := \int_\Omega f \, d\mu$$

is a bounded linear functional. Indeed,

$$|\varphi(f)| \le \int_\Omega |f| \, d\mu \le \int_\Omega |f| \, d\mu + \int_\Omega |f| \, d\nu = \int_\Omega |f| \, d\rho \le \rho(\Omega)^{\frac{1}{2}} \|f\|_{L^2(\Omega,\Sigma,\rho)} \, .$$

By Theorem 3.2.9, there exists $g \in L^2(\Omega, \Sigma, \rho)$ with $\varphi(f) = \int_\Omega fg \, d\rho$, i.e.

$$\int_\Omega f(1-g) \, d\mu = \int_\Omega fg \, d\nu \, . \tag{3.2}$$

We now proceed in several steps.

Step 1: $g \ge 0$ μ-a.e. and ν-a.e.

Indeed, putting $A := \{g < 0\} \in \Sigma$ and $f = \mathbb{1}_A$, equation (3.2) implies

$$0 \le \int_A (1-g) \, d\mu = \int_A g \, d\nu = -\int_\Omega g^- \, d\nu$$

which yields $g^- = 0$ ν-a.e., hence $\nu(A) = 0$. Since $\mu \ll \nu$, we obtain $\mu(A) = 0$.

Step 2: $g < 1$ μ-a.e. and ν-a.e.

Indeed, putting $A := \{g \ge 1\} \in \Sigma$ and $f = \mathbb{1}_Q$, equation (3.2) implies

$$0 \ge \int_A (1-g) \, d\mu = \int_A g \, d\nu \ge \int_A 1 \, d\nu = \nu(A)$$

hence $\nu(A) = 0$. Thus, since $\mu \ll \nu$ also $\mu(A) = 0$.

Step 3: Equation (3.2) is true for every f with $f \ge 0$ ν-a.e.

Indeed, putting $f_n := f \wedge n \in L^2(\Omega, \Sigma, \rho)$ we have $f_n(1-g) \ge 0$ μ-a.e. and $f_n(1-g) \uparrow f(1-g)$ and $f_n g \ge 0$ ν-a.e. and $f_n \uparrow fg$. Hence, by monotone convergence and (3.2)

$$\int_\Omega f(1-g) \, d\mu = \sup_{n \in \mathbb{N}} \int_\Omega f_n(1-g) \, d\mu = \sup_{n \in \mathbb{N}} \int_\Omega fg \, d\nu = \int_\Omega fg \, d\nu \, .$$

Step 4: We finish the proof.

Put $h = \frac{g}{1-g}$. Note that $h \ge 0$ ν-a.e. by Steps 1 and 2. Now, for $A \in \Sigma$, put $f = (1-g)^{-1} \mathbb{1}_A$. Then $f \ge 0$ ν-a.e., hence, by Step 3

$$\mu(A) = \int_\Omega f \, (1-g) \, d\mu = \int_\Omega fg \, d\nu = \int_A h \, d\nu \, .$$

□

Let us present an application of the Radon-Nikodym theorem.

Theorem 3.3.2. *Let (Ω, Σ, ν) be a finite measure space, $1 < p < \infty$ and $\varphi \in L^p(\Omega, \Sigma, \nu)^*$ be such that $\varphi(f) \geq 0$ whenever $f \geq 0$ almost everywhere. Let $q \in (1, \infty)$ be such that $\frac{1}{q} + \frac{1}{p} = 1$. Then there exists $g \in L^q(\Omega, \Sigma, \nu)$ such that*

$$\varphi(f) = \int_\Omega fg \, d\nu \,.$$

Proof. Put $\nu(A) = \varphi(\mathbb{1}_A)$. Then ν is a measure. Indeed, $\nu(\emptyset) = \varphi(0) = 0$. Moreover, if $(A_k) \subset \Sigma$ and $A = \bigcup_{k\in\mathbb{N}} A_k$, observe that with $f_n := \sum_{k=1}^n \mathbb{1}_{A_k} = \mathbb{1}_{\bigcup_{k=1}^n A_k}$ we have $f_n \to \mathbb{1}_A$ pointwise. Since $f_n \leq \mathbb{1}_\Omega \in L^1(\Omega, \Sigma, \nu)$, we have $f_n \to \mathbb{1}_A$ in $L^1(\Omega, \Sigma, \nu)$ by dominated convergence. Using linearity and continuity of φ,

$$\nu(\bigcup_{n\in\mathbb{N}} A_k) = \varphi(\mathbb{1}_A) = \lim_{n\to\infty} \varphi(f_n) = \lim_{n\to\infty} \sum_{k=1}^n \nu(A_k) = \sum_{k=1}^\infty \nu(A_k) \,.$$

Moreover, $\mu \ll \nu$ since $\nu(A) = 0$ implies $\mathbb{1}_A = 0$ ν-a.e. hence $\mathbb{1}_A = 0$ in $L^1(\Omega, \Sigma, \nu)$, whence $\mu(A) = \varphi(0) = 0$.

By Theorem 3.3.1, there exists $g \in L^1(\Omega, \Sigma, \nu)$ with $g \geq 0$ such that

$$\varphi(\mathbb{1}_A) = \int_\Omega \mathbb{1}_A g \, d\nu \,.$$

Using linearity of φ and the integral, we see that for a nonnegative simple function $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$ we have

$$\varphi(f) = \sum_{k=1}^n a_k \varphi(\mathbb{1}_{A_k}) = \sum_{k=1}^n a_k \int_\Omega \mathbb{1}_{A_k} g \, d\nu = \int_\Omega fg \, d\nu \,.$$

Now an approximation argument yields that this equality is also true for arbitrary positve, bounded measureable functions f.

Let us now prove that $g \in L^q(\Omega, \Sigma, \nu)$. To that end, we would like to take $f := \frac{|g|^q}{g}$ with the convention that $\frac{0}{0} = 0$, as in this case $fg = |g|^q = |f|^p$. However, it is not clear, whether $f \in L^p$.

Putting $A_n := \{g \leq n\}$, we have that $\mathbb{1}_{A_n} f$ is bounded. Consequently,

$$\int_{A_n} |g|^q \, d\nu = \int_\Omega (\mathbb{1}_{A_n} f) g \, d\nu = \varphi(\mathbb{1}_{A_n} f) \leq \|\varphi\| \|\mathbb{1}_{A_n} f\|_p = \Big(\int_{A_n} |g|^q \, d\nu \Big)^{\frac{1}{p}} \,.$$

It follows that

$$\Big(\int_{A_n} |g|^q \, d\nu \Big)^{\frac{1}{q}} \leq \|\varphi\| < \infty \,.$$

Monotone convergence yields $g \in L^q$ and $\|g\|_q \leq \|\varphi\|$.

Taking this information into account, the equality

$$\varphi(f) = \int_\Omega fg \, d\nu$$

can now be extended to arbitrary positve $f \in L^p$ using the dominated convergence theorem. Splitting f into positive and negative part (and then, if $\mathbb{K} = \mathbb{C}$ into real and imaginary part), it holds for all $f \in L^p$.

□

Remark 3.3.3. Theorem 3.3.2 remains valid without the assumption that $\varphi(f) \geq 0$ for all $f \geq 0$. In the proof one either uses that arbitrary $\varphi \in (L^p)^*$ can be decomposed into positive functionals or uses directly a variant of the Radon-Nikodym theorem for so-called signed-measures (i.e. the measure may take also negative or even complex values). We do not go into details here and refer the reader to the literature for more information.

3.4 Orthonormal Bases

Definition 3.4.1. Let H be a Hilbert space. An *orthonormal system* is a subset $S \subset H$ such that (i) $\|x\| = 1$ for all $x \in S$ (i.e. every vector in S is *normalized*) and (ii) for every $x, y \in S$ with $x \neq y$ we have $(x \,|\, y) = 0$ (i.e. distinct vectors are *orthogonal*).

An *orthonormal basis* of H is an orthonormal system S such that $\overline{\operatorname{span}} S = H$.

Example 3.4.2. Let $H = \ell^2$. Then $S := \{\mathbf{e}_j : j \in \mathbb{N}\}$, where $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots)$ and 1 is at position j, is an orthonormal basis of H. Indeed, $\|\mathbf{e}_j\| = 1$ and $(\mathbf{e}_i \,|\, \mathbf{e}_j) = 0$ for $i \neq j$. Finally, let $\mathbf{x} = (x_1, x_2, \ldots) \in \ell^2$. Then $\mathbf{y}_n := (x_1, \ldots, x_n, 0, 0, \ldots) \in \operatorname{span} S$. Moreover, $\|\mathbf{x} - \mathbf{y}_n\|^2 = \sum_{k=n+1}^{\infty} |x_k|^2 \to 0$ as $n \to \infty$, proving that $\operatorname{span} S$ is dense in H.

Example 3.4.3. Let $H = L^2((0, 2\pi), \mathscr{B}((0, 2\pi)), \lambda)$ with $(f \,|\, g) = \int_0^{2\pi} f(t)\overline{g(t)}\, dt$. Then $S := \{e_k : k \in \mathbb{Z}\}$ where $e_k : [0, 2\pi] \to \mathbb{C}$ is given by $e_k(t) := (2\pi)^{-\frac{1}{2}} e^{ikt}$ is an orthonormal basis of H. Indeed

$$\|e_k\|^2 = \frac{1}{2\pi} \int_0^{2\pi} e^{ikt} e^{-ikt}\, dt = \frac{1}{2\pi} \int_0^{2\pi} 1\, dt = 1$$

and, for $k \neq l$, we have

$$(e_k \,|\, e_l) = \frac{1}{2\pi} \int_0^{2\pi} e^{i(k-l)t}\, dt = \frac{1}{2\pi} \Big[\frac{e^{i(k-l)t}}{i(k-l)} \Big|_0^{2\pi} = 0\,,$$

since the exponential function is $2\pi i$-periodic. To see that $\overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\} = L^2(0, 2\pi)$, first observe that all trigonometric polynomials belong to $\overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\}$. Hence, by Corollary 1.7.16, all continuous functions $f : [0, 2\pi] \to \mathbb{C}$ with $f(0) = f(2\pi)$ belong to $\overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\}$. But then every $f \in C[0, 2\pi]$ belongs to $\overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\}$. Indeed, for $n \in \mathbb{N}$, let $\varphi_n : [0, 2\pi] \to [0, 1]$ be a continuous function such that $\varphi_n(t) = 1$ for all $2n^1 - 1 \leq t \leq 2\pi - 2n^{-1}$ and $\varphi_n(t) = 0$ for $0 \leq t \leq n^{-1}$ and for $2\pi - n^{-1} \leq t \leq 2\pi$. Then $\varphi_n f$ is a continuous periodic function and hence belongs to $\overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\}$. Since $\varphi_n f \to f$ pointwise almost everywhere and $|\varphi_n f| \leq |f|$, the dominated convergence theorem yields that $f \in \overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\}$.

Since $C([0, 2\pi])$ is dense in $L^2(0, 2\pi)$, it follows that $\overline{\operatorname{span}}\{e_k : k \in \mathbb{Z}\} = L^2(0, 2\pi)$.

Exercise 3.4.4. Let γ be standard Gaussian measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, i.e.

$$\gamma(A) = \int_A e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}}\,.$$

We consider the Hilbert space $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \gamma)$ with the canonical inner product $(f \,|\, g) := \int_{\mathbb{R}} f \bar{g}\, d\gamma$. The *Hermite polynomials* H_n are defined as

$$H_n(t) = (-1)^n e^{\frac{t^2}{2}} \Big(\frac{d^n}{dt^n} e^{-\frac{t^2}{2}} \Big)\,.$$

Show that $\mathscr{H} := \{H_n : n \in \mathbb{N}\}$ is an orthonormal system in $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \gamma)$.

Lemma 3.4.5. *Let S be an orthonormal system and $e_1, \ldots, e_n \in S$ be distinct. If for scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ we have $\sum_{j=1}^n \lambda_j e_j = 0$, then $\lambda_1 = \ldots = \lambda_n = 0$, i.e. S is linearly independent in the meaning of linear algebra.*

Proof. If $\sum_{j=1}^n \lambda_j e_j = 0$, then for $1 \le i \le n$, we have

$$0 = (0 \,|\, e_i) = \Big(\sum_{j=1}^n \lambda_j e_j, e_i \Big) = \sum_{j=1} \lambda_j (e_j \,|\, e_i) = \lambda_i$$

since S is orthonormal. □

Theorem 3.4.6. *(Gram-Schmidt procedure)*

Let $(x_k)_{k\in\mathbb{N}}$ be a linearly independent sequence in an inner product space H. Then there exists an orthonormal system $(e_k)_{k\in\mathbb{N}}$ such that $\operatorname{span}\{x_1, \ldots, x_n\} = \operatorname{span}\{e_1, \ldots, e_n\}$ *for all $n \in \mathbb{N}$. An analogous result holds for finite independent sets.*

Proof. Since (x_k) is independent, $x_1 \neq 0$. We may thus define $e_1 := \|x_1\|^{-1} x_1$. Then $\|e_1\| = 1$ and $\operatorname{span}\{x_1\} = \operatorname{span}\{e_1\}$.

Now suppose that we have already constructed orthonormal vectors $e_1, \ldots, e_n$ with $\operatorname{span}\{x_1, \ldots, x_k\} = \operatorname{span}\{e_1, \ldots, e_k\}$ for all $1 \le k \le n$. Put $\tilde{e}_{n+1} := x_{n+1} - \sum_{j=1}^n (x_{n+1} \,|\, e_j) e_j$. Since $x_1, \ldots, x_{n+1}$ are linearly independent, $\tilde{e}_{n+1} \neq 0$. Hence, we may define $e_{n+1} := \|\tilde{e}_{n+1}\|^{-1} \tilde{e}_{n+1}$.

By construction, $\|e_{n+1}\| = 1$ and

$$\begin{aligned}(e_{n+1} \,|\, e_k) &= \|\tilde{e}_{n+1}\|^{-1} \Big((x_{n+1} \,|\, e_k) - \sum_{j=1}^n (x_{n+1} \,|\, e_j)(e_j \,|\, e_k) \Big) \\ &= \|\tilde{e}_{n+1}\|^{-1} ((x_{n+1} \,|\, e_k) - (x_{n+1} \,|\, e_k)) = 0\,.\end{aligned}$$

Moreover, $\operatorname{span}\{x_1, \ldots, x_{n+1}\} = \operatorname{span}\{e_1, \ldots, e_{n+1}\}$. Proceeding by induction, the claim follows.

□

Corollary 3.4.7. *Every separable Hilbert space has a countable orthonormal basis.*

Proof. If $\overline{\operatorname{span}} S = H$, then, by linear algebra, we may pick a linearly independent sequence from $\operatorname{span} S$ whose span is also dense in H (this sequence may be finite). We then apply the Gram-Schmidt procedure to this sequence to obtain an orthonormal system whose span is dense. □

Example 3.4.8. Consider the Hilbert space $L^2(-1, 1)$. We apply the Gram-Schmidt procedure to the independent vectors f_0, f_1, f_2, given by $f_j(t) = t^j$.

We have $\|f_0\|^2 := \int_{-1}^1 1\, dt = 2$. We hence put $e_0(t) \equiv \frac{1}{\sqrt{2}}$.

Next observe that

$$(f_1 \,|\, e_0) = \frac{1}{\sqrt{2}} \int_{-1}^1 t\, dt = \frac{1}{\sqrt{2}} \Big[\frac{1}{2} t^2 \Big|_{-1}^1 = 0.$$

Hence $\tilde{e}_1 = f_1 - 0e_0$. Since $\|\tilde{e}_1\|^2 = \int_{-1}^1 t^2\, dt = \frac{2}{3}$, we have $e_1(t) = \frac{\sqrt{3}}{\sqrt{2}} t$.

As for f_2, we have

$$(f_2 \,|\, e_0) = \int_{-1}^1 \frac{1}{\sqrt{2}} t^2\, dt = \frac{\sqrt{2}}{3}$$

and

$$(f_2 \,|\, e_1) = \int_{-1}^{1} \frac{\sqrt{3}}{\sqrt{2}} t^3 \, dt = 0 \,.$$

Thus $\tilde{e}_2 = f_2 - 0 \cdot e_1 - \frac{\sqrt{2}}{3} e_0$, hence $\tilde{e}_2 = t^2 - \frac{1}{3}$. Since $\|\tilde{e}_2\|^2 = \int_{-1}^{1} (t^2 - \frac{1}{3})^2 \, dt = \frac{2}{5}$, we obtain $e_2(t) = \sqrt{\frac{5}{2}} (t^2 - \frac{1}{3})$.

Corollary 3.4.9. *(Orthogonal Projection onto a finite dimensional space)*
Let H be a Hilbert space and $\{e_1, \dots, e_n\}$ be an orthonormal system.
Then, for $K := \operatorname{span}\{e_1, \dots, e_n\}$, the orthogonal projection P_K onto K is given by

$$P_K x = \sum_{j=1}^{n} (x \,|\, e_j) e_j \,.$$

Proof. In the proof of Theorem 3.4.6, it was seen that $x - \sum_{j=1}^{n} (x \,|\, e_j) e_j \perp e_k$ for $k = 1, \dots, n$ and thus $x - \sum_{j=1}^{n} (x \,|\, e_j) e_j \in \operatorname{span}\{e_1, \dots, e_n\}^{\perp}$. By Proposition 3.2.6, $P_K x = \sum_{j=1}^{n} (x \,|\, e_j) e_j$. □

Example 3.4.10. On $L^2(-1,1)$, the orthogonal projection onto $K := \operatorname{span}\{1, t, t^2\}$ is given by

$$P_k f := \sum_{j=0}^{2} (f \,|\, e_j) e_j$$

since e_0, e_1, e_2 is an orthonormal basis of K.

Lemma 3.4.11. *(Bessel's inequality)*
Let H be a Hilbert space and $(e_n)_{n \in \mathbb{N}}$ be an orthonormal system. Then, for every $x \in H$,

$$\sum_{n \in \mathbb{N}} |(x \,|\, e_n)|^2 \leq \|x\|^2 \,.$$

Proof. For $N \in \mathbb{N}$, let $K_N := \operatorname{span}\{e_1, \dots, e_N\}$ and P_N be the orthogonal projection onto K_N. By Corollary 3.4.9, $P_N x = \sum_{k=1}^{N} (x \,|\, e_k) e_k$. Now Pythagoras Theorem yields

$$\|x\|^2 = \|P_N x\|^2 + \|x - P_N\|^2 \geq \|P_N x\|^2 = \sum_{k=1}^{N} |(x \,|\, e_k)|^2 \,.$$

Since N was arbitrary, the claim follows. □

We now can describe also orthogonal projections onto infinite-dimensional subspaces:

Proposition 3.4.12. *Let H be a Hilbert space and K be a closed linear subspace of H. If $(e_n)_{n \in \mathbb{N}}$ is an orthonormal basis of K, then the series $\sum_{n \in \mathbb{N}} (s \,|\, e_n) e_n$ converges for every $x \in H$ and the orthogonal projection onto K is given by*

$$P_K x := \sum_{n \in \mathbb{N}} (x \,|\, e_n) e_n \,.$$

Proof. For $1 \le m \le n$, we have by Pythagoras theorem

$$\Big\| \sum_{k=1}^{n} (x \,|\, e_k)e_k - \sum_{k=1}^{m} (x \,|\, e_k)e_k \Big\|^2 = \sum_{k=m+1}^{n} |(x \,|\, e_k)|^2 .$$

As a consequence of Bessel's inequality, the latter converges to 0 as $n, m \to \infty$: Thus $\sum_{k=1}^{n} (x \,|\, e_k)e_k$ is a Cauchy sequence in H and hence convergent. This proves the first assertion.

For the second, note that $y = \sum_{k=1}^{\infty} (x \,|\, e_k)e_k \in K = \overline{\operatorname{span}}\{e_k : k \in \mathbb{N}\}$ and $x - y \perp e_j$ for all $j \in \mathbb{N}$, which follows from the fact that $x - \sum_{k=1}^{N} (x \,|\, e_k)e_k \perp e_j$ (see Corollary 3.4.9) and the continuity of the inner product. □

Theorem 3.4.13. *Let H be a separable Hilbert space, $(e_n)_{n\in\mathbb{N}}$ be an orthonormal system. The following are equivalent.*

(a) $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis of H.

(b) $x \perp e_n$ for all $n \in \mathbb{N}$ implies that $x = 0$.

(c) $x = \sum_{k=1}^{\infty} (x \,|\, e_k)e_k$.

(d) $(x \,|\, y) = \sum_{j=1}^{\infty} (x \,|\, e_k)(e_k \,|\, y)$ for all $x, y \in H$.

(e) (Parseval's identity) For all $x \in H$ we have

$$\|x\|^2 = \sum_{k=1}^{\infty} |(x \,|\, e_k)|^2 .$$

Proof. (a) ⇒ (b): If $x \perp e_k$ for all $k \in \mathbb{N}$, then $x \perp \overline{\operatorname{span}}\{e_n\}$. If $\overline{\operatorname{span}}\{e_n\} = h$, it follows that $x \perp x$ and thus $x = 0$ by (IP1).

(b) ⇒ (c): We have $x - \sum_{k\in\mathbb{N}} (x \,|\, e_k)e_k \perp e_n$ for all $n \in \mathbb{N}$. Thus, by (b) $x = \sum_{k\in\mathbb{N}} (x \,|\, e_k)e_k$.

(c) ⇒ (d): By (c),

$$\begin{aligned}(x \,|\, y) &= \Big(\sum_{n\in\mathbb{N}} (x \,|\, e_n)e_n, \sum_{m\in\mathbb{N}} (y \,|\, e_m)e_n \Big) \\ &= \sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} (x \,|\, e_n)\overline{(y \,|\, e_m)}(e_n \,|\, e_m) \\ &= \sum_{n\in\mathbb{N}} (x \,|\, e_n)(e_n \,|\, y)\end{aligned}$$

where we used the continuity of the inner product in the second step and the orthonormality of the e_n in the third.

(d) ⇒ (e): Put $x = y$.

(e) ⇒ (a): If $x \in \operatorname{span}\{e_n\}^\perp$, then $\|x\|^2 = 0$ by Parseval's identity. Thus $\operatorname{span}\{e_n\}^\perp = \{0\}$ and hence $H = \{0\}^\perp = (\operatorname{span}\{e_n : n \in \mathbb{N}\}^\perp)^\perp = \overline{\operatorname{span}}\{e_n : n \in \mathbb{N}\}$.

□

We rephrase Theorem 3.4.13 in a slightly different way.

Definition 3.4.14. Let $(H, (\cdot\,|\,\cdot)_H)$ and $(U, (\cdot\,|\,\cdot))$ be Hilbert spaces and $U \in \mathscr{L}(H, V)$. Then U is called an *isometry* if $\|Ux\|_V = \|x\|_H$. It is called *unitary* if $(Ux\,|\,Uy)_V = (x\,|\,y)_H$. It is called an *isometric isomorphism* if it is an isomorphism and surjective.

Remark 3.4.15. Clearly, an isometry is injective and every unitary operator is an isometry. It can be proved that every isometry between Hilbert spaces is unitary.

Corollary 3.4.16. *Let H be a separable Hilbert space and $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of H. Then $U : H \to \ell^2$, defined by*

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

is a (unitary) isometric isomorphism.

Proof. Linearity of U is clear. That it is isometric resp. unitary follows from Theorem 3.4.13 (e) resp. (d). It thus only remains to show that U is surjective. To that end, given $\mathbf{a} \in \ell^2$, put $x = \sum_{k=1}^\infty a_k e_k$. This series converges in H. Indeed, putting $x_N := \sum_{k=1}^N a_k e_k$, we have, for $N > M$, that $\|x_N - X_M\|_H = \sum_{k=M+1}^N |a_k|^2 \to 0$ as $M \to \infty$. Hence x_N is a Cauchy sequence hence convergent.

Clearly, $Ux = \mathbf{a}$, proving surjectivity. □

Remark 3.4.17. Corollary 3.4.16 show that by fixing an orthonormal basis of H we can identify H with ℓ^2. Hence we can translate problems (which only depend on the Hilbert space structure) in an arbitrary (infinite dimensional) separable Hilbert space H into an equivalent problem in ℓ^2.

Definition 3.4.18. Let H be an infinite dimensional, separable Hilbert space. For $x \in H$, the series

$$\sum_{k=1}^\infty (x\,|\,e_k)e_k$$

is called the *Fourier series of* x. The vector $((x\,|\,e_k))_{k\in\mathbb{N}} \in \ell^2$ is called the *Fourier coefficients of* x.

Exercise 3.4.19. Classical Fourier series arise by considering the Hilbert space $L^2(0, 2\pi)$, endowed with the inner product $(f\,|\,g) := \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\,dt$, and the orthonormal basis $e_k : t \mapsto e^{ikt}$ for $k \in \mathbb{Z}$. Then, for $u \in L^2(0, 2\pi)$ the number $\hat{u}(k) := (u\,|\,e_k)$ is called the k-th Fourier coefficient.

(a) Show that if u is continuously differentiable with $u(0) = u(2\pi)$, then $\widehat{u'}(k) = ik\hat{u}(k)$.

(b) compute the Fourier coefficients of $u : t \mapsto \frac{1}{4}(\pi - t)^2$.

(c) Show that

$$\sum_{k=1}^\infty \frac{1}{k^2} = \frac{\pi^2}{6}\,.$$

3.5 Conditional Expectation

Definition 3.5.1. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathscr{G} \subset \Sigma$ be a sub-σ-algebra. A *conditional expectation given* $\mathscr{G}$ is a bounded linear operator $T \in \mathscr{L}(L^1(\Omega, \Sigma, \mathbb{P}), L^1(\Omega; \mathscr{G}, \mathbb{P}))$ such that

$$\int_A TX \, d\mathbb{P} = \int_A X \, d\mathbb{P} \tag{3.3}$$

for all $X \in L^1(\Omega, \Sigma, \mathbb{P})$ and $A \in \mathscr{G}$.

Lemma 3.5.2. *(Uniqueness)*
Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathscr{G} \subset \Sigma$. if T and S are two conditional expectations given $\mathscr{G}$, then $TX = SX$ almost surely, for all $X \in L^1(\Omega, \Sigma, \mathbb{P})$. Thus conditional expectation, if it exists, is unique.

Proof. We have $\{TX \le SX\} \in \mathscr{G}$. Thus

$$0 \ge \int_{\{TX \le SX\}} TX - SX \, d\mathbb{P} = \int_{\{TX \le SX\}} X - X \, d\mathbb{P} = 0 \,.$$

By Corollary 2.5.10, $TX = SX$ almost surely on $\{TX \le SX\}$. Similarly, one sees that $TX = SX$ almost surely on $\{TX \ge SX\}$ and thus $TX = SX$ almost surely. ☐

Definition 3.5.3. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathscr{G} \subset \Sigma$ be a σ-algebra. If there exists a (By Lemma 3.5.2 necessarily unique) conditional expectation given $\mathscr{G}$, we denote it by $\mathbb{E}(\,\cdot\,|\mathscr{G})$. If $\mathscr{G}$ is generated by the random variables $(X_i)_{i\in I}$, we also write $\mathbb{E}(\,\cdot\,|(X_i)_{i\in I})$.

Theorem 3.5.4. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathscr{G} \subset \Sigma$ be a σ-algebra. Then there exists a unique conditional expectation given $\mathscr{G}$. Moreover, if $X \le Y$ almost surely, then $\mathbb{E}(X|\mathscr{G}) \le \mathbb{E}(Y|\mathscr{G})$ almost surely.*

Proof. By Theorem 2.7.7, $L^2(\Omega, \mathscr{G}, \mathbb{P})$ is complete, hence, by Lemma 1.5.6, it is a closed subspace of $L^2(\Omega, \Sigma, \mathbb{P})$. Let T_2 be the orthogonal projection onto $L^2(\Omega, \mathscr{G}, \mathbb{P})$ which exists by Theorem 3.2.4.

Then, for $A \in \mathscr{G}$ we have $\mathbb{1}_A \in L^2(\Omega, \mathscr{G}, \mathbb{P})$ and thus $(\mathbb{1}_A \,|\, X - T_2X) = 0$ by Proposition 3.2.6. Hence (3.3) holds for $X \in L^2(\Omega, \Sigma, \mathbb{P})$.

Let us now prove the monotonicity assertion for L^2-functions. By linearity, we may assume that $Y = 0$. Thus, let $X \le 0$ almost surely. Put $A_n = \{T_2X \ge n^{-1}\} \in \mathscr{G}$, by the $\mathscr{G}$-measurability of T_2X. Then

$$\frac{1}{n}\mathbb{P}(A_n) \le \int_{A_n} T_2X \, d\mathbb{P} = \int_{A_n} X \, d\mathbb{P} = 0$$

since $X \le 0$ almost surely. It follows that $\mathbb{P}(A_n) = 0$. Consequently, $\mathbb{P}(T_2X > 0) = \mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) = 0$ and thus $T_2X \le 0$ almost surely as claimed.

It now follows that $|T_2X| \le T_2|X|$. Indeed, $X \le |X|$ and hence, by monotonicity, $T_2X \le T_2|X|$ and, similarly, $-X \le |X|$ yields $-T_2X \le T_2|X|$. This estimate implies that for $X \in L^2(\Omega, \Sigma, \mathbb{P}) \subset L^1(\Omega, \Sigma, \mathbb{P})$ we have

$$\begin{aligned}\int_\Omega |T_2X| \, d\mathbb{P} &\le \int_\Omega T_2|X| \, d\mathbb{P} = \int_\Omega T_2(X^+ + X^-) \, d\mathbb{P} = \int_\Omega T_2X^+ \, d\mathbb{P} + \int_\Omega T_2X^- \, d\mathbb{P} \\ &= \int_\Omega X^+ \, d\mathbb{P} + \int_\Omega X^- \, d\mathbb{P} = \int_\Omega |X| \, d\mathbb{P} \,.\end{aligned}$$

Thus, $\|T_2X\|_1 \le \|X\|_1$ for all $X \in L^2(\Omega,\Sigma,\mathbb{P})$, i.e. T_2 is a bounded linear operator on $L^2(\Omega,\Sigma,\mathbb{P})$ with respect to the norm $\|\cdot\|_1$. By Proposition 1.9.10, T_2 has a unique extension to a bounded linear operator T_1 on $L^1(\Omega,\Sigma,\mathbb{P})$ of norm at most 1.

Now fix $X \in L^1(\Omega,\Sigma,\mathbb{P})$. Then there exists a sequence $X_n \in L^2(\Omega,\Sigma,\mathbb{P})$ with $|X_n| \le |X|$ and $|X_n| \to X$ pointwise (we may even pick X_n as a simple function). By dominated convergence, $X_n \to X$ in $L^1(\Omega,\Sigma,\mathbb{P})$, and thus $T_1X_n \to T_1X$. By Lemma 2.7.8, we may assume that $T_1X_n \to T_1X$, passing to a subsequence if necessary.

Then, for $A \in \mathscr{G}$, we have

$$\int_A T_1X\,d\mathbb{P} = \lim_{n\to\infty}\int_A T_1X_n\,d\mathbb{P} = \lim_{n\to\infty}\int_A X_n\,d\mathbb{P} = \int_A X\,d\mathbb{P}\,,$$

where we have used dominated convergence in the first step (note that $|T_1X_n| \le T_1|X_n| \le T_1|X| \in L^1$), (3.3) for elements of $L^2(\Omega,\Sigma,\mathbb{P})$ in the second and dominated convergence in the last equality.

If $X \ge 0$, then X_n may be chosen nonnegative and increasing and thus, by monotonicity, $0 \le T_2X_n \uparrow T_1X$, proving the monotonicity also in the general case. □

Let us note some important properties of conditional expectation.

Proposition 3.5.5. *Let $(\Omega,\Sigma,\mathbb{P})$ be a probability space, $\mathscr{G} \subset \Sigma$ be a σ-algebra and $X, Y : \Omega \to \mathbb{C}$ be integrable random variables.*

(a) If Y is bounded and $\mathscr{G}$-measurable, then $\mathbb{E}(YX|\mathscr{G}) = Y\mathbb{E}(X|\mathscr{G})$.

(b) If Y is independent of $\mathscr{G}$, then $\mathbb{E}(Y|\mathscr{G}) = \mathbb{E}(Y)$.

Proof. (a) Clearly, $Y\mathbb{E}(X|\mathscr{F})$ is $\mathscr{F}$-measurable. It remains to prove that

$$\int_A Y\mathbb{E}(X|\mathscr{G})\,d\mathbb{P} = \int_A YX\,d\mathbb{P}$$

for all $A \in \Sigma$. If $Y = \mathbb{1}_B$ for $B \in \mathscr{G}$, then this is true by the definition of conditional expectation. By linearity, it is true if Y is a simple function. Approximating a general measurable function with a sequence of simple functions, the general result follows.

(b) First, let $Y = \mathbb{1}_B$ be an indicator function. Then,

$$\int_A Y\,d\mathbb{P} = \int_{A\cap B} d\mathbb{P} = \mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B) = \int_A \mathbb{E}Y\,d\mathbb{P}$$

by independence. By linearity, this generalizes to simple functions and an approximation argument finishes the proof. □

Exercise 3.5.6. Let $(\Omega,\Sigma,\mathbb{P})$ be a probability space and A_k be a sequence of disjoint sets in Σ whose union is Ω. Show that for $\mathscr{G} = \sigma(A_k : k \in \mathbb{N})$, we have

$$\mathbb{E}(X|\mathscr{G}) = \sum_{k=1}^{\infty}\frac{1}{\mathbb{P}(A_k)}\int_{A_k} X\,d\mathbb{P}\cdot\mathbb{1}_{A_k}\,.$$

Chapter 4

Characteristic Functions

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.
John von Neumann

In the last chapter, we have seen that a function $u \in L^2(0, 2\pi)$ is uniquely determined by its Fourier coefficients

$$\hat{u}(k) = \int_0^{2\pi} u(x)e^{-ikx}\,dx \quad (k \in \mathbb{Z})\,.$$

It is also possible to define a Fourier transform from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$, by setting, for $u \in L^2(\mathbb{R})$ and $t \in \mathbb{R}$

$$\hat{u}(t) = \int_{\mathbb{R}} u(x)e^{-itx}\,dx\,.$$

It can be proved that this (modulo a constant) establishes an isometric isomorphism from $L^2(\mathbb{R})$ onto itself.

Identifying u with the measure with density $udx =: \mu$, we could equivalently have written

$$\hat{u}(t) = \int_{\mathbb{R}} e^{-itx}\,d\mu(x)\,.$$

In this chapter we extend the notion of Fourier transform from functions to general finite measures μ on $\mathbb{R}$, actually, we will always consider $\mathbb{R}^d$ instead of $\mathbb{R}$.

The price to pay for allowing measures instead of functions is that we cannot use Hilbert space methods to prove results any more.

4.1 Definition and Elementary Properties

Definition 4.1.1. The *characteristic function* (or *Fourier transform*) of a finite measure μ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ is the map $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$, given by

$$\hat{\mu}(t) := \int_{\mathbb{R}^d} e^{itx}\,d\mu(x)\,,$$

where we write shorthand tx for $\sum_{j=1}^d t_j x_j$.

If $(\Omega, \Sigma, \mathbb{P})$ is a probability space and $X : \Omega \to \mathbb{R}^d$ is a Random vector, then the characteristic function of X is the characteristic function of its distribution μ_X. We often write φ_X instead of $\hat{\mu}_X$. Note that

$$\varphi_X(t) = \mathbb{E}(e^{itX}) .$$

Example 4.1.2. (a) The characteristic function of a Dirac-measure δ_{x_0} is given by

$$\hat{\delta}_{x_0}(t) = \int_{\mathbb{R}^d} e^{itx}\, d\delta_{x_0}(x) = e^{itx_0}$$

(b) If $\mu = \sum_{n\in\mathbb{N}} a_k \delta_{x_k}$, where $0 \leq a_k$ for all $k \in \mathbb{N}$ and $\sum_{k\in\mathbb{N}} a_k < \infty$, then

$$\hat{\mu}(t) = \int_{\mathbb{R}^d} e^{itx}\, d\mu(x) = \sum_{k\in\mathbb{N}} a_k e^{itx_k} .$$

In particular, the characteristic function of the *Poisson Distribution* pois_λ with intensity $\lambda > 0$, i.e. the measure pois_λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given by $\mathsf{pois}_\lambda = e^{-\lambda}\sum_{k=0}^\infty \frac{\lambda^k}{k!}\delta_k$ is given by

$$\widehat{\mathsf{pois}_\lambda}(t) = e^{-\lambda t}\sum_{k=0}^\infty \frac{\lambda^k}{k!} e^{itk} = e^{\lambda(e^{it}-1)} .$$

We next want to compute the characteristic function of Gaussian measure. To that end, it is convenient to first establish some properties of characteristic functions.

Proposition 4.1.3. *Let μ be a finite measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Then*

(a) $\hat{\mu}$ is equicontinuous.

(b) $|\hat{\mu}(t)| \leq \mu(\Omega) = \hat{\mu}(0)$.

(c) $\hat{\mu}$ is positive definite, i.e. for all $n \in \mathbb{N}$, points $t_1, \cdot t_n \in \mathbb{R}^d$ and complex numbers $\lambda_1, \ldots, \lambda_n$, we have

$$\sum_{k,l=1}^n \lambda_k \bar{\lambda}_l \hat{\mu}(t_k - t_l) \geq 0 .$$

Proof. (a) Fix $\varepsilon > 0$. Then there exists an $n \in \mathbb{N}$ such that $\mu(B(0,n)^c) \leq \varepsilon$. Indeed, we have $\mathbb{1}_{B(0,n)^c} \to 0$ pointwise as $n \to \infty$. Since $\mathbb{1}_{B(0,n)^c} \leq \mathbb{1}_{\mathbb{R}^d} \in L^1(\mu)$, the Dominated Convergence Theorem 2.5.16 yields $\mu(B(0,n)^c) = \int_{\mathbb{R}^d} \mathbb{1}_{B(0,n)^c)}\, d\mu \to 0$.

Next note that $|e^{itx} - e^{isx}| = |1 - e^{i(t-s)x}|$. The function $f : \mathbb{R}^{2d} \to \mathbb{C}$, given by $f(r,x) = e^{irx}$ is continuous, hence, by Theorem 1.6.12, equicontinuous on the compact set $\bar{B}(0,1) \times \bar{B}(0,n)$. Thus, we find $\delta > 0$ such that $|t - s| \leq \delta$ implies $|e^{itx} - e^{isx}| \leq \varepsilon$ for all $x \in \bar{B}(0,n)$. Consequently, for $t, s \in \mathbb{R}^d$ with $|t - s| \leq \delta$, we have

$$\begin{aligned}
|\hat{\mu}(t) - \hat{\mu}(s)| &\leq \int_{\mathbb{R}^d} |e^{itx} - e^{isx}|\, d\mu(x) \\
&= \int_{\bar{B}(0,n)} |e^{itx} - e^{isx}|\, d\mu(x) + \int_{\bar{B}(0,n)^c} |e^{itx} - e^{isx}|\, d\mu(x) \\
&\leq \int_{\bar{B}(0,n)} \varepsilon\, d\mu(x) + \int_{\bar{B}(0,n)^c} 1\, d\mu(x) \\
&\leq \varepsilon\mu(\Omega) + \varepsilon .
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this proves the equicontinuity of $\hat{\mu}$.

(b) Is immediate since $|e^{itx}| \leq 1 = e^{i0x}$.

(c) By linearity,

$$\sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l \hat{\mu}(t_k - t_l) = \int_{\mathbb{R}^d} \sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l e^{i(t_k - t_l)x} \, d\mu(x) \, .$$

Noting that

$$\begin{aligned} \sum_{k,l=1}^{n} \lambda_k \bar{\lambda}_l e^{i(t_k - t_l)x} &= \sum_{k=1}^{n} \lambda e^{it_k x} \overline{\sum_{l=1}^{n} \lambda_l e^{it_l x}} \\ &= \Big| \sum_{k=1}^{n} e^{it_k x} \Big|^2 \geq 0 \end{aligned}$$

the positive definiteness follows. □

Remark 4.1.4. An important result of Bochner asserts that any function $\varphi : \mathbb{R}^d \to \mathbb{C}$ which satisfies conditions (a), (b) and (c) of Proposition 4.1.3 is the characteristic function of a finite measure μ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$.

We next study the differentiability of characteristic functions. Let us first introduce some notations

Definition 4.1.5. Let d be a natural number. A *multi-index* is an element $\alpha = (\alpha_1, \ldots, \alpha_d)$ of $\mathbb{N}_0^d$. For a multi-index α, we write $|\alpha| := \alpha_1 + \cdots + \alpha_d$ and

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} \, .$$

Finally, for $t \in \mathbb{R}^d$, we write $t^\alpha := t_1^{\alpha_1} \cdot \ldots \cdot t_d^{\alpha_d}$ and $|t|^\alpha := |t_1|^{\alpha_1} \cdot \ldots \cdot |t_d|^{\alpha_d}$.

We say that a measure μ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ has finite moments of order $k \in \mathbb{N}$, if $t \mapsto |t|^\alpha \in L^1(\mu)$ for all α with $|\alpha| \leq k$.

Proposition 4.1.6. *If μ is a finite measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with finite moments of order k, then $\hat{\mu} \in C^k(\mathbb{R}^d)$ and*

$$D^\alpha \hat{\mu}(t) = i^{|\alpha|} \int_{\mathbb{R}^d} x^\alpha e^{itx} \, d\mu(x) \, .$$

In particular,

$$D^\alpha \hat{\mu}(0) = i^{|\alpha|} \int_{\mathbb{R}^d} x^\alpha \, d\mu(x) \, .$$

Proof. Note that $t \mapsto e^{itx}$ has partial derivatives of all orders and

$$D^\alpha e^{itx} = (ix)^\alpha e^{itx} = i^{|\alpha|} x^\alpha e^{itx} \, .$$

Moreover, $|D^\alpha e^{itx}| \leq |x|^\alpha$. Thus the results follows inductively from Proposition 2.6.2. □

We can now determine the characteristic function of standard Gaussian measure.

Lemma 4.1.7. *It is $\hat{\gamma}(t) = e^{-\frac{t^2}{2}}$.*

Proof. Since γ has finite moments of all orders, it follows from Proposition 4.1.6, that $\hat{\gamma}(t)$ is differentiable and, in particular, we have

$$\hat{\mu}'(t) = i\int_{\mathbb{R}} xe^{itx}e^{-\frac{x^2}{2}}\,\frac{dx}{\sqrt{2\pi}} = i\Big[-e^{itx}e^{-\frac{x^2}{2}}\Big|_{-\infty}^{\infty} + i\int_{\mathbb{R}} ite^{itx}e^{-\frac{x^2}{2}}\,\frac{dx}{\sqrt{2\pi}} = -t\hat{\mu}(t)\,.$$

Here we have used integration by parts. Thus $\hat{\mu}$ solves the differential equation $u'(t) = -tu(t)$. Since $\hat{\mu}(t) = 1$ by Proposition 4.1.3, we may conclude that $\hat{\mu}(t) = e^{-\frac{t^2}{2}}$, since this differential equation has a unique solution as a consequence of Theorem 1.8.9. □

Let us end this section by establishing some easy yet useful properties of characteristic functions of random variables.

Proposition 4.1.8. *Let X and Y be independent random variables, defined on a probability space $(\Omega, \Sigma, \mathbb{P})$. Then*

(a) $\varphi_{a+bX}(t) = e^{iat}\varphi_X(bt)$ *for all* $t \in \mathbb{R}$.

(b) $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

Proof. (a) By definition,

$$\varphi_{a+bX}(t) = \mathbb{E}e^{it(a+bX)} = e^{ita}\mathbb{E}e^{itbX} = e^{ita}\varphi(bt)\,.$$

(b) By Theorem 2.10.5, the distribution $\mu_{(X,Y)}$ of the vector (X, Y) is the product of the distributions μ_X of X and μ_Y of Y. Hence, using Fubini's theorem,

$$\begin{aligned}\varphi_{X+Y}(t) = \mathbb{E}e^{it(X+Y)} &= \int_{\mathbb{R}^2} e^{it(x+y)}\,d\mu_{(X,Y)}(x,y) = \int_{\mathbb{R}^2} e^{it(x+y)}\,d\mu_X\otimes\mu_Y(x,y)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{itx}e^{ity}\mu_X(x)\,d\mu_Y(y) = \varphi_X(t)\varphi_Y(t)\,.\end{aligned}$$

□

The great usefulness of characteristic functions lies in Property (b) in Proposition 4.1.8. An important application is the study of sums of independent random variables, where this property allows us to determine the distribution of such a sum and to determine the asymptotic behavior of such sums.

On the side of measures, the distribution of $X + Y$, where X and Y are independent is called the *convolution* of the distribution of X with the distribution of Y. Thus the Fourier-transform maps convolutions into products.

4.2 Uniqueness of Characteristic Functions

We now justify the name "characteristic function" by proving that a measure is uniquely determined by its characteristic function.

Theorem 4.2.1. *Let μ, ν be finite measures on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with $\hat{\mu} = \hat{\nu}$. Then $\mu = \nu$.*

Proof. First note that $\mu(\Omega) = \hat{\mu}(0) = \hat{\nu}(0) = \nu(\Omega)$. Since $\mathbb{R}^d$ is the countable union of compact sets, the compact subsets of $\mathbb{R}^d$ are a generator of $\mathscr{B}(\mathbb{R}^d)$, see Proposition 2.1.9, which is obviously stable under intersections. Thus, by Theorem 2.2.10, to prove $\mu = \nu$ it suffices to prove $\mu(K) = \nu(K)$ for all compact subsets of $\mathbb{R}^d$.

Since every compact set is the pointwise limit of a bounded sequence of functions with compact support, by dominated convergence it suffices to prove that $\int_\Omega f\,d\mu = \int_\Omega f\,d\nu$ for all continuous functions with compact support.

Let such a function f be given. Then there exists $k \in \mathbb{N}$ such that $f(x) = 0$ for all $x \notin (-2\pi k, 2\pi k)^d$. Given $\varepsilon > 0$, by the Stone-Weierstrass theorem there exists a trigonometric polynomial $p : x \mapsto \sum_{|\alpha_k| \leq n} a_k e^{i\alpha_k x}$ with $\sup\{|f(x) - p(x)| \,:\, x \in [-2\pi k, 2\pi k]^d\} \leq \varepsilon$. We write $K := [-2\pi k, 2\pi k]^d$. Passing to a larger k if necessary, we may assume that $\mu(K^c) \leq \varepsilon$ and $\nu(K^c) \leq \varepsilon$.

We note that

$$\int_{\mathbb{R}^d} p\,d\mu = \sum_{|\alpha_k| \leq n} a_k \hat{\mu}(\alpha_k) = \sum_{|\alpha_k| \leq n} a_k \hat{\nu}(\alpha_k) = \int_{\mathbb{R}^d} p\,d\nu\,.$$

Moreover,

$$\begin{aligned}
\Big|\int_{\mathbb{R}^d} f\,d\mu - \int_{\mathbb{R}^d} f\,d\nu\Big| &\leq \Big|\int_{\mathbb{R}^d} f - p\,d\mu\Big| + \Big|\int_{\mathbb{R}^d} p\,d\mu - \int_{\mathbb{R}^d} p\,d\nu\Big| + \int_{\mathbb{R}^d} p - f\,d\nu\Big| \\
&\leq \int_K |f - p|\,d\mu + \int_{K^c} |p|\,d\mu + 0 + \int_K |p - f|\,d\nu + \int_{K^c} |p|\,d\nu \\
&\leq \varepsilon(\mu(\Omega) + \nu(\Omega)) + 2\varepsilon(\|f\|_\infty + \varepsilon)\,.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\int_{\mathbb{R}^d} f\,d\mu = \int_{\mathbb{R}^d} f\,d\nu$. This finishes the proof. □

The uniqueness Theorem 4.2.1 can be used to compute distributions of random variables. It is of particular importance, when determining the distribution of sums of independent random variables.

Example 4.2.2. The symmetric stable distribution with parameter α is the measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with characteristic function $e^{-|t|^\alpha}$. These distributions have the following property: If $X_1, \ldots, X_n$ are independent random variables which all have as distribution the symmetric stable distribution with parameter α, then also $Y := \frac{1}{n^\alpha}\sum_{j=1}^n X_j$ has this distribution. Indeed, by Proposition 4.1.8,

$$\varphi_Y(t) = \big(\varphi_{X_j}(\frac{t}{n^\alpha})\big)^n = e^{-n|\frac{t}{n^\alpha}|^\alpha} = e^{-|t|^\alpha}\,.$$

Exercise 4.2.3. The Exponential distribution with parameter λ is the measure exp_λ on the measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ defined by

$$\mathsf{exp}_\lambda(A) := \int_A \mathbb{1}_{[0,\infty)}(t)\lambda e^{-\lambda t}\,dt\,.$$

The Laplace distribution with parameter λ is the measure L_λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ defined by

$$\mathsf{L}_\lambda(A) := \int_A \frac{1}{2\lambda} e^{-\frac{|t|}{\lambda}}\,dt\,.$$

Compute the characteristic functions of exp_λ and L_λ and prove that if X and Y are independent and have distribution exp_λ, then $X - Y$ has distribution $\mathsf{L}_{\frac{1}{\lambda}}$.

We can now give a more detailed account of independent random variables.

Theorem 4.2.4. *Let $(X_j)_{j\in J}$ be a family of random variables defined on a probability space $(\Omega, \Sigma, \mathbb{P})$. The following are equivalent:*

(a) The family $(X_j)_{j\in J}$ is independent.

(b) For all $n \in \mathbb{N}$, indices $j_1, \ldots, j_n \in J$ and functions $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{C}$ which are bounded and measurable we have

$$\mathbb{E}\big[f_1(X_{j_1}) \cdot \ldots \cdot f_n(X_{j_n})\big] = \mathbb{E}\big[f_1(X_{j_1})\big] \cdot \ldots \cdot \mathbb{E}\big[f_n(X_{j_n})\big] .$$

(c) For all $n \in \mathbb{N}$ and $j_1, \ldots, j_n \in J$ we have

$$\varphi_{(X_{j_1},\ldots,X_{j_n})}(t_1, \ldots, t_n) = \varphi_{X_{j_1}}(t_1) \cdot \ldots \cdot \varphi_{X_{j_n}}(t_n) .$$

(d) For all $n \in \mathbb{N}$ and $j_1, \ldots, j_n \in J$ we have

$$\mu_{(X_{j_1},\ldots,X_{j_n})} = \bigotimes_{k=1}^{n} \mu_{X_{j_k}} .$$

Proof. (a) $\Rightarrow$ (b): By the definition of independence, (b) holds whenever each f_k is an indicator function. By linearity, (b) holds whenever each f_k is a simple functions and then an approximation argument shows that (b) holds in full generality.

(b) $\Rightarrow$ (c): Pick $f_k : r \mapsto e^{it_k r}$.

(c) $\Rightarrow$ (d): Follows from Theorem 4.2.1 since $\varphi_{(X_{j_1},\ldots,X_{j_n})}(t) = \int_{\mathbb{R}^d} e^{itx}\, d\mu_{(X_{j_1},\ldots,X_{j_n})}$ and $\varphi_{X_{j_1}}(t_1) \cdot \ldots \cdot \varphi_{X_{j_n}}(t_n) = \int_{\mathbb{R}} e^{itx} d \bigotimes_{k=1}^{n} \mu_{X_{j_k}}$.

(d) $\Rightarrow$ (a): Was proved in Theorem 2.10.5. □

As an application of the uniqueness result, we discuss Gaussian random vectors.

Definition 4.2.5. A random variable X is called *Gaussian*, if for some $q \geq 0$ we have

$$\varphi_X(t) = e^{-\frac{q}{2}t^2} .$$

A random vector $X = (X_1, \ldots, X_d)$ is called *Gaussian* if for all $z \in \mathbb{R}^d$ the random variable $zX = \sum_{j=1}^{d} z_j X_j$ is Gaussian, i.e. there is a $q(z) \geq 0$ such that

$$\varphi_{zX}(t) = e^{-\frac{q(z)}{2}t^2} .$$

We will express this sometimes also by saying that the random variables $X_1, \ldots, X_d$ are *jointly Gaussian.*

Remark 4.2.6. It is easy to see that if X is a Gaussian random variable, then $\mu_X = \delta_0$ if $q = 0$, i.e. $X = 0$ almost surely, whereas in the case where $q > 0$ the distribution μ_X is given by

$$\mu_X(A) = \int_A e^{-\frac{x^2}{2q}} \frac{dt}{\sqrt{2\pi q}} .$$

In particular, a Gaussian random variable has finite moments of all orders, as is easy to see. Moreover, $\mathbb{E}X = i\varphi_X'(0) = 0$ and $\mathbb{E}X^2 = -\varphi_X''(0) = q$.

It should be noted, that sometimes in the literature a different terminology is used and it is allowed that a Gaussian random variable has nonzero mean. In this case, our notion of Gaussian random variable is more appropriately called *centered Gaussian random variable.*

Lemma 4.2.7. *X is a Gaussian random vector if and only if there exists a symmetric, positive semidefinite matrix Q such that*

$$\varphi_X(t) = e^{-\frac{1}{2}(Qt\,|\,t)}\,.$$

Proof. First assume that X is a Gaussian random vector, defined on some probability space $(\Omega, \Sigma, \mathbb{P})$. Taking $y = e_j$ in the definition of Gaussian random vector, we see that X_j is a Gaussian random variable, in particular $X_j \in L^2(\Omega, \Sigma, \mathbb{P})$, for all $1 \leq j \leq d$. By Hölder's inequality, $q_{ij} := \mathbb{E}X_iX_j$ exists. Clearly $q_{ij} = q_{ji}$ hence $Q := (q_{ij})$ is a symmetric matrix. Moreover,

$$(Qt\,|\,t) = \sum_{i,j=1}^d t_it_jq_{ij} = \sum_{i,j=1}^d t_it_j\mathbb{E}X_iX_j = \mathbb{E}\Big(\sum_{j=1}^d t_jX_j\Big)^2 \geq 0$$

proving that Q is positive semidefinite. By assumption zQ is a Gaussian random variable for all $z \in \mathbb{R}^d$ and $\mathbb{E}(zX)^2 = \mathbb{E}\big(\sum_{j=1}^d z_jX_j\big)^2 = (Qz\,|\,z)$. Hence, for $t \in \mathbb{R}$,

$$\varphi_X(zt) = \varphi_{zX}(t) = e^{-\frac{1}{2}(Qz\,|\,z)t^2} = e^{-\frac{1}{2}(Qzt\,|\,zt)}\,.$$

Conversely, if $\varphi_X(t) = e^{-\frac{1}{2}(Qt\,|\,t)}$ for all $t \in \mathbb{R}$, then, given $z \in \mathbb{R}^d$, we have

$$\varphi_{zX}(r) = \mathbb{E}e^{izX\cdot r} = \mathbb{E}e^{i(X\,|\,zr)} = \varphi_X(zr) = e^{-\frac{1}{2}(Qz\,|\,z)r^2}$$

proving that X is Gaussian. □

Definition 4.2.8. The matrix Q from the previous Lemma is called the *covariance matrix* of the Gaussian vector X.

Exercise 4.2.9. Let X be a d-dimensional Gaussian vector with covariance matrix Q and $T \in \mathbb{R}^{m\times d}$. Show that TX is a m-dimensional Gaussian vector with covariance matrix TQT^t, where T^t is the transposed matrix of T.

Proposition 4.2.10. *Let $X_1, \ldots, X_d$ be jointly Gaussian vectors. The following are equivalent:*

(a) The random variables $X_1, \ldots, X_d$ are uncorrelated, i.e. $\mathbb{E}X_iX_j = 0$ whenever $i \neq j$.

(b) The random variables $X_1, \ldots, X_d$ are independent.

Proof. (a) ⇒ (b): If $X_1, \ldots X_d$ are uncorrelated, then the covariance matrix Q has diagonal form, i.e. $q_{ij} = 0$ whenever $i \neq j$. Hence

$$\varphi_{(X_1,\ldots,X_d)}(t_1, \ldots, t_d) = e^{-\frac{1}{2}(Qt\,|\,t)} = e^{-\frac{1}{2}\sum_{j=1}^d q_{jj}t_j^2} = \prod_{j=1}^d e^{-\frac{q_{jj}}{2}t_j^2} = \prod_{j=1}^d \varphi_{X_j}(t_j)\,.$$

Hence $X_1, \ldots X_d$ are independent by Theorem 4.2.4.

(b) ⇒ (a): By Theorem 4.2.4, we have $\mathbb{E}f_1(X_i)f_2(X_j) = \mathbb{E}f_1(X_i) \cdot \mathbb{E}f_2(X_j)$ for all bounded and measurable $f_1, f_2 : \mathbb{R} \to \mathbb{C}$. By an approximation argument, we see that we can allow $f_1 = f_2 : t \mapsto t$ and this shows that X_i and X_j are uncorrelated. □

Exercise 4.2.11. Show that in Proposition 4.2.10 one cannot drop the assumption that the random variables $X_1, \ldots, X_d$ are *jointly* Gaussian.

More precisely, let X be a standard Gaussian random variable and r be an independent random variable with $\mathbb{P}(r = -1) = \mathbb{P}(r = 1) = \frac{1}{2}$. Show that $Y := rX$ is Gaussian and that X and Y are uncorrelated but not independent. Conclude that X and Y are not jointly Gaussian.

Proposition 4.2.12. *Let $(X_1, \ldots, X_d)$ be a Gaussian random vector and $Y_1, \ldots, Y_k \in \operatorname{span}\{X_1, \ldots, X_d\} =: H$. Moreover, let $K := \operatorname{span}\{Y_1, \ldots, Y_k\}$. Then the conditional expectation $\mathbb{E}(\,\cdot\,|Y_1, \ldots, Y_k)$ equals on H the orthogonal projection onto K.*

Proof. Replacing $Y_1, \ldots, Y_k$ with an orthonormal basis of K, we may assume that that the $Y_1, \ldots, Y_k$ are orthonormal. Note that this does not change the conditional expectation, as $\sigma(Y_1, \ldots, Y_k) = \sigma(\operatorname{span}\{Y_1, \ldots, Y_k\})$, where $\subset$ is trivial and $\supset$ follows from Proposition 2.4.3.

If $Z \in H$, then $Z = \sum_{j=1}^d a_j X_j$ for a suitable $a \in \mathbb{R}^d$, hence Z is a Gaussian random variable. We note that since $P_K Z$ is a linear combination of $Y_1, \ldots, Y_k$, it is $\sigma(Y_1, \ldots, Y_k)$-measurable. Moreover, $Y_1, \ldots, Y_k, Z - P_K Z$ are orthogonal (i.e. uncorrelated). Since these vectors are linear combinations of the X_j's, we have $(Y_1, \ldots, Y_k, Z - P_K Z) = T(X_1, \ldots, X_d)$ for a suitable linear map T. Thus $Y_1, \ldots, Y_k, Z - P_K Z$ are jointly Gaussian by Exercise 4.2.9 and now Proposition 4.2.10 shows that these random variables are independent. □

Exercise 4.2.13. Let $(X_1, \ldots, X_4)$ be a Gaussian random vector with covariance matrix

$$Q = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 4 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Put $Z = X_1 - X_2 + 2X_4$ and $Y_1 := X_1 - X_3, Y_2 := X_2 + X_3$. Determine $\mathbb{E}(Z|Y_1, Y_2)$.

— END OF SEMESTER —

4.3 Convergence in Distribution

Definition 4.3.1. Let (M,d) be a metric space and μ_n, for $n \in \mathbb{N}$, and μ be finite measures on $(M, \mathscr{B}(M))$.

We say that μ_n *converges weakly* to μ, and write $\mu_n \rightharpoonup \mu$ if

$$\int_M f\,d\mu_n \to \int_M f\,d\mu$$

for all $f \in C_b(M)$.

If X_n, for $n \in \mathbb{N}$ and X are M-valued random elements defined on probability spaces $(\Omega_n, \Sigma_n, \mathbb{P}_n)$ and $(\Omega, \Sigma, \mathbb{P})$ respectively, we say that X_n *converges to* X *in distribution* and write $X_n \stackrel{d}{\to} X$, if $\mu_{X_n} \rightharpoonup \mu_X$.

Example 4.3.2. Let (M,d) be a metric space and $x_n \to x$. Then $\delta_{x_n} \rightharpoonup \delta_x$. Indeed, for every bounded and continuous function $f : M \to \mathbb{K}$, we have

$$\int_M f\,d\delta_{x_n} = f(x_n) \to f(x) = \int_M f\,d\delta_x\,.$$

Example 4.3.3. Consider $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Let $\mu_n = 2n\mathbb{1}_{[-n^{-1},n^{-1}]}d\lambda$. Then $\mu_n \rightharpoonup \delta_0$. Indeed, for $f \in C_b(\mathbb{R})$ we find, given $\varepsilon > 0$ a $\delta > 0$ such that $|t| \leq \delta$ implies that $|f(t) - f(0)| \leq \varepsilon$. Now pick n_0 such that $n_0^{-1} < \delta$. Then

$$\Big|\int_{\mathbb{R}} f\,d\mu_n - \int_{\mathbb{R}} f\,d\delta_0\Big| = \Big|2n\int_{-\frac{1}{n}}^{\frac{1}{n}} f(t)\,dt - f(0)\,dt\Big| \leq \int_{-\frac{1}{n}}^{\frac{1}{n}} 2n|f(t) - f(0)|\,dt \leq \varepsilon$$

for all $n \geq n_0$.

Let us now study the connection of convergence in distribution with other modes of convergence (almost surely, in L^p, in measure) of random elements. Note that in this case, the random variables *have* to be defined on the *same* probability space (since otherwise almost sure convergence, L^p-convergence and convergence in measure are not defined). We also assume here that $M = \mathbb{K}$.

Lemma 4.3.4. *Let X_n, X be $\mathbb{K}$-valued random variables defined on a probability space $(\Omega, \Sigma, \mathbb{P})$. Consider the statements*

(a) $X_n \stackrel{P}{\to} X$.

(b) $X_n \stackrel{d}{\to} X$.

Then (a) ⇒ (b). If X is almost surely constant, then (b) ⇒ (a).

Proof. (a) ⇒ (b): Suppose that $\int_{\mathbb{K}} f\,d\mu_{X_n} = \int_\Omega f(X_n)\,d\mathbb{P} \not\to \int_\Omega f(X)\,d\mathbb{P}$. Then, passing to a subsequence, we have that $|\int_\Omega f(X_n) - f(X)\,d\mathbb{P}| \geq \varepsilon > 0$ for all $n \in \mathbb{N}$ and a suitable ε. Since $X_n \stackrel{P}{\to} X$, Theorem 2.8.5 yields that, passing to a further subsequence, we may assume that $X_n \to X$ almost surely. But then, by the continuity of f, we have $f(X_n) \to f(X)$ almost surely. Since f is bounded, the dominated convergence theorem yields $\int_\Omega f(X_n)\,d\mathbb{P} \to \int_\Omega f(X)\,d\mathbb{P}$ — a contradiction.

(b) $\Rightarrow$ (a): Now we additionally assume that $X = c$ almost surely for some $c \in \mathbb{K}$. Define $f : x \mapsto |x - c| \wedge 1$. Then $f \in C_b(\mathbb{K})$ and hence, by (b),

$$\int_\Omega |X_n - c| \wedge 1 \, d\mathbb{P} = \int_\mathbb{K} f \, d\mu_{X_n} \to \int_\mathbb{K} f \, d\mu_X = 0$$

proving that $X_n \overset{P}{\to} X$. □

Example 4.3.5. To see that (b) $\Rightarrow$ (a) does not hold in general, consider on $\Omega = \{-1, 1\}$, endowed with the power set $\mathscr{P}(\Omega)$ as σ-algebra, the measure $\mathbb{P} = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Moreover, let $X_n(x) = x$ for all $n \in \mathbb{N}$ and $X(x) = -x$. Then $X_n \overset{d}{\to} X$ but X_n does not converge in probability to X.

We next establish an equivalent description of weak convergence for probability measures.

Theorem 4.3.6. *(Portmanteau theorem)*
Let (M, d) be a metric space and μ_n, for $n \in \mathbb{N}$, and μ be probability measures on $(M, \mathscr{B}(M))$. The following are equivalent:

(a) $\mu_n \rightharpoonup \mu$.

(b) $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$ *for any open set* U.

(c) $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$ *for any closed set* C.

(d) $\mu_n(B) \to \mu(B)$ *for all* $B \in \mathscr{B}(M)$ *with* $\mu(\partial B) = 0$.

Proof. (a) $\Rightarrow$ (b): Given an open set U, consider a bounded, continuous function f with $0 \leq f \leq \mathbb{1}_U$. Then $\int_M f \, d\mu_n \leq \mu_n(U)$. Upon $n \to \infty$, it follows that $\int_M f \, d\mu = \lim_{n\to\infty} \int_M f \, d\mu_n \leq \liminf_{n\to\infty} \mu_n(U)$. Now consider a sequence f_k of bounded, continuous functions with $f_k \uparrow \mathbb{1}_U$. Then, by monotone convergence,

$$\mu(U) = \int_M \mathbb{1}_U \, d\mu = \sup_k \int_M f_k \, d\mu \leq \liminf_{n\to\infty} \mu_n(U)$$

by the above. This shows (b).

(b) $\Rightarrow$ (c): Since $\mu(A) = 1 - \mu(A^c)$ for all $A \in \mathscr{B}(M)$, this is clear by taking complements.

(b) and (c) $\Rightarrow$ (d) Let $B \in \mathscr{B}(M)$ with $\mu(\partial B) = 0$. Then

$$\mu(B^\circ) \leq \liminf_{n\to\infty} \mu_n(B^\circ) \leq \limsup_{n\to\infty} \mu_n(\bar{B}) \leq \mu(\bar{B}) = \mu(\partial B) + \mu(B^\circ) = \mu(B^\circ)$$

since $\bar{B} = B^\circ \cup \partial B$. Thus $\lim_{n\to\infty} \mu_n(B^\circ) = \lim_{n\to\infty} \mu_n(\bar{B}) = \mu(B)$.

(d) $\Rightarrow$ (a) Let $f \in C_b(M)$ be nonnegative. Since $\partial\{f \geq t\} \subset \{f = t\}$ and the latter sets are disjoint, it follows that for every $k \in \mathbb{N}$ there are at most finitely many t such that $\mu(\{f = t\}) \geq k^{-1}$. Consequently, for at most countably many t we can have $\mu(\partial\{f \geq t\}) > 0$.

Thus, by (d), we have $\lim_{n\to\infty} \mu_n(\{f \geq t\}) = \mu(\{f \geq t\})$ for all but countably many t, in particular for λ-almost all t. Hence, by dominated convergence and Proposition 2.9.4,

$$\int_M f \, d\mu_n = \int_0^{\|f\|_\infty} \mu(f \geq t) \, dt = \lim_{n\to\infty} \int_0^{\|f\|_\infty} \mu_n(f \geq t) \, dt = \lim_{n\to\infty} \int_M f \, d\mu_n \, .$$

For a general $f \in C_b(M)$, we split $f = f^+ - f^-$. Then the above yields $\lim_{n\to\infty} \int_M f \, d\mu_n = \int_M f \, d\mu$. □

Example 4.3.7. Here is a typical application of Theorem 4.3.6:

Suppose that X_n is a sequence of random variables which converges in distribution to a standard Gaussian random variable. Since $\gamma(\{c\}) = 0$ for all $c \in \mathbb{R}$, we have $\gamma(\partial I) = 0$ for any interval I. Hence, by Theorem 4.3.6,

$$\mathbb{P}(X_n \in I) = \mu_{X_n}(I) \to \mu_X(I) = \int_I e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}}.$$

Thus, asymptotically (i.e. for $n \to \infty$) the probability that $X_n \in I$ is known.

4.4 The Lévy Continuity Theorem

In this section, we prove the following theorem due to Lévy:

Theorem 4.4.1. *Let μ_n, for $n \in \mathbb{N}$ and μ be finite measures on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Then we have that $\mu_n \rightharpoonup \mu$ if and only if $\hat{\mu}_n \to \hat{\mu}$ pointwise. In this case, $\hat{\mu}_n$ converges to $\hat{\mu}$ uniformly on every compact subset of $\mathbb{R}^d$.*

We prepare the proof of Theorem 4.4.1 through a series of Lemmas. The first one is used to establish the uniform convergence on compact sets.

Lemma 4.4.2. *Let μ_k be a sequence of finite measures on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ with converges weakly to the finite measure μ. Then the characteristic functions $\hat{\mu}$ are uniformly equicontinuous, i.e. given $\varepsilon > 0$, there exists a $\delta > 0$ such that $|\hat{\mu}_k(t) - \hat{\mu}_k(s)| \le \varepsilon$ for all $t, s \in \mathbb{R}^d$ with $|t - s| \le \delta$ and all $k \in \mathbb{N}$.*

Proof. The proof is a slight variation of Proposition 4.1.3. Let us first assume that all measures μ_n and μ are probability measures.

As in the proof of that proposition, we find an $n \in \mathbb{N}$ with $\mu(B(0,n)^c) \le \varepsilon/2$. By the Portmanteau theorem 4.3.6 (here we use that we have probability measures), we have $\limsup_{n\to\infty} \mu_k(B(0,n)^c) \le \varepsilon/2$ since $B(0,n)^c$ is closed. It follows that $\mu_k(B(0,n)^c) \le \varepsilon$ except for at most finitely many k. By possibly enlarging n, we may assume that we have this inequality for *all* $k \in \mathbb{N}$. Now the uniform equicontinuity follows by repeating the computation in the proof of Proposition 4.1.3.

For the general case, note that $\mu_k(\mathbb{R}^d) = \int_{\mathbb{R}^d} \mathbb{1}\, d\mu_k \to \int_{\mathbb{R}^d} \mathbb{1}\, d\mu = \mu(\mathbb{R}^d)$. From this we conclude that $\mu_k, \mu \in [\varepsilon, M]$ for certain $0 < \varepsilon < M < \infty$. Defining $\nu_k := \mu_n(\mathbb{R}^d)^{-1}\mu_n$ and $\nu := \mu(\mathbb{R}^d)^{-1}\mu$, it follows that $\nu_n \rightharpoonup \nu$ and hence the first part yield that $\hat{\nu}_k$ are uniformly equicontinuous. However, since $\hat{\mu}_k = \mu_k(\mathbb{R}^d)^{-1}\hat{\nu}_k$, it follows that also the $\hat{\mu}_k$ are uniformly equicontinuous. ☐

We next establish an estimate for the "tails" of the measure μ on $\mathbb{R}$ in terms of the characteristic function. This is the crucial ingredient for our proof of Theorem 4.4.1.

Lemma 4.4.3. *Let μ be a finite measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Then, for every $r > 0$ we have*

$$\mu(\{|x| \ge r\}) \le \frac{r}{2} \int_{-\frac{2}{r}}^{\frac{2}{r}} \mu(\mathbb{R}) - \hat{\mu}(t)\, dt.$$

Proof. Fix $c > 0$. Then, by Fubini's theorem,

$$\int_{-c}^{c} \mu(\mathbb{R}) - \hat{\mu}(t)\,dt = \int_{-c}^{c}\int_{\mathbb{R}} 1 - e^{itx}\,d\mu(x)\,dt = \int_{\mathbb{R}}\int_{-c}^{c} 1 - e^{itx}\,dt\,d\mu(x)$$
$$= \int_{\mathbb{R}} 2c - \frac{e^{icx} - e^{-icx}}{ix}\,d\mu(x) = 2c\int_{\mathbb{R}} 1 - \frac{\sin cx}{cx}\,d\mu(x)\,.$$

Now note that for $t \geq 2$, we have $\sin t \leq t/2$ and thus $1 - t^{-1}\sin t \geq \frac{1}{2}$ for such t. Since $\sin -t = -\sin t$, the same estimate holds for $t \leq -2$. Since $|t^{-1}\sin t| \leq 1$ for all $0 \neq |t| \leq 2$, we can estimate the last integral in the above computation and obtain

$$\int_{-c}^{c} 1 - \hat{\mu}(t)\,dt \geq 2c\int_{\{|cx|\geq 2\}} 1 - \frac{\sin cx}{cx}\,d\mu(x) \geq c\int_{\{|cx|\geq 2\}} 1\,d\mu(x) = c\mu(\{|cx| \geq 2\})\,.$$

Picking $c = \frac{2}{r}$, the assertion follows. □

We are now ready for

Proof of Theorem 4.4.1. First assume that $\mu_n \rightharpoonup \mu$. Since $e^{itx} \in C_b(\mathbb{R}^d)$, we have $\hat{\mu}_n(t) = \int_{\mathbb{R}^d} e^{itx}\,d\mu_n \to \int_{\mathbb{R}^d} e^{itx}\,d\mu = \hat{\mu}(t)$ for all $t \in \mathbb{R}^d$. Thus, $\hat{\mu}_n$ converges to $\hat{\mu}$ pointwise. By Lemma 4.4.2, given $\varepsilon > 0$, we find $\delta > 0$ such that $|t - s| \leq \delta$ implies $|\hat{\mu}_n(t) - \hat{\mu}_n(s)| \leq \varepsilon$ and also $|\hat{\mu}(t) - \hat{\mu}(s)| \leq \varepsilon$. For fixed t, pick $n(t)$ such that $|\hat{\mu}_n(t) - \hat{\mu}(t)| \leq \varepsilon$ for all $n \geq n(t)$. Then for $s \in B(t, \delta)$, we have

$$|\hat{\mu}_n(s) - \hat{\mu}(s)| \leq |\hat{\mu}_n(s) - \hat{\mu}_n(t)| + |\hat{\mu}_n(t) - \hat{\mu}(t)||\hat{\mu}(t) - \hat{\mu}(s)| \leq 3\varepsilon\,.$$

Now let a compact set $K \subset \mathbb{R}^d$ be given. Then K is covered by the balls $(B(t,\delta))_{t\in K}$. By compactness, K is covered by finitely many of these balls, say $B(t_1,\delta), \ldots, B(t_k,\delta)$. But then $|\hat{\mu}_n(s) - \hat{\mu}(s)| \leq 3\varepsilon$ for all $n \geq \max\{n(t_1), \ldots, n(t_k)\}$. This proves that $\hat{\mu}_n$ converges to $\hat{\mu}$ uniformly on the compact subsets of $\mathbb{R}^d$.

Let us now prove the converse and assume that $\hat{\mu}_n \to \hat{\mu}$ pointwise.

For fixed $z \in \mathbb{R}^d$ and $r > 0$, we obtain by Lemma 4.4.3 that

$$\limsup_{n\to\infty} \mu_n(\{|zx| > r\}) \leq \limsup_{n\to\infty} \frac{r}{2}\int_{-\frac{2}{r}}^{\frac{2}{r}} \hat{\mu}_n(\mathbf{0}) - \hat{\mu}_n(tz)\,dt = \frac{r}{2}\int_{-\frac{2}{r}}^{\frac{2}{r}} \hat{\mu}(\mathbf{0}) - \hat{\mu}(tz)\,dt$$

Since $\hat{\mu}$ is continuous at 0, the last integral converges to 0 as $r \to \infty$, cf. Example 4.3.3.

Choosing $a = e_j$, we find, given $\varepsilon > 0$ numbers r_j such that

$$\limsup_{n\to\infty} \mu_n(\{|x_j| > r_j\}) \leq \varepsilon/2d\,.$$

As in the proof of Lemma 4.4.2, we infer that, possibly enlarging r_j, we have $\mu_n(\{|x_j| > r_j\}) \leq \varepsilon/d$ for all $n \in \mathbb{N}$ and $1 \leq j \leq d$. We define $r := \max\{r_j : 1 \leq j \leq d\}$. Then $([-r,r]^d)^c \subset \{|x_1| > r\} \cup \cdots \cup \{|x_d| > r\}$ and hence $\mu_n(([-r,r]^d)^c) \leq \varepsilon$. Possibly enlarging r again, we assume that $r = 2\pi k$ for some $k \in \mathbb{N}$.

We now proceed similar to the proof of Theorem 4.2.1:

Fix a function $f \in C_b(\mathbb{R}^d)$. Given $\varepsilon > 0$, we find a trigonometric polynomial p such that $\sup\{|f(x) - p(x)| : x \in [-r,r]^d\} \leq \varepsilon$. We also put $M := \sup_{n\in\mathbb{N}} \mu_n(\mathbb{R}^d)$. We then find

$$\Big|\int_{\mathbb{R}^d} f\,d\mu_n - \int_{\mathbb{R}^d} f\,d\mu\Big| \leq \Big|\int_{\mathbb{R}^d} p\,d\mu_n - \int_{\mathbb{R}^d} p\,d\mu\Big| + 2M\varepsilon + 2\varepsilon(\|f\|_\infty + \varepsilon)\,.$$

Since $\int_{\mathbb{R}^d} p\, d\mu_n \to \int_{\mathbb{R}^d} p\, d\mu$ as $n \to \infty$, which follows from the pointwise convergence of the characteristic function and linearity, we infer that

$$\limsup_{n\to\infty} \Big| \int_{\mathbb{R}^d} f\, d\mu_n - \int_{\mathbb{R}^d} f\, d\mu \Big| \leq 2M\varepsilon + 2\varepsilon(\|f\|_\infty + \varepsilon)$$

and hence, since $\varepsilon > 0$ was arbitrary, that $\limsup_{n\to\infty} \big| \int_{\mathbb{R}^d} f\, d\mu_n - \int_{\mathbb{R}^d} f\, d\mu \big| = 0$. This is equivalent to $\int_{\mathbb{R}^d} f\, d\mu_n \to \int_{\mathbb{R}^d} f\, d\mu$. Since $f \in C_b(\mathbb{R}^d)$ was arbitrary, it follows that $\mu_n \rightharpoonup \mu$. □

Here is a first application:

Proposition 4.4.4. *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability space and* X_n *be a sequence of Gaussian random variables which converges in probability to a random variable* X. *Then* X *is a Gaussian random variable.*

Proof. By assumption, $\varphi_{X_n}(t) = e^{-\frac{q_n}{2}t^2}$ for certain $q_n \in [0, \infty)$. By Lemma 4.3.4, $X_n \to X$ in distribution and thus, by Theorem 4.4.1, $\varphi_{X_n}(t) \to \varphi_X(t)$. Taking logarithms, it follows that q_n converges to some $q \in [0, \infty]$. However, $q = \infty$ cannot happen, since otherwise, $\varphi_{X_n}(t) \to \mathbb{1}_{\{0\}}$ which is not a characteristic function, since it is not continuous at 0.

Thus, $q_n \to q \in [0, \infty)$. It follows that $\varphi_X(t) = e^{-\frac{q}{2}t^2}$ and hence X is a Gaussian random variable. □

We will give some more applications of Theorem 4.4.1 in the next section.

Exercise 4.4.5. The binomial distribution $\mathsf{b}_{n,p}$ with parameters p and n is given as

$$\mathsf{b}_{n,p} := \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k \,.$$

Show that if $np_n \to \lambda \in (0, \infty)$ as $n \to \infty$, then $\mathsf{b}_{n,p_n} \to \mathsf{pois}_\lambda$, where the Poisson distribution is given by

$$\mathsf{pois}_\lambda = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \,.$$

4.5 Limit Theorems

We now apply the Lévy continuity theorem to prove two central results of probability theory, namely the *(weak) law of large numbers* and the *central limit theorem.*

The weak law of large numbers is concerned with the asymptotic behavior of *averages*

$$S_n := \frac{1}{n} \sum_{k=1}^{n} X_k$$

where X_n is a sequence of independent and identically distributed random variables.

The weak law of large numbers asserts that if the common distribution of the X_k has finite expectation (i.e. $\mathbb{E}|X_1| < \infty$), then S_n converges in distribution to $\mathbb{E}X_1$.

To get a feeling for this result, consider the following experiment:

Assume we are given a biased coin which comes up head with a probability p which is unknown to us. Interpreting “head” as 1 and “tail” as 0, the experiment of flipping the coin is described by a random variable X which has distribution $\mu_X = (1-p)\delta_0 + p\delta_1$. Then $\mathbb{E}X = 0 \cdot (1-p) + 1 \cdot p = p$.

Suppose now, we are interested to figure out the value of p. A rather natural approach is then to flip the coin a "large number of times" n, count how many times the coin comes up head, say k times, and then use $\frac{k}{n}$ as an estimate for p.

Mathematically, the repeated flipping of a coin is modeled by a sequence of identically distributed, independent (this is the mathematical interpretation of "the coin does not remember what came up beforehand") random variables X_n. The average $\frac{k}{n}$ is then exactly the value of the random variable S_n. Thus, the weak law of large numbers says that if we toss "often enough", then the (empirical) average $\frac{k}{n}$ converges (in distribution) to the true probability p.

This gives a justification for many statistical methods.

The central limit theorem is concerned with slightly different averages, namely

$$G_n := \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$$

which may be though of rescaling the averages S_n to obtain variance 1 (of course this results in additional assumptions on the X_k's). The central limit theorem then gives the asymptotic distribution of the G_n. This has various applications in statistics.

In the proof of both theorem, the Theorem of Taylor plays a role, so let us recall it briefly.

Lemma 4.5.1. *(Taylor)*
If $f : (a,b) \to \mathbb{K}$ is n-times continuously differentiable and $t_0 \in (a,b)$, then

$$f(t_0 + h) = \sum_{k=1}^n \frac{f^{(k)}(t_0)}{k!} h^k + r(h)$$

where r is a function satisfying $|h|^{-k} r(h) \to 0$ as $h \to 0$.

Theorem 4.5.2. *(Law of large numbers)*
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables defined on a common probability space $(\Omega, \Sigma, \mathbb{P})$. Furthermore, assume that $\mathbb{E}X_1 =: m (\equiv \mathbb{E}X_n)$ exists.
Then $S_n := \frac{1}{n}\sum_{k=1}^n X_n \xrightarrow{P} m$.

Proof. We denote the (common) characteristic function of the X_n's by φ. By Proposition 4.1.6, φ is differentiable and thus we have, by Taylor's theorem, $\varphi(t) = \varphi(0) + t\varphi'(0) + r(t) = 1 + itm + r(t)$, where $\frac{r(t)}{|h|} \to 0$ as $t \to 0$.

Hence, by Proposition 4.1.8,

$$\varphi_{S_n}(t) = \Big[\varphi\Big(\frac{t}{n}\Big)\Big]^n = \Big(1 + \frac{imt + nr(n^{-1}t)}{n}\Big)^n.$$

Thus, since $z_n \to z$ implies that $(1 + \frac{z_n}{n})^n \to e^z$, it follows that $\varphi_{S_n}(t) \to e^{itm}$ which is the characteristic function of δ_m. It follows that S_n converges in distribution to m. Since m is almost surely constant, $S_n \xrightarrow{P} m$ by Lemma 4.3.4. □

In Theorem 4.5.2, we cannot drop the assumption that $X_n \in L^1$ and replace the mean m with the median as the following example shows:

Example 4.5.3. The *Cauchy distribution* C is given by

$$\mathsf{C}(A) = \frac{1}{\pi}\int_A \frac{1}{1+x^2}\,dx\,.$$

The characteristic function of C is given by $\hat{\mathsf{C}}(t) = e^{-|t|}$ (the proof of this requires more refined tools than we have at hand, so we skip it). Note that $\hat{\mathsf{C}}$ is not differentiable, hence a Cauchy distributed random variable does not have finite expectation (this can also be seen directly from the fact that $\int_0^\infty \frac{x}{1+x^2}\,dx = \infty$.

Now if X_n is a sequence of independent random variables with Cauchy distribution, then

$$\varphi_{S_n}(t) = (e^{-|\frac{t}{n}|})^n = e^{-|t|}$$

so that S_n is also Cauchy distributed. In particular, the distribution does not converge to any degenerate distribution.

We now turn to the central limit theorem:

Theorem 4.5.4. *Let X_n be a sequence of independent random variables defined on a common probability space $(\Omega,\Sigma,\mathbb{P})$. Moreover, assume that $X_1 \in L^2(\Omega,\Sigma,\mathbb{P})$ with $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = \mathsf{Var}(X_1) = 1$.*

Then $G_n := \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$ converges in distribution to standard Gaussian measure.

Proof. We denote the common characteristic function of the X_n's by φ and note that, by Proposition 4.1.6, φ is twice differentiable. Hence, by Taylor's theorem, $\varphi(t) = 1 - \frac{t^2}{2} + r(t)$ where $\frac{r(t)}{t^2} \to 0$ as $t \to 0$.

We have

$$\varphi_{G_n}(t) = \Big[\varphi\Big(\frac{t}{\sqrt{n}}\Big)\Big]^n = \Big(1 - \frac{t^2}{2n} + r(\frac{t}{\sqrt{n}}\Big)^n = \Big(1 + \frac{-\frac{t^2}{2} + nr\big(\frac{t}{\sqrt{n}}\big)}{n}\Big)^n$$

which converges to $e^{-\frac{t^2}{2}}$ as $n \to \infty$. But that is the characteristic function of standard Gaussian measure. □

Further Reading

The material found in these lecture notes are treated in many textbooks.

For Chapter 1, we recommend the books by Lang [6] and Jost [5]. More on measure theory may be found in the book by Bartle [1]; in Germany, also the book by Bauer [2] is used as a standard introduction to the topic. The already cited book [5] also contains a chapter on measure and integration. As for Hilbert spaces, we recommend the books by Conway [4] and Werner [8]. The book [6] contains a chapter focused on Fourier series. More information about characteristic functions can be found in basically every book on probability theory. We have benefited from the books of Varadhan [7] and Bauer [3].

Bibliography

[1] Robert G. Bartle. *The Elements of Integration and Lebesgue measure.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1995.

[2] Heinz Bauer. *Maß- und Integrationstheorie.* Walter de Gruyter & Co., Berlin, second edition, 1992.

[3] Heinz Bauer. *Wahrscheinlichkeitstheorie.* Walter de Gruyter & Co., Berlin, fifth edition, 2002.

[4] John B. Conway. *A Course in Functional Analysis*, volume 96. Springer-Verlag, New York, second edition, 1990.

[5] Jürgen Jost. *Postmodern Analysis.* Universitext. Springer-Verlag, Berlin, third edition, 2005.

[6] Serge Lang. *Undergraduate Analysis.* Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1997.

[7] S. R. S. Varadhan. *Probability Theory*, volume 7 of *Courant Lecture Notes in Mathematics.* New York University Courant Institute of Mathematical Sciences, New York, 2001.

[8] Dirk Werner. *Funktionalanalysis.* Springer-Verlag, Berlin, extended edition, 2000.