Notes on Integration Theory

April 6, 2017

Jon Johnsen

ABSTRACT. The present set of notes are written to support our students at the mathematics 6 level, in the study of Lebesgue integration and set-theoretic measure theory.

Contents

CHAPTER 1

Measure of a set

In the following we shall describe precisely what is meant by the measure of a set. Examples are many of this notion: length of an interval I on the real axis, area of a rectangle R in the Euclidean plane or space; or the number of elements in the set; or even the probability of an event represented by the set.

In general a measure μ on a set X is a mapping going from a system, $\mathbb{E}$, of subsets of X to the positive extended real numbers,

$$\mu\colon \mathbb{E} \to [0,\infty]. \tag{1.0.1}$$

The first requirement here is that the domain of μ, which is the system $\mathbb{E}$ of subsets, has to form a σ-algebra:

DEFINITION 1.0.1. A family $\mathbb{E}$ of subsets of X is said to be a σ-algebra in X if

(i) $X \in \mathbb{E}$;
(ii) $\complement E \in \mathbb{E}$ whenever $E \in \mathbb{E}$;
(iii) $\bigcup_{n\in\mathbb{N}} E_n \in \mathbb{E}$ whenever E_1, E_2,... are in $\mathbb{E}$.

In view of the first two points above, when $\mathbb{E}$ is a σ-algebra, then $\emptyset = \complement X$ is a member of $\mathbb{E}$ too. This enters the formal definition of a measure, as does the third point above:

DEFINITION 1.0.2. A mapping $\mu\colon \mathbb{E} \to [0,\infty]$, defined on a σ-algebra $\mathbb{E}$ in X, is said to be a measure on X if

(i) $\mu(\emptyset) = 0$;
(ii) $\mu(\bigcup_{n\in\mathbb{N}} E_n) = \sum_{n\in\mathbb{N}} \mu(E_n)$ whenever the sequence of sets $E_n \in \mathbb{E}$ are pairwise disjoint.

Further facts on these fundamental notions are developed in the next sections.

1.1. Measurable sets

When a σ-algebra $\mathbb{E}$ in X is given, it is customary to designate the sets $E \in \mathbb{E}$ as the ($\mathbb{E}$-)mesurable sets, as it were if a measure was defined on $\mathbb{E}$. Moreover, a pair $(X,\mathbb{E})$ consisting of a set X and a σ-algebra $\mathbb{E}$ in X is often referred to as a measurable space.

Among the basic facts on σ-algebras one has:

$$\left.\begin{aligned} A\cup B &\in \mathbb{E}\\ A\cap B &\in \mathbb{E}\\ A\setminus B &\in \mathbb{E}\end{aligned}\right\} \quad \text{whenever } A,B \in \mathbb{E}; \tag{1.1.1}$$

$$\bigcap_{n\in\mathbb{N}} A_n \in \mathbb{E} \quad \text{whenever } A_1, A_2, \dots \in \mathbb{E}. \tag{1.1.2}$$

These claims are seen at once, since

$$A \cup B = A \cup B \cup \emptyset \cup \dots \cup \emptyset \cup \dots; \tag{1.1.3}$$

$$A \cap B = \complement(\complement A \cup \complement B); \tag{1.1.4}$$

$$A \setminus B = A \cup \complement B; \tag{1.1.5}$$

$$\bigcap_{n\in\mathbb{N}} A_n \in \mathbb{E} = \complement(\bigcup_{n\in\mathbb{N}} \complement A_n). \tag{1.1.6}$$

The power set $\mathbb{P}(X)$, consisting of all subsets of X, is of course always a σ-algebra in X; and evidently the largest possible one (in the ordering given by inclusion). The system $\{X, \emptyset\}$ is clearly the smallest σ-algebra in X.

It is immediately seen that an intersection of given σ-algebras $\mathbb{E}_i$ in X,

$$\bigcap_{i\in I} \mathbb{E}_i = \{ A \subset X \mid \forall i\colon A \in \mathbb{E}_i \}, \tag{1.1.7}$$

constitutes another σ-algebra in X. This leads to the fact that each system $\mathbb{D}$ of subsets of X is contained in smallest a σ-algebra, namely the one called $\sigma(\mathbb{D})$ in

LEMMA 1.1.1. *To each system $\mathbb{D}$ of subsets of X there exists a smallest σ-algebra $\sigma(\mathbb{D})$ in X that contains $\mathbb{D}$. That is,*

- *$\sigma(\mathbb{D})$ is a σ-algebra in X satisfying $\mathbb{D} \subset \sigma(\mathbb{D})$;*
- *$\sigma(\mathbb{D}) \subset \mathbb{F}$ for every σ-algebra $\mathbb{F}$ in X satisfying $\mathbb{D} \subset \mathbb{F}$.*

PROOF. Clearly $\mathbb{P}(X)$ is a σ-algebra containing $\mathbb{D}$, so the intersection of all the σ-algebras $\mathbb{F}$ such that $\mathbb{D} \subset \mathbb{F}$ gives a non-empty collection $\mathbb{E}$ of subsets, which contains $\mathbb{D}$ and is a σ-algebra by the remark given prior to the lemma, cf. (1.1.7). □

One calls $\sigma(\mathbb{D})$ the σ-algebra *generated* by $\mathbb{D}$. And when $\mathbb{E} = \sigma(\mathbb{D})$, then $\mathbb{D}$ is said to be a generating system for the σ-algebra $\mathbb{E}$. For a system $\mathbb{D} = \{D_1, D_2, \dots, D_n\}$ of n subsets of X it can be shown inductively that $\sigma(\mathbb{D})$ contains at most 2^{2^n} sets.

Note that the above is a pure existence proof. In general there is no explicit criterion for given set $A \subset X$ to belong to $\sigma(\mathbb{D})$, which is one of the inconveniences in integration theory.

1.2. Borel algebras

For a metric space (X, d), the system $\mathbb{G}$ of open sets generates a σ-algebra $\sigma(\mathbb{G})$, which is the so-called Borel algebra of X, that is,

$$\mathbb{B}(X) = \sigma(\mathbb{G}). \tag{1.2.1}$$

It is seen at once that $\mathbb{B}(X) = \sigma(\mathbb{F})$, when $\mathbb{F}$ denotes the system of closed sets in X, for the inclusions $\mathbb{F} \subset \sigma(\mathbb{G})$ and $\mathbb{G} \subset \sigma(\mathbb{F})$ are obvious; whence $\sigma(\mathbb{G}) = \sigma(\mathbb{F})$.

This applies especially to the Euclidean spaces $\mathbb{R}^d$ of dimension $d \geq 1$, where we write $\mathbb{B}_d = \mathbb{B}(\mathbb{R}^d)$, and $\mathbb{B} = \mathbb{B}_1$ for simplicity. In this case, $\mathbb{G}_d$ and $\mathbb{F}_d$ denote the systems of open and closed sets, respectively.

By denoting the collection of compact sets in $\mathbb{R}^d$ by $\mathbb{K}_d$, every $F \in \mathbb{F}_d$ is a countable union of compact sets, namely $\bigcup_N (F \cap \bar{B}(0,N))$, so it follows that $\mathbb{K}_d$ also generates the Borel sets in $\mathbb{R}^d$,

$$\mathbb{B}_d = \sigma(\mathbb{K}_d). \tag{1.2.2}$$

However, it is important to obtain further convenient generating systems for $\mathbb{B}_d$. One choice could be the following type of d-dimensional rectangels induced by real numbers $a_i < b_i$ for $i = 1, \dots, d$, which is referred to here as *standard intervals*:

$$I =]a_1, b_1] \times \dots \times]a_d, b_d] = \{ x = (x_1, \dots, x_d) \in \mathbb{R}^d \bigm| \forall i\colon a_i < x_i \leq b_i \}. \tag{1.2.3}$$

The system of such standard intervals I is denoted by $\mathbb{I}_d$.

One obvious interest of the standard intervals $\mathbb{I}_d$ is the classical notion of the d-dimensional volume $v_d(I)$ associated to each $I \in \mathbb{I}_d$,

$$v_d(I) = (b_1 - a_1)\dots(b_d - a_d). \tag{1.2.4}$$

We shall later see that this definition induces a unique measure m_d on the Borel algebra $\mathbb{B}_d$ such that $m_d(I) = v_d(I)$ for all $I \in \mathbb{I}_d$. Here m_d is the Lebesgue measure on $\mathbb{R}^d$.

As not all sets are standard intervals, we may rethorically pose the following didactic question:

> Why does the unit ball $B(0,1) = \{\, x \in \mathbb{R}^d \mid x_1^2 + \dots + x_d^2 < 1 \,\}$ have a volume?–or rather: why is the ball $\mathbb{B}_d$-measurable: why does the Lebesgue measure $m_d(B(0,1))$ exist?

A key ingredient in the understanding of this question, and of understanding the extension of v_d on $\mathbb{I}_d$ to the measure m_d on $\mathbb{B}_d$, is the fact that also the standard intervals generate the Borel algebra,

$$\mathbb{B}_d = \sigma(\mathbb{I}_d). \tag{1.2.5}$$

Indeed, every $I \in \mathbb{I}_d$ is a countable intersection of open sets, since

$$I =]a_1, b_1] \times \dots \times]a_d, b_d] = \bigcap_{n\in\mathbb{N}}]a_1, b_1 + \frac{1}{n}[\times \dots \times]a_d, b_d + \frac{1}{n}[. \tag{1.2.6}$$

Being a σ-algebra, $\sigma(\mathbb{G}) = \mathbb{B}_d$ is stable under such intersections, so the above shows that $I \in \mathbb{B}_d$. Since I is arbitrary, $\sigma(\mathbb{I}_d) \subset \mathbb{B}_d$. As for the converse inclusion, it is seen analogously that it suffices to show that $\sigma(\mathbb{I}_d)$ contains any given open set in $\mathbb{R}^d$:

LEMMA 1.2.1. *Every open set $G \neq \emptyset$ in $\mathbb{R}^d$ is a countable union of disjoint cubes in $\mathbb{I}_d$.*

PROOF. We consider the cube $C_{k,p} \in \mathbb{I}_d$ consisting of the $x \in \mathbb{R}^d$ for which $k_i 2^{-p} < x_i \le k_i 2^{p+1}$ for $i = 1, \dots, d$. First we let O_1 be the union of all the cubes $C_{k,1}$ that are contained in G; inductively we let O_p denote the union of the cubes $C_{k,p}$ that are contained in $G \setminus (O_1 \cup \dots \cup O_{p-1})$. This gives a countable union $\bigcup_{p\in\mathbb{N}} O_p \subset G$, where equality moreover holds because every x in G is an inner point. □

Summing up we have,

$$\mathbb{B}_d = \sigma(\mathbb{G}_d) = \sigma(\mathbb{F}_d) = \sigma(\mathbb{K}_d) = \sigma(\mathbb{I}_d). \tag{1.2.7}$$

For example, a countable set $\{\, x_n \in \mathbb{R}^d \mid n \in \mathbb{N} \,\}$ (a sequence) is a Borel set, since it is a countable union of the singletons $\{x_n\}$, that are closed.

For $d = 1$, further generating systems for $\mathbb{B}$ can be introduced in terms of half-lines. For example, it is an exercise to derive that

$$\mathbb{B} = \sigma(\{\,]a, \infty[\mid a \in \mathbb{R} \,\}). \tag{1.2.8}$$

On the extended real line $\bar{\mathbb{R}}$ there is a metric given by $d(x,y) = |\arctan x - \arctan y|$, using the convention $\arctan(\pm\infty) = \pm\frac{\pi}{2}$. When restricted to $\mathbb{R}$, this induces the usual topology (i.e. system of open sets) on the real line. The associated Borel algebra $\mathbb{B}(\bar{\mathbb{R}}) = \bar{\mathbb{B}}$ is also generated by a family of half-lines,

$$\bar{\mathbb{B}} = \sigma(\{\,]a, \infty] \mid a \in \mathbb{R} \,\}). \tag{1.2.9}$$

This is related to the usual Borel algebra $\mathbb{B}$ by the fact that $A \in \bar{\mathbb{B}}$ if and only if $A \cap \mathbb{R} \in \mathbb{B}$.

1.3. Measures

A measure space is a triple $(X, \mathbb{E}, \mu)$ consisting of a set X and a fixed σ-algebra $\mathbb{E}$ in X together with a measure μ defined on X, having $\mathbb{E}$ as its domain:

$$\mu\colon \mathbb{E} \to [0, \infty]. \tag{1.3.1}$$

Cf. Definition 1.0.2 for this.

Given a measure μ on X, the number $\mu(E)$ is referred to as the measure of E for any measurable set $E \subset X$, i.e. for $E \in \mathbb{E}$. Intuitively it may be useful to think of μ as a kind mass distribution in in X. When $\mu(X) < \infty$, then μ is termed *finite*; in case $\mu(X) = 1$, the measure μ is called a *probability measure* or a (probability) *distribution*.

According to Definition 1.0.2 a measure has to be *denumerably* additive. On the one hand, this property is decisive for the strong limit theorems for the Lebegue integral, we shall meet later. On the other hand, it easily implies the (more naive property of) finite additivity, which is the first of the following basic facts (I)–(VI) on measures:

(I) $\mu\Big(\bigcup_{j=1}^n E_j\Big) = \sum_{j=1}^n \mu(E_j)$ for pairwise disjoint sets $E_1, \dots, E_n \in \mathbb{E}$.
(II) $\mu(E) \le \mu(F)$ whenever $E \subset F$ for $E, F \in \mathbb{E}$.
(III) $\mu(F \setminus E) = \mu(F) - \mu(E)$ whenever $E \subset F$ and $\mu(E) < \infty$ for $E, F \in \mathbb{E}$.
(IV) $\mu\Big(\bigcup_{j=1}^\infty E_j\Big) \le \sum_{j=1}^\infty \mu(E_j)$ for arbitrary $E_1, E_2 \dots$ in $\mathbb{E}$.
$\mu\Big(\bigcup_{j=1}^n E_j\Big) \le \sum_{j=1}^n \mu(E_j)$ for arbitrary $E_1, E_2 \dots, E_n$ in $\mathbb{E}$.
(V) $\mu(E_n) \nearrow \mu\big(\bigcup_{n=1}^\infty E_n\big)$ whenever $E_1 \subset E_2 \subset \dots \subset E_n \subset \dots$ for $E_n \in \mathbb{E}$.
(VI) $\mu(E_n) \searrow \mu\big(\bigcap_{n=1}^\infty E_n\big)$ whenever $\mu(E_1) < \infty$ and $E_1 \supset E_2 \supset \dots \supset E_n \supset \dots$ for $E_n \in \mathbb{E}$.

In fact, (I) can be seen from $\mu(E_1 \cup \dots \cup E_n \cup \emptyset \cup \emptyset \dots) = \mu(E_1) + \dots + \mu(E_n) + 0 + 0 + \dots$. Both (II) and (III) follow from the consequence of (I) that $\mu(F) = \mu(F \setminus E) + \mu(E)$.

Moreover, (III) is based on the trick that the $\mathbb{E}$-measurable sets

$$F_1 = E_1, \qquad F_j = E_j \setminus (\bigcup_{k<j} E_k) \quad \text{for } j \ge 2, \tag{1.3.2}$$

are pairwise disjoint. Clearly $\bigcup_{j\in\mathbb{N}} F_j = \bigcup_{j\in\mathbb{N}} E_j$, as to every $x \in \bigcup_{j\in\mathbb{N}} E_j$ there is a minimal index k such that $x \in E_k$, and hence $x \in F_k$. Consequently (II) gives that $\mu\big(\bigcup_{j\in\mathbb{N}} E_j\big) = \mu\big(\bigcup_{j\in\mathbb{N}} F_j\big) = \sum_{j\in\mathbb{N}} \mu(F_j) \le \sum_{j\in\mathbb{N}} \mu(E_j)$. In case $E_j = \emptyset$ holds eventually, the second part of (IV) follows readily.

Property (V) reduces to convergence of an infinite series via the disjoint sets F_j in (IV), which yield $\mu(E_n) = \sum_{j=1}^n \mu(F_j) \nearrow \sum_{j=1}^\infty \mu(F_j) = \mu\big(\bigcup_{n=1}^\infty E_n\big)$.

Finally, in (VI), setting $D_n = E_1 \setminus E_n$ gives $D_1 \subset D_2 \subset \dots$ and $\bigcup_n D_n = E_1 \setminus \bigcap_n E_n$ so that (III) and (V) entail

$$\mu(E_1) - \mu(E_n) = \mu(D_n) \nearrow \mu(\bigcup_n D_n) = \mu(E_1) - \mu(\bigcap_n E_n).$$

Using continuity of multiplication by -1 and of addition of $\mu(E_1)$, one arrives at (VI).

Though the theory of measures is rich, we shall at this point just proceed to give some uncomplicated examples.

EXAMPLE 1.3.1 (Lebesgue measure). On the real axis there is, as we shall see later, a unique measure $m\colon \mathbb{B} \to [0, \infty]$, the Lebesgue measure, which is defined on the collection $\mathbb{B}$ of all Borel sets $B \subset \mathbb{R}$ and has the property that $m(\,]a, b]) = b - a$ whenever $a < b$.

The classical Riemann integral $\int_a^b f(x)\,dx$ of a continuous function $f\colon [a, b] \to \mathbb{R}$ is equal to the Lebegue integral $\int_{[a,b]} f\,dm$—but the interest of this lies in the strong results, say on limits of integrals, which are available for the Lebesgue integral.

The Lebesgue measure m also yields an example of the necessity of the assumption in property (III) that $\mu(E) < \infty$: for $F =]0,\infty[$ and $E =]1,\infty[$ one has $m(F \setminus E) = 1$, which cannot be found as $M(F) - M(E)$ [not even if $\infty - \infty$ were ascribed the value 0].

Likewise it is necessary in (VI) above that $\mu(E_1) < \infty$: if $E_n =]n,\infty[$ for every $n \in \mathbb{N}$, then $m(\cap_n E_n) = m(\emptyset) = 0$, but this is clearly not the limit of $m(E_n) = \infty$ for $n \to \infty$.

EXAMPLE 1.3.2 (Counting measure). The function μ defined on the power set $\mathbb{P}(X)$ of an arbitrary set X (possibly uncountable) by the rule

$$\mu(E) = \begin{cases} \text{number of elements in } E, & \text{for finite subsets } E \subset X, \\ \infty, & \text{for infinite subsets } E \subset X, \end{cases} \tag{1.3.3}$$

is a measure on X, known as the *counting measure*.

EXAMPLE 1.3.3 (Measure concentrated in a subset). Every measurable subset $A \in \mathbb{E}$ in a measure space $(X, \mathbb{E}, \mu)$ induces a another measure on X

$$E \mapsto \mu(E \cap A), \quad E \in \mathbb{E}, \tag{1.3.4}$$

which is *concentrated* in A in the sense that it is zero on every measurable subset disjoint from A.

EXAMPLE 1.3.4 (The convex cone of measures). On a measurable space $(X, \mathbb{E})$ the measures form a cone, since the product of a measure and a positive number yields another measure; and the cone is convex since the set of measures is stable under addition.

Indeed, for any (finite or infinite) family of measures $(\mu_j)_{j\in J}$, and given numbers $a_j \in \bar{\mathbb{R}}_+$ for $j \in J$, also the map

$$\mu(E) = \sum_{j\in J} a_j \mu_j(E), \quad E \in \mathbb{E}, \tag{1.3.5}$$

is a measure on $\mathbb{E}$. In fact, for any sequence $E_1, E_2, \dots$ of disjoint sets in $\mathbb{E}$,

$$\mu\big(\bigcup_n E_n\big) = \sum_j (a_j \sum_n \mu_j(E_n)) = \sum_{(j,n)} a_j \mu_j(E_n) = \sum_n \sum_j a_j \mu_j(E_n) = \sum_n \mu(E_n). \tag{1.3.6}$$

EXAMPLE 1.3.5 (Dirac measure). In an arbitrary set X there is to each element $a \in X$ a measure ε_a defined on $\mathbb{P}(X)$ by

$$\varepsilon_a(E) = \begin{cases} 1, \text{ for } E \ni a, \\ 0, \text{ for } E \not\ni a. \end{cases} \tag{1.3.7}$$

This probability measure is the Dirac measure at a, also known as the *point measure* at a.

CHAPTER 2

Measurable maps

In this chapter we shall study the measurability of a map $f\colon X \to Y$. Basically this is a property that ascertains that f is compatible with given σ-algebras in X and Y.

2.1. Measurable preimages

In the following we consider measurable spaces $(X,\mathbb{E})$, $(Y,\mathbb{F})$ and $(Z,\mathbb{G})$ together with two mappings

$$X \xrightarrow{f} Y \xrightarrow{g} Z. \tag{2.1.1}$$

Measurability of such maps are defined in terms of preimages, in analogy with continuity:

DEFINITION 2.1.1. The map $f\colon X \to Y$ is said to be measurable, or more precisely $\mathbb{E}$-$\mathbb{F}$-measurable, if its preimages of $\mathbb{F}$-measurable sets are $\mathbb{E}$-measurable, that is,

$$\forall F \in \mathbb{F}\colon\ f^{-1}(F) \in \mathbb{E}. \tag{2.1.2}$$

$\mathbb{F}$-$\mathbb{G}$-measurability of g is defined analogously.

Since $(g \circ f)^{-1}(G) = f^{-1}(g^{-1}(G))$, it is clear that $(g \circ f)^{-1}(G) \in \mathbb{E}$ for every $G \in \mathbb{G}$ whenever f and g are measurable. This proves

PROPOSITION 2.1.2. *When f and g as above are measurable, then the composite map $g \circ f$ is $\mathbb{E}$-$\mathbb{G}$-measurable.*

Since the sets in the σ-algebra $\mathbb{F}$ can be difficult to describe, Definition 2.1.1 is in general somewhat impractical as it stands. However, as a cornerstone it suffices to test the condition on the preimage for the members of a generating system:

PROPOSITION 2.1.3. *Let $(X,\mathbb{E})$ and $(Y,\mathbb{F})$ be measurable spaces and $f\colon X \to Y$ a given map. When $\mathbb{E} = \sigma(\mathbb{D})$, then f is $\mathbb{E}$-$\mathbb{F}$-measurable if and only if*

$$f^{-1}(D) \in \mathbb{E} \quad \textit{for all sets } D \in \mathbb{D}. \tag{2.1.3}$$

PROOF. The necessity of the condition is trivial. To prove its sufficiency we consider the auxiliary system

$$\mathbb{H} = \left\{ F \subset Y \;\middle|\; f^{-1}(F) \in \mathbb{E} \right\}. \tag{2.1.4}$$

The aim is to prove the inclusion $\mathbb{F} \subset \mathbb{H}$. By the assumption on f it holds true that $\mathbb{D} \subset \mathbb{H}$. Moreover, $\mathbb{H}$ is itself a σ-algebra, for $Y \in \mathbb{H}$ is trivial and $\complement F \in \mathbb{H}$ holds for all $F \in \mathbb{H}$ since $f^{-1}(\complement F) = X \setminus f^{-1}(F) \in E$; whilst $f^{-1}(\bigcup_n F_n) = \bigcup_n f^{-1}(F_n)$ shows that $\mathbb{H}$ is stable under union of countably many disjoint sets in $\mathbb{H}$ (notice that the $f^{-1}(F_n)$ are disjoint members of $\mathbb{E}$). Hence $\mathbb{F} = \sigma(\mathbb{D}) \subset \mathbb{H}$, as desired. $\square$

The attentive reader will have noticed that the above proof contains an important technique: given the task of proving a statement for all sets in a given σ-algebra, it suffices to prove that the statement is true for the sets in *some* σ-algebra $\mathbb{H}$, provided the latter contains a generating system for the former.

In case X and Y are metric spaces, a map $F\colon X \to Y$ is simply said to be Borel measurable, if it is $\mathbb{B}(X)$-$\mathbb{B}(Y)$-measurable. For $Y = \mathbb{R}^d$ such a map is referred to as a Borel function.

Using Proposition 2.1.3 with $\mathbb{D}$ as the system $\mathbb{G}_Y$ of open sets in Y, it is seen at once that continuity implies Borel measurability:

PROPOSITION 2.1.4. *When (X,d_X) and $(Y;d_Y)$ are metric spaces, then every continuous map $f\colon X\to Y$ is Borel measurable.*

Thus there exists an abundance of Borel functions $f\colon X\to\mathbb{R}$ on every metric space (X,d_X), as any pair of closed sets in X can be separated by a continuous function (X is a normal space).

As another application of Proposition 2.1.3, it is seen from (1.2.8) that a criterion for Borel measurability is that $f^{-1}(]a,\infty[)\in\mathbb{E}$ for every $a\in\mathbb{R}$. For functions $f\colon X\to\bar{\mathbb{R}}$ one may use (1.2.9) instead to reduce Borel mesurability to a test of whether $f^{-1}(]a,\infty])\in\mathbb{E}$. This may be formulated in an elegant way as

PROPOSITION 2.1.5. *For a measurable space $(X,\mathbb{E})$ a function $f\colon X\to\mathbb{R}$ is Borel measurable if and only if*

$$\forall a\in\mathbb{R}\colon \big\{\,x\in X \;\big|\; f(x)>a\,\big\}\in\mathbb{E}. \tag{2.1.5}$$

The same criterion applies to functions $f\colon X\to\bar{\mathbb{R}}$.

The above is useful also for functions of the form $f\colon X\to\mathbb{R}^d$, for here the Borel measurability of $f(x)=(f_1(x),\dots,f_d(x))$ holds precisely when all the f_j are Borel functions:

PROPOSITION 2.1.6. *On a measurable space $(X,\mathbb{E})$ a function $f\colon X\to\mathbb{R}^d$ is Borel measurable if and only if the coordinate function f_j is measurable for $j=1,\dots,d$.*

PROOF. According to Proposition 2.1.5 the coordinate functions f_j are all measurable if and only if for every $(a_1,\dots,a_d)\in\mathbb{R}^d$ the σ-algebra $\mathbb{E}$ contains the sets

$$\big\{\,x\in X \;\big|\; f_j(x)>a_j\,\big\}=f^{-1}(\{\,y\in\mathbb{R}^d\mid y_j>a_j\,\}),\quad j=1,\dots,d. \tag{2.1.6}$$

But this property is by Proposition 2.1.3 equivalent to the measurability of f itself, if it can be shown that the system $\mathbb{D}$ of sets of the form $\{\,y\in\mathbb{R}^d\mid y_j>a_j\,\}$ constitute a generating system for $\mathbb{B}_d$.

However, it is clear that $\sigma(\mathbb{D})\subset\mathbb{B}_d$, for each set in $\mathbb{D}$ is open. Conversely every standard interval $]a_1,b_1]\times\dots\times]a_d,b_d]$ is a member of $\sigma(\mathbb{D})$, for it is an intersection of the d sets

$$\{\,y\in\mathbb{R}^d\mid a_j<y_j\le b_j\,\}=\{\,y\in\mathbb{R}^d\mid a_j<y_j\,\}\setminus\{\,y\in\mathbb{R}^d\mid b_j<y_j\,\}\in\sigma(\mathbb{D}). \tag{2.1.7}$$

Hence $\mathbb{B}_d=\sigma(\mathbb{I}_d)\subset\sigma(\mathbb{D})$. Altogether $\mathbb{D}$ is shown to generate $\mathbb{B}_d$, as desired. □

As a special case of this one has for $d=2$, as $\mathbb{C}$ identifies with the metric space $\mathbb{R}^2$:

PROPOSITION 2.1.7. *A complex function $f\colon X\to\mathbb{C}$ on a mesurable space $(X,\mathbb{E})$ is measurable if and only if* $\operatorname{Re} f$ *and* $\operatorname{Im} f$ *both are measurable maps $X\to\mathbb{R}$.*

EXAMPLE 2.1.8. The Dirichlét function $1_{\mathbb{Q}}\colon\mathbb{R}\to\mathbb{R}$ is discontinuous at every point in $\mathbb{R}$, but nonetheless it is a Borel function. Indeed, for every $a\in\mathbb{R}$ one has

$$\big\{\,x\in\mathbb{R} \;\big|\; 1_{\mathbb{Q}}(x)>a\,\big\}=\begin{cases}\emptyset & \text{for } a\ge 1,\\ \mathbb{Q} & \text{for } 0\le a<1,\\ \mathbb{R} & \text{for } a<0,\end{cases} \tag{2.1.8}$$

and the sets $\emptyset$, $\mathbb{Q}$, $\mathbb{R}$ are all Borel sets; cf. Proposition 2.1.5.

2.2. Limits of measurable functions

In the following $\mathbb{E}$ denotes a σ-algebra in a set $X \neq \emptyset$.

PROPOSITION 2.2.1. *Whenever* $f_1, f_2, \dots$ *is a sequence of* $\mathbb{E}$*-measurable functions* $X \to \bar{\mathbb{R}}$*, then also* $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ *and* $\liminf_n f_n$ *are* $\mathbb{E}$*-measurable.*

PROOF. To show the $\mathbb{E}$-$\bar{\mathbb{B}}$-measurability of $\sup_n f_n(x) = \sup\{ f_n(x) \mid n \in \mathbb{N} \}$ it suffices by Proposition 2.1.5 to consider an arbitrary $a \in \mathbb{R}$ and note that

$$\{ x \in X \mid \sup_n f_n(x) > a \} = \bigcup_{n=1}^{\infty} \{ x \in X \mid f_n(x) > a \} \in \mathbb{E}. \tag{2.2.1}$$

Similarly it is seen for every $a \in \mathbb{R}$ that $A = \{ x \in X \mid \inf_n f_n(x) < a \} \in \mathbb{E}$, whence the inequality $\inf_n f_n(x) \geq a$ holds in $\complement A \in \mathbb{E}$; which by passing to a union of such sets yields that also $\{ x \in X \mid \inf_n f_n(x) > a \}$ is in $\mathbb{E}$. Therefore $\inf_n f_n$ is measurable.

Using the above successively on the functions

$$\limsup_n f_n = \inf_{p \geq 1} (\sup_{n \geq p} f_n), \qquad \liminf_n f_n = \sup_{p \geq 1} (\inf_{n \geq p} f_n), \tag{2.2.2}$$

the measurability also follows for $\limsup_n f_n$ and $\liminf_n f_n$. $\square$

It is well known that the class of continuous functions on $\mathbb{R}$ is too small to be stable under passage to pointwise limits. E.g. the functions

$$f_n(x) = \max(0, \min(\frac{x}{n}, 1)) \tag{2.2.3}$$

converge pointwise to the discontinuous $f = 1_{]0,\infty[}$. Moreover, the differentiable functions $g_n(x) = \sqrt{\frac{1}{n} + x^2}$ converge pointwise to the non-smooth function $|x|$.

However, the class of Borel functions *is* large enough to be stable under pointwise convergence. The is first shown for extended real functions.

THEOREM 2.2.2. *When a sequence* $f_1, f_2, \dots$ *of* $\mathbb{E}$*-measurable functions* $f_n \colon X \to \bar{\mathbb{R}}$ *is pointwise convergent in* $\bar{\mathbb{R}}$*, then also the limit function* $f = \lim_n f_n$ *is* $\mathbb{E}$*-measurable.*

PROOF. According to the assumption, $(f_n(x))$ converges in $\bar{\mathbb{R}}$ for every $x \in X$, so

$$f(x) = \liminf_n f_n(x) = \limsup_n f_n(x) \qquad \text{for all } x \in X. \tag{2.2.4}$$

Hence f inherits the measurability from, say $\liminf_n f_n$; cf. Proposition 2.2.1. $\square$

For real and complex functions the corresponding result is also valid:

THEOREM 2.2.3. *When a sequence* $f_1, f_2, \dots$ *of* $\mathbb{E}$*-measurable functions* $f_n \colon X \to \mathbb{C}$ *is pointwise convergent in* $\in \mathbb{C}$*, then also the limit function* $f = \lim_n f_n$ *is* $\mathbb{E}$*-measurable.*

PROOF. Clearly $f(x) = \lim_n f_n(x)$ has its real and imaginary parts given by the functions $\lim_n \operatorname{Re} f_n(x)$ and $\lim_n \operatorname{Im} f_n(x)$. These are $\mathbb{E}$-$\bar{\mathbb{B}}$-measurable by the above, and also $\mathbb{E}$-$\mathbb{B}$-measurable in view of Proposition 2.1.5. Hence f is $\mathbb{E}$-measurable. $\square$

This theorem is noteworthy inasmuch as it is not every day (!) one encounters a class of functions, which is stable under pointwise convergence. But it is also a most useful result, since measurability is the basic requirement for a function f to be integrable.

2.3. Rules of calculus

For brevity it is customary to form new functions $f \wedge g$ and $f \vee g$ from given ones $f, g \colon X \to \mathbb{R}$ by setting

$$f \wedge g(x) = \min(f(x), g(x)), \qquad f \vee g(x) = \max(f(x), g(x)). \tag{2.3.1}$$

For these and the more usual constructions based on f, g one has:

PROPOSITION 2.3.1. *When $f,g\colon X\to\mathbb{R}$ are $\mathbb{E}$-measurable and $c\in\mathbb{R}$, then also the functions*

$$|f|,\ cf,\ f+g,\ f\wedge g,\ f\vee g,\ fg \tag{2.3.2}$$

are $\mathbb{E}$-measurable.

PROOF. The vector function $\varphi=(f,g)$ is $\mathbb{E}$-measurable as a map $X\to\mathbb{R}^2$ according to Proposition 2.1.6. Therefore the claim follows by composing this with the following maps, which are continuous $\mathbb{R}^2\to\mathbb{R}$ and hence Borel measurable,

$$(y_1,y_2)\mapsto y_1+y_2 \quad \text{or, respectively, } y_1\wedge y_2,\ y_1\vee y_2 \text{ and } y_1y_2. \tag{2.3.3}$$

Note that cf is covered via the case $g\equiv c$, whence $|f|=f\vee(-f)$ gives the rest. □

The case of a rational function $f(x)/g(x)$ requires a special consideration, which makes it better placed in the complex context:

PROPOSITION 2.3.2. *For functions $f,g\colon X\to\mathbb{C}$ and $c\in\mathbb{C}$ the $\mathbb{E}$-measurability carries over to the functions*

$$|f|,\ \operatorname{Re}f,\ \operatorname{Im}f,\ \bar{f},\ cf,\ f+g,\ fg. \tag{2.3.4}$$

If in addition $g(x)\neq 0$ for all $x\in X$, the same is true for $\frac{f(x)}{g(x)}$.

PROOF. The function $|f|$ is a composite with the continuous map $z\mapsto|z|$, $z\in\mathbb{C}$. Both $\operatorname{Re}f,\operatorname{Im}f$ are by definition $\mathbb{E}$-measurable as f is so. Then Proposition 2.3.1 implies that $\bar{f}=\operatorname{Re}f-\mathrm{i}\operatorname{Im}f$ is measurable. Similarly for $fg=(\operatorname{Re}f\operatorname{Re}g-\operatorname{Im}f\operatorname{Im}g)+\mathrm{i}(\operatorname{Re}f\operatorname{Im}g+\operatorname{Im}f\operatorname{Re}g)$. The sum $f+g$ is a little easier. The rational function

$$\frac{f(x)}{g(x)}=f(x)\bar{g}(x)\frac{1}{|g(x)|^2} \tag{2.3.5}$$

is treated in an exercise. □

EXAMPLE 2.3.3. Given two $\mathbb{E}$-measurable functions $f,g\colon X\to\mathbb{R}$, it is always the case that the σ-algebra $\mathbb{E}$ contains the sets

$$\big\{\,x\in X \bigm| f(x)<g(x)\,\big\},\qquad \big\{\,x\in X \bigm| f(x)\le g(x)\,\big\},\qquad \big\{\,x\in X \bigm| f(x)=g(x)\,\big\}. \tag{2.3.6}$$

Indeed, by setting $\varphi=g-f$ the sets are equal to the preimages $\varphi^{-1}(\,]0,\infty[\,)$, $\varphi^{-1}([0,\infty[\,)$ and $\varphi^{-1}(\{0\})$, respectively; and these belong to $\mathbb{E}$ since φ is $\mathbb{E}$-measurable by Proposition 2.3.1.

2.4. Subspaces

Each non-empty subset A of a measurable space $(X,\mathbb{E})$ inherits a σ-algebra, which is denoted by $\mathbb{E}_A$ and given by

$$\mathbb{E}_A=\big\{\,A\cap E \bigm| E\in\mathbb{E}\,\big\}. \tag{2.4.1}$$

Indeed, $A=A\cap X\in\mathbb{E}_A$ and the formula $A\setminus(A\cap E)=A\cap(X\setminus E)$ shows that $\mathbb{E}_A$ is stable under passage to complements; finally $\bigcup_{n=1}^\infty(A\cap E_n)=A\cap(\bigcup_{n=1}^\infty E_n)$ belongs to $\mathbb{E}_A$ when the E_n are in $\mathbb{E}$.

The inherited σ-algebra $\mathbb{E}_A$ is also called the *induced* σ-algebra. The measurable space $(A,\mathbb{E}_A)$ is the *subspace* determined by A and $\mathbb{E}$.

In case $A\subset X$ is a measurable subset, i.e. $A\in\mathbb{E}$, then $A\cap E$ is in $\mathbb{E}$ for every $E\in\mathbb{E}$, so $\mathbb{E}_A\subset\mathbb{E}$. The converse is clear, so

$$\mathbb{E}_A\subset\mathbb{E} \iff A\in\mathbb{E}. \tag{2.4.2}$$

In the affirmative case $\mathbb{E}_A=\{\,E\in\mathbb{E}\mid E\subset A\,\}$, as every such E fulfills $E=A\cap E$.

The inclusion map $i=i_{A,X}\colon A\to X$, given by $i(x)=x$, is always $\mathbb{E}_A$-$\mathbb{E}$-measurable, since $i^{-1}(E)=A\cap E$ for every $E\in\mathbb{E}$. Moreover, any σ-algebra in A that makes i measurable must contain the intersections $A\cap E$, $E\in\mathbb{E}$. This proves

LEMMA 2.4.1. *On every subset $A \neq \emptyset$, the induced σ-algebra $\mathbb{E}_A$ is the* smallest *σ-algebra in A, which makes the inclusion map i measurable.*

In the situation $\varphi\colon X \to Y$ is a map between measurable spaces $(X,\mathbb{E})$ and $(Y,\mathbb{F})$, one may consider its restriction $\varphi|_A$ to a non-empty subset $A \subset X$, yielding the commutative diagram:

$$\begin{array}{ccc} A & \xrightarrow{\varphi|_A} & Y \\ {\scriptstyle i}\big\downarrow & & \big\downarrow{\scriptstyle i} \\ X & \xrightarrow{\varphi} & Y \end{array} \tag{2.4.3}$$

Since the restriction fulfills $\varphi|_A = \varphi \circ i$, it is always $\mathbb{E}_A$-$\mathbb{E}$-measurable.

Dual to this situation, one can always endow a map with a larger codomain, and this does not affect the measurability either, provided the smaller codomain has the σ-algebra, which is induced by the larger. In fact, when $\varphi(X) \subset B$ for some (necessarily non-empty) subset $B \subset Y$, there is a map $\tilde{\varphi}\colon X \to B$ acting like φ and a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\tilde{\varphi}} & B \\ {\scriptstyle i}\big\downarrow & & \big\downarrow{\scriptstyle i} \\ X & \xrightarrow{\varphi} & Y \end{array} \tag{2.4.4}$$

Here $\tilde{\varphi}$ is $\mathbb{E}$-$\mathbb{F}_B$-measurable if and only if φ is $\mathbb{E}$-$\mathbb{F}$-measurable, as $\tilde{\varphi}^{-1}(B \cap F) = \varphi^{-1}(F)$.

Building on these considerations, it is also possible to show that φ is measurable whenever it is pieced together from measurable pieces:

PROPOSITION 2.4.2. *Let $\varphi\colon X \to Y$ be given as*

$$\varphi(x) = \begin{cases} \varphi_1(x) & \text{for } x \in A_1, \\ \varphi_2(x) & \text{for } x \in A_2, \\ \dots & \\ \varphi_n(x) & \text{for } x \in A_n, \end{cases} \tag{2.4.5}$$

whereby $X = A_1 \cup A_2 \cup \dots \cup A_n$ is a partition of X into disjoint non-empty sets $A_i \in \mathbb{E}$ and each φ_i is a map $A_i \to Y$. If φ_i is $\mathbb{E}_{A_i}$-$\mathbb{F}$-measurable for each $i \in \{1,2,\dots,n\}$, then φ is $\mathbb{E}$-$\mathbb{F}$-measurable.

PROOF. For every set $F \in \mathbb{F}$ we have

$$\varphi^{-1}(F) = (\bigcup_{i=1}^{n} A_i) \cap \varphi^{-1}(F) = \bigcup_{i=1}^{n} (A_i \cap \varphi^{-1}(F)) = \bigcup_{i=1}^{n} \varphi_i^{-1}(F). \tag{2.4.6}$$

Here the set on the right-hand side is in $\mathbb{E}$, because $\varphi_i^{-1}(F) \in \mathbb{E}_{A_i} \subset \mathbb{E}$. □

EXAMPLE 2.4.3. There is a Borel function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0 & \text{for } x < 0, \\ \cos x & \text{for } 0 \le x < 2\pi, \\ \log x & \text{for } x \ge 2\pi. \end{cases} \tag{2.4.7}$$

Indeed, $\cos\colon \mathbb{R} \to \mathbb{R}$ is continuous hence Borel, so by (2.4.3) its restriction to $A_2 = [0,2\pi[$ is $\mathbb{B}_{A_2}$-measurable (cf. Proposition 2.4.2). Via the trick of noting the continuity $\mathbb{R} \to \mathbb{R}$ of $0 \vee \log$, it is similarly seen that $\log$ is $\mathbb{B}_{A_3}$-measurable for $A_3 = [2\pi, \infty[$.

EXAMPLE 2.4.4. For two $\mathbb{E}$-measurable functions $f,g\colon X \to [0,\infty]$, it is also always the case that the σ-algebra $\mathbb{E}$ contains the sets

$$\{\, x \in X \mid f(x) < g(x) \,\}, \qquad \{\, x \in X \mid f(x) \le g(x) \,\}, \qquad \{\, x \in X \mid f(x) = g(x) \,\}. \tag{2.4.8}$$

However, as the difference $g-f$ may be undefined in this case, another argument than that in Example 2.3.3 is required.

Using the measurability of f and g it is straightforward to verify from (1.2.9) that $\mathbb{E}$ contains the sets

$$\begin{aligned} F &= \{\, x \in X \mid f(x) < \infty \,\} = f^{-1}([0,\infty[), \\ G &= \{\, x \in X \mid g(x) < \infty \,\} = g^{-1}([0,\infty[), \\ F_\infty &= \{\, x \in X \mid f(x) = \infty \,\} = f^{-1}(\{\infty\}), \\ G_\infty &= \{\, x \in X \mid g(x) = \infty \,\} = g^{-1}(\{\infty\}). \end{aligned} \tag{2.4.9}$$

So to verify that e.g. $A = \{\, x \in X \mid f(x) < g(x) \,\}$ belongs to $\mathbb{E}$ one may note that

$$A = (F \cap G_\infty) \cup \{\, x \in F \cap G \mid f(x) < g(x) \,\} \tag{2.4.10}$$

and that the last of these sets belongs to $\mathbb{E}_{F\cap G}$ according to (2.4.3) and Example 2.3.3; since $\mathbb{E}_{F\cap G} \subset \mathbb{E}$ this implies that $A \in \mathbb{E}$. The two other sets are treated analogously.

CHAPTER 3

Lebesgue integral

CHAPTER 4

Uniqueness theorem for measures

CHAPTER 5

Products of sets and measures

In this chapter we shall develop the fact that, under some liberal conditions, one can interchange the order of integration because both iterated integrals identify with the integral over the product set $X \times Y$,

$$\int_X \big(\int_Y f(x,y)\,d\nu(y) \big)\,d\mu(x) = \int_{X\times Y} f\,d\mu\otimes\nu = \int_Y \big(\int_X f(x,y)\,d\mu(x) \big)\,d\nu(y). \quad (5.0.11)$$

This is a cornerstone of the whole theory, an area where the benefit of the Lebesgue integral is very apparent. In fact, the technical difficulties are moved away from designing partitions of X and Y and into the construction of the central subject: the product measure $\mu\otimes\nu$.

5.1. Products of measure spaces

TBA

5.2. Theorems of Tonelli and Fubini

Using the construction of the product measure $\mu\otimes\nu$ on the product set $X\times Y$ of two σ-finite measure spaces, one can now derive the following result on the interchange of the order of integration for functions $f(x,y)$ in $\mathscr{M}^+$:

THEOREM 5.2.1 (Tonelli). *Let $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ be two σ-finite measure spaces. For every function $f\colon X\times Y\to[0,\infty]$ in $\mathscr{M}^+(X\times Y,\mathbb{E}\otimes\mathbb{F})$ one has:*

(i) *the function $x\mapsto \int_Y f(x,\cdot)\,d\nu$ is in $\mathscr{M}^+(X,\mathbb{E})$;*
(ii)
$$\int_X \Big(\int_Y f(x,y)\,d\nu(y) \Big)\,d\mu(x) = \int_{X\times Y} f\,d\mu\otimes\nu. \quad (5.2.1)$$

The analogous results are valid for the function $y\mapsto \int_X f(\cdot,y)\,d\mu$.

PROOF. TBA □

It is remarkable that the theorem holds under a natural set of assumptions, as no other requirement has been made than $f(x,y)$ should be positive and measurable. From the contents of the theorem it is seen that both iterated integrals in formula (5.0.11) make sense and are equal to $\int_{X\times Y} f\,d\mu\otimes\nu$.

Let again $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ be two σ-finite measure spaces, and suppose there is given a function

$$f\colon X\times Y\to\mathbb{C}. \quad (5.2.2)$$

When f is $\mathbb{E}\otimes\mathbb{F}$-measurable, then the induced function $f(x,\cdot)$ is $\mathbb{F}$-measurable for every $x\in X$. This was shown previously as a property of the product σ-algebra.

For the purposes of the Fubini theorem we derive the following result, which states that by integrating one variable out, one obtains a measurable function on the set where this integration is well defined:

LEMMA 5.2.2. *Let $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ be two σ-finite measure spaces and suppose $f\colon X\times Y\to\mathbb{C}$ is $\mathbb{E}\otimes\mathbb{F}$-measurable. Then there is a measurable set $A\subset X$ given by*

$$A=\big\{\,x\in X\;\big|\; f(x,\cdot)\in\mathscr{L}(Y,\mathbb{F},\nu)\,\big\},\tag{5.2.3}$$

and if $A\neq\emptyset$ the function $g\colon A\to\mathbb{C}$ given by

$$g(x)=\int_Y f(x,\cdot)\,d\nu,\qquad x\in A,\tag{5.2.4}$$

is $\mathbb{E}$-measurable (as $\mathbb{E}_A\subset\mathbb{E}$).

PROOF. The statement is a straightforward consequence of the case that f has real values, which therefore is assumed. According to Tonelli's theorem, applied to $f^+,f^-\in\mathscr{M}^+$, there are two functions p,n belonging to $\mathscr{M}^+(X,\mathbb{E})$ defined by the expressions

$$p(x)=\int_Y f^+(x,\cdot)\,d\nu,\qquad n(x)=\int_Y f^-(x,\cdot)\,d\nu.\tag{5.2.5}$$

Moreover, since $f^\pm(x,\cdot)=f(x,\cdot)^\pm$, we get from the definition of Lebesgue integrability of $f(x,\cdot)$ that

$$A=\big\{\,x\in X\;\big|\;p(x)<\infty\,\big\}\bigcap\big\{\,x\in X\;\big|\;n(x)<\infty\,\big\}.\tag{5.2.6}$$

To see that $A\in\mathbb{E}$ one can apply Example 2.4.4.

When $A\neq\emptyset$, then there is a function $g=p|_A-n|_A$, which is E_A-measurable by the rules of calculus for measurable functions. □

The content of this lemma is of some independent interest. But it is maily because of its proof that it is useful below.

Indeed, by adding an assumption of integrability of $f(x,y)$ one now arrives at the famous Fubini's Theorem:

THEOREM 5.2.3 (Fubini). *Let $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ be two σ-finite measure spaces with $\mu(X)>0$, $\nu(Y)>0$. For every function $f\colon X\times Y\to\mathbb{C}$ in $\mathscr{L}(X\times Y,\mathbb{E}\otimes\mathbb{F},\mu\otimes\nu)$ one has:*

(i) *the set $A=\big\{\,x\in X\;\big|\;f(x,\cdot)\in\mathscr{L}(Y,\mathbb{F},\nu)\,\big\}$ belongs to $\mathbb{E}$ and $\mu(X\setminus A)=0$;*
(ii) *the function $x\mapsto\int_Y f(x,\cdot)\,d\nu$ is μ-integrable on A;*
(iii)

$$\int_{X\times Y} f\,d\mu\otimes\nu=\int_A\Big(\int_Y f(x,y)\,d\nu(y)\Big)\,d\mu(x).\tag{5.2.7}$$

The analogous results are valid for the function $y\mapsto\int_X f(\cdot,y)\,d\mu$.

PROOF. Obviously the real-valued case will imply the complex valued statement without difficulties. Continuing from the proof of the lemma, we note that Tonelli's theorem in addition to (5.2.5) gives that

$$\int_X p\,d\mu=\int_{X\times Y} f^+\,d\mu\otimes\nu,\qquad \int_X n\,d\mu=\int_{X\times Y} f^-\,d\mu\otimes\nu\tag{5.2.8}$$

These integrals are both finite, since $f\in\mathscr{L}(\mu\otimes\nu)$. Hence $0=\mu(\{x\in X\mid p(x)=\infty\}$ and $0=\mu(\{x\in X\mid n(x)=\infty\}$, as p, $n\in\mathscr{M}^+$; cf. the proof of Lemma 5.2.2. The union of these sets yield a measurable null-set of non-integrability of $f(x,\cdot)$. That is, A is in $\mathbb{E}$ and $\mu(X\setminus A)=0$, as claimed.

From the assumption $\mu(X)>0$ it follows that $A\neq\emptyset$. Therefore $p|_A$ and $n|_A$ are well-defined functions in $\mathscr{L}(A,\mu)$, in view of the above finiteness; and so is $g=p|_A-n|_A$, cf. (ii). For this one finds

$$\int_A g\,d\mu=\int_X 1_A g\,d\mu=\int_X p\,d\mu-\int_X n\,d\mu=\int_{X\times Y} f\,d\mu\otimes\nu,\tag{5.2.9}$$

by using (5.2.8). Inserting the expression for g one arrives at (iii). □

It may seem disappointing that the integral over $X \times Y$ in (iii) only was identified with the iterated integral $\int_A \Big(\dots \Big) \, d\mu$. *Post festum*, however, one may change the outer integral over A to one over X, simply by integrating $1_A(x) \int_Y f(x,\cdot)\, d\nu$. Here the value 0 on the complement $X \setminus A$ is immaterial, because this set is a null set according to (i). With this understanding it is usually sufficient to abbreviate the result in Fubini's theorem to the simpler formula:

$$\int_X \big(\int_Y f(x,\cdot)\, d\nu \big)\, d\mu = \int_{X \times Y} f \, d\mu \otimes \nu = \int_Y \big(\int_X f(\cdot, y)\, d\mu \big)\, d\nu. \tag{5.2.10}$$

However, in special circumstances one may need the full statement in Theorem 5.2.3.

CHAPTER 6

The Lebesgue spaces L_p

Foreløbig dansk tekst:

Som supplement til bogens kapitel 7 om funktionsrum kommer her en oversigt fra en anden synsvinkel.

Først og fremmest ønsker vi at måle *graden* af integrabilitet. Dette gøres ved at indføre klassen $\mathscr{L}_p(X,\mathbb{E},\mu)$ af målelige funktioner $f\colon X\to\mathbb{C}$, som for et givet $p\in[1,\infty[$ opfylder

$$\int |f|^p\,d\mu<\infty. \tag{6.0.11}$$

Motivationen er ret ligetil, hvis μ er et sandsynlighedsmål: Da er $f\in\mathscr{L}_1(\mu)$ hvis og kun den stokastiske variable f har middelværdi, mens $g\in\mathscr{L}_2(\mu)$ gælder netop når den stokastiske variable g har varians. I matematisk analyse spiller (den nedenfor beskrevne variant) $L_2(\mu)$ en afgørende rolle som et grundlæggende Hilbertrum.

Det er ret ligetil at se, at klassen $\mathscr{L}_p(\mu)$ er et vektorrum. Jvf. sætning 7.4. Nu kunne man ønske sig at vise, at vektorrummet $\mathscr{L}_p(\mu)$ endda har en norm givet ved udtrykket

$$\|f\|_p=(\int |f|^p\,d\mu)^{\frac{1}{p}}. \tag{6.0.12}$$

En brugbar konsekvens af dette ville så være, at rummet ville blive et metrisk rum med metrikken $d(f,g)=\|f-g\|_p$. Dernæst kunne man såf.eks. undersøge om rummet er et fuldstændigt metrisk rum.

Som første del af normegenskaben ses at $\|cf\|_p=(\int |c|^p|f|^p\,d\mu)^{1/p}=|c|\|f\|_p$ for enhver skalar $c\in\mathbb{C}$. Næste del kunne være trekantsuligheden, som indebærer at der for alle $f,g\in\mathscr{L}_p(\mu)$ gælder

$$(\int |f+g|^p\,d\mu)^{\frac{1}{p}}\le(\int |f|^p\,d\mu)^{\frac{1}{p}}+(\int |g|^p\,d\mu)^{\frac{1}{p}}. \tag{6.0.13}$$

Dette er kendt som Minkowskis ulighed, men for $p>1$ er den lidt krævende at vise, så det er en større sætning i integrationsteorien. Jvf. sætning 7.8.

Imidlertid er $\|\cdot\|_p$ generelt kun en *seminorm* på $\mathscr{L}_p$. Der gælder nemlig

$$\|f\|_p=0 \iff \int |f|^p\,d\mu=0 \iff |f|^p=0\ \mu\text{-n.o.} \iff f=0\ \mu\text{-n.o.} \tag{6.0.14}$$

Da nulvektoren i $\mathscr{L}_p$ er funktionen $f\equiv 0$, så er dette altså generelt utilstrækkeligt til at sikre at $\|\cdot\|_p$ er en norm. (Der er dog tale om en norm, hvis den tomme mængde er den eneste nulmængde; som f.eks. er tilfældet for tællemålet.) Derved bliver $d(f,g)$ kun en såkaldt pseudometrik på $\mathscr{L}_p$.

Den bredt accepterede udvej i denne situation er at opgive den strenge skelnen mellem funktioner, der kun er forskellige på en μ-nulmængde. Den opblødning har vi under alle omstændigheder, i og med at sådanne funktioner vil have samme integral.

Mere præcist indebærer dette, at vi kalder f, $g\colon X\to\mathbb{C}$ ækvivalente, og skriver $f\sim g$, dersom $f=g$ μ-n.o. Det er oplagt at $\sim$ er en ækvivalensrelation. Vi fører dernæst $\mathscr{L}_p(\mu)$ over i mængden af ækvivalensklasser, kaldet $L_p(\mu)$:

$$[f]=\{g\mid g\sim f\},\qquad L_p(X,\mathbb{E},\mu)=\{[f]\mid f\in\mathscr{L}_p(X,\mathbb{E},\mu)\}. \tag{6.0.15}$$

Algebraisk set bliver også $L_p(\mu)$ et vektorrum med kompositionerne

$$[f]+[g]=[f+g], \qquad c[f]=[cf]. \tag{6.0.16}$$

Klasserne på højresiderne ses nemlig let at være uafhængige af valget af repræsentanter på venstre side. (Prøv efter!) Alle 8 aksiomer for et vektorrum ses såumiddelbart at være opfyldt, idet $[0]$ hhv. $[-f]$ virker som nulvektor hhv. modsat vektor.

Hvad såmed normen ? Det simplest mulige ville være blot at anvende $\mathscr{L}_p$-seminormen påen repræsentant:

$$\big\|[f]\big\|_p=\Big(\int |f|^p\,d\mu\Big)^{\frac{1}{p}}=\|f\|_p. \tag{6.0.17}$$

Dette giver faktisk en afbildning $L_p(\mu)\to[0,\infty[$, for hvis $g\sim f$ for $f,g\in\mathscr{L}_p(\mu)$, dvs. $[f]=[g]$, ja såer $|g|^p=|f|^p$ μ-n.o., hvilket pga. bemærkning 4.16 giver $\int|g|^p\,d\mu=\int|f|^p\,d\mu$, og derfor at værdien i (6.0.17) ikke afhænger af valget af repræsentant for akvivalensklassen $[f]$.

I det praktiske arbejde noteres $[f]$ blot som f, idet man lader ækvivalensklassen med f som repræsentant være underforstået. (Med mindre man for præcisionens skyld vil understrege, at man betragter en ækvivalensklasse af funktioner der er ens μ-n.o.) F.eks. skrives i stedet for $\|[f]\|_p$ blot $\|f\|_p$, og 0 i stedet for $[0]$.

Afbildningen $\|\cdot\|_p\colon L_p(\mu)\to[0,\infty[$ er faktisk en norm, idet den opfylder

$$\|cf\|_p=|c|\|f\|_p \tag{6.0.18}$$

$$\|f+g\|_p\le\|f\|_p+\|g\|_p \tag{6.0.19}$$

$$\|f\|_p=0 \iff f=0 \tag{6.0.20}$$

Thi trekantsuligheden i (6.0.19) er en direkte konsekvens af Minkowskis ulighed i (6.0.13), hvor vi kan læse venstre- og højresiderne som funktionsværdierne af $\|\cdot\|_p$ i $[f+g]$, $[f]$ og $[g]$; jvf. (6.0.17). Tilsvarende aflæses (6.0.18) af observationen foran (6.0.13). Endelig vises (6.0.20) af biimplikationerne i (6.0.14).

Vektorrummet $L_p(\mu)$ har derfor en metrik givet ved $d(f,g)=\|f-g\|_p$ for $1\le p<\infty$. Som et meget tilfredstillende resultat er disse metriske rum altid fuldstændige. Dette er kendt som Fischers fuldstændighedssætning, jvf. sætning 7.18, som er en hjørnesten i integrationsteorien og dens anvendelser.

Fuldstændige normerede vektorrum betegnes i litteraturen som Banachrum efter Stefan Banach, som lavede en omfattende analyse af slige rum i slutningen af 1920'erne. Hovedeksemplet påBanachrum er $L_p(X,\mathbb{E},\mu)$ med $1\le p<\infty$ for et vilkårligt målrum $(X,\mathbb{E},\mu)$.

Dog er tilfældet $p=2$ specielt, fordi normen på$L_2(\mu)$ er såvenlig at udspringe af det indre produkt, som for vilkårlige $f,g\in L_2(\mu)$ er givet ved

$$(f|g)=\int f(x)\overline{g(x)}\,d\mu(x). \tag{6.0.21}$$

Selvom det formelt er klart at $(f|f)=\int|f|^2d\mu=\|f\|_2^2$, såer det ikke uden videre klart, at integranden $f\overline{g}$ i det indre produkt overhovedet er integrabel for $f,g\in L_2(\mu)$. Dog giver banaliteten $(a-b)^2\ge 0$ for $a,b\in\mathbb{R}$ at

$$2ab\le a^2+b^2, \tag{6.0.22}$$

hvoraf man ser at $|f\overline{g}|=|f||g|\le|f|^2+|g|^2$ og udleder at $\int|f\overline{g}|\,d\mu<\infty$. I øvrigt vil integrabiliteten ogsåblive en nem konsekvens af Hölders ulighed, jvf. sætning 7.5, som vi blandt andet skal udnytte til at vise Minkowskis ulighed.

Fordi det indre produkt inducerer en norm (som inducerer en metrik), der er fuldstændig, såbetegnes $L_2(X,\mathbb{E},\mu)$ som et Hilbertrum til minde om David Hilbert, der omkring 1910 udførte omfattende analyser af spektralteori pådens slags vektorrum.

6.1. Fischer's completeness theorem

In the normed vector space $L_p(\mu)$, $1 \le p < \infty$ there is a version of majorised convergence, which one could conveniently refer to as the theorem on *L_p-majorised convergence:*

THEOREM 6.1.1. *Suppose that a sequence (f_n) is given in $L_p(\mu)$ for some $p \in [1, \infty[$ and that $f \colon X \to \mathbb{C}$ is such that*

$$f = \lim_{n\to\infty} f_n \quad \mu\text{-a.e.} \tag{6.1.1}$$

When there exists a function $g \in \mathscr{M}^+(X, \mathbb{E})$ such that (with $\infty^p = \infty$)

$$\forall n \in \mathbb{N} \colon |f_n| \le g \quad \mu\text{-a.e.}, \qquad \int g^p \, d\mu < \infty, \tag{6.1.2}$$

then $f \in L_p(\mu)$ and

$$\|f_n - f\|_p \to 0 \tag{6.1.3}$$

for $n \to \infty$.

PROOF. In the situation described in the statement, we may arrange that convergence holds everywhere, as all functions may be multiplied by $1_{X \setminus N}$ for some measurable null set N. Thus f can be assumed measurable. Moreover, $\int |f|^p \, d\mu \le \int g^p \, d\mu < \infty$, so that $f \in L_p(\mu)$ as stated.

Clearly both $|f_n(x) - f(x)|^p \to 0$ and $|f_n - f|^p \le 2^p g^p$ hold μ-a.e., so from the extension of the Majorised Convergence Theorem it follows that

$$\int |f_n - f|^p \, d\mu \xrightarrow[n\to\infty]{} \int 0 \, d\mu = 0. \tag{6.1.4}$$

The proof is complete. □

Completeness of a normed vector space V can be rephrased in a way that is most convenient for the study of $L_p(\mu)$. Indeed, a series $\sum_{n=1}^\infty x_n$ of vectors in V is said to be *absolutely convergent* if it has a finite norm series, that is, if $\sum_{n=1}^\infty \|x_n\| < \infty$. This notion enters

LEMMA 6.1.2. *A normed vector space V is complete if and only if every absolutely convergent series $\sum_{n=1}^\infty x_n$ in V is converging to some vector x in V.*

PROOF. Given a Cauchy series (x_n) in V, there are indices $n_1 < n_2 < \dots$ such that $\|x_n - x_m\| \le 2^{-k}$ whenever $n, m \ge n_k$. In particular $\|x_{n_{k+1}} - x_{n_k}\| \le 2^{-k}$, whence

$$\sum_{k=1}^\infty \|x_{n_{k+1}} - x_{n_k}\| \le 1. \tag{6.1.5}$$

So when absolute convergence implies convergence, the series $x_{n_1} + \sum_{k=1}^\infty (x_{n_{k+1}} - x_{n_k})$ converges to some vector x in V. As the series is telescopic, this means that $x = \lim_{k\to\infty} x_{n_k}$. Since moreover

$$\|x - x_n\| \le \|x - x_{n_k}\| + \|x_{n_k} - x_n\|, \tag{6.1.6}$$

it follows that the given sequence (x_n) converges to x as well.

The converse conclusion follows by applying the triangle inequality to a difference $s_{N+p} - s_N$ of two partial sums of an arbitrary absolutely convergent series $\sum_{n=1}^\infty x_n$ in V. □

Thus prepared, we proceed to state and prove the fundamental fact about the Lebesgue spaces $L_p(\mu)$:

THEOREM 6.1.3 (Fischer's Completeness Theorem). *The normed space $L_p(X, \mathbb{E}, \mu)$ is complete for any measure space $(X, \mathbb{E}, \mu)$ and $1 \le p < \infty$. In other words, $L_p(X, \mathbb{E}, \mu)$ is a Banach space.*

PROOF. Invoking Lemma 6.1.2, we let a series $\sum_{k=1}^{\infty} g_k$ of functions $g_k \in \mathscr{L}_p(\mu)$ be given such that

$$S := \sum_{k=1}^{\infty} \|g_k\|_p < \infty. \tag{6.1.7}$$

We shall determine a function $f \in \mathscr{L}_p(\mu)$ such that $\|f - \sum_{k=1}^{n} g_k\|_p \to 0$ holds for $n \to \infty$. (More precisely, this will show that $[f] = \lim_{n\to\infty} \sum_{k=1}^{n} [g_k]$ holds in $L_p(\mu)$, as desired.) Actually it turns out that the pointwise sum of $\sum_{k=1}^{\infty} g_k(x)$ exists a.e., and that this works as the function f.

1° Note first that there is a function $h \in \mathscr{M}^+(X, \mathbb{E})$ given by the formula

$$h(x) = \sum_{k=1}^{\infty} |g_k(x)|. \tag{6.1.8}$$

Clearly this auxiliary function fulfils $h(x) < \infty$ at $x \in X$ if, and only, if the series $\sum_{k=1}^{\infty} g_k(x)$ converges (absolutely) in $\mathbb{C}$.

2° The function h is a possible L_p-majorant, since $\int h^p \, d\mu < \infty$: obviously the convention $\infty^p = \infty$ gives for $n \to \infty$ that

$$(\sum_{k=1}^{n} |g_k(x)|)^p \nearrow h(x)^p, \tag{6.1.9}$$

so by the Monotone Convergence Theorem we have

$$\| \sum_{k=1}^{n} |g_k| \|_p^p = \int (\sum_{k=1}^{n} |g_k|)^p \, d\mu \nearrow \int h^p \, d\mu. \tag{6.1.10}$$

This yields that $\int h^p \, d\mu \in [0, S^p]$, hence is finite by (6.1.7), as the triangle inequality gives

$$0 \le \| \sum_{k=1}^{n} |g_k| \|_p \le \sum_{k=1}^{n} \| |g_k| \|_p \le \sum_{k=1}^{\infty} \|g_k\|_p = S. \tag{6.1.11}$$

The function h is indeed an L_p-majorant for the sequence given by $f_n = \sum_{k=1}^{n} g_k(x)$, for the inequality $|f_n| \le h$ holds on X for every n as an immediate consequence of (6.1.9).

3° We may define a measurable function by $f = \sum_{k=1}^{\infty} g_k 1_{X\setminus N}$, whereby $N \in \mathbb{E}$ is a null set chosen, as we may, so that $h(x) = \infty$ only holds in N. Indeed, the series converges pointwise also at every $x \notin N$, cf. 1°.

According to Theorem 6.1.1 we have that $f \in \mathscr{L}_p$ and that $\|f - f_n\|_p \to 0$ for $n \to \infty$. The proof is complete. $\square$

The attentive readerr will have noticed that the proof gave a bit more than stated:

COROLLARY 6.1.4. *A series $\sum_{k=1}^{\infty} g_k$ of functions satisfying $\sum_{k=1}^{\infty} \|g_k\|_p < \infty$ (for some $p \in [1, \infty[$) converges absolutely μ-a.e. on X as well as in p-mean to a function $f \in \mathscr{L}_p(\mu)$ satisfying*

$$\|f\|_p \le \sum_{k=1}^{\infty} \|g_k\|_p. \tag{6.1.12}$$

In the situation of this corollary, $f = \sum_{k=1}^{\infty} g_k$ holds in $L_p(\mu)$, so it is tempting to insert this in (6.1.12) to obtain

$$\| \sum_{k=1}^{\infty} g_k \|_p \le \sum_{k=1}^{\infty} \|g_k\|_p. \tag{6.1.13}$$

Although this is an abuse of notation, the above does appear as a convenient generalisation of Minkowski's inequality.

At the basic level, the best relation between pointwise convergence and convergence in p-mean is the following:

COROLLARY 6.1.5. *Whenever $f_1, f_2, \dots$ is a sequence in $L_p(\mu)$, for some $1 \le p < \infty$, that converges to some f in $L_p(\mu)$, then there is a subsequence $f_{n_1}, f_{n_2}, \dots$ converging pointwise to f μ-a.e. It is moreover possible to obtain an L_p-majorant for (f_{n_k}), that is some $g \in \mathscr{M}^+$ fulfilling $|f_{n_k}| \le g$ for all k and $\int g^p \, d\mu < \infty$.*

PROOF. This is a corollary to the proofs of Lemma 6.1.2 and Theorem 6.1.3. First $n_1 < n_2 < \dots$ are chosen so that

$$\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty. \tag{6.1.14}$$

The subsequence $f_{n_1}, f_{n_2}, \dots$ is the sequence of partial sums of $f_{n_1} + \sum_{k=1}^{\infty}(f_{n_{k+1}} - f_{n_k})$, which by construction has a finite norm series, so Corollary 6.1.4 yields that it converges both pointwise a.e. and in p-mean to some $\tilde{f}$ in $L_p(\mu)$. By hypothesis it also converges to f in the normed space $L_p(\mu)$, and therefore $f = \tilde{f}$ in $L_p(\mu)$; so f is also an a.e. pointwise limit of the subsequence.

Going back to the proof given for Fischer's theorem, the function h there is in the present case given by

$$g(x) = |f_{n_1}(x)| + \sum_{k=1}^{\infty} |f_{n_{k+1}}(x) - f_{n_k}(x)|, \tag{6.1.15}$$

so this is a possible L_p-majorant here. $\square$

COROLLARY 6.1.6. *If a sequence $f_1, f_2, \dots$ in $L_p(\mu)$ converges in p-mean to some $\varphi \in L_p(\mu)$ as well as pointwise to some function $\psi \colon X \to \mathbb{C}$, then $\varphi = \psi$ holds μ-a.e.*

Indeed, according to Corollary 6.1.5 a suitable subsequence (f_{n_k}) converges a.e. to φ, and of course also to w a.e.

6.2. Density of nice functions

Simple functions and C_c.

THEOREM 6.2.1. *The space $C_c(\mathbb{R}^d)$ of continuous functions with compact support is dense in $L_p(\mathbb{R}^d)$ whenever $1 \le p < \infty$.*

CHAPTER 7

Convolution

7.1. Convolution of Borel functions

7.2. The Banach algebra $L_1(\mathbb{R}^d)$

7.3. Strong convergence of translation

As a convenient notation for a function f defined on $D(f) \subset \mathbb{R}^d$, we shall denote the translated function $f(x-a)$ by $\tau_a f$ for $a \in \mathbb{R}^d$; that is,

$$\tau_a f(x) = f(x-a). \tag{7.3.1}$$

Here $\tau_a f$ is defined on the subset $a + D(f)$, in general. This is redundant of course if $D(f) = \mathbb{R}^d$. In particular this is so when $f \in L_p(\mathbb{R}^d)$, and then it is clear that also $\tau_a f \in L_p(\mathbb{R}^d)$, for because of the translation invariance of the Lebesgue measure we may note once and for all that

$$\|\tau_a f\|_p = (\int_{\mathbb{R}^d} |f(x-a)|^p \, dx)^{1/p} = \|f\|_p. \tag{7.3.2}$$

In many cases it is a useful result that $\tau_a f \to f$ for $a \to 0$ in $\mathbb{R}^d$, when $f \in L_p(\mathbb{R}^d)$ is fixed. The basic result in this direction is

PROPOSITION 7.3.1. *When $f \colon \mathbb{R}^d \to \mathbb{C}$ belongs to $L_p(\mathbb{R}^d)$ for $1 \le p < \infty$, then*

$$\|\tau_a f - f\|_p = (\int_{\mathbb{R}^d} |f(x-a) - f(x)|^p \, dx)^{\frac{1}{p}} \to 0 \qquad \text{for } a \to 0. \tag{7.3.3}$$

PROOF. Given $\varepsilon > 0$ there exists by the density of $C_c(\mathbb{R}^d)$ in $L_p(\mathbb{R}^d)$ a function $g \in C_c$ such that $\|f - g\|_p \le \varepsilon/3$; cf. Theorem 6.2.1. Because of the translation invariance of the Lebesgue measure this gives via the triangle inequality

$$\|\tau_a f - f\|_p \le \|\tau_a(f-g)\|_p + \|\tau_a g - g\|_p + \|g - f\|_p \le \frac{2\varepsilon}{3} + \|\tau_a g - g\|_p. \tag{7.3.4}$$

The function $g \in C_c$ is uniformly continuous on $\mathbb{R}^d$ since $g \equiv 0$ outside a ball $B(0,R)$ containing $\operatorname{supp} g$. In fact, to $\varepsilon' = \varepsilon / 3 m_d(B(0,R+1))^{\frac{1}{p}}$ there exists by its uniform continuity on $\bar{B}(0,N+1)$ some $\delta \in]0,1[$ such that

$$|\tau_a g(x) - g(x)| \le \varepsilon' 1_{B(0,R+1)}(x) \quad \text{for } |a| < \delta,\, x \in \mathbb{R}^d. \tag{7.3.5}$$

From this we obtain

$$\|\tau_a g - g\|_p \le \varepsilon' m_d(B(0,R+1))^{\frac{1}{p}} = \varepsilon. \tag{7.3.6}$$

Consequently $\|\tau_a f - f\|_p \le \varepsilon$ holds for $|a| < \delta$. □

The result is known as so-called strong convergence of translation to the identity, which is written $\tau_a \to I$ strongly for $a \to 0$.

It is noteworthy from the proof how the density of the continuous functions with compact support, i.e. of C_c, gave a reduction to such functions. And that the property was relatively straightforward to obtain for the elements in the dense subset.

It is easy to see that the strong convergence $\tau_a \to I$ cannot be extended to hold in (the norm of) the space $L_\infty(\mathbb{R}^d)$.

7.4. Approximative units in the Lebesgue spaces

The next result describes a sequence of functions h_n which seemingly approaches a unit, i.e. a neutral element of the convolution in the Banach algebra $L_1(\mathbb{R}^d)$. It may be seen in various ways, however, that no such unit exists. The following is therefore a substitute.

THEOREM 7.4.1. *When $(h_n)_{n\in\mathbb{N}}$ is a sequence in $\mathscr{L}(\mathbb{R}^d)$ such that*

(i) $\forall n \in \mathbb{N}\colon\ h_n \geq 0$,
(ii) $\forall n \in \mathbb{N}\colon\ \int h_n \, dm_d = 1$,
(iii) $\forall \delta > 0\colon\ \int_{|x|>\delta} h_n(x)\,dx \to 0$ *for* $n \to \infty$,

then it holds true for every $f \in \mathscr{L}(\mathbb{R})$ that

$$\|f * h_n - f\|_1 = \int_{\mathbb{R}^d} |f * h_n(x) - f(x)|\,dx \to 0 \quad \text{for } n \to \infty. \tag{7.4.1}$$

That is, such a sequence (h_n) is an approximative unit *for the Banach algebra $L_1(\mathbb{R}^d)$.*

PROOF. Since (ii) gives $f(x) = f(x)\int h_n(y)\,dy$, we obtain for every $x \in D(f * k_n)$, hence almost everywhere,

$$|f * h_n(x) - f(x)| \leq \int |f(x-y) - f(x)| h_n(y)\,dy. \tag{7.4.2}$$

Here $(x,y) \mapsto |f(x-y) - f(x)|h_n(y)$ is a Borel function on $\mathbb{R}^{2d}$, so from Tonelli's theorem we obtain

$$\begin{aligned} \|f * h_n - f\|_1 &\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y) - f(x)| h_n(y)\,dydx \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y) - f(x)| h_n(y)\,dxdy = \int_{\mathbb{R}^d} \|\tau_y f - f\|_1 h_n(y) dy. \end{aligned} \tag{7.4.3}$$

Now we may to any given $\varepsilon > 0$ fix $\delta > 0$ so that $\|\tau_y f - f\|_1 \leq \varepsilon/2$ for $|y| \leq \delta$, and since $\|\tau_y f - f\|_1 \leq 2\|f\|_1$ by the translations invariance, we obtain

$$\int_{|y|\leq\delta} \|\tau_y f - f\|_1 h_n(y) dy \leq \int_{|y|<\delta} \frac{\varepsilon}{2} h_n(y) dy \leq \frac{\varepsilon}{2}, \tag{7.4.4}$$

$$\int_{|y|>\delta} \|\tau_y f - f\|_1 h_n(y) dy \leq 2\|f\|_1 \int_{|y|\geq\delta} h_n(y) dy. \tag{7.4.5}$$

According to (iii) there is some N such the last term is less than $\varepsilon/2$ for $n > N$, so it follows that $\|f * h_n - f\|_1 \leq \varepsilon$ for such n. □

The attentive reader may have noticed that the existence of an approximative unit still remains to be shown. But any integrable Borel function $h \geq 0$ for which $\int_{\mathbb{R}^d} h\,dm_d = 1$ induces a sequence fulfilling (i), (ii) and (iii) via the formula

$$h_n(x) = n^d h(nx), \quad n \in \mathbb{N}. \tag{7.4.6}$$

Indeed, the integral in (iii) may for $z = \frac{1}{n}y$ be written $\int_{\mathbb{R}^d} 1_{|z|>n\delta}(z) h(z)\,dz$, which obviously goes to 0 by the Majorised Convergence Theorem.

As a simple example there is $h = 1_{[0,1]^d}$. To give an example with a function in C^∞ for $d = 1$ one may consider $h(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, so that

$$h_n(x) = \frac{1}{\pi}\frac{n}{1+n^2x^2}. \tag{7.4.7}$$

Obviously the peak at $x = 0$ becomes increasingly more pronounced as $n \to \infty$.

The content of the theorem extends readily to the analogous situation of a family $(h_t)_{t>0}$ in $\mathscr{L}$, which also fulfils (i)–(iii). In fact, such a family may be obtained as above by letting $h_t(x) = t^d h(tx)$. But for simplicity we shall just consider approximative units that are sequences.

For the Lebesgue spaces $L_p(\mathbb{R}^d)$ with $1 < p < \infty$ the situation is different, since these are not convolution algebras. Moreover, the above approximative units are not members of L_p for such p. Nevertheless there are similar, important results that we now describe.

The fact that the Lebesgue spaces $L_p(\mathbb{R}^d)$ with $1 \le p < \infty$ are invariant under convolution by an integrable function is a consequence of the next result:

THEOREM 7.4.2. *When $f \in \mathscr{L}_p(\mathbb{R}^d)$ for $1 \le p < \infty$ and $g \in \mathscr{L}_1(\mathbb{R}^d)$, then $f * g$ is defined almost everywhere in $\mathbb{R}^d$. The induced equivalence class is also denoted by $f * g$, it belongs to $L_p(\mathbb{R}^d)$ and fulfils*

$$\|f * g\|_p \le \|f\|_p \|g\|_1. \tag{7.4.8}$$

PROOF. As the case $p = 1$ was covered previously, we assume $1 < p < \infty$ and determine its dual exponent q from $p + q = pq$.

According to Hölder's inequality we have, with integrals over $\mathbb{R}^d$,

$$\begin{aligned} \int |f(x-y)g(y)|\,dy &= \int |f(x-y)||g(y)|^{\frac{1}{p}}|g(y)|^{\frac{1}{q}}\,dy \\ &\le \Big(\int |f(x-y)|^p|g(y)|\,dy\Big)^{\frac{1}{p}}\Big(\int |g(y)|\,dy\Big)^{\frac{1}{q}} < \infty \end{aligned} \tag{7.4.9}$$

whenever $\int |f(x-y)|^p|g(y)|\,dy < \infty$, that is, for all $x \in D(|f|^p * |g|)$. Because of this inequality, we have the inclusion $D(|f|^p * |g|) \subset D(f * g)$. As the former fills $\mathbb{R}^d$ except for a nullset, since $|f|^p, |g| \in \mathscr{L}$, so does the latter. Hence $f * g$ is almost everywhere defined.

For $x \in D(f*g)$ we have $|f*g(x)| \le \int |f(x-y)g(y)|\,dy$, so the above inequality gives

$$|f * g(x)|^p \le |f|^p * |g|(x)\|g\|_1^{p-1}. \tag{7.4.10}$$

By integrating this inequality and using that $L_1(\mathbb{R}^d)$ is a convolution algebra we obtain

$$\int |f*g(x)|^p\,dx \le \int |f|^p * |g|\,dx \|g\|_1^{p-1} \le (\||f|^p\|_1\|g\|_1)\|g\|_1^{p-1} = \|f\|_p^p\|g\|_1^p, \tag{7.4.11}$$

from where the stated inequality follows at once. □

Thus prepared we turn to approximation of functions in $\mathscr{L}_p$ by convolutions:

THEOREM 7.4.3. *When $(h_n)_{n\in\mathbb{N}}$ is an approximative unit in $\mathscr{L}(\mathbb{R}^d)$ and $f \in \mathscr{L}_p(\mathbb{R}^d)$ for some $p \in [1, \infty[$, then*

$$\|f * h_n - f\|_p \to 0 \qquad \text{for } n \to \infty. \tag{7.4.12}$$

PROOF. The case $p = 1$ was covered in Theorem 7.4.1, so we may assume $1 < p < \infty$ and determine its dual exponent q from $p + q = pq$.

The functions $f * h_n(x)$ and $|f|^p * h_n(x)$ are both defined outside a certain nullset, and for such x we get from Hölder's inequality,

$$\begin{aligned} |f * h_h(x) - f(x)| &\le \int |f(x-y) - f(x)|h_n(y)\,dy \\ &= \int |f(x-y) - f(x)|h_n(y)^{\frac{1}{p}}h_n(y)^{\frac{1}{q}}\,dy \\ &\le \Big(\int |f(x-y) - f(x)|^p h_n(y)\,dy\Big)^{\frac{1}{p}} \cdot 1. \end{aligned} \tag{7.4.13}$$

Using Tonelli's theorem this implies

$$\|f * h_n - f\|_p^p \le \int\int |f(x-y) - f(x)|^p h_n(y)\,dxdy = \int \|\tau_y f - f\|_p^p h_n(y)\,dy. \tag{7.4.14}$$

From this inequality, the proof can be completed analogously to the proof of Theorem 7.4.1, using that $\tau_y \to I$ strongly on L_p for $y \to 0$ and that (iii) holds for the h_n. □

cartoon on existence of functions in

$$C_c^\infty = C_c \bigcap C^\infty. \tag{7.4.15}$$

(Cf. section 8.4 in the book, p. 189–190)

In view of this, it is clear that an appoximative unit $(h_n)_{n\in\mathbb{N}}$ can be chosen (in many ways) so that it satisfies

$$h_n(x) = n^d h(nx), \quad h \in C_c^\infty(\mathbb{R}^d), \int_{\mathbb{R}^d} h\,dx = 1, \tag{7.4.16}$$

$$h_n(x) = 0 \text{ for } |x| > \frac{1}{n}, \quad h \geq 0, \quad h_n(x) = 1 \text{ for } |x| < \frac{1}{2n}. \tag{7.4.17}$$

An elegant way of stating the last line could be that $1_{B(0,1/2n)} \leq h \leq 1_{B(0,1/n)}$. Such a choice of (h_n) is understood in the following.

Thus prepared, one may obtain the next result describing, during the course of the proof, how any $f \in L_p(\mathbb{R}^d)$ can be approximated by a sequence of C_c^∞-functions in a convenient way:

THEOREM 7.4.4. *When $f \in L_p(\mathbb{R}^d)$ for $1 \leq p < \infty$ there is a sequence of functions $g_m \in C_c^\infty(\mathbb{R}^d)$ such that*

$$\|f - g_m\|_p \to 0 \quad \text{for } m \to \infty. \tag{7.4.18}$$

In case $f \in L_p(\mathbb{R}^d) \bigcap L_q(\mathbb{R}^d)$ for $p, q \in [1, \infty[$ the above sequence $(g_m)_{m\in\mathbb{N}}$ can be so chosen that

$$\|f - g_m\|_q \to 0 \quad \text{for } m \to \infty \tag{7.4.19}$$

also holds.

PROOF. For f given as in the theorem, the g_n are chosen from the family

$$(f 1_{B(0,N)}) * h_n(x) = \int_{|y|<N} f(y) h_n(x-y)\,dy, \quad N, n \in \mathbb{N}. \tag{7.4.20}$$

First we note that each of these functions is in C^∞, since the differential operator ∂^α can be applied under the integral sign, using $|f(y)| n^{d+|\alpha|} \sup |\partial^\alpha h|$ as the majorant on $\mathbb{R}^d$. Its support is compact, in fact contained in $\bar{B}(0, N+1)$, so $(f 1_{B(0,N)}) * h_n$ belongs to C_c^∞.

For each $\varepsilon = 2^{-m}$, $m \in \mathbb{N}$, we observe the inequality

$$\|f - (f 1_{B(0,N)}) * h_n\|_p \leq \|f - f 1_{B(0,N)}\|_p + \|f 1_{B(0,N)} - (f 1_{B(0,N)}) * h_n\|_p. \tag{7.4.21}$$

The first term on the right-hand side is less than $\varepsilon/2$ for some N_m, as can be seen from the Majorised Convergence Theorem. The second term is with $N = N_m$ also less than $\varepsilon/2$ when the index is chosen as some suitable n_m; as $f 1_{B(0,N_m)}$ belongs to L_p this is a consequence of Theorem 7.4.3. Hence $g_m = (f 1_{B(0,N_m)}) * h_{n_m}$ achieves that $g_m \in C_c^\infty$ and

$$\|f - g_m\|_p \leq 2^{-m}. \tag{7.4.22}$$

If $f \in L_q$ holds too, one can arrange that also $\|f - f 1_{B(0,N)}\|_q \leq \varepsilon/2$ by taking N_m suitably larger (if necessary). Then $\|f 1_{B(0,N_m)} - (f 1_{B(0,N_m)}) * h_{n_m}\| \leq \varepsilon/2$ holds both in L_p and in L_q for some sufficiently large n_m. Thus $\|f - g_m\| < \varepsilon$ holds in both spaces. (Obviously one can even arrange that $n_1 < n_2 < \dots$ and $N_1 < N_2 < \dots$, when useful.) □

7.5. Parseval–Plancherel's theorem

Because of the obvious inclusion $C_c^\infty(\mathbb{R}^d) \subset \mathscr{S}(\mathbb{R}^d)$, it is clear from Theorem 7.4.4 that the Schwartz space $\mathscr{S}$ is dense in $L_P(\mathbb{R}^d)$ for each $p \in [1, \infty[$.

This density may now be used to extend the Fourier transformation $\mathscr{F}$ on $\mathscr{S}(\mathbb{R}^d)$ and its bijectiveness to the setting of $L_2(\mathbb{R}^d)$.

To define $\mathscr{F}$ on any given $f \in L_2(\mathbb{R}^d)$ it suffices to take, as we may, a sequence (f_n) in $\mathscr{S}$ such that $f_n \to f$ in L_2 for $n \to \infty$ and then define the extended Fourier tranformation $\mathscr{F}_2$ on f to be

$$\mathscr{F}_2 f = \lim_{n\to\infty} \mathscr{F} f_n. \tag{7.5.1}$$

Indeed, it is first of all clear that this limit exists in L_2, for Parseval's formula for Schwartz functions shows at once that $(\mathscr{F} f_n)$ is a Cauchy sequence in L_2,

$$\|\mathscr{F} f_n - \mathscr{F} f_m\|_2 = \|\mathscr{F}(f_n - f_m)\|_2 = (2\pi)^d \|f_n - f_m\|_2. \tag{7.5.2}$$

Secondly $\lim_{n\to\infty} \mathscr{F} f_n$ does not depend on the particular choice of the Schwartz functions f_n, for if also $\|f - g_n\|_2 \to 0$ for $g_n \in \mathscr{S}$, then the interlaced sequence

$$f_1, g_1, f_2, g_2, \dots, f_n, g_n, \dots \tag{7.5.3}$$

is another Cauchy sequence, which $\mathscr{F}$ by (7.5.2) sends to a Cauchy sequence in L_2—but since a Cauchy sequence cannot have more than one cluster point, the two obvious cluster points $\lim_{n\to\infty} \mathscr{F} f_n$ and $\lim_{n\to\infty} \mathscr{F} g_n$ are equal. Hence $\mathscr{F}_2$ is a well-defined map

$$\mathscr{F}_2 \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d) \tag{7.5.4}$$

Thirdly, $\mathscr{F}_2 \psi = \mathscr{F} \psi$ for every $\psi \in \mathscr{S}$, for then $f_n = \psi$ for every n will do. Hence $\mathscr{F}_2$ coincides with $\mathscr{F}$ in the dense subset $\mathscr{S}$.

Finally it follows from the calculus of limits that $\mathscr{F}_2 \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ is a linear map. For when $z, w \in \mathbb{C}$ and $f, g \in L_2$ are approached in L_2-norm by $\varphi_n, \psi_n \in \mathscr{S}$, respectively, then

$$z\mathscr{F}_2 f + w\mathscr{F}_2 g = z \lim_n \mathscr{F} \varphi_n + w \lim_n \mathscr{F} \psi_n \tag{7.5.5}$$

$$= \lim_n (z\mathscr{F} \varphi_n + w\mathscr{F} \psi_n) = \lim_n \mathscr{F}(z\varphi_n + w\psi_n) = \mathscr{F}_2(zf + wg). \tag{7.5.6}$$

It is a fundamental result that the Fourier transformation $\mathscr{F}_2$, defined on L_2 as above, actually is an *isometry* when using the trick of invoking a suitably weighted Lebesgue measures on $\mathbb{R}^d$. This is also known as the Parseval–Plancherel formula:

THEOREM 7.5.1. *The Fourier transformation $\mathscr{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ extends in a unique way to a continuous, linear, bijective isometry*

$$\mathscr{F}_2 \colon L_2(m_d) \to L_2((2\pi)^{-d} m_d). \tag{7.5.7}$$

In particular it holds for all $f, g \in L_2(\mathbb{R}^d)$ that

$$\int_{\mathbb{R}^d} |f(x)|^2 \, dx = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\mathscr{F}_2 f(\xi)|^2 \, d\xi, \tag{7.5.8}$$

$$\int_{\mathbb{R}^d} f(x)\overline{g(x)} \, dx = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathscr{F}_2 f(\xi) \overline{\mathscr{F}_2 g(\xi)} \, d\xi. \tag{7.5.9}$$

Analogously the Fourier co-transformation $\overline{\mathscr{F}}$ on $\mathscr{S}(\mathbb{R}^d)$ has an extension $\overline{\mathscr{F}}_2$ with the same properties as a map $\overline{\mathscr{F}}_2 \colon L_2(\mathbb{R}^d, m_d) \to L_2(\mathbb{R}^d, (2\pi)^{-d} m_d)$, and

$$\mathscr{F}_2^{-1} = (2\pi)^{-d} \overline{\mathscr{F}}_2. \tag{7.5.10}$$

(Fourier's' inversion formula for $\mathscr{F}_2$.)

PROOF. Injectivity of $\mathscr{F}_2$ is immediate from the isometric property $\|\mathscr{F}_2 f\| = \|f\|$, which holds for the norms in (7.5.7) because of (7.5.8), which in its turn follows by taking $g = f$ in (7.5.9).

The formula (7.5.9) is easily carried over from the corresponding fact for $\mathscr{F}$ on Schwartz functions, for with the $\varphi_n, \psi_n \in \mathscr{S}$ used prior to the theorem we may first infer that the inner product on the Hilbert space $L_2(m_d)$ is continuous in the two entries jointly: we have

$$(\varphi_n \mid \psi_n) - (f \mid g) = (\varphi_n - f \mid \psi_n - g) + (f \mid \psi_n - g) + (\varphi_n - f \mid g), \tag{7.5.11}$$

which via the triangle inequality implies that $(\varphi_n \mid \psi_n) \to (f \mid g)$ for $n \to \infty$. Similarly the inner product $((\cdot \mid \cdot))$ on $L_2((2\pi)^{-d} m_d)$ is jointly continuous. Using this we have

$$\begin{aligned}\int_{\mathbb{R}^d} f(x)\overline{g(x)}\,dx = (f \mid g) &= \lim_n (\varphi_n \mid \psi_n) \\ &= \lim_n ((\mathscr{F}\varphi_n \mid \mathscr{F}\psi_n)) = ((\lim_n \mathscr{F}\varphi_n \mid \lim_n \mathscr{F}\psi_n)) \\ &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \mathscr{F}_2 f(\xi)\overline{\mathscr{F}_2 g(\xi)}\,d\xi.\end{aligned} \tag{7.5.12}$$

Moreover, it is easy to see from (7.5.8) that $\mathscr{F}_2$ is continuous (cf. (7.5.2)).

To show that $\mathscr{F}_2$ also is surjective, note that its range $\mathscr{F}_2(L_2)$ contains the dense subset $\mathscr{F}(\mathscr{S}) = \mathscr{S}$. In addition its range is closed in L_2 as $\mathscr{F}_2$ is an isometry, for if $\mathscr{F}_2 f_n \to h$ in L_2, then $(\mathscr{F}_2 f_n)$ is a Cauchy sequence in $L_2((2\pi)^{-d} m_d)$, and hence (f_n) is so in $L_2(m_d)$; that is $f_n \to g$ in L_2, so that $h = \lim_n \mathscr{F}_2 f_n = \mathscr{F}_2 g$ by the continuity of $\mathscr{F}_2$. Altogether

$$\mathscr{F}_2(L_2) = \overline{\mathscr{F}_2(L_2)} \supset \overline{\mathscr{S}} = L_2, \tag{7.5.13}$$

and since the converse inclusion is trivial, $\mathscr{F}_2$ is surjective. The results so far carry over to the Fourier co-transformation by (temporarily) setting $\overline{\mathscr{F}}_2 g = \overline{\mathscr{F}_2 \overline{g}}$.

The uniqueness of $\mathscr{F}_2$ follows from its continuity, for if $\widetilde{\mathscr{F}_2}$ is any extension of $\mathscr{F}$ having the properties shown for $\mathscr{F}_2$, then for every $g \in L_2$ we have

$$\widetilde{\mathscr{F}_2} g = \lim_n \mathscr{F}\psi_n = \mathscr{F}_2 g. \tag{7.5.14}$$

Likewise the continuity of $\overline{\mathscr{F}}_2$ implies its uniqueness; so an application of the limit procedure prior to the theorem to $\overline{\mathscr{F}}$ would have given the same map $\overline{\mathscr{F}}_2$.

Finally, Fourier's' inversion formula on $\mathscr{S}$ gives the identities

$$(2\pi)^{-d}\overline{\mathscr{F}}_2\mathscr{F}_2\psi_n = \psi_n = \mathscr{F}_2(2\pi)^{-d}\overline{\mathscr{F}}_2\psi_n \tag{7.5.15}$$

so by passing to the limit the continuity of $\mathscr{F}_2$ and $\overline{\mathscr{F}}_2$ gives

$$(2\pi)^{-d}\overline{\mathscr{F}}_2\mathscr{F}_2 g = g = \mathscr{F}_2(2\pi)^{-d}\overline{\mathscr{F}}_2 g. \tag{7.5.16}$$

As $g \in L_2(\mathbb{R}^d)$ is arbitrary here, this proves the inversion formula for $\mathscr{F}_2$. □

The map $\mathscr{F}_2$ is sometimes called the Fourier–Plancherel transformation.

In order to drop the tedious distinction between $\mathscr{F}$, as defined on $L_1(\mathbb{R}^d)$, and the map $\mathscr{F}_2$ defined on $L_2(\mathbb{R}^d)$ in the complicated way above, it is convenient to show that they give the same result result on the functions f on which they are both defined.

Indeed, to this end we may apply the fine result in the second part of Theorem 7.4.4. This states that there exists a sequence $\psi_n \in C_c^\infty \subset \mathscr{S}$ converging to f in *both* L_1 and L_2, and because of the continuity of $\mathscr{F}\colon L_1 \to C_b$ and $\mathscr{F}_2\colon L_2 \to L_2$ we see that the sequence $\mathscr{F}_2\psi_n = \mathscr{F}\psi_n$ for $n \to \infty$ fulfils

$$\|\mathscr{F}_2 f - \mathscr{F}\psi_n\|_2 \to 0, \qquad \sup_{\xi\in\mathbb{R}^d} |\mathscr{F}f(\xi) - \mathscr{F}\psi_n(\xi)| \to 0 \tag{7.5.17}$$

However, when a sequence such as $\mathscr{F}\psi_n$ converges both pointwise and in quadratic mean, then the two limit functions coincide. Therefore $\mathscr{F}_2 f(\xi) = \mathscr{F}f(\xi)$ for all $\xi \in \mathbb{R}^d$, so these considerations may be celebrated with the following diagram:

$$\begin{array}{ccc} \mathscr{F}_2\psi_n(\xi) & = & \mathscr{F}\psi_n(\xi) \\ \big\downarrow & & \big\downarrow \\ \mathscr{F}_2 f(\xi) & = & \int e^{-\mathrm{i}x\cdot\xi} f(x)\,dx \end{array} \tag{7.5.18}$$

Summing up we have shown:

PROPOSITION 7.5.2. *$\mathscr{F}_2 f = \mathscr{F} f$ for every $f \in L_1(\mathbb{R}^d) \bigcap L_2(\mathbb{R}^d)$, so for such f we have*

$$\mathscr{F}_2 f(\xi) = \int_{\mathbb{R}^d} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx. \tag{7.5.19}$$

Because of the above result, it is now safe to simplify the notation from $\mathscr{F}_2$ to $\mathscr{F}$. By doing so, the Fourier transformation is easily seen to give a surjective linear isometry between the ordinary Lebesgue spaces:

$$(2\pi)^{-d/2}\mathscr{F} \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d). \tag{7.5.20}$$

Here the inverse is $(2\pi)^{-d/2}\overline{\mathscr{F}}$.

REMARK 7.5.3. The above discussion may be completed by the following classical fact on how the Fourier transformed function $\mathscr{F} f$ can be computed for any $f \in L_2(\mathbb{R}^d)$. In fact, for any such f it is clear that $f1_{B(0,N)}$ belongs to the intersection $L_1 \cap L_2$ because the ball $B(0,N)$ has finite measure. So according to Proposition 7.5.2 we have

$$\mathscr{F}(f1_{B(0,N)})(\xi) = \int_{|x|<N} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx. \tag{7.5.21}$$

Here the function on the right-hand side can be seen as a *truncated* Fourier integral, but it belongs in fact to $C_b(\mathbb{R}^d)$ as $f1_{B(0,N)}$ is in L_1. Since we have $f1_{B(0,N)} \to f$ in L_2, these continuous functions converge in $L_2(\mathbb{R}^d)$ for $N \to \infty$ to the function $\mathscr{F} f$; whence a subsequence converges pointwise (a.e. to a representative of) $\xi \mapsto \mathscr{F} f(\xi)$.

REMARK 7.5.4. Further applications of the Schwartz space $\mathscr{S}(\mathbb{R}^d)$, which was introduced ca. 1950 by L. Schwartz, can be found in his fundamental book [**Sch66**].

Epilogue

In 1872, K. Weierstrass presented his famous example of a *nowhere* differentiable, yet continuous function W on the real line $\mathbb{R}$. In terms of two real parameters $b \geq a > 1$, this may be written as

$$W(t) = \sum_{j=0}^{\infty} \frac{\cos(b^j t)}{a^j}, \quad t \in \mathbb{R}. \tag{0.0.22}$$

With elementary considerations, Weierstrass proved that W is continuous at every $t_0 \in \mathbb{R}$, but not differentiable at any $t_0 \in \mathbb{R}$ in case

$$\frac{b}{a} > 1 + \frac{3\pi}{2}, \quad b \text{ is an odd integer.} \tag{0.0.23}$$

Subsequently several mathematicians attempted to relax the unnatural condition (0.0.23), but with limited luck. And much later G. H. Hardy [**Har16**] was able to remove it by proving the following result:

THEOREM 0.0.5 (Hardy 1916). *For every real number $b \geq a > 1$ the functions*

$$W(t) = \sum_{j=0}^{\infty} a^{-j}\cos(b^j t), \qquad S(t) = \sum_{j=0}^{\infty} a^{-j}\sin(b^j t), \tag{0.0.24}$$

are bounded and continuous on $\mathbb{R}$, but have no points of differentiability.

Here the assumption $b \geq a$ is optimal for every $a > 1$, for W is in $C^1(\mathbb{R})$ whenever $\frac{b}{a} < 1$, due to uniform convergence of the derivatives. (Strangely, this was not observed in [**Har16**, Sect. 1.2], where Hardy tried to justify the sufficient condition $b \geq a$ as being more natural than e.g. (0.0.23).) Hardy also proved that $S'(0) = +\infty$ for

$$1 < a \leq b < 2a - 1, \tag{0.0.25}$$

so then the graph of $S(t)$ is not rough at $t = 0$ (similarly $W'(\pi/2) = +\infty$ if $b \in 4\mathbb{N} + 1$).

However, Hardy's treatment is not entirely elementary and yet it fills ca. 15 pages. It is perhaps partly for this reason that several attempts have been made over the years to find other examples. These have often involved a replacement of the sine and cosine above by a function with a zig-zag graph; the first one was due to T. Takagi [**Tak03**] who considered $t \mapsto \sum_{j=0}^{\infty} 2^{-j}\operatorname{dist}(2^j t, \mathbb{Z})$.

But as a drawback, the partial sums are not C^1 for such series of zig-zag functions. And due to the dilations every $x \in \mathbb{R}$ is a limit $x = \lim r_N$ where each $r_N \in \mathbb{Q}$ is a point at which the N^{th} partial sum has no derivatives; whence nowhere-differentiability of the sum function is less startling in this case. Even so, a fine example of this sort was given in just 13 lines by J. McCarthy [**McC53**].

However, there is an equally short proof of nowhere-differentiability, using a few basics of integration theory. This is well within reach in these lecture notes.

REMARK 0.0.6. By a well-known heuristic reasoning, $W(t)$ is nowhere-differentiable since the j^{th} term cannot cancel the oscillations of the previous ones: it is out of phase with the previous terms as $b > 1$ and the amplitudes moreover decay exponentially since $\frac{1}{a} < 1$. As $b \geq a > 1$ the combined effect is large enough (vindicated by the optimality of $b \geq a$ noted after Theorem 0.0.5).

To present the ideas in a clearer way we consider the following function f_θ, which may serve as a typical nowhere differentiable function,

$$f_\theta(t) = \sum_{j=0}^{\infty} 2^{-j\theta} e^{\mathrm{i}2^j t}, \qquad 0 < \theta \le 1. \tag{0.0.26}$$

It is convenient to choose an auxiliary function $\chi\colon \mathbb{R} \to \mathbb{C}$ thus: the Fourier transformed function $\mathscr{F}\chi(\tau) = \overset{\wedge}{\chi}(\tau) = \int_{\mathbb{R}} e^{-\mathrm{i}t\tau}\chi(t)\,dt$ is chosen as a C^∞-function fulfilling

$$\overset{\wedge}{\chi}(1) = 1, \qquad \overset{\wedge}{\chi}(\tau) = 0 \text{ for } \tau \notin]\tfrac{1}{2}, 2[; \tag{0.0.27}$$

for example by setting

$$\overset{\wedge}{\chi}(\tau) = 1_{]\frac{1}{2},2[}(\tau)\cdot \exp\Big(2 - \frac{1}{(2-\tau)(\tau-1/2)}\Big). \tag{0.0.28}$$

Since $\overset{\wedge}{\chi} \in C_c^\infty(\mathbb{R}) \subset \mathscr{S}(\mathbb{R})$, the fact that $\mathscr{F}$ maps $\mathscr{S}(\mathbb{R})$ bijectively to itself yields that also χ belongs to the Schwartz space $\mathscr{S}(\mathbb{R})$. And clearly $\int \chi\,dt = \overset{\wedge}{\chi}(0) = 0$.

With this preparation, the function f_θ is particularly simple to treat, using only common exercises in integration theory: First one may introduce the convolution

$$2^k\chi(2^k\cdot) * f_\theta(t_0) = \int_{\mathbb{R}} 2^k\chi(2^k t) f_\theta(t_0 - t)\,dt, \tag{0.0.29}$$

which is in $L_\infty(\mathbb{R})$ since $f_\theta \in L_\infty(\mathbb{R})$ and $\chi \in L_1(\mathbb{R})$. Secondly this will be analysed in two different ways in the proof of

PROPOSITION 0.0.7. *For $0 < \theta \le 1$ the function $f_\theta(t) = \sum_{j=0}^{\infty} 2^{-j\theta} e^{\mathrm{i}2^j t}$ is a continuous 2π-periodic, hence bounded function $f_\theta\colon \mathbb{R} \to \mathbb{C}$ without points of differentiability.*

PROOF. By uniform convergence, f_θ is a continuous 2π-periodic and bounded function for each $\theta > 0$. This follows from Weierstrass's majorant criterion as $\sum 2^{-j\theta} < \infty$.

Inserting the series defining f_θ into (0.0.29), the Majorised Convergence Theorem allows the sum and integral to be interchanged (e.g. with $\frac{2^k}{1-2^{-\theta}}|\chi(2^k t)|$ as a majorant),

$$\begin{aligned} 2^k\chi(2^k\cdot) * f_\theta(t_0) &= \lim_{N\to\infty} \sum_{j=0}^{N} 2^{-j\theta} \int_{\mathbb{R}} 2^k \chi(2^k t) e^{\mathrm{i}2^j(t_0 - t)}\,dt \\ &= \sum_{j=0}^{\infty} 2^{-j\theta} e^{\mathrm{i}2^j t_0} \int_{\mathbb{R}} e^{-\mathrm{i}z2^{j-k}} \chi(z)\,dz \\ &= 2^{-k\theta} e^{\mathrm{i}2^k t_0} \overset{\wedge}{\chi}(1) = 2^{-k\theta} e^{\mathrm{i}2^k t_0}. \end{aligned} \tag{0.0.30}$$

Here it was tacitly used that $\overset{\wedge}{\chi}(2^{j-k}) = 1$ for $j = k$, and that it equals 0 for $j \ne k$.

Moreover, since $f_\theta(t_0)\int_{\mathbb{R}} \chi\,dz = 0$ (cf. the note prior to the proposition) this gives

$$2^{-k\theta} e^{\mathrm{i}2^k t_0} = 2^k\chi(2^k\cdot) * f_\theta(t_0) = \int_{\mathbb{R}} \chi(z)(f_\theta(t_0 - 2^{-k}z) - f_\theta(t_0))\,dz. \tag{0.0.31}$$

So in case f_θ were differentiable at t_0, $F(h) := \frac{1}{h}(f_\theta(t_0 + h) - f_\theta(t_0))$ would define a function in $C_b(\mathbb{R})$ for which $F(0) = f'(t_0)$, and the Majorised Convergence Theorem, with $|z\chi(z)|\sup_{\mathbb{R}}|F|$ as the majorant, would imply that

$$\begin{aligned} 2^{(1-\theta)k} e^{\mathrm{i}2^k t_0} = \int (-z)F(-2^{-k}z)\chi(z)\,dz \xrightarrow[k\to\infty]{} -f'(t_0)\int_{\mathbb{R}} z\chi(z)\,dz& \\ = -f'(t_0)\,\mathrm{i}\,\frac{d\overset{\wedge}{\chi}}{d\tau}(0)& \\ = 0.& \end{aligned} \tag{0.0.32}$$

Hence $1 - \theta < 0$ would hold; and this would contradict the assumption that $\theta \le 1$. □

By now this argument is of course of a classical nature, as the Majorised Convergence Theorem is from 1908, cf. [**Leb08**].

Bibliography

[Har16] G. H. Hardy, *Weierstrass's non-differentiable function*, Trans. Amer. Math. Soc. **17** (1916), no. 3, 301–325.

[Leb08] H. Lebesgue, *Sur la méthode de M. Goursat pour la résolution de l'équation de Fredholm*, Bull. Soc. Math. France **36** (1908), 3–19.

[McC53] J. McCarthy, *An everywhere continuous nowhere differentiable function*, Amer. Math. Monthly **60** (1953), 709.

[Sch66] L. Schwartz, *Théorie des distributions*, revised and enlarged ed., Hermann, Paris, 1966.

[Tak03] T. Takagi, *A simple example of the continous function without derivative*, Phys.-Math. Soc. Japan **1** (1903), 176–177, *The Collected Papers of Teiji Takagi*, S. Kuroda, Ed.,Iwanami (1973), 5-6.