Notes on Integration Theory

April 20, 2018

Jon Johnsen

ABSTRACT. The present set of lecture 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 family, $\mathbb{E}$, of subsets of X to the positive extended real numbers,

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

The first requirement on μ is that the domain of μ, which is $\mathbb{E}$, 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 ($\mathbb{E}$-)measurable 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

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

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

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

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

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.

As for operations on σ-algebras, one may for a family of given σ-algebras $\mathbb{E}_i$ in X, whereby i runs through an arbitrary index set I, consider the intersection of the family, namely

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

It is immediately seen that this 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. This is usually called $\sigma(\mathbb{D})$:

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 a classical exercise to see that $\mathbb{B}(X)=\sigma(\mathbb{F})$, when $\mathbb{F}$ denotes the system of closed sets in X. Indeed, the inclusions $\mathbb{F}\subset\sigma(\mathbb{G})$ and $\mathbb{G}\subset\sigma(\mathbb{F})$ are obvious; whence $\sigma(\mathbb{F})\subset\sigma(\mathbb{G})$ and $\sigma(\mathbb{G})\subset\sigma(\mathbb{F})$. Altogether $\sigma(\mathbb{G})=\sigma(\mathbb{F})$.

Especially the above applies 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*:

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

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

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 $B(0,1) = \{\, x \in \mathbb{R}^d \mid x_1^2 + \cdots + x_d^2 < 1 \,\}$ 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

$$I =]a_1, b_1] \times \cdots \times]a_d, b_d] = \bigcap_{n\in\mathbb{N}} \left( ]a_1, b_1 + \tfrac{1}{n}[\times \cdots \times]a_d, b_d + \tfrac{1}{n}[\right). \tag{1.2.6}$$

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 in $\mathbb{R}^d$ is a countable union of disjoint cubes in $\mathbb{I}_d$.*

PROOF. For $G \neq \emptyset$ we consider the cube $C_{k,p} \in \mathbb{I}_d$ consisting of the $x \in \mathbb{R}^d$ for which $k_i 2^{-p} < x_i \leq (k_i + 1)2^{-p}$ 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 $G \setminus (O_1 \cup \cdots \cup O_{p-1})$. 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. $\square$

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, using the convention $\arctan(\pm\infty) = \pm\frac{\pi}{2}$, a metric given by

$$d(x,y) = |\arctan x - \arctan y|. \tag{1.2.9}$$

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,

$$\bar{\mathbb{B}} = \sigma(\{\,]a, \infty] \mid a \in \mathbb{R} \,\}). \tag{1.2.10}$$

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, (IV) is based on the trick that there are pairwise disjoint $\mathbb{E}$-measurable sets

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

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 ascertaining 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

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

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 (1.2.10) 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*

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

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

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

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 $]a_1,b_1]\times\cdots\times]a_d,b_d]$ 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 measurable 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 Dirichlet 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

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

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

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

Similarly it is seen for every $a \in \mathbb{R}$ that $A = \big\{\, x \in X \;\big|\; \inf_n f_n(x) < a \,\big\} \in \mathbb{E}$, 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 $\big\{\, x \in X \;\big|\; \inf_n f_n(x) > a \,\big\}$ 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 $\liminf_n f_n$. □

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 continuous functions

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

converge pointwise to $f = 1_{]0,\infty[}$, which is discontinuous. 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. □

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. □

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
$$\begin{aligned}&\{\,x\in X \mid f(x)<g(x)\,\},\\ &\{\,x\in X \mid f(x)\leq g(x)\,\},\\ &\{\,x\in X \mid f(x)=g(x)\,\}.\end{aligned} \tag{2.3.6}$$
Indeed, for $\varphi=g-f$ the sets are equal to the preimages $\varphi^{-1}(]0,\infty[)$, $\varphi^{-1}([0,\infty[)$ and $\varphi^{-1}(\{0\})$, respectively; 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
$$\mathbb{E}_A=\{\,A\cap E \mid E\in\mathbb{E}\,\}. \tag{2.4.1}$$
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 measurable 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

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

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

First, by (2.4.4) the $\mathbb{E}$-measurability of f, g means that they are measurable as maps $X \to \bar{\mathbb{R}}$. Secondly, using this it is straightforward to verify from (1.2.10) that $\mathbb{E}$ contains the sets

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

So to verify that e.g. $A = \big\{\, x \in X \;\big|\; f(x) < g(x) \,\big\}$ 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 in (2.4.8) are treated analogously.

CHAPTER 3

The Lebesgue integral

EXAMPLE 3.0.5. The function $f(x) = \frac{\sin x}{1+x^2}$ belongs to $\mathscr{L}(\mathbb{R}, \mathbb{B}, m)$: It is continuous, hence a Borel function on $\mathbb{R}$. Introducing the auxiliary function $g(x) = \frac{1}{1+x^2}$ in $\mathscr{M}^+(\mathbb{R}, \mathbb{B})$ we have $\int_{\mathbb{R}} g\,dm = \lim_{n\to\infty} \int g 1_{]-n,n]}\,dm = \lim_{n\to\infty}(\arctan(n) - \arctan(-n)) = \pi < \infty$, and g is a majorant for f as

$$|f(x)| = |\sin x|\left|\frac{1}{1+x^2}\right| \leq 1 \cdot \frac{1}{1+x^2} = g(x). \tag{3.0.11}$$

Consequently $f(x) = \frac{\sin x}{1+x^2}$ is Lebesgue integrable on the real axis $\mathbb{R}$.

CHAPTER 4

Fourier transformation, the naive approach

CHAPTER 5

Uniqueness theorem for measures

In the formal construction of the Lebesgue measure m_d on $\mathbb{R}^d$, one arrives at the following result: for each Borel set $B \in \mathbb{B}_d$,

$$m_d(B) = \inf\{ \sum_{n\in\mathbb{N}} v_d(I_n) \mid B \subset \bigcup_{n\in\mathbb{N}} I_n, \quad \forall n\colon I_n \in \mathbb{I}_d \}. \tag{5.0.12}$$

in rough terms, this means that the Lebesgue measure of a Borel set B is the total volume of the "most economical" covering of B by standard intervals.

DEFINITION 5.0.6. A measure $\mu\colon \mathbb{E} \to [0,\infty]$ is said to be finite if $\mu(X) < \infty$, and it is called *σ-finite* when $X = \bigcup_{n\in\mathbb{N}} A_n$ for a sequence of sets $A_n \in \mathbb{E}$ satisfying $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$. The same terminology applies to the measure space $(X, \mathbb{E}, \mu)$.

In the affirmative case one can arrange, if practical, that $A_1 \subset A_2 \subset \dots$, for A_n can be replaced by $A'_n = A_1 \cup \dots \cup A_n$, as $\mu(A'_n) < \infty$. Or conversely, the A_n can be redefined so that they are pairwise disjoint.

A main example is the Lebesgue measure m_d on $\mathbb{R}^d$, which is σ-finite since obviously $\mathbb{R}^d = \bigcup_{n\in\mathbb{N}}[-n,n]^d$.

CHAPTER 6

Product measures

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

$$\int_X \Big(\int_Y f(x,y)\,d\nu(y) \Big)\,d\mu(x) = \int_{X\times Y} f\,d\mu\otimes\nu = \int_Y \Big(\int_X f(x,y)\,d\mu(x) \Big)\,d\nu(y). \quad (6.0.13)$$

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 $\mu\otimes\nu$.

6.1. Products of measure spaces

In the following we consider two measure spaces $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$. The task is then to introduce the derived measure space $(X\times Y,\mathbb{E}\otimes\mathbb{F},\mu\otimes\nu)$.

This is build on the Cartesian product $X\times Y = \{\,(x,y) \mid x\in X,\ y\in Y\,\}$. The product σ-algebra $\mathbb{E}\otimes\mathbb{F}$ is straightforward to discuss—but construction of the product measure $\mu\otimes\nu$ is non-trivial and requires σ-finiteness of X and Y.

6.1.1. Measurability on a Cartesian product. As a guidance for the Cartesian product $X\times Y$, we consider the two projections

$$\pi_1 : X\times Y \to X, \qquad \pi_2 : X\times Y \to Y. \quad (6.1.1)$$

If $\mathbb{G}$ is a given σ-algebra on $X\times Y$ which makes both π_1, π_2 measurable, then $\mathbb{G}$ contains $\pi_1^{-1}(A) = A\times Y$ and $\pi_2^{-1}(B) = X\times B$ for all $A\in\mathbb{E}$ and $B\in\mathbb{F}$, so

$$\pi_1^{-1}(A)\bigcap\pi_2^{-1}(B) = A\times B \in \mathbb{G}. \quad (6.1.2)$$

Therefore such a σ-algebra $\mathbb{G}$ must necessarily contain the collection of $\mathscr{R}$ of *measurable rectangles*,

$$\mathscr{R} = \{\,A\times B \mid A\in\mathbb{E},\ B\in\mathbb{F}\,\}. \quad (6.1.3)$$

Consequently $\sigma(\mathscr{R})\subset\mathbb{G}$ then holds. Conversely, it is seen from (6.1.2) that π_1, π_2 are in fact measurable with respect to $\sigma(\mathscr{R})$. This leads to

DEFINITION 6.1.1. The *product σ-algebra* $\mathbb{E}\otimes\mathbb{F}$ in the Cartesian product $X\times Y$ is the smallest making π_1, π_2 measurable. That is, $\mathbb{E}\otimes\mathbb{F} = \sigma(\mathscr{R})$.

Obviously there will in general be many sets in $\mathbb{E}\otimes\mathbb{F}$ that are not rectangles in $\mathscr{R}$.

There is one rule for measurability we need to add for the product set $X\times Y$. This concerns the *tensor product* $g\otimes h$, which is given by

$$(x,y)\mapsto g(x)h(y) \qquad \text{for } (x,y)\in X\times Y. \quad (6.1.4)$$

Hereby g and h are functions defined on X and Y, respectively, having values in the same set in $\{\mathbb{R},\bar{\mathbb{R}},\mathbb{C}\}$. E.g., when $A\subset X$, $B\subset Y$, then $1_A\otimes 1_B = 1_{A\times B}$.

The tensor product $g\otimes h$ inherits measurability thus:

LEMMA 6.1.2. *When g is $\mathbb{E}$-measurable on X and h is $\mathbb{F}$-measurable on Y, then the tensor product $g\otimes h$ is measurable with respect to the product σ-algebra $\mathbb{E}\otimes\mathbb{F}$ on $X\times Y$.*

PROOF. Using the definition, we may write $g \otimes h(x,y) = g(x) \cdot h(y) = g \otimes 1_Y(x,y) \cdot 1_X \otimes h(x,y)$ with the understanding that tensor products precede multiplication in the hierarchy. Hence it suffices to obtain measurability of the tensor products $g \otimes 1_Y$ and $1_X \otimes h$.

In the first case we obtain for every Borel set D in the codomain ($\mathbb{R}$, $\bar{\mathbb{R}}$ or $\mathbb{C}$),

$$(g \otimes 1_Y)^{-1}(D) = \{ (x,y) \mid g(x) \in D \} = g^{-1}(D) \times Y \in \mathbb{E} \otimes \mathbb{F}. \tag{6.1.5}$$

The second function $1_X \otimes h$ is proved measurable in an analogous way. □

The two sets X and Y identify with subsets of $X \times Y$ in multiple ways. In fact, it is convenient to introduce *injections* $j_x \colon Y \to X \times Y$ and $j_y \colon X \to X \times Y$ parametrised by the members $x \in X$ and $y \in Y$, respectively. These maps are given by

$$j_x(y) = (x,y), \qquad j_y(x) = (x,y). \tag{6.1.6}$$

Here each j_x is $\mathbb{F}$-$\mathbb{E} \otimes \mathbb{F}$-measurable, since for each generating set $A \times B \in \mathscr{R}$,

$$j_x^{-1}(A \times B) = \begin{cases} B & \text{in case } A \ni x, \\ \emptyset & \text{in case } A \not\ni x. \end{cases} \tag{6.1.7}$$

Similarly each j_y is $\mathbb{E}$-$\mathbb{E} \otimes \mathbb{F}$-measurable.

These injections are useful for the discussion of *sections* of sets $G \subset X \times Y$ and of functions $f \colon X \times Y \to \mathbb{R}$, say. Indeed, G has sections G_x and G^y parametrised by the $x \in X$ and $y \in Y$, namely

$$G_x = \{ y \in Y \mid (x,y) \in G \}, \qquad G^y = \{ x \in X \mid (x,y) \in G \}. \tag{6.1.8}$$

The function f gives rise to the two sections $f(x,\cdot) \colon Y \to \mathbb{R}$ and $f(\cdot,y) \colon X \to \mathbb{R}$, which are obtained by freezing the first and the second entry, respectively.

Such sections are highly compatible with product σ-algebra measurability:

PROPOSITION 6.1.3. *Let $x \in X$ and $y \in Y$ be fixed. Then $G_x \in \mathbb{E}$ and $G^y \in \mathbb{F}$ whenever $G \in \mathbb{E} \otimes \mathbb{F}$. For every $\mathbb{E} \otimes \mathbb{F}$-measurable function f (having values in $\mathbb{R}$, $\bar{\mathbb{R}}$ or $\mathbb{C}$) the section $f(x,\cdot)$ is $\mathbb{F}$-measurable and the section $f(\cdot,y)$ is $\mathbb{E}$-measurable.*

PROOF. The statements are immediate consequences of the four obvious formulas $G_x = j_x^{-1}(G)$ and $G^y = j_y^{-1}(G)$ and, respectively, $f(x,\cdot) = f \circ j_x$ and $f(\cdot,y) = f \circ j_y$. □

To give a main example, note that $\mathbb{R}^p \times \mathbb{R}^q$ identifies with $\mathbb{R}^{p+q}$ by interpreting (x,y) with $x = (x_1, \dots, x_p)$ and $y = (y_1, \dots, y_q)$ as the element $(x_1, \dots, x_p, y_1, \dots, y_q)$ in $\mathbb{R}^{p+q}$. Hereby $\mathbb{B}_p \otimes \mathbb{B}_q$ emerges as a candidate for a σ-algebra in $\mathbb{R}^{p+q}$:

PROPOSITION 6.1.4. *The Borel algebra $\mathbb{B}_{p+q}$ coincides with the product σ-algebra $\mathbb{B}_p \otimes \mathbb{B}_q$; that is $\mathbb{B}_{p+q} = \mathbb{B}_p \otimes \mathbb{B}_q$.*

PROOF. Setting $d = p+q$, the projections $\pi_1 \colon \mathbb{R}^d \to \mathbb{R}^p$ and $\pi_2 \colon \mathbb{R}^d \to \mathbb{R}^q$ are continuous, hence Borel maps. Since $\mathbb{B}_p \otimes \mathbb{B}_q$ is the smallest σ-algebra with respect to which they are both measurable, this yields $\mathbb{B}_p \otimes \mathbb{B}_q \subset \mathbb{B}_d$.

The converse inclusion $\mathbb{B}_d \subset \mathbb{B}_p \otimes \mathbb{B}_q$ results because any standard interval $I \in \mathbb{I}_d$ can be written $I = I_1 \times I_2$ for standard intervals $I_1 \in \mathbb{I}_p \subset \mathbb{B}_p$ and $I_2 \in \mathbb{I}_q \subset \mathbb{B}_q$; whence $\mathbb{B}_d = \sigma(\mathbb{I}_d) \subset \mathbb{B}_p \otimes \mathbb{B}_q$. □

As a corollary to the inclusion $\mathbb{B}_{p+q} \subset \mathbb{B}_p \otimes \mathbb{B}_q$ and Proposition 6.1.3 it should be observed that every section of a Borel set (or of a Borel function) in $\mathbb{R}^d$ is another Borel set (Borel function).

And as a corollary to the inclusion $\mathbb{B}_p \otimes \mathbb{B}_q \subset \mathbb{B}_{p+q}$ it is noted that $A \times B$ is a Borel set in $\mathbb{R}^{p+q}$ whenever A and B are Borel sets in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and that Borel functions g and h on $\mathbb{R}^p$ and $\mathbb{R}^q$ give rise to the Borel function $g \otimes h$ on $\mathbb{R}^{p+q}$.

6.1.2. Product measures. In the construction of the product measure $\mu\otimes\nu$ it is crucial to require that both μ and ν are σ-finite. But in this situation, it is easy to show that there is at most one measure on $\mathbb{E}\otimes\mathbb{F}$ giving the "product of the side lengths" on the measurable rectangles—this follows from the Uniqueness Theorem for Measures:

PROPOSITION 6.1.5. *When $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ are σ-finite measure spaces, then there is at most one measure $\pi\colon \mathbb{E}\otimes\mathbb{F}\to[0,\infty]$ satisfying*

$$\pi(A\times B)=\mu(A)\nu(B)\quad \textit{for all } A\in\mathbb{E},\ B\in\mathbb{F}. \tag{6.1.9}$$

In the affirmative case the measure space $(X\times Y,\mathbb{E}\otimes\mathbb{F},\pi)$ is also σ-finite.

PROOF. Suppose both ρ and π are two measures satisfying (6.1.9), that is,

$$\pi(A\times B)=\mu(A)\nu(B)=\rho(A\times B)\quad \text{for all } A\in\mathbb{E},\ B\in\mathbb{F}. \tag{6.1.10}$$

The measurable rectangles $\mathscr{R}=\{\,A\times B \mid A\in\mathbb{E},\ B\in\mathbb{F}\,\}$ is by definition a generating system for the product σ-algebra $\mathbb{E}\otimes\mathbb{F}$, and $\mathscr{R}$ is obviously stable under intersections as

$$(A\times B)\cap(A'\times B')=(A\cap A')\times(B\cap B')\in\mathscr{R}. \tag{6.1.11}$$

By the assumed σ-finiteness, there are sequences (A_n) and (B_n) in $\mathbb{E}$ and $\mathbb{F}$, respectively, such that $A_1\subset A_2\subset\dots$ and $X=\bigcup_{n\in\mathbb{N}}A_n$ while $\mu(A_n)<\infty$ for all n, and $B_1\subset B_2\subset\dots$ and $Y=\bigcup_{n\in\mathbb{N}}B_n$ while $\nu(B_n)<\infty$ for all n. Using this to set $K_n=A_n\times B_n$, we have $K_n\in\mathscr{R}\subset\mathbb{E}\otimes\mathbb{F}$ for $n\in\mathbb{N}$ and

$$K_1\subset K_2\subset\dots\subset\bigcup_{n\in\mathbb{N}}K_n=X\times Y. \tag{6.1.12}$$

Moreover, (6.1.10) yields that $\pi(K_n)=\mu(A_n)\nu(B_n)<\infty$ for all $n\in\mathbb{N}$, so the Uniqueness Theorem for Measures states that $\rho(G)=\pi(G)$ for all $G\in\mathbb{E}\otimes\mathbb{F}$; that is, $\rho=\pi$.

When such a measure π exists, it is seen at once from the above considerations on the K_n that π is σ-finite. □

Existence of a measure $\pi\colon\mathbb{E}\otimes\mathbb{F}\to[0,\infty]$ having property (6.1.9) is a more demanding matter. However, in principle the situation is simple inasmuch as one can write down the following expressions for its action on any set $\mathbb{G}\in\mathbb{E}\otimes\mathbb{F}$:

$$\pi(G)=\int_X \nu(G_x)\,d\mu(x)=\int_Y \mu(G^y)\,d\nu(y). \tag{6.1.13}$$

These identities are natural because they show that measures can be found by integration after "cutting into slices".

Whilst $\nu(G_x)\ge 0$ and $\mu(G^y)\ge 0$ are obvious, it is non-trivial that these functions are $\mathbb{E}$- and $\mathbb{F}$-measurable, which is what remains to make the above integrals defined. For simplicity we give the details for $\nu(G_x)$.

As a convenient notation we may introduce the function $\varphi_G(x)=\nu(G_x)$. It is straightforward to see that

$$\varphi_{A\times B}=\nu(B)1_A\qquad\text{for } A\times B\in\mathscr{R}, \tag{6.1.14}$$

$$\varphi_{\bigcup_{n\in\mathbb{N}}G_n}=\sum_{n\in\mathbb{N}}\varphi_{G_n}\qquad\text{for pairwise disjoint sets } G_n\in\mathbb{E}\otimes\mathbb{F}. \tag{6.1.15}$$

However, we shall utilise σ-classes to obtain the measurability:

LEMMA 6.1.6. *The function $\varphi_G(x)=\nu(G_x)$ is $\mathbb{E}$-measurable whenever $G\in\mathbb{E}\otimes\mathbb{F}$.*

PROOF. The lemma is first shown under the additional assumption that $\nu(Y)<\infty$. Here we see at once from (6.1.14) that there is an inclusion $\mathscr{R}\subset\mathbb{H}$, namely, the measurable rectangles in $\mathscr{R}$ all belong to the system of subsets

$$\mathbb{H}=\{\,G\in\mathbb{E}\otimes\mathbb{F}\mid \varphi_G\in\mathscr{M}^+(X,\mathbb{E})\,\}. \tag{6.1.16}$$

In particular $X \times Y \in \mathbb{H}$. And if $G_1, G_2, \dots \in \mathbb{H}$ are pairwise disjoint, then $G = \bigcup_{n\in\mathbb{N}} G_n$ also belongs to $\mathbb{H}$: when $\varphi_{G_n} \in \mathscr{M}^+(X,\mathbb{E})$ for $n \in \mathbb{N}$ the rules of calculus of measurable functions and formula (6.1.15) yield that $\varphi_G = \sum_n \varphi_{G_n} \in \mathscr{M}^+(X,\mathbb{E})$. Finally $G \in \mathbb{H}$ implies $\complement G \in \mathbb{H}$, for (6.1.14) and (6.1.15) entail that

$$\varphi_G + \varphi_{\complement G} = \varphi_{X\times Y} = \nu(Y) < \infty; \tag{6.1.17}$$

so by subtraction of real numbers we have $\varphi_{\complement G} = \nu(Y) - \varphi_G$, where the left-hand side is positive and the right-hand side is $\mathbb{E}$-measurable. Altogether $\mathbb{H}$ is a σ-class.

From this it follows that $\mathbb{H} = \mathbb{E}\otimes\mathbb{F}$ (as desired), for the generating set $\mathscr{R}$ is stable under intersections (cf. the proof of Proposition 6.1.5), which by the fundamental lemma gives

$$\mathbb{D}(\mathscr{R}) \subset \mathbb{H} \subset \mathbb{E}\otimes\mathbb{F} = \sigma(\mathscr{R}) = \mathbb{D}(\mathscr{R}). \tag{6.1.18}$$

In general, when ν is merely σ-finite, the above is exploited thus: first we fix sets $B_1 \subset B_2 \subset \dots$ in $\mathbb{F}$ such that $Y = \bigcup_n B_n$ and $\nu(B_n) < \infty$ for all n; then we introduce the necessarily finite measures $\nu_n(B) = \nu(B \cap B_n)$ for $n \in \mathbb{N}$. So for any given $G \in \mathbb{E}\otimes\mathbb{F}$, the above yields $\varphi_G^{(n)}(x) = \nu_n(G_x) \in \mathscr{M}^+(X,\mathbb{E})$. Then a general property of measures gives

$$\varphi_G^{(n)}(x) = \nu(G_x \cap B_n) \nearrow \nu(G_x) = \varphi_G(x), \tag{6.1.19}$$

which implies the measurability of φ_G. Hence $\varphi_G \in \mathscr{M}^+(X,\mathbb{E})$. □

Thus prepared, we arrive at the main theorem on product measures:

THEOREM 6.1.7. *When $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ are σ-finite measure spaces, then there is a uniquely determined measure $\pi\colon \mathbb{E}\otimes\mathbb{F} \to [0,\infty]$ for which*

$$\pi(A\times B) = \mu(A)\nu(B) \quad \textit{for all } A \in \mathbb{E},\ B \in \mathbb{F}. \tag{6.1.20}$$

The product measure $\pi = \mu\otimes\nu$ is given by the formulas in (6.1.13)*, and the measure space $(X\times Y, \mathbb{E}\otimes\mathbb{F}, \mu\otimes\nu)$ is σ-finite.*

PROOF. From the measurability in Lemma 6.1.6 it is now clear that the first integral in (6.1.13) makes sense; so does the second as it is only notation that differs (the order of $\mathbb{E}$ and $\mathbb{F}$ in $\mathbb{E}\otimes\mathbb{F}$ is immaterial). It therefore suffices to show that the first integral defines a measure π fulfilling (6.1.20); the argument then shows the same thing for the second integral, and they give the same measure because of the uniqueness result in Proposition 6.1.5.

Now, by (6.1.14) we have $\pi(A\times B) = \int_X \varphi_{A\times B}\,d\mu = \nu(B)\int_X 1_A\,d\mu = \mu(A)\nu(B)$ for all $A\times B \in \mathscr{R}$, so the map π satisfies (6.1.20). Hence $\pi(\emptyset) = \pi(\emptyset\times\emptyset) = \mu(\emptyset)\nu(\emptyset) = 0$.

Finally, whenever we have pairwise disjoint sets $G_n \in \mathbb{E}\otimes\mathbb{F}$, then (6.1.15) gives

$$\pi\Big(\bigcup_{n\in\mathbb{N}} G_n\Big) = \int_X \varphi_{\bigcup_{n\in\mathbb{N}} G_n}\,d\mu = \sum_{n\in\mathbb{N}} \int_X \varphi_{G_n}\,d\mu = \sum_{n\in\mathbb{N}} \pi(G_n). \tag{6.1.21}$$

Therefore the map π is a measure on the product σ-algebra $\mathbb{E}\otimes\mathbb{F}$, as claimed. □

The general result above gives the following reassuring result for the Euclidean spaces:

THEOREM 6.1.8. *By identifying $\mathbb{R}^p\times\mathbb{R}^q$ with $\mathbb{R}^{p+q}$, the Lebesgue measures m_p, m_q and m_{p+q} defined on the Borel algebras $\mathbb{B}_p$, $\mathbb{B}_q$ and $\mathbb{B}_{p+q}$, respectively, satisfy*

$$m_p\otimes m_q = m_{p+q}. \tag{6.1.22}$$

PROOF. Since $\mathbb{B}_p\otimes\mathbb{B}_q = \mathbb{B}_{p+q}$ was shown in Proposition 6.1.4, and since m_p and m_q are σ-finite, the product measure $m_p\otimes m_q$ is defined on the Borel algebra $\mathbb{B}_{p+q}$. Its value is $m_p(I)m_q(J) = v_p(I)v_q(J)$ for every standard interval $I\times J = I_1\times\dots I_p\times J_1\times\dots J_q$ in $\mathbb{I}_{p+q}$; that is, the value is the product of the side lengths. But this property characterises the Lebesgue measure m_{p+q} on $\mathbb{B}_{p+q}$. □

6.2. 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 below Theorem 6.2.2 on reiterated integration of functions $f(x,y)$ in $\mathscr{M}^+$.

However, it is instructive first to study a typical example:

EXAMPLE 6.2.1. The function $f(x,y) = \frac{1}{1+(xy)^2}$ is continuous, hence in $\mathscr{M}^+(\mathbb{R}^2, \mathbb{B}_2)$. By integrating y out, one obtains a function $I(x)$ given by

$$\begin{aligned} I(x) &= \frac{1}{x}\int_{\mathbb{R}} \frac{x}{1+(xy)^2}\,dy = \frac{\pi}{|x|}, \qquad \text{for } x \neq 0,\\ I(0) &= \int_{\mathbb{R}} 1\,dx = m(\mathbb{R}) = \infty. \end{aligned} \tag{6.2.1}$$

So obviously integration theory contains functions that attain the value ∞ for good reasons.

The fact that $I(x)$ in the above example belongs to $\mathscr{M}^+(\mathbb{R}, \mathbb{B})$ (even though $I(0) = \infty$) illustrates part (i) in the following general result:

THEOREM 6.2.2 (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)
$$\int_X \Big(\int_Y f(x,y)\,d\nu(y)\Big)\,d\mu(x) = \int_{X\times Y} f\,d\mu \otimes \nu. \tag{6.2.2}$$

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

PROOF. As f is assumed $\mathbb{E} \otimes \mathbb{F}$-measurable, the section $f(x,\cdot)$ is in $\mathscr{M}^+(Y, \mathbb{F})$ for every $x \in X$. Therefore $g(x) = \int_Y f(x,\cdot)\,d\nu$ is a well-defined function on X. To show its measurability and the formula for its integral, we proceed in three steps.

In case $f(x,y) = 1_G(x,y)$ for some $G \in \mathbb{E} \otimes \mathbb{F}$, the claims are essentially shown previously: $1_G(x,\cdot) = 1_{G_x}$ for $x \in X$, whence $g(x) = \int_Y 1_G(x,\cdot)\,d\nu = \int_Y 1_{G_x}\,d\nu = \nu(G_x)$ is $\mathbb{E}$-measurable according to Lemma 6.1.6; cf. (i). Moreover, its integral is by (6.1.13) the product measure of G; that is, $\int g\,d\mu = \mu \otimes \nu(G) = \int 1_G\,d\mu \otimes \nu$, obtaining (ii).

Now the two properties extend to every simple function $f = c_1 1_{G_1} + \cdots + c_n 1_{G_n}$ on $X \times Y$ such that $0 < c_j < \infty$ and $G_j \in \mathbb{E} \otimes \mathbb{F}$ for each $j \in \{1, \dots, n\}$. For it is straightforward to see that the set of functions in $\mathscr{M}^+(X \times Y, \mathbb{E} \otimes \mathbb{F})$ satisfying (i) and (ii) is stable under addition and under multiplication by scalars in $\,]0, \infty[\,$.

Finally, to each $f \in \mathscr{M}^+(X \times Y, \mathbb{E} \otimes \mathbb{F})$ there is a sequence of simple $\mathbb{E} \otimes \mathbb{F}$-measurable functions f_n such that $f_n \nearrow f$. By monotone convergence and the already proved part of the theorem,

$$\int_{X\times Y} f\,d\mu \otimes \nu = \lim_{n\to\infty} \int_{X\times Y} f_n\,d\mu \otimes \nu = \lim_{n\to\infty} \int_X \big(\int_Y f_n(x,\cdot)\,d\nu\big)\,d\mu. \tag{6.2.3}$$

Since $f_n(x,\cdot) \nearrow f(x,\cdot)$ holds for the sections for every $x \in X$, a second application of the monotone convergence theorem yields

$$g_n(x) = \int f_n(x,\cdot)\,d\nu \nearrow \int f(x,\cdot)\,d\nu = g(x). \tag{6.2.4}$$

Here the g_n are $\mathbb{E}$-measurable, as (i) is shown for the simple functions f_n, so this yields that g is $\mathbb{E}$-measurable; i.e. (i) holds for f. A third use of monotone convergence now gives

$$\lim_{n\to\infty} \int_X g_n\,d\mu = \int_X g\,d\mu. \tag{6.2.5}$$

Insertion of this into the above shows that also (ii) is valid for $f \in \mathscr{M}^+(X \times Y, \mathbb{E} \otimes \mathbb{F})$. $\square$

It is noteworthy that the theorem holds under natural assumptions, inasmuch as no other requirement has been made on $f(x,y)$ than it should be positive and measurable. From the contents of the theorem it is seen then that both reiterated integrals in formula (6.0.13) make sense and are equal to $\int_{X\times Y} f\,d\mu\otimes\nu$. The necessesity of the σ-finiteness of X, Y is seen in an exercise.

EXAMPLE 6.2.3. Tonelli's theorem allows us to calculate

$$I=\int_0^1(\int_y^1 \frac{1}{1+x^4}\,dx)\,dy=\frac{\pi}{8}. \tag{6.2.6}$$

Indeed, via the triangle $T=\{\,(x,y)\in\mathbb{R}^2\mid 0\le y\le x\le 1\,\}$ we may consider the function

$$f(x,y)=1_T(x,y)\frac{1}{1+x^4}\in\mathscr{M}^+(\mathbb{R}^2,\mathbb{B}_2). \tag{6.2.7}$$

One section of this is $f(\cdot,y)=1_{[0,1]}(y)1_{[y,1]}(\cdot)\frac{1}{1+(\cdot)^4}$, and by Tonelli's theorem we may integrate x out (for fixed $y\in\mathbb{R}$) and follow up by integration with respect to y. This yields

$$\int_{\mathbb{R}^2} f\,dm_2=\int_{\mathbb{R}}1_{[0,1]}(y)(\int_{\mathbb{R}}1_{[y,1]}(x)\frac{1}{1+x^4}\,dx)\,dy=I. \tag{6.2.8}$$

But the other section of f is given by $f(x,\cdot)=1_{[0,1]}(x)1_{[0,x]}(\cdot)\frac{1}{1+x^4}$, so by integrating in the opposite order, as we may (cf. the last part of Tonelli's theorem), we get

$$\begin{aligned}\int_{\mathbb{R}^2} f\,dm_2&=\int_{\mathbb{R}}1_{[0,1]}(x)\big(\frac{1}{1+x^4}\int_{\mathbb{R}}1_{[0,x]}(y)\,dy\big)\,dx=\int_0^1\frac{1}{1+x^4}x\,dx\\&=\frac12\int_0^1\arctan'(x^2)(x^2)'\,dx=\frac12\left[\arctan(x^2)\right]_0^1=\frac{\pi}{8}.\end{aligned} \tag{6.2.9}$$

It follows that $I=\frac{\pi}{8}$ as claimed.

EXAMPLE 6.2.4 (Integration in polar coordinates). Each point $(x,y)\in\mathbb{R}^2$ has polar coordinates (r,θ), where $r=\sqrt{x^2+y^2}$ and $\theta\in\mathbb{R}$ fulfils $(x,y)=r(\cos\theta,\sin\theta)$. Since points $(x,y)\ne(0,0)$ only have θ determined modulo 2π, we may introduce the map

$$\begin{aligned}\varphi\colon &X\to Y,\\ &X=]-\pi,\pi[\times\mathbb{R}_+,\\ &Y=\mathbb{R}^2\setminus\{\,(x,0)\mid x\le 0\,\}.\end{aligned} \tag{6.2.10}$$

This is a bijection with both φ, φ^{-1} belonging to C^∞, so Jacobi's transformation theorem applies. Now

$$\det D\varphi(r,\theta)=\begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix}=r, \tag{6.2.11}$$

so since the excluded halfline is a nullset in the plane, we get the general formula for a positive or integrable Borel function $f\colon\mathbb{R}^2\to\mathbb{R}$, using the theorem of Tonelli or Fubini,

$$\begin{aligned}\int_{\mathbb{R}^2} f\,dm_2=\int_Y f(x,y)\,dm_2(x,y)&=\int_X f(r\cos\theta,r\sin\theta)r\,dm_2(r,\theta)\\&=\int_0^\infty\Big(\int_{-\pi}^{\pi}f(r\cos\theta,r\sin\theta)\,d\theta\Big)r\,dr.\end{aligned} \tag{6.2.12}$$

As an example of this one has, on the one hand, that

$$\int_{\mathbb{R}^2}e^{-x^2-y^2}\,dm_2(x,y)=\int_0^\infty\Big(\int_{-\pi}^{\pi}e^{-r^2}\,d\theta\Big)r\,dr=\pi\int_0^\infty 2re^{-r^2}\,dr=\pi. \tag{6.2.13}$$

On the other hand a direct application of Tonelli's theorem yields

$$\int_{\mathbb{R}^2}e^{-x^2-y^2}\,dm_2(x,y)=\int_{\mathbb{R}}(\int_{\mathbb{R}}e^{-x^2-y^2}\,dy)\,dx=(\int_{\mathbb{R}}e^{-x^2}\,dx)^2. \tag{6.2.14}$$

By taking the square root of these identities, one arrives at the well-known fact that

$$\int_{\mathbb{R}} e^{-x^2}\,dx = \sqrt{\pi}. \tag{6.2.15}$$

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

$$f\colon X\times Y\to\mathbb{C}. \tag{6.2.16}$$

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 that by integrating one variable out, one obtains a measurable function on the set where this integration is well defined:

LEMMA 6.2.5. *Let $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ be two σ-finite measure spaces, and let $f\colon X\times Y\to\mathbb{C}$ be $\mathbb{E}\otimes\mathbb{F}$-measurable. Then there is a measurable set $A\subset X$ given by*

$$A=\big\{\,x\in X \,\big|\, f(x,\cdot)\in\mathscr{L}(Y,\mathbb{F},\nu)\,\big\}, \tag{6.2.17}$$

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

$$g(x)=\int_Y f(x,\cdot)\,d\nu, \qquad x\in A, \tag{6.2.18}$$

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

REMARK 6.2.6. For the function in Example 6.2.1 it is clear that $A=\,]-\infty,0[\,\cup\,]0,\infty[\,$.

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

$$p(x)=\int_Y f^+(x,\cdot)\,d\nu, \qquad n(x)=\int_Y f^-(x,\cdot)\,d\nu. \tag{6.2.19}$$

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

$$A=\big\{\,x\in X \,\big|\, p(x)<\infty\,\big\}\bigcap\big\{\,x\in X \,\big|\, n(x)<\infty\,\big\}. \tag{6.2.20}$$

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 $\mathbb{E}_A$-measurable by the rules of calculus for measurable functions. □

The content of this lemma is of some independent interest. But it is mainly 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 6.2.7 (Fubini). *Let $(X,\mathbb{E},\mu)$ and $(Y,\mathbb{F},\nu)$ be two σ-finite measure spaces. For every function $f\colon X\times Y\to\mathbb{C}$ 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)

$$\int_{X\times Y} f\,d\mu\otimes\nu=\int_A\Big(\int_Y f(x,y)\,d\nu(y)\Big)\,d\mu(x). \tag{6.2.21}$$

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

PROOF. There is only something to show if $\mu(X)>0$, $\nu(Y)>0$; which therefore is assumed.

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 (6.2.19) gives that

$$\int_X p\,d\mu = \int_{X\times Y} f^+\,d\mu\otimes\nu, \qquad \int_X n\,d\mu = \int_{X\times Y} f^-\,d\mu\otimes\nu \tag{6.2.22}$$

These integrals are both finite, since $f\in\mathscr{L}(\mu\otimes\nu)$. Since p, $n\in\mathscr{M}^+$, cf. the proof of Lemma 6.2.5, this yields $\mu(P)=0=\mu(N)$ for the two sets $P=\{x\in X\mid p(x)=\infty\}$ and $N=\{x\in X\mid n(x)=\infty\}$. Now, $P\cup N$ is a measurable null-set with its complement $A=X\setminus(P\cup N)$ equal to the set of x for which $f(x,\cdot)$ is ν-integrable. That is, $A\in\mathbb{E}$ 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, which belong to $\mathscr{L}(A,\mu)$ in view of the above finiteness; and so is $g=p|_A-n|_A$, cf. (ii). Using the definition of the Lebesgue integral one now finds

$$\int_A g\,d\mu = \int_X p\,d\mu - \int_X n\,d\mu = \int_{X\times Y} f\,d\mu\otimes\nu, \tag{6.2.23}$$

because of (6.2.22). 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 in $\mathbb{E}$ according to (i); cf. (2.4.2). With this understanding it is usually sufficient to abbreviate the result in Fubini's theorem to the simpler formula:

$$\int_X\Big(\int_Y f(x,\cdot)\,d\nu\Big)\,d\mu = \int_{X\times Y} f\,d\mu\otimes\nu = \int_Y\Big(\int_X f(\cdot,y)\,d\mu\Big)\,d\nu. \tag{6.2.24}$$

However, in special circumstances one may need the full statement in Theorem 6.2.7.

Often the theorems of Tonelli and Fubini are applied in succession, as we now explain:

EXAMPLE 6.2.8. When $f\in\mathscr{L}(X,\mu)$ and $g\in\mathscr{L}(Y,\nu)$ for σ-finite measures, then $f\otimes g\in\mathscr{L}(\mu\otimes\nu)$ and

$$\int_{X\otimes Y} f\otimes g\,d\mu\otimes\nu = \int_X f\,d\mu\int_Y g\,d\nu. \tag{6.2.25}$$

Indeed, $f\otimes g$ and $|f\otimes g|$ are $\mathbb{E}\otimes\mathbb{F}$-measurable, so by Tonelli's theorem,

$$\int_{X\otimes Y}|f|\otimes|g|\,d\mu\otimes\nu = \int_X|f|(\int_Y|g|\,d\nu)\,d\mu = \int_X|f|\,d\mu\int_Y|g|\,d\nu<\infty. \tag{6.2.26}$$

This shows that $f\otimes g\in\mathscr{L}(\mu\otimes\nu)$. Hence Fubini's theorem applies to $f\otimes g$ that once again gives (6.2.26), only with $|f|$, $|g|$ replaced by the functions f, g themselves; which shows (6.2.25). Hereby also the remark in (6.2.24) is invoked.

EXAMPLE 6.2.9. Let $A\subset\mathbb{R}^d$ be a Borel set and $f,g\colon A\to\mathbb{R}$ be to Borel functions such that $f\le g$ on A. The "sandwich set"

$$G=\big\{\,(x,y)\in A\times\mathbb{R}\;\big|\;f(x)\le y\le g(x)\,\big\} \tag{6.2.27}$$

is then Borel measurable, i.e. $G\in\mathbb{B}_{d+1}$ and its Lebesgue measure is given by

$$m_{d+1}(G)=\int_A(g(x)-f(x))\,dx \tag{6.2.28}$$

This may be shown via the auxiliary functions $\varphi,\psi\colon\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}$,

$$\begin{aligned}\varphi(x,y)&=1_{A\times\mathbb{R}}(x,y)(g(x)-y)-1_{(\complement A)\times\mathbb{R}}(x,y),\\ \psi(x,y)&=1_{A\times\mathbb{R}}(x,y)(y-f(x))-1_{(\complement A)\times\mathbb{R}}(x,y).\end{aligned} \tag{6.2.29}$$

Obviously

$$G = \varphi^{-1}([0,\infty[) \cap \psi^{-1}([0,\infty[), \tag{6.2.30}$$

so $G \in \mathbb{B}_{d+1}$ follows if $\varphi^{-1}([0,\infty[)$ and $\psi^{-1}([0,\infty[)$ belong to $\mathbb{B}_{d+1}$. By the rules of measurability this boils down, for φ, to the fact that $g(x) - y$ is $\mathbb{B}_{A\times\mathbb{R}}$-measurable; which may be seen for y by restricting the $\mathbb{B}_{d+1}$-measurable projection π_2 to the Borel set $A \times \mathbb{R}$, and by regarding $g(x)$ as the composition of g and π_1, which is $\mathbb{B}_{A\times\mathbb{R}}$-$\mathbb{B}_A$-measurable. Similarly one can show that ψ is measurable.

Exploiting the construction of the product measure $m_{d+1} = m_d \otimes m$, we now get

$$\int_{\mathbb{R}^d} m(G_x)\, dm_d(x) = \int_{\mathbb{R}^d} 1_A(x) m([f(x), g(x)])\, dm_d(x), \tag{6.2.31}$$

and we can identify the left- and right-hand sides with those in (6.2.28), respectively.

Moreover, when $\Phi \colon A \to \mathbb{R}$ is an integrable or a positive Borel function, then the section $(1_G\Phi)_x$ is given by

$$y \mapsto \begin{cases} 1_{[f(x),g(x)]}(y)\Phi(x,y) & \text{for } x \in A, \\ 0 & \text{for } x \in \mathbb{R}^d \setminus A. \end{cases} \tag{6.2.32}$$

so the theorems of Tonelli and Fubini give the well-known formula

$$\int_G \Phi\, dm_{d+1} = \int_A \Big(\int_{f(x)}^{g(x)} \Phi(x,y)\, dy \Big)\, dx. \tag{6.2.33}$$

CHAPTER 7

Classical inequalities

In the following a measure space $(X, \mathbb{E}, \mu)$ is thought to be given such that $X \neq \emptyset$.

7.1. Seminormed Lebesgue spaces

It is a classical task to quantify different *degrees* of integrability of functions f defined on a measure space $(X, \mathbb{E}, \mu)$. This is done via the introduction of the *Lebesgue* space $\mathscr{L}_p(\mu) = \mathscr{L}_p(X, \mathbb{E}, \mu)$ consisting of the measurable functions $f \colon X \to \mathbb{C}$, which for some given $p \in]0, \infty[$ satisfy

$$\int |f|^p \, d\mu < \infty. \tag{7.1.1}$$

The motivation for this is easy in case μ is a probability measure: then $f \in \mathscr{L}_1(\mu)$ if and only if the stochastic variable f has a mean value, whilst $g \in \mathscr{L}_2(\mu)$ holds precisely when the stochastic variable g has variance. In mathematical analysis (the variant) $L_2(\mu)$ is a main source of so-called Hilbert spaces, which are decisive in many investigations.

Obviously the case $p = 1$ just gives back the set $\mathscr{L}(\mu) = \mathscr{L}(X, \mathbb{E}, \mu)$ of μ-integrable functions on X; i.e. $\mathscr{L}_1(\mu) = \mathscr{L}(\mu)$. In general $\mathscr{L}_p(\mu)$ is said to consist of the functions f that are p times integrable on X, though they are simply called quadratically integrable on X for $p = 2$.

In case μ is the counting measure on an index set J, the above amounts to the set of p times summable families $(a_j)_{j \in J}$, which are defined by the condition $\sum_{j \in J} |a_j|^p < \infty$. The set of such families is denoted by $\ell_p(J)$, and for $J = \mathbb{N}$ this is abbreviated to ℓ_p.

It is elementary to see that the set $\mathscr{L}_p(\mu)$ has the structure of a vector space:

LEMMA 7.1.1. *The set $\mathscr{L}_p(X, \mathbb{E}, \mu)$ is a vector space for $0 < p < \infty$.*

PROOF. Given functions $f, g \in \mathscr{L}_p(\mu)$ and a scalar $c \in \mathbb{C}$, it is clear from the calculus that cf and $f + g$ are $\mathbb{E}$-measurable; we have $\int |cf|^p \, d\mu = |c|^p \int |f|^p \, d\mu < \infty$ and

$$|f+g|^p \le (|f|+|g|)^p \le 2^p(|f| \vee |g|)^p \le 2^p(|f|^p \vee |g|^p) \le 2^p(|f|^p + |g|^p), \tag{7.1.2}$$

which by integration of both sides yields that $\int |f+g|^p \, d\mu < \infty$. □

In order to focus on the simplest cases wee now specialise to the the cases in which $1 \le p < \infty$.

It would then be desirable to show that the vector space $\mathscr{L}_p(\mu)$ has a norm given by the expression

$$\|f\|_p = \Big(\int |f|^p \, d\mu \Big)^{\frac{1}{p}}. \tag{7.1.3}$$

This would be the case if, for all $f, g \in \mathscr{L}_p(\mu)$ and $c \in \mathbb{C}$,

$$\begin{aligned} &\text{(i)} && \|f\|_p \ge 0, \text{ with equality only for } f = 0, \\ &\text{(ii)} && \|cf\|_p = |c| \|f\|_p, \\ &\text{(iii)} && \|f+g\|_p \le \|f\|_p + \|g\|_p. \end{aligned} \tag{7.1.4}$$

However, $\|\cdot\|_p$ is in general only a *seminorm* on $\mathscr{L}_p(\mu)$, which means that $\|f\|_p = 0$ may hold for some $f \neq 0$; cf. (i). Indeed, although $f \equiv 0$ is the zero vector in $\mathscr{L}_p$, we have

$$\|f\|_p = 0 \iff \int |f|^p \, d\mu = 0 \iff |f|^p = 0 \ \mu\text{-a.e.} \iff f = 0 \ \mu\text{-a.e.} \tag{7.1.5}$$

So it is only if $\emptyset$ is the only null set that $\|\cdot\|_p$ is a norm. This includes the case of the counting measure on X, so the above spaces $\ell_p(J)$ are normed by the expression

$$\|a_j\|_p = \Big(\sum_{j\in J} |a_j|^p\Big)^{\frac{1}{p}}. \tag{7.1.6}$$

Of course the conditions (ii) and (iii) must also be rigorously verified. But (ii) follows easily since $\|cf\|_p = (\int |c|^p |f|^p \, d\mu)^{1/p} = |c| (\int |f|^p \, d\mu)^{1/p}$ for every scalar $c \in \mathbb{C}$. However, the trangle inequality in (iii) means that all $f, g \in \mathscr{L}_p(\mu)$ should fulfill

$$\Big(\int |f+g|^p \, d\mu\Big)^{\frac{1}{p}} \leq \Big(\int |f|^p \, d\mu\Big)^{\frac{1}{p}} + \Big(\int |g|^p \, d\mu\Big)^{\frac{1}{p}}. \tag{7.1.7}$$

This is known as *Minkowski's inequality*, and while it is trivial to obtain for $p = 1$, it is rather more demanding for $p > 1$, in which case it is a main result in integration theory.

For the proof of Minkowski's inequality we shall obtain two other inequalities, which are of independent interest. They are both related to the notion of *dual exponents*. These are numbers p, q in $]1, \infty[$ that satisfy the basic relation

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{7.1.8}$$

This can be written in many equivalent ways, as e.g.

$$\frac{1}{p} + \frac{1}{q} = 1 \iff p + q = pq \iff p = (p-1)q \iff q = \frac{p}{p-1}. \tag{7.1.9}$$

When $p > 1$ is fixed, then it is customary to designate the $q > 1$ fulfilling (7.1.8) as the conjugate or dual exponent to p, and to denote it by p' instead of q.

LEMMA 7.1.2 (Young's inequality). *When $p > 1$ and $q > 1$ are dual exponents, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$uv \leq \frac{u^p}{p} + \frac{v^q}{q} \qquad \text{for all } u \geq 0,\ v \geq 0. \tag{7.1.10}$$

Equality holds if and only if the numerators are equal, i.e. when $u^p = v^q$.

PROOF. The auxiliary function $F(u,v) = \frac{u^p}{p} + \frac{v^q}{q} - uv$ satisfies, for any fixed $v \geq 0$, that $F(0,v) \geq 0$ whilst $F(u,v) = \frac{v^q}{q} + u(\frac{u^{p-1}}{p} - v) \to \infty$ for $u \to \infty$. Now $\frac{\partial F}{\partial u} = u^{p-1} - v$ has the uniquely given zero $u_0 = v^{1/(p-1)}$, where $u_0^p = v^q$, so F attains its minimum at (u_0, v) with $F(u_0, v) = v^q(\frac{1}{p} + \frac{1}{q} - 1) = 0$. Hence $F(u,v) \geq 0$ whenever $u \geq 0$, $v \geq 0$. □

A standard application of Young's inequality gives the next result, which is Hölder's inquality. But first we take the opportunity to free the discussion from functions known a priori to belong to $\mathscr{L}_p(\mu)$. Indeed, the integral in (7.1.1) makes sense for every measurable function $f \colon X \to \mathbb{C}$, and it is therefore customary to extend the definition of $\|\cdot\|_p$ as

$$\|f\|_p = \begin{cases} (\int |f|^p \, d\mu)^{\frac{1}{p}} & \text{for } f \in \mathscr{L}_p(\mu), \\ \infty & \text{for } f \notin \mathscr{L}_p(\mu). \end{cases} \tag{7.1.11}$$

The second line is of a consequence of the first with the common convention that $\infty^p = \infty$ for any $p > 0$; which we adopt throughout.

THEOREM 7.1.3 (Hölder's inequality). *When $p>1$, $q>1$ are dual exponents, that is $\frac{1}{p}+\frac{1}{q}=1$, and $f,g\colon X\to\mathbb{C}$ are $\mathbb{E}$-measurable, then*

$$\|fg\|_1=\int|fg|\,d\mu\le\|f\|_p\|g\|_q. \tag{7.1.12}$$

Consequently, the product fg is μ-integrable when $f\in\mathscr{L}_p(\mu)$ and $g\in\mathscr{L}_q(\mu)$.

PROOF. By the calculus fg is measurable, and there is nothing to show if $\|f\|_p=0$ or $\|g\|_q=0$, for then $fg=0$ μ-a.e. In the remaining cases we may derive from Young's inequality (7.1.10) that

$$\frac{|f|}{\|f\|_p}\frac{|g|}{\|g\|_q}\le\frac{1}{p}\frac{|f|^p}{\|f\|_p^p}+\frac{1}{q}\frac{|g|^q}{\|g\|_q^q}. \tag{7.1.13}$$

By integrating both sides we get $\frac{1}{\|f\|_p\|g\|_q}\int|fg|\,d\mu\le\frac{1}{p}+\frac{1}{q}=1$, whence (7.1.12) follows. When moreover $f\in\mathscr{L}_p(\mu)$ and $g\in\mathscr{L}_q(\mu)$, then the right-hand side of (7.1.12) is finite, so that $fg\in\mathscr{L}_1(\mu)$. □

In case $p=2=q$, the above result is usually called Cauchy–Schwarz' inequality. It is often useful in the following form:

COROLLARY 7.1.4. *When $f,g\in\mathscr{L}_2(\mu)$, then $fg\in\mathscr{L}(\mu)$ and*

$$|\int fg\,d\mu|\le\|fg\|_1=\int|fg|\,d\mu\le\|f\|_2\|g\|_2. \tag{7.1.14}$$

We are now ready to prove the inequality attributed to Minkowski:

THEOREM 7.1.5 (Minkowski's inequality). *If $1\le p<\infty$, then all $\mathbb{E}$-measurable functions $f,g\colon X\to\mathbb{C}$ fulfill*

$$\|f+g\|_p\le\|f\|_p+\|g\|_p. \tag{7.1.15}$$

Consequently $f\mapsto\|f\|_p$ is a seminorm on the functions f in $\mathscr{L}_p(\mu)$ for such p.

PROOF. The problem is to show the inequality for $p>1$ (as $p=1$ is trivial), but it suffices to cover the cases where $0<\|f+g\|_p<\infty$, as for $\|f+g\|_p=0$ the inequality is trivial, while for $\|f+g\|_p=\infty$ also the right-hand side of (7.1.15) is infinity (both functions cannot be in $\mathscr{L}_p$ then, cf. Lemma 7.1.1).

A constructive use of the reduction to $p>1$ is to make a splitting of the integrand:

$$\|f+g\|_p^p=\int|f+g||f+g|^{p-1}\,d\mu\le\int|f||f+g|^{p-1}\,d\mu+\int|g||f+g|^{p-1}\,d\mu. \tag{7.1.16}$$

As $|f+g|^{p-1}$ is $\mathbb{E}$-measurable, we get for the (prospective) seminorm corresponding to the dual exponent q from (7.1.9) that

$$\||f+g|^{p-1}\|_q=\Big(\int|f+g|^p\,d\mu\Big)^{\frac{1}{q}}=\|f+g\|_p^{p/q}=\|f+g\|_p^{p-1}. \tag{7.1.17}$$

Hölder's inequality therefore gives

$$\int|f||f+g|^{p-1}\,d\mu\le\|f\|_p\|f+g\|_p^{p-1}. \tag{7.1.18}$$

Since f and g are arbitrary measurable functions, they can switch roles; whence the same inequality with $|f|$ replaced by $|g|$ is obtained. By insertion into (7.1.16) this gives

$$\|f+g\|_p^p\le\big(\|f\|_p+\|g\|_p\big)\|f+g\|_p^{p-1}, \tag{7.1.19}$$

which yields Minkowski's inequality (7.1.15) after multiplication by $\|f+g\|_p^{1-p}$ (using that $0<\|f+g\|_p<\infty$).

The last statement is seen from the considerations on $\|\cdot\|_p$ prior to the theorem, as (7.1.15) yields that $f+g$ is $\mathscr{L}_p(\mu)$ when f, g are so; and then (7.1.15) is the triangle inequality for $\mathscr{L}_p(\mu)$. □

As another application of Hölder's inequality, one may obtain a hirarchy among the $\mathscr{L}_p$-seminorms when μ is a finite measure:

PROPOSITION 7.1.6. *When $\mu(X) < \infty$ and $r > p \geq 1$, then any measurable $f\colon X \to \mathbb{C}$ fulfils*

$$\|f\|_p \leq \mu(X)^{\frac{1}{p}-\frac{1}{r}} \|f\|_r. \tag{7.1.20}$$

Especially when $\mu(X) = 1$ it holds that $\|f\|_p \leq \|f\|_r$.

PROOF. Since the numbers $\frac{r}{p} > 1$ and $\frac{r}{r-p} > 1$ are dual exponents, Hölder's inequality applied to $|f|^p$ and the constant function 1 gives

$$\int |f|^p \, d\mu \leq \Big(\int |f|^{p\frac{r}{p}} \, d\mu \Big)^{\frac{p}{r}} \Big(\int 1 \, d\mu \Big)^{\frac{r-p}{r}} \leq \|f\|_p^r \cdot \mu(X)^{1-\frac{p}{r}}. \tag{7.1.21}$$

By taking p'th roots on both sides, the claim follows at once. □

As a consequence of the above, there are *embeddings* among the $\mathscr{L}_p(\mu)$ spaces:

$$\mu(X) < \infty, \ r > p \geq 1 \implies \mathscr{L}_r(\mu) \hookrightarrow \mathscr{L}_p(\mu). \tag{7.1.22}$$

Such results do not extend to cases with $\mu(X) = \infty$, since already for the Lebesgue measure on the real line $\mathbb{R}$ is easy to provide examples of functions belonging to $\mathscr{L}_p \setminus \mathscr{L}_r$, $\mathscr{L}_p \cap \mathscr{L}_r$ and $\mathscr{L}_r \setminus \mathscr{L}_p$ for $r > p \geq 1$.

EXAMPLE 7.1.7. Given functions $f \in \mathscr{L}_p(X, \mathbb{E}, \mu)$ and $g \in \mathscr{L}_p(Y, \mathbb{F}, \nu)$ for $1 \leq p < \infty$, whereby μ and ν are σ-finite measures on sets $X \neq \emptyset$ and $Y \neq \emptyset$, respectively, then the tensor product $f \otimes g$ belongs to $\mathscr{L}_p(\mu \otimes \nu)$ and

$$\|f \otimes g\|_p = \|f\|_p \|g\|_p. \tag{7.1.23}$$

Indeed, as $f \otimes g$ is $\mathbb{E} \otimes \mathbb{F}$-measurable, Tonelli's theorem yields that

$$\begin{aligned} \int_{X \times Y} |f \otimes g|^p \, d\mu \otimes \nu &= \int_X \Big(\int_Y |f(x)|^p |g(y)|^p \, d\nu(y) \Big) \, d\mu(x) \\ &= \|g\|_p^p \int |f(x)|^p \, d\mu(x) = \|f\|_p^p \|g\|_p^p. \end{aligned} \tag{7.1.24}$$

EXAMPLE 7.1.8 (Important cases). (a) Instead of $\mathscr{L}_p(\mathbb{R}^d, \mathbb{B}_d, m_d)$ we briefly write $\mathscr{L}_p(\mathbb{R}^d)$, and for any Borel set $A \subset \mathbb{R}^d$ we simplify $\mathscr{L}_p(A, m_d)$ to $\mathscr{L}_p(A)$. In both cases the Lebesgue measure is tacitly understood.

(b) For $J = \mathbb{N}$ we have the sequence space $\ell_p = \ell_p(\mathbb{N})$. In other words, it is given by $\ell_p = \{ (a_1, a_2, \dots) \mid \sum_{j=1}^\infty |a_j|^p < \infty \}$ whilst $\|(a_1, a_2, \dots)\|_p = \Big(\sum_{j=1}^\infty |a_j|^p \Big)^{\frac{1}{p}}$.

For this there is Hölder's inequality, if $\frac{1}{p} + \frac{1}{q} = 1$,

$$\sum_{j=1}^\infty |a_j b_j| \leq \Big(\sum_{j=1}^\infty |a_j|^p \Big)^{\frac{1}{p}} \Big(\sum_{j=1}^\infty |b_j|^q \Big)^{\frac{1}{q}}. \tag{7.1.25}$$

Minkowski's inequality takes the form, for $1 \leq p < \infty$,

$$\Big(\sum_{j=1}^\infty |a_j + b_j|^p \Big)^{\frac{1}{p}} \leq \Big(\sum_{j=1}^\infty |a_j|^p \Big)^{\frac{1}{p}} + \Big(\sum_{j=1}^\infty |b_j|^p \Big)^{\frac{1}{p}}. \tag{7.1.26}$$

(c) The space $\ell_p(\{1, 2, \dots, d\})$ is just $\mathbb{C}^d$ endowed with the norm $\|(z_1, \dots, z_d)\|_p = \Big(\sum_{j=1}^d |z_j|^p \Big)^{\frac{1}{p}}$. In this case the inequalities of Hölder and Minkowski are analogous to the above, only with summation for $1 \leq j \leq d$ instead. These were the inequalities put forward by Otto Hölder (1889) and Hermann Minkowski (1896).

7.2. The generalised Hölder inequality*

First of all it is not difficult to see that Hölder's inequality can be freed from the restriction to dual exponents: it suffices to have $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$. But even if $p \geq 1$ and $q \geq 1$ we may be forced to make do with $r>0$:

COROLLARY 7.2.1. *When $p_0>0$, $p_1>0$ and we introduce $r>0$ so that $\frac{1}{p_0}+\frac{1}{p_1}=\frac{1}{r}$, then all $\mathbb{E}$-measurable $f_0, f_1\colon X\to\mathbb{C}$ fulfil $\|f_0f_1\|_r=\big(\int|f_0f_1|^r\,d\mu\big)^{\frac{1}{r}}\leq\|f_0\|_{p_0}\|f_1\|_{p_1}$.*

This is immediately seen from the relations $\frac{1}{p_0/r}+\frac{1}{p_1/r}=1$ and $|fg|^r=\big||f|^r|g|^r\big|$.

Using the above corollary twice, there is also a straightforward proof by induction of an extension to products of many functions:

THEOREM 7.2.2. *When $p_0,\ p_1,\dots,\ p_n\in]0,\infty[$ and we introduce $r>0$ such that*

$$\frac{1}{p_0}+\frac{1}{p_1}+\cdots+\frac{1}{p_n}=\frac{1}{r}, \tag{7.2.1}$$

then all $\mathbb{E}$-measurable $f_0, f_1,\dots,f_n\colon X\to\mathbb{C}$ fulfil

$$\|f_0f_1\dots f_n\|_r=\Big(\int|f_0f_1\dots f_n|^r\,d\mu\Big)^{\frac{1}{r}}\leq\|f_0\|_{p_0}\|f_1\|_{p_1}\dots\|f_n\|_{p_n}. \tag{7.2.2}$$

Consequently $f_0f_1\dots f_n\in L_r(\mu)$ for this r, whenever $f_j\in\mathscr{L}_{p_j}(\mu)$ for $j\in\{0,1,\dots,n\}$.

The corollary also gives a result on intersections of Lebesgue spaces, where in addition there is a basic case of *interpolated* norms:

PROPOSITION 7.2.3. *If $0<p_0<p_1$ there is an inclusion*

$$\mathscr{L}_{p_0}(\mu)\cap\mathscr{L}_{p_1}(\mu)\subset\bigcap_{p_0\leq q\leq p_1}L_q(\mu), \tag{7.2.3}$$

and for any $\mathbb{E}$-measurable function $f\colon X\to\mathbb{C}$,

$$\|f\|_q\leq\|f\|_{p_0}^{1-\theta}\|f\|_{p_1}^{\theta} \tag{7.2.4}$$

when $\theta\in[0,1]$ is chosen so that $\frac{1}{q}=\frac{1-\theta}{p_0}+\frac{\theta}{p_1}$.

PROOF. Applying Corollary 7.2.1 to $|f|=|f|^{1-\theta}|f|^{\theta}$ one arrives at the inequality; thence the inclusion. $\square$

7.3. Jensen's inequality*

THEOREM 7.3.1 (Jensen's inequality). *Let $(X,\mathbb{E},\mu)$ be a measure space for which $\mu(X)=1$, that is, a probability space. Suppose $\phi\colon\ ]a,b[\to\mathbb{R}$ is a convex function ($a<b$ in $[-\infty,\infty]$). When $f\in\mathscr{L}(\mu)$ is a real Borel function and $a<f<b$ on X, then*

$$\phi\Big(\int f\,d\mu\Big)\leq\int\phi\circ f\,d\mu. \tag{7.3.1}$$

Then same inequality is valid if $f\in\mathscr{M}^+(X,\mathbb{E})$ satisfies $a<f<b$ on X and $\int f\,d\mu<\infty$.

REMARK 7.3.2. The right-hand side of Jensen's inequality may be ∞ in case the real function $\phi\circ f$ is integrable in the extended sense.

PROOF. When $f\in\mathscr{L}(\mu)$ we may set $t=\int f\,d\mu$. As μ is a probability measure $a<t<b$ follows, since e.g. for $a>-\infty$ one has that $f(x)-a>0$ and that $f-a$ as a member of $\mathscr{M}^+$ can be approximated from below by a sequence s_n of simple positive measurable functions. ($\int s_n\,d\mu\geq 0$ for all n, and necessarily with sharp inequality eventually as X is not a null set.)

The convex function ϕ is subdifferentiable at $t\in]a,b[$, so for certain $h\in\mathbb{R}$ it holds for $a<s<b$ that

$$\phi(s)\geq\phi(t)+h(s-t). \tag{7.3.2}$$

Taking $s = f(x)$ this becomes

$$\phi(f(x)) - \phi(t) - h(f(x) - t) \geq 0, \tag{7.3.3}$$

so that integration of this function in $\mathscr{M}^+$ gives

$$0 \leq \int (\phi(f(x)) - \phi(t) - h(f(x) - t))\,d\mu \leq \infty. \tag{7.3.4}$$

Now, as constant functions are μ-integrable, by adding the trivial identities $\int \phi(t)\,d\mu = \phi(t) = \phi(\int f\,d\mu)$ and $h\int(f-t)\,d\mu = 0$, we arrive at once at the inequality in the theorem.

If $f \in \mathscr{M}^+$ satisfies the conditions in the theorem, $f^{-1}(\{\infty\})$ is a nullset $N \in \mathbb{E}$, so $\tilde{f} = f 1_{X\setminus N} + c1_N$ is μ-integrable with $\int \tilde{f}\,d\mu = t$, and for $a < c < b$ it moreover satisfies $\int \phi \circ \tilde{f}\,d\mu = \int \phi \circ f\,d\mu$. Thus Jensen's inequality is carried over from $\tilde{f}$ to f. $\square$

Since $\exp\colon \mathbb{R} \to \mathbb{R}$ is convex, there is the obvious example that for $f \in \mathscr{L}(X, \mathbb{E}, \mu)$,

$$\exp\Big(\int_X f\,d\mu\Big) \leq \int_X e^{f(x)}\,d\mu(x), \tag{7.3.5}$$

or the equally non-trivial result that, when $f \geq 1$,

$$\exp\Big(\int_X \log f\,d\mu\Big) \leq \int_X f(x)\,d\mu(x). \tag{7.3.6}$$

In case of the finite set $X = \{1, 2, \dots, n\}$ and $\mu(\{i\}) = \frac{1}{n}$ for all i, then if we denote $f(i) = x_i$ in $\mathbb{R}$, the first of these inequalities becomes

$$e^{\frac{1}{n}(x_1 + x_1 + \cdots + x_n)} \leq \frac{1}{n}(e^{x_1} + e^{x_2} + \cdots + e^{x_n}). \tag{7.3.7}$$

But now a substitution of $y_i = e^{x_i}$ yields the classical inequality between the so-called *geometric and aritmetic means*:

$$\sqrt[n]{y_1 y_2 \dots y_n} \leq \frac{1}{n}(y_1 + y_2 + \cdots + y_n). \tag{7.3.8}$$

More generally, given any numbers $\alpha_i > 0$ such that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1, \tag{7.3.9}$$

then $\mu(\{i\}) = \alpha_i$ gives a probability measure μ on $X = \{1, 2, \dots, n\}$. For any convex function $\varphi\colon\ ]a, b[\to \mathbb{R}$ in Theorem 7.3.1, this leads to the original form of Jensen's inequality

$$\varphi\Big(\sum_{i=1}^{n} \alpha_i x_i\Big) \leq \sum_{i=1}^{n} \alpha_i \varphi(x_i). \tag{7.3.10}$$

Taking again $\varphi = \exp$, we get a general weighted version of the above inequality between geometric and aritmetic means, for $y_i \geq 0$,

$$y_1^{\alpha_1} y_2^{\alpha_2} \dots y_n^{\alpha_n} \leq \alpha_1 y_1 + \alpha_2 y_2 + \cdots + \alpha_n y_n. \tag{7.3.11}$$

By setting $p_i = \frac{1}{\alpha_i}$ and $u_i = y_i^{1/p_i}$, one finds for $u_i \geq 0$ and $p_i > 1$ with $\frac{1}{p_1} + \cdots + \frac{1}{p_n} = 1$,

$$u_1 u_2 \dots u_n \leq \frac{u_1^{p_1}}{p_1} + \frac{u_2^{p_2}}{p_2} + \cdots + \frac{u_n^{p_n}}{p_n}. \tag{7.3.12}$$

This is also known as *Young's inequality*.

The above examples elucidate the scope of the abstract result on Jensens's inequality given in Theorem 7.3.1.

CHAPTER 8

Lebesgue spaces

To improve on the spaces $\mathscr{L}_p(\mu)$, which in Section 7.1 were shown to be only seminormed, we here introduce their normed equivalents, the Lebegue spaces written $L_p(\mu)$.

8.1. The normed Lebesgue spaces L_p

The widely accepted remedy for the seminormed Lebesgue spaces $\mathscr{L}_p(\mu)$ is to stop the strict distinguishing of functions that only differ on a μ-nullset. This softening is already present as such functions will have the same integral.

More precisely, this means that we call f, $g\colon X \to \mathbb{C}$ *equivalent*, and write $f \sim g$, whenever $f = g$ μ-a.e. It is obvious that $\sim$ is an equivalence relation, that is, $f \sim f$, $f \sim g \iff g \sim f$ and $f \sim g \wedge g \sim h \implies f \sim h$ (the reflexive, symmetric and transitive property, respectively).

Thereafter we transfer $\mathscr{L}_p(\mu)$ to the set of equivalence classes, denoted by $L_p(\mu)$:

$$[f] = \{\, g \mid g \sim f \,\}, \qquad L_p(\mu) = L_p(X,\mathbb{E},\mu) = \left\{\, [f] \;\middle|\; f \in \mathscr{L}_p(X,\mathbb{E},\mu) \,\right\}. \tag{8.1.1}$$

Algebraically also $L_p(\mu)$ is a vector space with the compositions

$$[f] + [g] = [f+g], \qquad c[f] = [cf]. \tag{8.1.2}$$

Indeed, the classes on the right-hand sides are readily seen to be independent of the choice of representatives made on the left. All 8 axioms for a vector space is easily seen to be fulfilled, when $[0]$ and $[-f]$ are used as the zero vector and the opposite vector, respectively.

For the norm on $L_p(\mu)$ the simple solution is just to let the seminorm of $\mathscr{L}_p$ act on a representative:

$$\big\|[f]\big\|_p = (\int |f|^p \, d\mu)^{\frac{1}{p}} = \|f\|_p. \tag{8.1.3}$$

Indeed, this is a map $L_p(\mu) \to [0,\infty[$, for even if $g \sim f$ for $\mathbb{E}$-measurable functions f, g, then we have $|g|^p = |f|^p$ μ-a.e., which because of Remark 4.16 gives $\int |g|^p \, d\mu = \int |f|^p \, d\mu$, whence the value in (8.1.3) is independent of the choice of representative.

In practice $[f]$ is often simply written as f, whereby it is tacitly understood that it is the equivalence class determined by f that is considered. For example $\|[f]\|_p$ is simplified to $\|f\|_p$, and 0 replaces $[0]$.

The map $\|\cdot\|_p\colon L_p(\mu) \to [0,\infty[$ is actually a norm whenever $1 \le p < \infty$; cf. (7.1.4). Indeed, the triangle inequality is a direct consequence of Minkowski's inequality in (7.1.7) or (7.1.15), where one can read the left- and right-hand sides as the values of $\|\cdot\|_p$ at $[f+g]$, $[f]$ og $[g]$; cf. (8.1.3). Similarly the positive homogeneity is inferred from the observation in front of (7.1.7). Finally (7.1.5) gives the crucial property that $\big\|[f]\big\|_p = 0$ holds if and only if $f \sim 0$, i.e. if and only if $[f] = 0$.

The vector space $L_p(\mu)$ therefore has a metric given by

$$d(f,g) = \|f-g\|_p = \big(\int |f-g|^p \, d\mu\big)^{\frac{1}{p}} \qquad \text{for } 1 \le p < \infty. \tag{8.1.4}$$

For (equivalence classes) f, $f_1, f_2, \dots$ in $L_p(\mu)$, one says if $d(f_n,f) = \|f_n - f\|_p \to 0$ for $n \to \infty$ that f_n converges to f in $L_p(\mu)$, or in L_p-norm, or that f_n converges to f in *p-mean* (in quadratic mean for $p = 2$). One motivation for this could be that for convergence in

1-norm, that is, $\int |f_n - f|\,d\mu \to 0$ for $n \to \infty$, the integral can be thought of as the mean deviation of f_n from f.

As a very satisfying result, the metric spaces $L_p(\mu)$ are always complete. This is known as Fischer's Completeness Theorem, cf. Theorem 8.2.3, which is a cornerstone in integration theory and its applications.

Complete normed vector spaces are referred to in the literature as Banach spaces after Stefan Banach, who made extensive investigations of such spaces in the 1920's.

DEFINITION 8.1.1. A Banach space is a pair $(B, \|\cdot\|)$ consisting of a vector space B (over $\mathbb{R}$ or $\mathbb{C}$) and a norm $\|\cdot\|$ on B, for which the induced metric $d(v,w) = \|v-w\|$ is complete in the sense that every Cauchy sequence converges in B.

If the norm on a Banach space B is induced by an inner product on B, then B is called a Hilbert space (and is usually denoted by H).

A main source of Banach spaces is the family $L_p(X, \mathbb{E}, \mu)$, with $1 \le p < \infty$, for an arbitrary measure space $(X, \mathbb{E}, \mu)$.

The case $p = 2$ is special, though, for the norm on $L_2(\mu)$ is induced by an inner product, which for arbitrary $f, g \in L_2(\mu)$ is given by

$$(f|g) = \int f(x)\overline{g(x)}\,d\mu(x). \tag{8.1.5}$$

Even though it is formally clear that $(f|f) = \int |f|^2 d\mu = \|f\|_2^2$, it is less obvious that the integrand $f\overline{g}$ in the inner product $(f|g)$ is integrable for $f, g \in L_2(\mu)$. But this follows at once from the Cauchy–Schwarz inequality, cf. Corollary 7.1.4. Thus the space $L_2(X, \mathbb{E}, \mu)$ is a main example of a Hilbert space (David Hilbert made fundamental investigations of spectral theory of quadratic forms on complete inner product spaces around 1910).

EXAMPLE 8.1.2. To elucidate on the space $L_p(\mathbb{R}^d)$, we first give a straightforward argument for the integrability of the continuous function $f_N(x) = (1+|x|)^{-N}$ when $N > d$: by Tonelli's theorem,

$$\begin{aligned}\int f_N\,dm_d &\le \int (1+|x_1|)^{-N/d} \dots (1+|x_d|)^{-N/d}\,dm_d = \Big(\int_{\mathbb{R}} (1+t)^{-N/d}\,dt\Big)^d \\ &\le 2^d \Big(1 + \lim_{k\to\infty} \Big[\frac{t^{1-N/d}}{1-N/d}\Big]_1^k\Big)^d = 2^d \big(1 + \frac{d}{N-d}\big)^d < \infty.\end{aligned} \tag{8.1.6}$$

Conversely, the multinomial formula gives, with a sum over $d = n_0 + n_1 + \dots + n_d$,

$$\begin{aligned}(1+|x|)^d &\le (1 + |x_1| + \dots + |x_d|)^d \\ &= \sum \frac{d!}{n_0! n_1! \dots n_d!} |x_1|^{n_1} \dots |x_d|^{n_d} \le d!(1+|x_1|)\dots(1+|x_d|).\end{aligned} \tag{8.1.7}$$

From this it follows that

$$\int f_d\,dm_d \ge \Big(\int_{\mathbb{R}} \frac{1}{1+|t|}\,dt\Big)^d / d! \ge \Big(\int_1^\infty \frac{1}{t}\,dt\Big)^d / d! = \infty. \tag{8.1.8}$$

Thus neither f_d nor f_N with $N < d$ belongs to $L_1(\mathbb{R}^d)$.

Replacing N by Np, one obtains from the above that

$$(1+|x|)^{-N} \in L_p(\mathbb{R}^d) \iff N > d/p. \tag{8.1.9}$$

8.2. 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 8.2.1. *If a sequence (f_n) is given in $L_p(\mu)$ for some $p \in [1, \infty[$ and the function $f\colon X \to \mathbb{C}$ is such that*

$$f = \lim_{n\to\infty} f_n \quad \mu\text{-a.e.}, \tag{8.2.1}$$

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

$$\forall n \in \mathbb{N}\colon |f_n| \le g \quad \mu\text{-a.e.}, \qquad \int g^p \, d\mu < \infty, \tag{8.2.2}$$

implies that (the class determined by) f belongs to $L_p(\mu)$ with $\|f\|_p \le (\int g^p \, d\mu)^{\frac{1}{p}}$ and

$$\|f_n - f\|_p \xrightarrow[n\to\infty]{} 0. \tag{8.2.3}$$

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

Now, both $|f_n(x) - f(x)|^p \to 0$ and $|f_n - f|^p \le 2^p g^p$ hold on X, so it follows that

$$\int |f_n - f|^p \, d\mu \xrightarrow[n\to\infty]{} \int 0 \, d\mu = 0, \tag{8.2.4}$$

using the Majorised Convergence Theorem. □

Completeness of a normed vector space V can be rephrased in a way that is rather useful 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 8.2.2. *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

$$\sum_{k=1}^{\infty} \|x_{n_{k+1}} - x_{n_k}\| \le 1. \tag{8.2.5}$$

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

$$\|x - x_n\| \le \|x - x_{n_k}\| + \|x_{n_k} - x_n\|, \tag{8.2.6}$$

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

The converse conclusion is seen by applying the triangle inequality to a difference $s_{N+p} - s_N$ of two partial sums of any given 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 8.2.3 (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 8.2.2, we let a series $\sum_{k=1}^{\infty} g_k$ of functions $g_k \in \mathscr{L}_p(\mu)$ be given such that

$$S := \sum_{k=1}^{\infty} \|g_k\|_p < \infty. \tag{8.2.7}$$

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 $\sum_{k=1}^{\infty} g_k(x)$ exists a.e., and that this works as the function f.

1° There is an auxiliary function $h \in \mathscr{M}^+(X,\mathbb{E})$ given by the formula

$$h(x) = \sum_{k=1}^{\infty} |g_k(x)|. \tag{8.2.8}$$

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

2° The function h is a possible L_p-majorant for the sequence given by $f_n = \sum_{k=1}^n g_k(x)$, for the convention $\infty^p = \infty$ gives for $n \to \infty$ that

$$(\sum_{k=1}^{n} |g_k(x)|)^p \nearrow h(x)^p, \tag{8.2.9}$$

so $|f_n| \le h$ holds on X for every n. Moreover, the Monotone Convergence Theorem entails

$$\|\sum_{k=1}^{n} |g_k|\|_p^p = \int (\sum_{k=1}^{n} |g_k|)^p \, d\mu \nearrow \int h^p \, d\mu, \tag{8.2.10}$$

where triangle inequality gives, for all $n \in \mathbb{N}$,

$$0 \le \|\sum_{k=1}^{n} |g_k|\|_p \le \sum_{k=1}^{n} \||g_k|\|_p \le \sum_{k=1}^{\infty} \|g_k\|_p = S. \tag{8.2.11}$$

This yields that $\int h^p \, d\mu \in [0, S^p]$, hence we have $\int h^p \, d\mu < \infty$ by (8.2.7).

3° We now define a measurable function by $f = \sum_{k=1}^{\infty} g_k 1_{X \setminus N}$, whereby $N = h^{-1}(\{\infty\})$ is in $\mathbb{E}$ with $\mu(N) = 0$ (the series converges pointwise in $X \setminus N$, cf. 1°). According to Theorem 8.2.1 the function f (or rather $[f]$) belongs to $L_p(\mu)$ with the norm estimate $\|f\|_p \le S$ and $\|f - f_n\|_p \to 0$ for $n \to \infty$. □

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

COROLLARY 8.2.4. *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 both (absolutely) μ-a.e. on X as well as in p-mean to the same function satisfying $f \in L_p(\mu)$ and*

$$\|f\|_p \le \sum_{k=1}^{\infty} \|g_k\|_p. \tag{8.2.12}$$

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 (8.2.12) to obtain a *formal* generalisation of Minkowski's inequality:

$$\|\sum_{k=1}^{\infty} g_k\|_p \le \sum_{k=1}^{\infty} \|g_k\|_p. \tag{8.2.13}$$

This is a dangerous abuse of notation, since when the right-hand side diverges there may be no limit function for the series $\sum_{k=1}^{\infty} g_k$, so that the left-hand side makes no sense.

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

COROLLARY 8.2.5. *Every sequence $f_1, f_2, \dots$ in $L_p(\mu)$ that converges to $f \in L_p(\mu)$ in p-mean has a subsequence $f_{n_1}, f_{n_2}, \dots$ converging pointwise to f μ-a.e. There is 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 8.2.2 and Theorem 8.2.3. First $n_1 < n_2 < \dots$ are chosen so that

$$\sum_{k=1}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p < \infty. \tag{8.2.14}$$

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 8.2.4 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 $L_p(\mu)$, and therefore $f = \tilde{f}$ in the normed space $L_p(\mu)$; so f is also an a.e. pointwise limit of the subsequence (f_{n_k}).

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

$$g(x) = |f_{n_1}(x)| + \sum_{k=1}^{\infty} |f_{n_{k+1}}(x) - f_{n_k}(x)|, \tag{8.2.15}$$

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

COROLLARY 8.2.6. *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 8.2.5 a suitable subsequence (f_{n_k}) converges a.e. to φ, and of course also to ψ a.e.

8.3. Density of nice functions

In general, for an arbitrary measure space $(X, \mathbb{E}, \mu)$, the equivalence classes in spaces like $L_p(\mu)$ may seem like rather weird objects. However, in some sense "most" of these spaces consist of "nice" functions.

To explain this, it is recalled that in a metric space (M,d), a subset A is (everywhere) *dense* in another subset $B \subset M$ if to every point $b \in B$ one can find points of A arbitrarily close to b. That is, A is dense in B if every open ball $B(b,\delta)$, for $b \in B$ and $\delta > 0$, satisfies $A \cap B(b,\delta) \neq \emptyset$. This can be phrased concisely by means of a reverse inclusion:

DEFINITION 8.3.1. For subsets A, B of M, A is *dense* in B if $B \subset \overline{A}$.

As examples, $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ are dense in one another (though $\mathbb{Q} \subset \overline{\mathbb{R} \setminus \mathbb{Q}}$ is less trivial); notice that these sets are disjoint and that the definition actually allows this.

By abuse of language, a sequence (x_n) in M is called dense if its range $\{\, x_n \mid n \in \mathbb{N} \,\}$ is dense in M. A metric space M is called *separable* if there is a dense sequence of points $x_n \in M$.

In a normed vector space V, a subset S is said to be *total* in V if it spans a dense subspace in V; that is, if $\operatorname{span} S = \{\, \lambda_1 s_1 + \dots + \lambda_n s_n \mid \forall j \le n\colon\ \lambda_j \in C,\ s_j \in S \,\}$ is dense in V. Or, in other words, S is total in V if $V = \overline{\operatorname{span}} S$.

In general the indicator functions form a total set in the Lebesgue spaces:

LEMMA 8.3.2. *For an arbitrary measure space $(X, \mathbb{E}, \mu)$ and $1 \le p < \infty$ the simple $\mathscr{L}_p$-functions are everywhere dense in $L_p(\mu)$.*

PROOF. First it is noted that for a simple $\mathbb{E}$-measurable function $g = \sum_j a_j 1_{A_j}$, written with its different non-zero values $a_1, \dots, a_N$, one has $\int |g|^p \, d\mu = \sum_j |a_j|^p \mu(A_j)$, so g belongs to $L_p(\mu)$ if and only if each $\mu(A_j) < \infty$; i.e., if and only if $\mu(\{\, x \mid g(x) \neq 0 \,\}) < \infty$.

Given $f \in L_p(\mu)$ and $\varepsilon > 0$ we must show the existence of such a function g with $\|f - g\|_p < \varepsilon$. We may assume that $f \ge 0$, for else we may combine the triangle inequality with the identity

$$f = (\operatorname{Re} f^+ - \operatorname{Re} f^-) + \mathrm{i}(\operatorname{Im} f^+ - \operatorname{Im} f^-). \tag{8.3.1}$$

There is a sequence $0 \le g_1 \le g_2 \le \dots$ of simple $\mathbb{E}$-measurable functions such that $g_n \nearrow f$, but since $g_n \le f$ (so that f is an L_p majorant) we have $g_n \in L_p(\mu)$ and $\|f - g_n\|_p \to 0$ for $n \to \infty$. So it suffices to take $g = g_n$ for some suitably large n. $\square$

In case of the Euclidean space $\mathbb{R}^d$ the above can be made a little more precise, because the measurable sets can be replaced by standard intervals. It is customary to refer to functions of the form $f = \sum_{j=1}^N c_j 1_{I_j}$, whereby each $I_j \in \mathbb{I}_d$, as a *step function*.

Using the construction of the Lebesgue measure m_d one can prove

PROPOSITION 8.3.3. *The step functions $\mathbb{R}^d \to \mathbb{C}$ are dense in $L_p(\mathbb{R}^d)$ for $1 \le p < \infty$. Phrased differently: The indicator functions for standard intervals are total in $L_p(\mathbb{R}^d)$.*

PROOF. For $f \in L_p(\mathbb{R}^d)$ and $\varepsilon > 0$ it must be shown that there is some step function g such that $\|f - g\|_p < \varepsilon$. (Notice that every step function belongs to $\mathscr{L}_p$.) Because of Lemma 8.3.2 it suffices to prove this in case f is a simple function $f = \sum b_j 1_{B_j}$ in L_p; hereby each B_j is a Borel set of finite measure. Via the triangle inequality this gives a further reduction to the case that $f = 1_B$ for some Borel set B with $m_d(B) < \infty$.

From the construction of the Lebesgue measure, cf. (5.0.12), it is seen that there is a sequence of standard intervals (I_n) in $\mathbb{I}_d$ such that

$$B \subset \bigcup_{n\in\mathbb{N}} I_n, \qquad \sum_{n\in\mathbb{N}} v_d(I_n) < m_d(B) + (\frac{\varepsilon}{2})^p. \tag{8.3.2}$$

(One may first arrange that the inclusion is strict.)

Now we set $g = 1_G$ where $G = \bigcup_{n\le N} I_n$ for N so large that $\sum_{n>N} v_d(I_n) < (\frac{\varepsilon}{2})^p$. Here

$$\begin{aligned} \|g - 1_B\|_p &\le \|1_{\bigcup_{n\in\mathbb{N}} I_n} - 1_B\|_p + \|g - 1_{\bigcup_n I_n}\|_p \\ &\le m_d(\bigcup_{n\in\mathbb{N}} I_n \setminus B)^{\frac{1}{p}} + m_d(\bigcup_{n>N} I_n)^{\frac{1}{p}} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \le \varepsilon. \end{aligned} \tag{8.3.3}$$

Finally, it is elementary to show that G is a finite union of some, possibly, smaller standard intervals E_j, which one may write up in terms of all the endpoints of the I_n, in all dimensions. Hence $g = \sum_j 1_{E_j}$ is a step function. □

For a general standard interval $I =]a_1, b_1] \times \cdots \times]a_d, b_d]$ it is clear that the volume $v_d(I)$ is a continuous function of the $2d$ variables $a_1, \dots, a_d, b_1, \dots, b_d$. The first part of the next result is therefore obvious:

PROPOSITION 8.3.4. *When $I =]a_1, b_1] \times \cdots \times]a_d, b_d] \neq \emptyset$ and $\varepsilon > 0$, then there are standard intervals $J =]c_1, d_1] \times \cdots \times]c_d, d_d]$ satisfying $c_j < a_j < b_j < d_j$ for all j and*

$$m_d(J \setminus I) < \varepsilon. \tag{8.3.4}$$

For each such standard interval J there is a function $g \in C^\infty(\mathbb{R}^d)$ such that

$$g(x) = 1 \text{ for } x \in \bar{I}, \qquad g(x) = 0 \text{ for } x \in \mathbb{R}^d \setminus J^\circ, \tag{8.3.5}$$

whilst $0 \le g \le 1$ on $\mathbb{R}^d$.

PROOF. To obtain the desired $g \in C^\infty(\mathbb{R}^d)$ in case $d = 1$, we first note that there is a C^∞-function defined by

$$f(x) = \begin{cases} e^{-1/x} & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases} \tag{8.3.6}$$

Indeed, it is C^∞ on $\mathbb{R} \setminus \{0\}$ and seen inductively to fulfil $f^{(k)} = p_k(1/x)e^{-1/x}$ on $\mathbb{R}_+$ for some polynomial p_k; so it follows that $f^{(k+1)}(0) = 0$ if $\lim_{x\to 0^+}(1/x)p(1/x)e^{-1/x} = 0$ for every polynomial p; which in its turn follows from the fact that $\lim_{x\to\infty} x^n e^{-x} = 0$.

Consequently the function $x \mapsto f(x)f(\delta - x)$ is C^∞ for each $\delta > 0$; it is positive for $0 < x < \delta$, zero otherwise.

This gives rise to another function h in $C^\infty(\mathbb{R})$ given by

$$h(x) = \int_0^x f(t)f(\delta - t)\,dt \Big/ \int_0^\delta f(t)f(\delta - t)\,dt. \tag{8.3.7}$$

This satisfies $h = 0$ for $x \le 0$, while $h(x) = 1$ for $x \ge \delta$; and in general $0 \le h \le 1$.

Given $c < a < b < d$ we may specify $\delta > 0$ such that $\delta \le a - c$ and $\delta \le d - b$ and form the product

$$g(x) = h(x - c)h(d - x). \tag{8.3.8}$$

This is also in $C^\infty(\mathbb{R})$ and fulfils $0 \le g \le 1$ as well as

$$g(x) = 1 \text{ for } x \in [a, b], \qquad g(x) = 0 \text{ for } x \notin [c, d]. \tag{8.3.9}$$

Finally, applying this construction in each dimension, we may for each $i \in \{1, \dots, d\}$ pick a function $g_i \in C^\infty(\mathbb{R})$ such that $g_i = 1$ on $[a_i, b_i]$ and $g = 0$ on $\complement]c_i, d_i[$ and set

$$g(x) = g_1 \otimes \cdots \otimes g_d(x) = g_1(x_1) \dots g_d(x_d). \tag{8.3.10}$$

This function has the stated properties. □

The construction in the above proof is noteworthy also in a broader context. Indeed, recalling that the support of a function $f\colon \mathbb{R}^d \to \mathbb{C}$ is defined as the closure of the set where it is non-zero,

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}}, \tag{8.3.11}$$

it is customary to let $C_0(\mathbb{R}^d)$ denote the space of continuous functions having compact support in $\mathbb{R}^d$. This is endowed with the sup-norm, that is, $\|f\| = \sup\{\,|f(x)| \mid x \in \mathbb{R}^d\,\}$ for $f \in C_0(\mathbb{R}^d)$, and this space is complete.

Further one introduces the space of smooth functions with compact support,

$$C_0^\infty(\mathbb{R}^d) = C^\infty(\mathbb{R}^d) \bigcap C_0(\mathbb{R}^d). \tag{8.3.12}$$

This is of course normed as a subspace of $C_0(\mathbb{R}^d)$, but it is not complete. In fact , it is not possible to norm $C_0^\infty(\mathbb{R}^d)$ such that the resulting metric space is complete—-this requires a more general topology, but we abstain from discussing this here (it is a major subject within the distribution theory of Laurent Schwarz [**Sch66**]).

Anyhow, it is of course crucial to know that the intersection (8.3.12) is not empty, and here Proposition 8.3.4 and its proof shows that there is an abundance of such functions. To emphasize the importance of the existence of such functions, let us depart from the above construction of $f(x)f(\delta - x)$ and write up a more general explicit expression for such a function:

$$\varphi(x) = \begin{cases} e^{\frac{d-c}{(x-c)(x-d)}} & \text{for } c < x < d, \\ 0 & \text{for } x \in \mathbb{R} \setminus]c,d[\,. \end{cases} \tag{8.3.13}$$

This is not as fine as the above $g(x)$, though, for there is no plateau where $g = 1$. Nevertheless, such C_0^∞-functions are useful for cut-off techniques in modern mathematical analysis.

The above considerations also add up to the following cornerstone in the theory of the Lebesgue spaces on Euclidean space:

THEOREM 8.3.5. *The space $C_0^\infty(\mathbb{R}^d)$ of continuous functions with compact support is dense in $L_p(\mathbb{R}^d)$ whenever $1 \leq p < \infty$.*

PROOF. For $f \in L_p(\mathbb{R}^d)$ and $\varepsilon > 0$ it must be shown that there is some $g \in C_0^\infty(\mathbb{R}^d)$ such that $\|f - g\|_p < \varepsilon$. Since indicator functions for standard intervals form a total set in $L_p(\mathbb{R}^d)$, cf. Proposition 8.3.3, it suffices to treat the case $f = 1_I$ for

$$I =]a_1, b_1] \times \cdots \times]a_d, b_d]. \tag{8.3.14}$$

According to Proposition 8.3.4 there is some $J \in \mathbb{I}_d$ such that $m_d(J \setminus I) < \varepsilon^p$ and a function $g \in C_0^\infty(\mathbb{R}^d)$ such that $0 \leq g \leq 1$ with $\operatorname{supp} g = \overline{J}$ and $g = 1$ on $\overline{I}$. Since

$$|1_I - g| \leq 1_{J \setminus I}, \tag{8.3.15}$$

clearly

$$\|1_I - g\|_p^p \leq \int 1_{J \setminus I}\, dm_d = m_d(J \setminus I) < \varepsilon^p. \tag{8.3.16}$$

This shows the claimed density. □

The theorem has an extension to arbitrary Radon measures on $\mathbb{R}^d$, but this has been omitted for the sake of simplicity. However, there is no chance to extend the result to $p = \infty$, as for example indicator functions cannot be uniformly approximated by continuous functions.

CHAPTER 9

Convolution

As a new operation on functions, the convolution $f * g$ of two functions is studied here. It appears as a useful tool when one needs to replace a rough function f by a function with a smooth graph; or more fundamentally as the probability distribution of the sum $X+Y$ of two stochastic variables X, Y having distributions f, g respectively.

9.1. Convolution of Borel functions

The convolution of two Borel functions $f, g\colon \mathbb{R}^d \to \mathbb{C}$ is defined (tentatively) by the formula

$$f * g(x) = \int_{\mathbb{R}^d} f(x-y) g(y)\, dy. \tag{9.1.1}$$

Once and for all one may here observe that $f(x-y)g(y)$ is a Borel function on $\mathbb{R}^{2d}$, since it is composed of $f \otimes g$ and the continuous map $(x,y) \mapsto (x-y, y)$. Hence the integrand is Borel measurable for each fixed $x \in \mathbb{R}^d$.

However, for the above formula to make sense, it is also necessary that the integrand is integrable, which may depend on the considered x. Therefore the domain of definition of $f * g$ is formally defined as follows:

$$\begin{aligned} D(f*g) &= \{\, x \in \mathbb{R}^d \mid f(x-\cdot)g(\cdot) \in \mathscr{L}(\mathbb{R}^d) \,\} \\ &= \{\, x \in \mathbb{R}^d \mid \int_{\mathbb{R}^d} |f(x-y)g(y)|\, dy < \infty \,\}. \end{aligned} \tag{9.1.2}$$

Here it is instructive to use the definition to show the following peculiar property of the convolution $f * g$:

$$\left.\begin{aligned} f &\equiv 0 \text{ in } \complement A \\ g &\equiv 0 \text{ in } \complement B \end{aligned}\right\} \implies f * g \equiv 0 \text{ in } \complement(A+B) \subset D(f*g). \tag{9.1.3}$$

Indeed, for the integrand in (9.1.1) it is clear that $f(x-y)g(y) \neq 0$ implies that $x - y \in A$ and $y \in B$, whence $x \in y + A \subset B + A$. So for $x \notin A+B$ the integrand is identically 0, hence integrable with $f * g(x) = 0$ for such x; so $\complement(A+B) \subset D(f*g)$ with $f*g \equiv 0$ in $\complement(A+B)$.

One obvious interpretation of (9.1.3), and of its proof, is that the problem with the convolution $f * g$ lies at the points $x \in \mathbb{R}^d$ where $f * g(x)$ has a chance of being non-zero; cf. the above $A+B$. This theme is met repeatedly in the theory.

Among the general properties of $f * g$, one has the strongest possible form of the commutativity:

$$D(f*g) = D(g*f), \qquad \text{and for } x \text{ herein } f*g(x) = g*f(x). \tag{9.1.4}$$

In fact, m_d is for each $x \in \mathbb{R}^d$ invariant under the isometry $\varphi(y) = x - y$, whence

$$\begin{aligned} \int |f(x-\cdot)g|\, dm_d &= \int |f(x-y)g(y)|\, d\varphi(m_d)(y) \\ &= \int |f(x-\varphi(y))g(\varphi(y))|\, dm_d(y) = \int |g(x-\cdot)f|\, dm_d. \end{aligned} \tag{9.1.5}$$

Here either both or none of the two sides are finite, so consequently $D(f*g) = D(g*f)$. For x in this common domain, the above argument without $|\cdot|$ yields $f*g(x) = g*f(x)$.

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

It is a classical fact that convolution $(f,g) \mapsto f*g$ of two integrable functions yields another integrable function, at least almost everywhere:

THEOREM 9.2.1. *For $f,g \in \mathscr{L}_1(\mathbb{R}^d)$ the convolution $f*g(x)$ is defined for almost every $x \in \mathbb{R}^d$ and induces an element—also denoted by $f*g$—in $L_1(\mathbb{R}^d)$ for which*

$$\|f*g\|_1 \le \|f\|_1 \|g\|_1. \tag{9.2.1}$$

PROOF. Since the Borel function $(x,y) \mapsto |f(x-y)g(y)|$ is positive, we may combine Tonelli's theorem with the translation invariance of the Lebesgue measure to obtain

$$\begin{aligned}\int_{\mathbb{R}^{2d}} |f(x-y)g(y)|\,d(x,y) &= \int |g(y)| \int |f(x-y)|\,dx\,dy \\ &= \int |g(y)| \int |f(x)|\,dx\,dy = \|f\|_1\|g\|_1 < \infty.\end{aligned} \tag{9.2.2}$$

Fubini's theorem now yields that $y \mapsto f(x-y)g(y)$ is integrable with respect to y for x outside a measurable nullset, and moreover that the almost everywhere defined function $x \mapsto \int f(x-y)g(y)\,dy$, which is $f*g(x)$, induces an element of $L_1(R^d)$.

Now, for $x \in D(f*g)$ one has the estimate

$$|f*g(x)| \le \int |f(x-y)g(y)|\,dy, \tag{9.2.3}$$

and integration of both sides gives

$$\|f*g\|_1 \le \int\int |f(x-y)g(y)|\,dy\,dx = \|f\|_1\|g\|_1, \tag{9.2.4}$$

by using Tonelli's theorem and invoking (9.2.2). □

As an addendum to the theorem, it is straightforward to see that for $f,g,h \in L_1(\mathbb{R}^d)$ the *distributive* laws hold almost everywhere, hence as elements in $L_1(\mathbb{R}^d)$, namely $(cf)*g = cf*g = f*(cg)$ for $c \in \mathbb{C}$ and

$$f*(g+h) = f*g + f*h, \qquad (f+g)*h = f*h + g*h. \tag{9.2.5}$$

Moreover there is *associativity* in $L_1(\mathbb{R})$,

$$f*(g*h)(x) = (f*g)*h(x). \tag{9.2.6}$$

Indeed, it is clear from Theorem 9.2.1 that both sides make sense; they may be shown to be equal by an extension of the argument for the commutativity. Details are left as an exercise.

A famous interpretation of the above is that the Banach space $L_1(\mathbb{R}^d)$ forms a commutative algebra, when the convolution $f*g$ is used as the multiplication (whereas $\mathscr{L}_1(\mathbb{R}^d)$ is not stable under convolution; cf. Theorem 9.2.1). Moreover, because of the inequality (9.2.1), the convolution is a *continuous* map

$$L_1(\mathbb{R}^d) \times L_1(\mathbb{R}^d) \to L_1(\mathbb{R}^d), \tag{9.2.7}$$

because whenever $f_n \to f$ and $g_n \to g$ in $L_1(\mathbb{R}^d)$ it follows that also $f_n * g_n \to f*g$, since

$$\begin{aligned}\|f_n*g_n - f*g\|_1 &= \|(f_n-f)*(g_n-g) + f*(g_n-g) + (f_n-f)*g\|_1 \\ &\le \|f_n-f\|_1\|g_n-g\|_1 + \|f\|_1\|g_n-g\|_1 + \|f_n-f\|_1\|g\|_1.\end{aligned} \tag{9.2.8}$$

These properties are summed up by referring to $L_1(\mathbb{R}^d)$ as a commutative *Banach algebra*.

Lebesgue spaces $L_p(\mathbb{R}^d)$ with $1 < p < \infty$ are not convolution algebras. But the fact that they are invariant under convolution by an integrable function is a consequence of the next result.

In its proof below it is instructive to note that the basic Theorem 9.2.1 is applied via a remarkable *pointwise* estimate:

$$|f*g(x)|^p \le c|f|^p * |g|(x). \tag{9.2.9}$$

In itself it is surprising that the convolution, with its definition, should satisfy a pointwise estimate at all (but note that (9.2.3) yields (9.2.9) for $p = 1$). However, this is obviously useful because both $|f|^p$, $|g|$ are integrable according to the assumptions in

THEOREM 9.2.2. *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*

$$\|f * g\|_p \le \|f\|_p \|g\|_1. \tag{9.2.10}$$

PROOF. As the case $p = 1$ was covered in Theorem 9.2.1, we assume $1 < p < \infty$ and determine the dual exponent from $p + q = pq$. With integrals over $\mathbb{R}^d$, Hölder's inequality gives

$$\begin{aligned}\int |f(x-y)g(y)|\,dy &= \int |f(x-y)||g(y)|^{\frac{1}{p}}|g(y)|^{\frac{1}{q}}\,dy \\ &\le \Big(\int |f(x-y)|^p |g(y)|\,dy\Big)^{\frac{1}{p}} \Big(\int |g(y)|\,dy\Big)^{\frac{1}{q}}.\end{aligned} \tag{9.2.11}$$

For each $x \in D(|f|^p * |g|)$ the last expression and consequently also the first integral is finite, so such x belong to $D(f * g)$. Hence we have

$$D(|f|^p * |g|) \subset D(f * g). \tag{9.2.12}$$

The former set fills $\mathbb{R}^d$ except for a nullset, as $|f|^p, |g| \in \mathscr{L}$; hence $f * g$ is defined a.e.

Now, the above inequality straightforwardly implies the pointwise estimate

$$|f * g(x)|^p \le |f|^p * |g|(x) \|g\|_1^{p-1} \qquad \text{for } x \in D(|f|^p * |g|). \tag{9.2.13}$$

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

$$\int |f * g(x)|^p\,dx \le \|g\|_1^{p-1} \int |f|^p * |g|\,dx \le \|g\|_1^{p-1} \big\||f|^p\big\|_1 \big\||g|\big\|_1 = \|f\|_p^p \|g\|_1^p. \tag{9.2.14}$$

This yields the stated inequality at once. ☐

While this result is true as stated also for $p = \infty$, there is a better result in this case, which is derived in the next section after a preparation of independent interest.

9.3. Strong convergence of translation

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

$$\tau_a f(x) = f(x - a). \tag{9.3.1}$$

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 for such f 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

$$\|\tau_a f\|_p = \Big(\int_{\mathbb{R}^d} |f(x-a)|^p\,dx\Big)^{1/p} = \|f\|_p. \tag{9.3.2}$$

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

PROPOSITION 9.3.1. *When $f \colon \mathbb{R}^d \to \mathbb{C}$ belongs to $L_p(\mathbb{R}^d)$ for $1 \le p < \infty$, then*

$$\|\tau_a f - f\|_p = \Big(\int_{\mathbb{R}^d} |f(x-a) - f(x)|^p\,dx\Big)^{\frac{1}{p}} \to 0 \qquad \textit{for } a \to 0. \tag{9.3.3}$$

PROOF. Given $\varepsilon > 0$ there exists by the density of $C_0(\mathbb{R}^d)$ in $L_p(\mathbb{R}^d)$ a function $g \in C_0$ such that $\|f-g\|_p \le \varepsilon/3$; cf. Theorem 8.3.5. Because of the translation invariance of the Lebesgue measure this gives via the triangle inequality

$$\|\tau_a f - f\|_p \le \|\tau_a(f-g)\|_p + \|\tau_a g - g\|_p + \|g-f\|_p \le \frac{2\varepsilon}{3} + \|\tau_a g - g\|_p. \tag{9.3.4}$$

The function $g \in C_0$ is uniformly continuous on $\mathbb{R}^d$ since $g \equiv 0$ outside a ball $B(0,R)$ containing $\operatorname{supp} g$. In fact, to $\varepsilon' = \varepsilon/3m_d(B(0,R+1))^{\frac{1}{p}}$ there exists by its uniform continuity on $\bar{B}(0,R+1)$ some $\delta \in]0,1[$ such that

$$|g(x-a)-g(x)| \le \varepsilon' 1_{B(0,R+1)}(x) \quad \text{for } |a| < \delta,\, x \in \mathbb{R}^d. \tag{9.3.5}$$

From this we obtain

$$\|\tau_a g - g\|_p \le \varepsilon' m_d(B(0,R+1))^{\frac{1}{p}} = \frac{\varepsilon}{3}. \tag{9.3.6}$$

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

The above result is known as *strong* convergence of translation τ_a to the identity I, which is written $\tau_a \to I$ strongly for $a \to 0$. (Strong convergence of operators is a subject within functional analysis.)

It is noteworthy from the proof how the density of the continuous functions with compact support, i.e. of C_0, 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 $\tau_a \to I$ does not hold in (the norm of) the space $L_\infty(\mathbb{R}^d)$.

As an application of the strong convergence $\tau_a \to I$, this property is now exploited in a proof of uniform continuity of convolutions involving dual exponents:

THEOREM 9.3.2. *When $f \in L_p(\mathbb{R}^d)$, $g \in L_q(\mathbb{R}^d)$ for $p,q \in [1,\infty]$ satisfying $\frac{1}{p}+\frac{1}{q}=1$, then $f*g$ belongs to $C_b(\mathbb{R}^d)$, it is uniformly continuous and*

$$\|f*g\|_\infty \le \|f\|_p\|g\|_q. \tag{9.3.7}$$

PROOF. Since also $y \mapsto f(x-y)$ belongs to $L_p(\mathbb{R}^d)$ for each $x \in \mathbb{R}^d$, it follows from Hölder's inequality that $f(x-\cdot)g(\cdot)$ is integrable and that

$$\forall x \in \mathbb{R}^d\colon\ |f*g(x)| \le \int |f(x-y)g(y)|\,dy \le \|f\|_p\|g\|_q. \tag{9.3.8}$$

This yields the desired estimate. For the uniform continuity we first assume $1 \le p < \infty$. Then the above inequality yields

$$|f*g(x-z) - f*g(x)| \le \int |(\tau_z f(x-y) - f(x-y))g(y)|\,dy \le \|\tau_z f - f\|_p\|g\|_q, \tag{9.3.9}$$

where the right-hand side, regardless of x, goes to 0 for $|z| \to 0$ according to Proposition 9.3.1. For $p = \infty$ one may exchange the roles of f and g in the argument. □

Theorem 9.3.2 reveals one of the fundamental properties of convolution: $f*g$ is always more regular than the two factors f and g. Indeed, even when both f and g in the theorem are discontinuous, their convolution is nonetheless continuous, and uniformly so.

9.4. 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 $L_1(\mathbb{R}^d)$. That would be a Borel function $u(x)$ such that $u*f = f$ would hold for all $f \in L_1(\mathbb{R}^d)$. But it may be seen in various ways, however, that no such unit u exists.

The following is therefore a substitute.

THEOREM 9.4.1. *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*

$$\|f * h_n - f\|_1 = \int_{\mathbb{R}^d} |f * h_n(x) - f(x)|\, dx \to 0 \quad \text{for } n \to \infty. \tag{9.4.1}$$

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,

$$|f * h_n(x) - f(x)| \leq \int |f(x-y) - f(x)| h_n(y)\, dy. \tag{9.4.2}$$

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

$$\begin{aligned} \|f * h_n - f\|_1 &\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y) - f(x)|h_n(y)\, dydx \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y) - f(x)|h_n(y)\, dxdy = \int_{\mathbb{R}^d} \|\tau_y f - f\|_1 h_n(y) dy. \end{aligned} \tag{9.4.3}$$

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 translation invariance, we obtain

$$\int_{|y|\leq\delta} \|\tau_y f - f\|_1 h_n(y) dy \leq \int_{|y|<\delta} \frac{\varepsilon}{2} h_n(y) dy \leq \frac{\varepsilon}{2}, \tag{9.4.4}$$

$$\int_{|y|>\delta} \|\tau_y f - f\|_1 h_n(y) dy \leq 2\|f\|_1 \int_{|y|\geq\delta} h_n(y) dy. \tag{9.4.5}$$

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

$$h_n(x) = n^d h(nx), \quad n \in \mathbb{N}. \tag{9.4.6}$$

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

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

$$h_n(x) = \frac{1}{\pi}\frac{n}{1+n^2x^2}. \tag{9.4.7}$$

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. Nevertheless there are similar, important results that we now describe.

First of all approximative units are members of $\mathscr{L}_1$, hence have well-defined convolutions $f * h_n$ with functions $f \in \mathscr{L}_p$ for $1 \leq p < \infty$ according to Theorem 9.2.2. Thus prepared we turn to approximation of functions in $\mathscr{L}_p$ by convolutions:

THEOREM 9.4.2. *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*

$$\|f*h_n-f\|_p\to 0 \qquad \text{for } n\to\infty. \tag{9.4.8}$$

PROOF. The case $p=1$ was covered in Theorem 9.4.1, 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, as $\int h_n\,dy=1$,

$$\begin{aligned}|f*h_h(x)-f(x)| &\le \int |f(x-y)-f(x)|h_n(y)^{\frac1p}h_n(y)^{\frac1q}\,dy\\ &\le \Big(\int |f(x-y)-f(x)|^p h_n(y)\,dy\Big)^{\frac1p}.\end{aligned} \tag{9.4.9}$$

This implies, by raising to the power p, integrating and applying Tonelli's theorem,

$$\|f*h_n-f\|_p^p\le\int\int|f(x-y)-f(x)|^p h_n(y)\,dxdy=\int\|\tau_yf-f\|_p^p h_n(y)\,dy. \tag{9.4.10}$$

From this inequality, the proof can be completed analogously to the proof of Theorem 9.4.1, using that $\tau_y\to I$ strongly on L_p for $y\to 0$ and that (iii) holds for the h_n. $\square$

This theorem cannot be extended to the case $p=\infty$, for since $f\in L_\infty$ and $h_n\in L_1$, Theorem 9.3.2 yields that $f*h_n$ is in $C_b(\mathbb{R}^d)$—so if it were true that $\|f-f*h_n\|_\infty\to 0$ for $n\to\infty$, it would follow that $(f*h_n)$ is a Cauchy sequence in the uniform norm, whence also $f\in C(\mathbb{R}^d)$; a contradiction of the fact that $C(\mathbb{R}^d)$ is a proper subspace of $L_\infty(\mathbb{R}^d)$.

But as a positive result in this direction one has:

PROPOSITION 9.4.3. *Let $f\in C_b(\mathbb{R}^d)$ and suppose $(h_n)_{n\in\mathbb{N}}$ is is an approximative unit. Then there is pointwise convergence for all $x\in\mathbb{R}^d$,*

$$f*h_n(x)\to f(x) \quad \text{for } n\to\infty. \tag{9.4.11}$$

The convergence is uniform if f is bounded and uniformly continuous.

The proof is left as an exercise.

9.5. Approximation by smooth functions with compact support

For a special purpose we shall return to the function g introduced in the proof of Proposition 8.3.4. In fact, for $a=1/2=-b$ and $c=1=-d$ in (8.3.9) we may write up a nice rotationally symmetric function $\chi\in C_0^\infty(\mathbb{R}^d)$,

$$\chi(x)=g(|x|). \tag{9.5.1}$$

Because the singularity of $|\cdot|$ at $x=0$ is mapped to the interior of a region where g is constant, this function χ is smooth as claimed. Moreover,

$$\operatorname{supp}\chi=\bar B(0,1), \qquad \chi(x)=1 \ \text{ for } x\in\bar B(0,1/2). \tag{9.5.2}$$

In view of this, it is clear that an appoximative unit $(h_n)_{n\in\mathbb{N}}$ via (9.4.6) can be chosen (in many ways) so that it satisfies

$$h_n(x)=n^dh(nx), \quad h\in C_0^\infty(\mathbb{R}^d), \quad \int_{\mathbb{R}^d}h\,dx=1, \tag{9.5.3}$$

$$h_n(x)=0 \text{ for } |x|>\frac1n, \quad 0\le h\le 1, \quad h_n(x)=1 \text{ for } |x|<\frac{1}{2n}. \tag{9.5.4}$$

An elegant way of stating the last line could be that

$$1_{B(0,1/2n)}\le h\le 1_{B(0,1/n)}. \tag{9.5.5}$$

Such a choice of (h_n) is understood in the following.

Thus prepared, one may obtain the next result, which describes during the course of the proof, how any function $f \in L_p(\mathbb{R}^d)$ can be approximated by a convenient sequence of C_0^∞-functions, chosen by truncation and smoothing:

THEOREM 9.5.1. *When $f \in L_p(\mathbb{R}^d)$ for $1 \le p < \infty$ there is a sequence of functions $g_m \in C_0^\infty(\mathbb{R}^d)$ such that*

$$\|f - g_m\|_p \to 0 \quad \text{for } m \to \infty. \tag{9.5.6}$$

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

$$\|f - g_m\|_q \to 0 \quad \text{for } m \to \infty \tag{9.5.7}$$

also holds.

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

$$(f 1_{B(0,N)}) * h_n(x) = \int_{|y|<N} f(y) h_n(x-y)\,dy, \quad N, n \in \mathbb{N}. \tag{9.5.8}$$

Indeed, each of these functions is in C^∞, since the differential operator ∂^α can be applied under the integral sign, using $1_{B(0,R)}|f| n^{d+|\alpha|} \sup|\partial^\alpha h|$ as the majorant on $\mathbb{R}^d$. The support of the convolution is compact, in fact by (9.1.3) it is contained in

$$\bar{B}(0,N) + \bar{B}(0,1) = \bar{B}(0,N+1). \tag{9.5.9}$$

So altogether $(f 1_{B(0,N)}) * h_n$ belongs to C_0^∞.

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

$$\|f - (f 1_{B(0,N)}) * h_n\|_p \le \|f - f 1_{B(0,N)}\|_p + \|f 1_{B(0,N)} - (f 1_{B(0,N)}) * h_n\|_p. \tag{9.5.10}$$

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 9.4.2. Hence $g_m = (f 1_{B(0,N_m)}) * h_{n_m}$ achieves that $g_m \in C_0^\infty$ and

$$\|f - g_m\|_p \le 2^{-m}. \tag{9.5.11}$$

If $f \in L_q$ holds too, one can arrange that also $\|f - f 1_{B(0,N)}\|_q \le \varepsilon/2$ by taking N_m suitably larger (if necessary). Then $\|f 1_{B(0,N_m)} - (f 1_{B(0,N_m)}) * h_{n_m}\| \le \varepsilon/2$ 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.) $\square$

9.6. Young's convolution inequality*

As a general result on convolution of functions in the Lebesgue spaces one has the next extension of Theorem 9.2.1 and Theorem 9.2.2. Like in the proof of the latter, the result follows by deducing a useful pointwise estimate of $f * g$.

THEOREM 9.6.1. *When $f \in \mathscr{L}_p(\mathbb{R}^d)$ and $g \in \mathscr{L}_q(\mathbb{R}^d)$ for $p, q \in [1, \infty]$ satisfying that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ for some $r \in [1, \infty]$, then $f * g$ is defined almost everywhere in $\mathbb{R}^d$, and the induced equivalence class belongs to $L_r(\mathbb{R}^d)$ and fulfils*

$$\|f * g\|_r \le \|f\|_p \|g\|_q. \tag{9.6.1}$$

Such a number r fulfils $r \ge \max(p,q)$, with equality if and only if $\min(p,q) = 1$.

PROOF. The last claim is clear. In view of the previous results, it remains to treat $1 < p < \infty$ and $1 < q < \infty$. As Theorem 9.3.2 applies for $r = \infty$, we may consider $r < \infty$.

We aim at obtaining a pointwise estimate of $|f * g|^r$ in terms of $|f|^p * |g|^q$. To do so we apply Hölder's inequality for three functions, using that

$$1 = \frac{1}{p} + \frac{1}{q} - \frac{1}{r} = \Big(\frac{1}{p} - \frac{1}{r}\Big) + \Big(\frac{1}{q} - \frac{1}{r}\Big) + \frac{1}{r} \tag{9.6.2}$$

where the terms are positive since $r > p$ and $r > q$. Indeed, in the manner described we get for $x \in D(f * g)$,

$$\begin{aligned}\Big|\int f(x-\cdot)g\,dy\Big| &\le \int |f(x-y)|^{1+\frac{p}{r}-\frac{p}{r}}|g(y)|^{1+\frac{q}{r}-\frac{q}{r}}\,dy\\ &= \int (|f(x-y)|^p|g(y)|^q)^{\frac{1}{r}}|f(x-y)|^{\frac{r-p}{r}}|g(y)|^{\frac{r-q}{r}}\,dy \\ &\le \Big(\int |f(x-\cdot)|^p|g|^q\,dy\Big)^{\frac{1}{r}}\Big(\int |f(x-\cdot)|^p\,dy\Big)^{\frac{r-p}{rp}}\Big(\int |g|^q\,dy\Big)^{\frac{r-q}{rq}}.\end{aligned} \tag{9.6.3}$$

For one thing this yields the pointwise estimate, for $x \in D(f * g)$,

$$|f * g(x)|^r \le |f|^p * |g|^q(x)\|f\|_p^{r-p}\|g\|_q^{r-q}. \tag{9.6.4}$$

Secondly, for $x \in D(|f|^p * |g|^q)$ the last and consequently also the second term in (9.6.3) is finite, so such x belong to $D(f * g)$. Hence $D(|f|^p * |g|^q) \subset D(f * g)$, where the former set fills $\mathbb{R}^d$ except for a nullset since $|f|^p, |g|^q \in \mathscr{L}$, and so $f * g$ is defined a.e.

By integrating (9.6.4), after extension by 0 to all $x \in \mathbb{R}^d$, and using that $L_1(\mathbb{R}^d)$ is a Banach algebra, we obtain

$$\begin{aligned}\int |f * g(x)|^r\,dx &\le \|f\|_p^{r-p}\|g\|_q^{r-q}\int |f|^p * |g|^q\,dx \\ &\le \|f\|_p^{r-p}\|g\|_q^{r-q}\big\||f|^p\big\|_1\big\||g|^q\big\|_1 = \|f\|_p^r\|g\|_q^r.\end{aligned} \tag{9.6.5}$$

This yields the stated inequality at once. □

Note that there is not complete freedom: the existence of the number r is part of the assumption in Theorem 9.6.1. This condition on p, q is attributable to the fact that the two factors f, g *combined* must deliver the integrability that makes $f * g$ defined.

To describe an interesting addendum to Young's inequality (9.6.3), one should note that it states that if $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ for some $r \in [1, \infty]$, then there exists a constant $C_{p,q}$ with the following property,

$$\|f * g\|_r \le C_{p,q}\|f\|_p\|g\|_q \qquad \text{for all } f \in L_p(\mathbb{R}^d),\ g \in L_q(\mathbb{R}^d). \tag{9.6.6}$$

Moreover, this constant may by Theorem 9.6.1 be taken as $C_{p,q} = 1$.

But (9.6.6) turns out to be true even for certain constants $C_{p,q} < 1$. Indeed, in 1975 the best constant was found explicitly by Beckner. His result was the following:

THEOREM 9.6.2. *When $p, q, r \in]1, \infty[$ with $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, then the smallest constant $C_{p,q}$ for which* (9.6.6) *holds true is given by*

$$C_{p,q} = \Big(\frac{C_pC_q}{C_r}\Big)^{\frac{d}{2}}, \tag{9.6.7}$$

whereby the generic constant C_p is given by the expressions

$$C_p = \frac{p^{1/p}}{p'^{1/p'}} = p^{\frac{1}{p}}(1-\frac{1}{p})^{1-\frac{1}{p}} \tag{9.6.8}$$

in terms of the dual exponent $p' = 1/(1-1/p)$.

A straightforward calculation shows that the best constant is given by the formula

$$C_{p,q} = \left(\frac{(1-\frac{1}{p})^{1-\frac{1}{p}}}{(\frac{1}{p})^{\frac{1}{p}}}\frac{(1-\frac{1}{q})^{1-\frac{1}{q}}}{(\frac{1}{q})^{\frac{1}{q}}}\frac{(\frac{1}{p}+\frac{1}{q}-1)^{\frac{1}{p}+\frac{1}{q}-1}}{(2-\frac{1}{p}-\frac{1}{q})^{2-\frac{1}{p}-\frac{1}{q}}}.\right)^{\frac{d}{2}}, \tag{9.6.9}$$

It is by no means obvious, of course, that $C_{p,q} < 1$ is valid under the given conditions on p, q, that is, when $1 < \frac{1}{p} + \frac{1}{q} < 2$.

Geometrically the inequality $C_{p,q}^{2/d} < 1$ is attributable to the fact that $f(x) = x^x$ on $[0,1]$ has its graph lying in a certain asymmetric way around the line $x = 1/2$. Analytically one may introduce the auxiliary function

$$g(x) = \frac{(1-x)^{1-x}}{x^x} = \frac{e^{(1-x)\log(1-x)}}{e^{x\log x}} \tag{9.6.10}$$

and observe that for $x = \frac{1}{p}$ and $y = \frac{1}{q}$ the inequality $C_{p,q}^{2/d} < 1$ is equivalent to

$$g(x)g(y) < g(x+y-1) \qquad \text{when } 1 < x+y < 2. \tag{9.6.11}$$

Which for $h(x) = \log g(x) = (1-x)\log(1-x) - x\log x$ also is equivalent to

$$F(x,y) := h(x+y-1) - h(x) - h(y) > 0. \tag{9.6.12}$$

Here one may first observe that $h(x) \to 0^{\pm}$ for $x \to 0^+$ and $x \to 1^-$, respectively, so that $F(x,y) \to 0$ both for $x \to 1$ and $y \to 1$. Moreover, writing out the full expression for $F(x,y)$, it is seen from a cancellation of terms that $F \to 0$ also for $x + y \to 1^+$. This means that F extends to a continuous function, which is defined on the compact triangle T given by $1 \le x+y \le 2$, $x \le 1$, $y \le 1$, and that this F vanishes at the boundary of T.

Insertion shows that $F(\frac{3}{4}, \frac{3}{4}) = 4\log 2 + \frac{3}{2}\log 3 > 0$, so the inequality $F > 0$ holds on the open triangle T° if the equation $\nabla F(x,y) = (0,0)$ only has a single solution in T°. But since $\nabla F = (h'(x+y-1) - h'(x), h'(x+y-1) - h'(y))$, whereby $h'(x) = -2 - \log(x - x^2)$, the critical points solve the equations

$$x - x^2 = x + y - 1 = y - y^2. \tag{9.6.13}$$

These imply that $|x - \frac{1}{2}| = |y - \frac{1}{2}|$, so any solution must satisfy $x = y$. But for $x = y$ one arrives at $x^2 + x - 1 = 0$, having the single solution $x = (\sqrt{5}-1)/2$ in $[0,1]$. Consequently (9.6.12) has been verified.

Altogether this substantiates the startling fact that the best constant in Young's inequality satisfies $C_{p,q} < 1$ for $1 < \frac{1}{p} + \frac{1}{q} < 2$.

To elucidate the nature of the constant $C_{p,q}$, one can simply verify that (9.6.6) is an *identity* in case of the two Gaussian functions, adapted to p and q,

$$f(x) = \exp(-p'|x|^2), \qquad g(x) = \exp(-q'|x|^2). \tag{9.6.14}$$

Setting $E = \int_{\mathbb{R}^d} e^{-|x|^2}\,dx$ for brevity (the value is unimportant), a substitution yields

$$\|f\|_p = (\int e^{-|y|^2}\,dy)^{\frac{1}{p}}(pp')^{-\frac{d}{2p}} = E^{\frac{1}{p}}\left(\frac{p'^{1/p'}}{p^{1/p}}\frac{1}{p'}\right)^{d/2} = E^{\frac{1}{p}}\left(p'C_p\right)^{-d/2}. \tag{9.6.15}$$

Similarly one has $\|g\|_q = E^{\frac{1}{q}}\left(q'C_q\right)^{-d/2}$.

The convolution is elementary to compute, and since $\frac{p'+q'}{p'q'} = \frac{1}{q'} + \frac{1}{p'} = \frac{1}{r'}$, one arrives via completion of a square and translation invariance of Lebesgue measure at

$$\begin{aligned} f * g(x) &= e^{-p'(1-\frac{p'}{p'+q'})|x|^2}\int \exp(-(p'+q')\big|y - \frac{p'}{p'+q'}x\big|^2)\,dy \\ &= e^{-\frac{p'q'}{p'+q'}|x|^2}\int \exp(-(p'+q')|y|^2)\,dy \\ &= e^{-r'|x|^2}(p'+q')^{-d/2}E. \end{aligned} \tag{9.6.16}$$

Since this function also is Gaussian, we get using (9.6.15) once more,

$$\|f * g\|_r = (p'+q')^{-d/2}E^{1+\frac{1}{r}}\left(r'C_r\right)^{-d/2}. \tag{9.6.17}$$

Therefore, when inserting the norms into (9.6.6), then all powers of E cancel since $1+\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, and thus we are left with the inequality, cf. (9.6.8),

$$\left(\frac{p'q'}{p'+q'}\frac{1}{r'}\right)^{d/2}(C_r)^{-d/2} \le C_{p,q}\,(C_p)^{-d/2}\,(C_q)^{-d/2}. \tag{9.6.18}$$

In view of the above formula for r', we may simplify to

$$(C_r)^{-d/2}\,(C_pC_q)^{d/2} \le C_{p,q}. \tag{9.6.19}$$

This is obviously an identity if we insert the expression for $C_{p,q}$ given in Theorem 9.6.2. Another way of expressing this is, of course, that the optimal constant $C_{p,q}$ cannot be smaller than the one given in the theorem.

For the fact that $C_{p,q}$ does satisfy (9.6.6), the reader is referred to Beckner's article.

REMARK 9.6.3. The relation $\frac{1}{p}+\frac{1}{q}=1+\frac{1}{r}$ might seem mysterious at first glance, but it is instructive to note that in terms of the dual exponents it is equivalent to $\frac{1}{p'}+\frac{1}{q'}=\frac{1}{r'}$; which shows that $\|vw\|_{r'}\le\|v\|_{p'}\|w\|_{q'}$ holds because of Hölder's inequality. Indeed, under the additional assumption that $r\ge 2$, or rather that $r'\le 2$, Young's convolution inequality (9.6.6) can be proved easily by combining this case of Hölder's inequality with the boundedness of the Fourier transformation $\mathscr{F}: L_p\to L_{p'}$ (the Hausdorff–Young theorem)—and the best constant $C_{p,q}$ is obtained in this way, as $C_p^{d/2}(2\pi)^{-d/p'}$ is the operator norm of $\mathscr{F}: L_p\to L_{p'}$. This profound fact was observed already by Beckner.

CHAPTER 10

The Fourier–Plancherel transformation

As a major application of the Lebesgue integral, and of the corresponding Lebesgue spaces $L_p(\mathbb{R}^d)$, we shall now study the Fourier transformation, which has fundamental applications in physics, chemistry and statistics as well as a tool for solving partial differential equations.

As usual $x\cdot\xi = x_1\xi_1 + \cdot + x_d\xi_d$ denotes the scalar produkt on $\mathbb{R}^d$, and $|x| = \sqrt{x\cdot x}$ is the induced norm.

DEFINITION 10.0.4. When $f\colon \mathbb{R}^d \to \mathbb{C}$ belongs to $\mathscr{L}_1(\mathbb{R}^d)$, then the Fourier transformed function $\hat{f}\colon \mathbb{R}^d \to \mathbb{C}$ is defined for $\xi \in \mathbb{R}^d$ by the integral

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx. \tag{10.0.20}$$

The map $f \mapsto \hat{f}$ is denoted by $\mathscr{F}$ and is called the *Fourier transformation*.

REMARK 10.0.5. The importance of the Fourier transformed function $\hat{f}$ was illustrated already by Fourier, who claimed for "every" function $f(x)$ that $\hat{f}(\xi)$ should be understood as the amplitude of the pure harmonic oscillation $e^{\mathrm{i}x\cdot\xi}$ in a synthesis of $f(x)$ itself, namely, it holds true that

$$f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{\mathrm{i}x\cdot\xi}\hat{f}(\xi)\,d\xi. \tag{10.0.21}$$

This is *Fourier's inversion formula*. However, the statement was too general to be entirely justifiable (and made in the era before the notion of a function was settled), but it is a virtue of the Lebesgue integral that nowadays one can completely clarify its validity.

First we note that $\hat{f}(\xi)$ is well defined whenever $f \in \mathscr{L}_1(\mathbb{R}^d)$, for the integrand is Borel measurable for each fixed ξ and $x \mapsto e^{-\mathrm{i}x\cdot\xi} f(x)$ has the integrable majorant $|f|$. However, this observation also implies that $\hat{f}(\xi)$ is a continuous function of $\xi \in \mathbb{R}^d$, and we moreover have for all $\xi \in \mathbb{R}^d$

$$\big|\hat{f}(\xi)\big| \le \int |e^{-\mathrm{i}x\cdot\xi} f(x)|\,dx = \|f\|_1. \tag{10.0.22}$$

This clearly shows that $\hat{f}\colon \mathbb{R}^d \to \mathbb{C}$ also is a bounded function, so $\hat{f} \in C_b(\mathbb{R}^d)$. Hence the Fourier transformation is a linear map $\mathscr{F}\colon \mathscr{L}_1(\mathbb{R}^d) \to \mathbb{C}_b(\mathbb{R}^d)$, and the above shows that

$$\sup_{\xi\in\mathbb{R}^d} |\hat{f}(\xi)| \le \|f\|_1. \tag{10.0.23}$$

Now, it is clear that the value of $\hat{f}(\xi)$ remains unchanged even if f is changed on a nullset, so $\hat{f}$ only depends on the equivalence class $[f]$. Consequently there is induced a map $L_1(\mathbb{R}^d) \to C_b(\mathbb{R}^d)$, also called the Fourier transformation, by the formula

$$\mathscr{F}([f]) = \hat{f}. \tag{10.0.24}$$

This shows the main parts of

PROPOSITION 10.0.6. *The Fourier transformation is a linear map*

$$\mathscr{F}\colon L_1(\mathbb{R}^d) \to C_b(\mathbb{R}^d), \tag{10.0.25}$$

which is bounded in the sense that

$$\sup_{\mathbb{R}^d} |\mathscr{F} f| \leq \|f\|_1. \tag{10.0.26}$$

Moreover, $\hat{f} = \mathscr{F} f$ *is uniformly continuous on* $\mathbb{R}^d$.

PROOF. For the uniform continuity we may assume that $\|f\|_1 > 0$ and observe that

$$|\hat{f}(\xi+\eta) - \hat{f}(\xi)| = |\int (e^{-\mathrm{i}x\cdot\eta} - 1)e^{-\mathrm{i}x\cdot\xi} f(x)\,dx| \leq \int_{\operatorname{supp} f} |e^{-\mathrm{i}x\cdot\eta} - 1||f(x)|\,dx. \tag{10.0.27}$$

The last integral is independent of ξ, so uniform continuity seems plausible at least.

If $f \in C_0(\mathbb{R}^d)$ there is some $R > 0$ such that $\operatorname{supp} f \subset B(0,R)$, due to the compact support; and by continuity there is to a given $\varepsilon > 0$ some $\delta > 0$ such that $|e^z - 1| \leq \varepsilon/\|f\|_1$ for $|z| < \delta$. Then the above implies that $|\hat{f}(\xi+\eta) - \hat{f}(\xi)| \leq \varepsilon$ for $|\eta| < \delta/R$, any $\xi \in \mathbb{R}^d$.

For general $f \in L_1(\mathbb{R}^d)$ there exists by Theorem 8.3.5 some function $g \in C_0(\mathbb{R}^d)$ satisfying $\|f - g\|_1 \leq \varepsilon$. For this we have

$$|\hat{f}(\xi+\eta) - \hat{f}(\xi)| \leq 2\varepsilon + |\hat{g}(\xi+\eta) - \hat{g}(\xi)|, \tag{10.0.28}$$

where the last term as above can be shown to be less than ε for $|\eta|$ sufficiently small. □

REMARK 10.0.7. $\mathscr{F}\colon L_1(\mathbb{R}^d) \to C_b(\mathbb{R}^d)$ is actually a *continuous* map between these normed vector spaces, because it is bounded as stated in (10.0.26): if $f_n \to f$ in L_1, then

$$0 \leq \sup |\hat{f}_n - \hat{f}| = \sup |\mathscr{F}(f_n - f)| \leq \|f_n - f\|_1 \to 0, \tag{10.0.29}$$

so $\hat{f}_n \to \hat{f}$ in C_b for $n \to \infty$. Continuity ought to be stated in Proposition 10.0.6, of course, but this transition from boundedness to continuity is just an elementary fact for linear maps.

Among the basic properties there is the addendum that $\hat{f}$ vanishes at infinity:

LEMMA 10.0.8 (Riemann–Lebesgue). *When* $f \in L_1(\mathbb{R}^d)$ *then* $\hat{f}(\xi) \to 0$ *for* $|\xi| \to \infty$.

We postpone the proof of this until later, when a trivial argument will be available.

Instead we proceed to show a fundamental fact about $\mathscr{F}$, namely that it *intertwines* differentiation and multiplication. For simplicity details are given in the 1-dimensional case. But first we need a small lemma, which is of interest in itself.

LEMMA 10.0.9. *When* $f\colon \mathbb{R} \to \mathbb{C}$ *is differentiable with both* f, f' *belonging to* $L_1(\mathbb{R})$, *then* $\lim_{x\to\pm\infty} f(x) = 0$.

PROOF. If $a = \lim_{x\to\infty} f(x)$ exists, then $a = 0$ follows because, for some $N \in \mathbb{R}$, the inequality $|f(x)| \geq \frac{|a|}{2} 1_{]N,\infty[}(x)$ will hold; which is impossible for $a \neq 0$ as it would contradict the integrability of f. Similarly the limit $\lim_{x\to-\infty} f(x)$ must equal 0 if it exists.

Since $f' \in L_1(\mathbb{R})$ it is seen by majorised convergence that $\int_{|x|>n} |f'|\,dx \to 0$ for $n \to \infty$. That is, for every $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that $\int_{|x|\geq N} |f'|\,dx \leq \varepsilon$; whence it holds for $x' < x'' < -N$ and for $N < x' < x''$ that

$$|f(x'') - f(x')| = \Big|\int_{x'}^{x''} f'(x)\,dx\Big| \leq \int_{|x|>N} |f'(x)|\,dx \leq \varepsilon. \tag{10.0.30}$$

Whenever $x_n \to \infty$ for $n \to \infty$ this implies (by substituting x_n, x_m for x', x'') that $(f(x_n))$ is a Cauchy sequence; hence $a = \lim_{n\to\infty} f(x_n)$ exists. For any other sequence $y_n \to \infty$

the same argument shows that $b = \lim_{n\to\infty} f(y_n)$ exists; but $a = b$ since both a and b are cluster points of the Cauchy sequence one arrives at (again via (10.0.30)) by interlacing,

$$f(x_1), f(y_1), f(x_2), f(y_2), \dots, f(x_n), f(y_n), \dots. \tag{10.0.31}$$

Since the sequences are arbitrary, it follows that $\lim_{x\to\infty} f(x)$ exists. Analogously it is shown that also $\lim_{x\to-\infty} f(x)$ exists. □

Now we are ready to prove the fundamental

THEOREM 10.0.10. *Let $f \in L_1(\mathbb{R})$ be given.*

(i) *If $x \mapsto xf(x)$ also belongs to $L_1(\mathbb{R})$, then $\hat{f} \in C^1(\mathbb{R})$ and*

$$\mathscr{F}(xf(x))(\xi) = \mathrm{i}\frac{d\hat{f}}{d\xi}(\xi). \tag{10.0.32}$$

(ii) *If $f \in C^1(\mathbb{R})$ and $f' \in L_1(\mathbb{R})$, then*

$$\mathscr{F}(\frac{df}{dx})(\xi) = \mathrm{i}\xi\hat{f}(\xi). \tag{10.0.33}$$

PROOF. In (i) the integrability of $xf(x)$ allows us to calculate its Fourier transformed function, and we get

$$-\mathrm{i}\mathscr{F}(xf(x))(\xi) = \int -\mathrm{i}xe^{-\mathrm{i}x\cdot\xi} f(x)\,dx = \int \frac{\partial}{\partial\xi}(e^{-\mathrm{i}x\cdot\xi} f(x))\,dx = \frac{d\hat{f}}{d\xi}(\xi). \tag{10.0.34}$$

Indeed, the differentiation with respect to ξ can be taken outside the integral because the functions

$$x \mapsto -\mathrm{i}xe^{-\mathrm{i}x\cdot\xi} f(x), \qquad \xi \in \mathbb{R}^d, \tag{10.0.35}$$

have $|xf(x)|$ as an integrable majorant. Recycling this majorisation, the expression for $\hat{f}{}'$ is also seen to be a continuous function of ξ.

For part (ii) we note that a partial integration for each natural number gives

$$\int_{-n}^{n} e^{-\mathrm{i}x\cdot\xi} f'(x)\,dx = e^{-\mathrm{i}n\xi} f(n) - e^{\mathrm{i}n\xi} f(-n) + \mathrm{i}\xi \int_{-n}^{n} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx. \tag{10.0.36}$$

Here $f(n)$, $f(-n)$ tend to 0 for $n \to \infty$, as by assumption $f, f' \in L_1(\mathbb{R})$; cf. Lemma 10.0.9. Using the Majorised Convergence Theorem we find, again since $f, f' \in L_1(\mathbb{R})$,

$$\mathscr{F}(f')(\xi) = \mathrm{i}\xi\mathscr{F}f(\xi) \tag{10.0.37}$$

by a comparison of the limits on the left- and right-hand side. □

As a final fact for the Fourier transformation of integrable functions, it is mentioned that it plays well together with the convolution studied in Chapter 9:

PROPOSITION 10.0.11. *For functions $f, g \in L_1(\mathbb{R}^d)$ one has that*

$$\mathscr{F}(f * g) = \mathscr{F}f \cdot \mathscr{F}g. \tag{10.0.38}$$

PROOF. As the integrability of $f(x-y)g(y)$ remains unchanged after multiplication by the function $e^{-\mathrm{i}x\cdot\xi}$ in L_∞, it follows from Fubini's theorem that

$$\begin{aligned}
\mathscr{F}(f * g) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} e^{-\mathrm{i}x\cdot\xi} f(x-y)g(y)\,d(x,y) \\
&= \int_{\mathbb{R}^d} g(y)\Big(\int_{\mathbb{R}^d} e^{-\mathrm{i}x\cdot\xi} f(x-y)\,dx\Big)\,dy \\
&= \int_{\mathbb{R}^d} g(y)\Big(\int_{\mathbb{R}^d} e^{-\mathrm{i}(x+y)\cdot\xi} f(x)\,dx\Big)\,dy \\
&= \int_{\mathbb{R}^d} g(y)e^{-\mathrm{i}y\cdot\xi}\Big(\int_{\mathbb{R}^d} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx\Big)\,dy = \mathscr{F}f(\xi)\cdot\mathscr{F}g(\xi).
\end{aligned} \tag{10.0.39}$$

Indeed, the translation invariance of the Lebesgue measure yields the third expression. □

Now, given three functions f, g, h in the Banach algebra $L_1(\mathbb{R}^d)$ both $f*(g*h)$ and $(f*g)*h$ are well defined convolution products in $L_1(\mathbb{R}^d)$, and it is seen from Proposition 10.0.11 that their Fourier transformed functions fulfil

$$\mathscr{F}(f*(g*h)) = \hat{f}\cdot\hat{g}\cdot\hat{h} = \mathscr{F}((f*g)*h). \tag{10.0.40}$$

It is therefore a compelling argument that we must have associativity, as claimed in (9.2.6), of the convolution in the Banach algebra $L_1(\mathbb{R}^d)$,

$$f*(g*h) = (f*g)*h. \tag{10.0.41}$$

This is true, in fact, but the argument requires that *injectivity* of $\mathscr{F}\colon L_1(\mathbb{R}^d)\to C_b(\mathbb{R}^d)$ has been rigorously proved. This we address in the next section, where the property comes well within reach of these notes.

REMARK 10.0.12. *Non-surjectivity* of $\mathscr{F}\colon L_1(\mathbb{R}^d)\to C_b(\mathbb{R}^d)$ requires a much more profound knowledge of the Fourier transformation. Indeed, one of the known proofs involves a consideration of the almost step-like function $g\in C_0(\mathbb{R})$ given in terms of a parameter $a>0$ by the following, where φ is so chosen that g is C^∞ for $\xi>0$,

$$g(\xi) = \begin{cases} \frac{1}{(-\log\xi)^a} & \text{for } 0<\xi<1/2,\\ \varphi(\xi) & \text{for } 1/2\le\xi<1,\\ 0 & \text{for } \xi\notin]0,1[. \end{cases} \tag{10.0.42}$$

This turns out to be the Fourier transformation, in some generalised sense, of a function $f(x)$ that behaves modulo some unimportant lower order terms as $\frac{1}{2\pi\mathrm{i}x(\log x)^a}$ for $x\to\infty$; this does not belong to $L_1(\mathbb{R})$ for $0<a\le 1$. But to conclude that $g\notin\mathscr{F}(L_1)$ one still needs to know that even the generalised Fourier transform is injective. These non-trivial considerations are outside the scope of the present notes (the reader may consult Exercise 7.1.13 in [**Hör85**]).

10.1. The Schwartz space of rapidly decreasing functions

PROOF OF RIEMANN–LEBESGUE'S LEMMA. Since $\mathscr{S}$ is dense in L_1, there is to any given $f\in L_1$ and $\varepsilon>0$ some $g\in\mathscr{S}$ such that $\|f-g\|_1\le\frac{\varepsilon}{2}$. Since $\mathscr{F}\colon L_1(\mathbb{R}^d)\to C_b(\mathbb{R}^d)$ is bounded, we obtain for all ξ,

$$|\hat{f}(\xi)| \le |\mathscr{F}(f-g)(\xi)| + |\hat{g}(\xi)| \le \frac{\varepsilon}{2} + |\hat{g}(\xi)|. \tag{10.1.1}$$

Here $|\hat{g}(\xi)|\le\frac{\varepsilon}{2}$ holds for all $|\xi|\ge R$ for a suitable $R>0$; this follows since $\hat{g}$ as a member of $\mathscr{S}$ is rapidly decreasing. □

The density of the Schwartz space $\mathscr{S}(\mathbb{R}^d)$ in $L_1(\mathbb{R}^d)$ is also useful e.g. for extending the Placherel formula to other settings, such as when $f\in L_1(\mathbb{R}^d)$ and $g\in\mathscr{S}$:

$$\int \hat{f}(\xi)\overline{\hat{g}(\xi)}\,d\xi = (2\pi)^d\int f(x)\overline{g(x)}\,dx. \tag{10.1.2}$$

Indeed, when $\varphi_n\in\mathscr{S}$ are chosen so that $\varphi_n\to f$ in L_1, as we may, then Plancherel's formula gives $\int\hat{\varphi}_n(\xi)\overline{\hat{\psi}(\xi)}\,d\xi = (2\pi)^d\int\varphi_n(x)\overline{\psi(x)}\,dx$. In the limit for $n\to\infty$ the above identity results, using majorised convergence on the left-hand side where $\hat{\varphi}_n(\xi)\to\hat{f}(\xi)$ holds uniformly by the continuity of $\mathscr{F}$ (cf. Remark 10.0.7) whilst $2\|f\|_1|\hat{g}(\xi)|$ is an integrable majorant for all sufficiently large n, since $\sup|\hat{\varphi}_n|\le\|\varphi_n\|_1\le 2\|f\|_1$ holds eventually in view of the continuity of $\|\cdot\|_1$—and Hölder's inequality for the dual pair $(1,\infty)$ on the right-hand side yields that $|\int(f-\varphi_n)\overline{g}\,dx|\le\|f-\varphi_n\|_1\|g\|_\infty\to 0$ for $n\to\infty$.

Thus prepared one may give a simple proof of the following basic result:

PROPOSITION 10.1.1. *$\mathscr{F}\colon L_1(\mathbb{R}^d)\to C_b(\mathbb{R}^d)$ is injective.*

PROOF. Assume that $\mathscr{F} f = 0$ in C_b. Then (10.1.2) gives for $g = \overline{h_n(y-x)}$, where $h_n \in C_0^\infty$ is the special approximative unit in (9.5.3) and $y \in \mathbb{R}^d$ is fixed but arbitrary,

$$0 = (2\pi)^{-d} \int \hat{f} \cdot \overline{\mathscr{F}\overline{h_n}(y-\cdot)}\, d\xi = \int f(x) h_n(y-x)\, dx = f * h_n(y). \tag{10.1.3}$$

Since the last expression is zero for every y, and yet converges to f in L_1 by Theorem 9.4.1, we conclude from Corollary 8.2.6 that $f = 0$ in L_1, as desired. □

10.2. Parseval–Plancherel's theorem

Because of the obvious inclusion $C_0^\infty(\mathbb{R}^d) \subset \mathscr{S}(\mathbb{R}^d)$, it is clear from Theorem 9.5.1 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

$$\mathscr{F}_2 f = \lim_{n\to\infty} \mathscr{F} f_n. \tag{10.2.1}$$

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,

$$\|\mathscr{F} f_n - \mathscr{F} f_m\|_2 = \|\mathscr{F}(f_n - f_m)\|_2 = (2\pi)^d \|f_n - f_m\|_2. \tag{10.2.2}$$

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

$$f_1, g_1, f_2, g_2, \dots, f_n, g_n, \dots \tag{10.2.3}$$

is another Cauchy sequence, which $\mathscr{F}$ by (10.2.2) 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

$$\mathscr{F}_2 \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d) \tag{10.2.4}$$

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 $f, g \in L_2$, we may take sequences $\varphi_n, \psi_n \in \mathscr{S}$ such that $\|f - \varphi_n\|_2 \to 0$ and $\|g - \psi_n\|_2 \to 0$ for $n \to \infty$, and then

$$z\mathscr{F}_2 f + w\mathscr{F}_2 g = z \lim_n \mathscr{F} \varphi_n + w \lim_n \mathscr{F} \psi_n \tag{10.2.5}$$

$$= \lim_n (z\mathscr{F} \varphi_n + w\mathscr{F} \psi_n) = \lim_n \mathscr{F}(z\varphi_n + w\psi_n) = \mathscr{F}_2(zf + wg). \tag{10.2.6}$$

Altogether this means that $\mathscr{F}_2 \colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ is a well-defined linear map by (10.2.1).

It is a fundamental result that the Fourier transformation $\mathscr{F}_2$ on L_2 actually is an *isometry* when using the trick of invoking a suitably weighted Lebesgue measure on $\mathbb{R}^d$. This is a well-known interpretation of the general Parseval–Plancherel formula:

THEOREM 10.2.1. *The Fourier transformation $\mathscr{F} \colon \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ extends in a unique way to a continuous, linear, bijective isometry*

$$\mathscr{F}_2 \colon L_2(m_d) \to L_2((2\pi)^{-d} m_d). \tag{10.2.7}$$

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

$$\int_{\mathbb{R}^d} |f(x)|^2\, dx = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\mathscr{F}_2 f(\xi)|^2\, d\xi, \tag{10.2.8}$$

$$\int_{\mathbb{R}^d} f(x)\overline{g(x)}\, dx = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathscr{F}_2 f(\xi)\overline{\mathscr{F}_2 g(\xi)}\, d\xi. \tag{10.2.9}$$

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

$$\mathscr{F}_2^{-1} = (2\pi)^{-d} \overline{\mathscr{F}}_2. \tag{10.2.10}$$

(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 (10.2.7) because of (10.2.8), which in its turn follows by taking $g = f$ in (10.2.9).

The formula (10.2.9) 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

$$(\varphi_n \mid \psi_n) - (f \mid g) = (\varphi_n - f \mid \psi_n - g) + (f \mid \psi_n - g) + (\varphi_n - f \mid g), \tag{10.2.11}$$

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

$$\begin{aligned} \int_{\mathbb{R}^d} f(x)\overline{g(x)}\,dx &= (f \mid g) = \lim_n (\varphi_n \mid \psi_n) \\ &= \lim_n ((\mathscr{F}\varphi_n \mid \mathscr{F}\psi_n)) = ((\lim_n \mathscr{F}\varphi_n \mid \lim_n \mathscr{F}\psi_n)) \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathscr{F}_2 f(\xi) \overline{\mathscr{F}_2 g(\xi)}\,d\xi. \end{aligned} \tag{10.2.12}$$

Moreover, it is easy to see from (10.2.8) that $\mathscr{F}_2$ is continuous (cf. (10.2.2)).

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

$$\mathscr{F}_2(L_2) = \overline{\mathscr{F}_2(L_2)} \supset \overline{\mathscr{S}} = L_2, \tag{10.2.13}$$

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

$$\widetilde{\mathscr{F}_2} g = \lim_n \mathscr{F}\psi_n = \mathscr{F}_2 g. \tag{10.2.14}$$

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

$$(2\pi)^{-d} \overline{\mathscr{F}}_2 \mathscr{F}_2 \psi_n = \psi_n = \mathscr{F}_2 (2\pi)^{-d} \overline{\mathscr{F}}_2 \psi_n \tag{10.2.15}$$

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

$$(2\pi)^{-d} \overline{\mathscr{F}}_2 \mathscr{F}_2 g = g = \mathscr{F}_2 (2\pi)^{-d} \overline{\mathscr{F}}_2 g. \tag{10.2.16}$$

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 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 9.5.1. This states that there exists a sequence $\psi_n \in C_0^\infty \subset \mathscr{S}$ 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

$$\|\mathscr{F}_2 f - \mathscr{F}\psi_n\|_2 \to 0, \qquad \sup_{\xi\in\mathbb{R}^d} |\mathscr{F}f(\xi) - \mathscr{F}\psi_n(\xi)| \to 0 \tag{10.2.17}$$

However, when a sequence such as $\mathscr{F}\psi_n$ converges both pointwise and in quadratic mean, then the two limit functions coincide a.e.; cf. Corollary 8.2.6. Therefore $\mathscr{F}_2 f(\xi) = \mathscr{F}f(\xi)$ for m_d-almost every $\xi \in \mathbb{R}^d$, so these considerations may be celebrated with the following diagram:

$$\begin{array}{ccc} \mathscr{F}_2\psi_n(\xi) & = & \mathscr{F}\psi_n(\xi) \\ \Big\downarrow & & \Big\downarrow \\ \mathscr{F}_2 f(\xi) & = & \int e^{-\mathrm{i}x\cdot\xi} f(x)\,dx \end{array} \tag{10.2.18}$$

Summing up we have shown:

PROPOSITION 10.2.2. *$\mathscr{F}_2 f = \mathscr{F}f$ for every $f \in L_1(\mathbb{R}^d)\bigcap L_2(\mathbb{R}^d)$, so for such f we have*

$$\mathscr{F}_2 f(\xi) = \int_{\mathbb{R}^d} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx. \tag{10.2.19}$$

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:

$$(2\pi)^{-d/2}\mathscr{F}\colon L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d). \tag{10.2.20}$$

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

REMARK 10.2.3. 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 10.2.2 we have

$$\mathscr{F}(f1_{B(0,N)})(\xi) = \int_{|x|<N} e^{-\mathrm{i}x\cdot\xi} f(x)\,dx. \tag{10.2.21}$$

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 10.2.4. 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 by 2^j, every $t \in \mathbb{R}$ is a limit $t = \lim r_N$ 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,

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

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, cf. (8.3.13), 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 $\overset{\wedge}{\chi} \in C_0^\infty(\mathbb{R}) \subset \mathscr{S}(\mathbb{R})$, 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 $L_\infty(\mathbb{R})$ according to Theorem 9.3.2, 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 $0 < \theta \leq 1$ 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 $j \neq k$.

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 $\theta \leq 1$. □

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.

[Hör85] L. Hörmander, *The analysis of linear partial differential operators*, Grundlehren der mathematischen Wissenschaften, Springer Verlag, Berlin, 1983, 1985.

[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.