Basic Measure Theory

Heinrich v. Weizsäcker

Revised Translation

2004/2008
Fachbereich Mathematik
Technische Universität Kaiserslautern
Preface

The aim of this text is to give a short introduction into the basic concepts of the Lebesgue theory of measure and integration on general spaces in order to enable the reader to follow Master courses in Probability and in Functional analysis. The main advantage of Lebesgue’s theory over more elementary concepts like the Riemann integral is its power in dealing with countable limit operations, both for the measure (viz. generalized volume) of sets, and for the integral of functions. These become essential features in higher real analysis and probability. The price is some additional work in constructing suitable domains of definition for measure and integral.

There are two main approaches to this theory. The first approach starts with measures and then constructs the integrals, in analogy to the classical passage from volume to integration. Conversely, the second approach constructs in several steps the space of integrable functions and then a set has finite measure simply if the indicator function of this set is integrable. In this text we follow the first line, even though it takes a little longer. The main reason for this choice is that in Probability the measurable sets (‘events’) have an important role by themselves and this path of arguments gives additional information about them. The common goal of both approaches are the limit theorems of chapter 4 (in particular the dominated convergence theorem). A reader who knows these limit theorems can start directly with chapter 5 and check the definitions and results of the first four chapters only when needed.

Up to minor modifications this text is a translation of the original German version [?]. Only chapter 2 has been revised more thoroughly. For most mathematical words and phrases we introduce both the English and German term.
Contents

1 Measures 5
  1.1 \( \sigma \)-Algebras 5
  1.2 Measures 7
  1.3 Null-Sets 8
  1.4 An example of a non-measurable set 9

2 Construction of measures 11
  2.1 Set algebras and contents 11
  2.2 Outer measures 14
  2.3 Proof of the Measure Extension Theorem 16
  2.4 Monotone classes and more on uniqueness 19

3 Measurable functions 21

4 The Integral 25
  4.1 The elementary integral 25
  4.2 The Integral for non-negative measurable and for integrable functions 26
  4.3 The Limit Theorems 30

5 Fubini's Theorem 33

6 Convergence in measure 37
  6.1 Comparison with other types of convergence 37
  6.2 A Metric for Local Convergence in Measure 39

7 The \( L^p \)-spaces 43

8 The Radon-Nikodym Theorem 47
  8.1 Absolute continuity of measures 47
  8.2 Lebesgue and Hahn-Jordan decomposition 49
Chapter 1

Measures

1.1 σ-Algebras

A measure measures the size of sets. Not all sets can be measured in this sense as we shall see in section 1.4. On the other hand the class of sets which are measurable should be sufficiently rich. In particular we want to keep measurability of sets if we perform simple operations like taking the complement or taking (countable) unions and intersections. This leads to the following definition.

**Definition 1.1** Let $S$ be a set and let $\mathcal{F}$ be a collection of subsets of $S$. Then $\mathcal{F}$ is called a **σ-algebra** over $S$, if the following holds:

a) $S \in \mathcal{F}$,

b) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$. \(^1\)

c) For every sequence $A_1, A_2, \cdots$ of elements of $\mathcal{F}$ the union $\bigcup_{n=1}^{\infty} A_n$ is also a member (i.e. an element) of $\mathcal{F}$.

**Remark:** 1. Properties a) and b) imply that the empty set $\emptyset$ belongs also to $\mathcal{F}$, since $\emptyset = S^c$.

2. Properties b) and c) imply that for every sequence $A_1, A_2, \cdots$ of elements of $\mathcal{F}$ the intersection $\bigcap_{n=1}^{\infty} A_n$ is in $\mathcal{F}$, because first $A_n^c \in \mathcal{F}$ for each $n$ due to b), hence $\bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F}$ due to c) and finally b) yields

$$\bigcap_{n=1}^{\infty} A_n = \left( \bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F}.$$ 

If we choose $A_n = \emptyset$ for $n \geq 3$ we see in particular that for every pair of sets $A, B \in \mathcal{F}$ the sets $A \cup B$ and $A \cap B$ and hence also the difference set $A \setminus B = A \cap B^c$ belong to $\mathcal{F}$.

3. The largest σ-algebras over a set $S$ is the **power set (Potenzmenge)** $2^S$ of all subsets of $S$. The smallest σ-algebra is the **trivial σ-algebra** which has only $\emptyset$ and $S$ as its elements.

The remainder of this section shows the typical procedure by which most σ-algebras are found. If one has several different σ-algebras over the same set

\(^1\) $A^c$ denotes the complement of $A$ in $S$. 
Let one can consider their intersection, i.e. the system of subsets of $S$ which are members of each of the given $\sigma$-algebras.

**Theorem 1.1** Let $(F_i)_{i \in I}$ be an arbitrary family of $\sigma$-algebras over $S$. Then their intersection $\bigcap_{i \in I} F_i$ is again a $\sigma$-algebra over $S$.

**Proof.** We should verify that the intersection has the three properties a)-c) in Definition 1.1. We restrict ourselves to condition c). Let $A_1, A_2, \ldots$ be a sequence of elements of $\bigcap_{i \in I} F_i$. Thus $A_n \in F_i$ for all $n \in \mathbb{N}$ and all $i \in I$. For fixed $i \in I$ it follows that $\bigcup_{n \in \mathbb{N}} A_n \in F_i$ since $F_i$ is a $\sigma$-algebra. This applies to each $i$, and hence $\bigcup_{n \in \mathbb{N}} A_n \in \bigcap_{i \in I} F_i$.

As a consequence for every system $E$ of subsets of $S$ there is a smallest $\sigma$-algebra which contains $E$.

**Definition 1.2** Let $E$ be any system of subsets of $S$. The intersection of the family of all $\sigma$-algebras which contain $E$ is called the $\sigma$-algebra generated by $E$. It will be denoted by $\sigma(E)$ and $E$ is called a generating system (Erzeugendensystem) of this $\sigma$-algebra.

**Remarks.** 1. Thus a set $A$ belongs to $\sigma(E)$ if and only every $\sigma$-algebra which contains all elements of $E$ must also contain $A$. This concept is similar to the concept of the linear hull of a set $E$ of vectors in a vector space $V$, which is the smallest vector space containing all vectors in $E$ or, the intersection of all linear subspaces of $V$ which contain $E$ as a subset. But while the elements of the linear hull can be described explicitly as the linear combinations of the elements of $E$, an analogous explicit representation of the elements of $\sigma(E)$ by the elements of $E$ is not possible in general.

2. The definition shows: If two systems of sets $E_1, E_2$ satisfy $E_1 \subset \sigma(E_2)$ then we have even $\sigma(E_1) \subset \sigma(E_2)$, because the $\sigma$-algebra $\sigma(E_2)$ is by assumption one of the $\sigma$-algebras whose intersection is $\sigma(E_1)$. In particular the larger a system of sets, the larger the $\sigma$-algebra it generates.

**Example 1.1** Let $Q$ be the system of all compact boxes $Q = [r_1, s_1] \times \cdots \times [r_m, s_m]$ in $\mathbb{R}^m$. Then every open set $U \subset \mathbb{R}^m$ is an element of the $\sigma$-algebra $\sigma(Q)$.

**Proof.** Each point $x \in U$ is in a cube of the form $Q = [r_1, s_1] \times \cdots \times [r_m, s_m]$ with rational numbers $r_i, s_i$, such that even the whole cube $Q$ is a subset of $U$. The set of all cubes with rational sides which are contained in $U$ is countable, hence there is a sequence $(Q_n)_n$ of compact cubes with $U = \bigcup_{n=1}^\infty Q_n$. Since $Q_n \in \sigma(Q)$ and $\sigma(Q)$ is a $\sigma$-Algebra, we get $U \in \sigma(Q)$.

Similarly we can consider the system $R_0$ of (half open) rectangular sets $R = (a_1, b_1] \times \cdots (a_d, b_d]$ in $\mathbb{R}^d$. Then each $R$ is the countable union of the compact boxes $[a_1 + \frac{1}{n}, b_1] \times \cdots [a_d + \frac{1}{n}, b_d]$ and conversely each compact box $[a_1, b_1] \times \cdots [a_d, b_d]$ is in the countable intersection of the half-open sets $R = (a_1 - \frac{1}{n}, b_1] \times \cdots (a_d - \frac{1}{n}, b_d]$ in $\mathbb{R}^d$. This implies that the two systems $Q$ and $R_0$ generate the same $\sigma$-algebra. There are many other generating systems of that $\sigma$-algebra.
Theorem 1.2 The following four systems of sets in \( \mathbb{R}^d \),
the system \( Q \) of all compact boxes,
the system \( R_0 \) of all half-open rectangular sets,
the system \( K \) of all compact subsets
the system \( \mathcal{O} \) of all open subsets
the system \( A \) of all closed subsets
all generate the same \( \sigma \)-algebra, i.e. \( \sigma(Q) = \sigma(R_0) = \sigma(K) = \sigma(\mathcal{O}) = \sigma(A) \).

Proof. The inclusion \( Q \subset K \subset A \) implies \( \sigma(Q) \subset \sigma(K) \subset \sigma(A) \). Since every closed set, being the complement of an open set, is in \( \sigma(\mathcal{O}) \), we have \( A \subset \sigma(\mathcal{O}) \) and the example above shows \( \mathcal{O} \subset \sigma(Q) \). These two observations yield the remaining inclusions \( \sigma(A) \subset \sigma(\mathcal{O}) \subset \sigma(Q) \).

Definition 1.3 Let \( (S, d) \) be a metric space. Let \( \mathcal{O}(S) \) be the system of all open subsets of \( S \). The \( \sigma \)-algebra generated by \( \mathcal{O}(S) \) is called the Borel \( \sigma \)-algebra of \( S \) and is denoted by \( B(S) \). Its elements are called the Borel sets of \( S \).

1.2 Measures

Definition 1.4 Let \( S \) be a set and let \( \mathcal{F} \) be a \( \sigma \)-algebra on \( S \). A map \( \mu : \mathcal{F} \to [0, \infty] \) is called a measure (Maß), if \( \mu(\emptyset) = 0 \) and if \( \mu \) is \( \sigma \)-additive, i.e. for every sequence of pairwise disjoint sets \( F_n \in \mathcal{F} \) one has

\[
\mu \left( \bigcup_{n=1}^{\infty} F_n \right) = \sum_{n=1}^{\infty} \mu(F_n). \tag{1.1}
\]

The simplest example of a measure is the counting measure (Zählmaß). It assigns to each finite set the cardinality of that set and to infinite sets the value \( \infty \). Another example is a point mass or Dirac measure \( \varepsilon_x \) which is defined by \( \varepsilon_x(A) := 1_A(x) \). Here \( 1_A \) is the indicator function of \( A \), i.e. \( \varepsilon_x(A) = 1 \) if \( x \in A \) and \( \varepsilon_x(A) = 0 \) otherwise. There are many more interesting measures, the most important one being Lebesgue measure: The \( d \)-dimensional Lebesgue-measure is a measure on a suitable \( \sigma \)-algebra over \( \mathbb{R}^d \) which coincides with the usual geometric volume for simple sets, e.g. for all boxes. But this fact that usual volume really defines a measure, requires some work. This will be done in the next chapter, in a more general setting.

Definition 1.5 If \( S \) is a set and \( \mathcal{F} \) is a \( \sigma \)-algebra on \( S \), then the pair \( (S, \mathcal{F}) \) is called a measurable space (meßbarer Raum). If moreover \( \mu \) is a measure on \( \mathcal{F} \), the triplet \( (S, \mathcal{F}, \mu) \) is a measure space (Maßraum). The elements of \( \mathcal{F} \) are called measurable sets (meßbare Mengen).

Let us collect some simple properties of measures.

Remarks 1. A measure is also finitely additive (endlich additiv): For disjoint measurable sets \( A \) and \( B \) we have

\[
\mu(A \cup B) = \mu(A) + \mu(B),
\]
just take $F_1 = A, F_2 = B$ and $F_n = 0$ for $n \geq 3$ in (1.1).

2. This implies that a measure is **monotone**, in fact if $A \subset B$ then

$$
\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)
$$

and in particular

$$
\mu(B \setminus A) = \mu(B) - \mu(A).
$$

3. A measure is **sub-additive**: For any two not necessarily disjoint sets $A, B$ in $\mathcal{F}$ one has

$$
\mu(A \cup B) \leq \mu(A) + \mu(B),
$$

because $\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \leq \mu(A) + \mu(B)$.

**Notation.** We write $A_n \uparrow A$ if the sequence $(A_n)$ of sets is monotonically increasing with union $A$; The notation $A_n \downarrow A$ is used similarly.

**Theorem 1.3** a) A measure is $\sigma$-continuous from below: For $A_n \in \mathcal{F}$ the condition $A_n \uparrow A$ implies $\mu(A_n) \rightarrow \mu(A)$.

b) A measure is $\sigma$-continuous from above: The conditions $B_n \downarrow B$ and $\mu(B_1) < \infty$ imply $\mu(B_n) \rightarrow \mu(B)$.

c) A measure is $\sigma$-sub-additive: For any sequence of (not necessarily disjoint) sets $A_n \in \mathcal{F}$

$$
\mu\left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu(A_n).
$$

**Proof.** a) Suppose $A_n \uparrow A$. Consider the sets $F_1 = A_1$ and $F_n = A_n \setminus A_{n-1}$ for $n \geq 2$. The $F_n$ are disjoint and $A_n = \bigcup_{k=1}^{n} F_k$ for every $n \in \mathbb{N}$, so by finite additivity $\mu(A_n) = \sum_{k=1}^{n} \mu(F_k)$. Then

$$
\mu(A) = \mu\left( \bigcup_{n=1}^{\infty} A_n \right) = \mu\left( \bigcup_{k=1}^{\infty} F_k \right) = \sum_{k=1}^{\infty} \mu(F_k)
$$

$$
= \lim_{n \to \infty} \sum_{k=1}^{n} \mu(F_k) = \lim_{n \to \infty} \mu(A_n).
$$

b) Suppose $B_n \downarrow B$ and $\mu(B_1) < \infty$. Choose $A_n = B_1 \setminus B_n$ and $A = B_1 \setminus B$. Then $A_n \uparrow A$ and in $B_1 = A_n \cup B_n = A \cup B$ both unions are disjoint. Thus $\mu(A) + \mu(B) = \mu(A_n) + \mu(B_n)$ for each $n$. According to a) $\mu(A) = \lim_{n \to \infty} \mu(A_n)$. Because of $\mu(B_1) < \infty$ and monotonicity all numbers are finite and we can conclude $\mu(B) = \lim_{n \to \infty} \mu(B_n)$.

c) Put $C_N = \bigcup_{n=1}^{N} A_n$. Sub-additivity gives $\mu(C_N) \leq \sum_{n=1}^{N} \mu(A_n)$. Part a), applied to the $C_N$, allows to pass to the limit as $N \to \infty$. ■

### 1.3 Null-Sets

Null-sets are very useful because for most purposes one can ignore everything that happens on them.

**Definition 1.6** We call a set $A \subset S$ **null-set (Nullmenge)** or, more precisely, a $\mu$-null-set if there is some $N \in \mathcal{F}$ with $\mu(N) = 0$ and $A \subset N$. 

By σ-sub-additivity the union of countably many μ-null sets is again a μ-null set.

Recall the operation of **symmetric difference** between two sets which is defined by

\[ A \triangle B = (A \setminus B) \cup (B \setminus A). \]  

(1.2)

It can also be written as

\[ A \triangle B = (A \cup B) \setminus (A \cap B). \]  

(1.3)

The symmetric difference consists of precisely those points at which the two sets differ, i.e. which are in one of the two sets but not in the other. If the symmetric difference \( A \triangle B \) of two sets is a μ-null-set then \( A \) and \( B \) are called μ-eqivalent. If moreover both sets are measurable then \( \mu(A \cup B) = \mu(A \cap B) = \mu(A) = \mu(B) \).

**Definition 1.7** A measure \( \mu \), and the associated measure space \( (S, \mathcal{F}, \mu) \) are **complete (vollständig)**, if every μ-null set belongs to the σ-algebra \( \mathcal{F} \).

We note in passing that every measure can be completed.

**Proposition 1.1** Let \( (S, \mathcal{F}, \mu) \) be a measure space. Let

\[ \mathcal{F}_\mu = \{ A' \subset S : A' \text{ is } \mu \text{-equivalent to some } A \in \mathcal{F} \}. \]

Extend \( \mu \) to \( \mathcal{F}_\mu \) by letting \( \mu(A') = \mu(A) \) if \( A \) and \( A' \) are μ-equivalent. Then \( (S, \mathcal{F}_\mu, \mu) \) is a complete measure space. It is called the **completion** of \( (S, \mathcal{F}, \mu) \), and the σ-algebra \( \mathcal{F}_\mu \) is the **μ-completion** of \( \mathcal{F} \).

**Proof.** The discussion preceding definition 1.7 shows that the extension of \( \mu \) to \( \mathcal{F}_\mu \) is unambiguous. Two sets \( A' \) and \( A \in \mathcal{F} \) differ at exactly the same points at which the complements \( A^{c} \) and \( A^{c} \) differ. This implies that \( \mathcal{F}_\mu \) is stable under complements. If \( (A'_n) \) is a sequence of sets in \( \mathcal{F}_\mu \) then there are sets \( A_n \in \mathcal{F} \) such that \( A'_n \triangle A_n \subset N_n \in \mathcal{F} \) where \( \mu(N_n) = 0 \) for each \( n \). All points at which the set \( A' = \bigcup_n A'_n \) differs from \( A = \bigcup_n A_n \in \mathcal{F} \) are contained in the μ-null-set \( N = \bigcup N_n \) and thus \( \bigcup_n A'_n \in \mathcal{F}_\mu \). So \( \mathcal{F}_\mu \) is a σ-algebra. If the \( A'_n \) are disjoint then the sets \( A_n \cap N^{c} = A'_n \cap N^{c} \) are also disjoint and hence

\[
\sum_{n=1}^{\infty} \mu(A'_n) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n \cap N^{c}) = \mu(A \cap N^{c}) = \mu(A) = \mu(A').
\]

Thus the extension of \( \mu \) to \( \mathcal{F}_\mu \) is a measure which clearly is complete. \( \blacksquare \)

### 1.4 An example of a non-measurable set

The concept of a σ-algebra serves primarily for having suitable domains of definition for measures. Why not define a measure \( \mu \) on all subsets \( A \) of the underlying space? In this section we want to show that this is not possible for 'Lebesgue measure' on the real line, more precisely for a finite measure on an interval with the very natural additional condition of translation-invariance.
Theorem 1.4 (Vitali set) There is a set $E \subset [0,1]$ with the following property: Let $\lambda$ be a measure on a $\sigma$-algebra $\mathcal{L}$ over $[0,1]$ such that $\lambda([0,1]) = 1$ and for all $F \in \mathcal{L}$ we have $F + x \mod 1 \in \mathcal{L}$ for $x \in [0,1]$ and
\begin{equation}
\lambda(F + x \mod 1) = \lambda(F) \tag{1.4}
\end{equation}
Then $E \notin \mathcal{L}$.

Proof. Call two numbers $x, y \in [0,1]$ similar if they differ by a rational number. Clearly this is an equivalence relation and thus we get a partition of the interval into disjoint equivalence classes. Now let $E$ be a set which contains exactly one element out of each equivalence class. In order to see that such a set cannot be in $\mathcal{L}$ denote for all rational numbers $r$ with $0 \leq r < 1$ by $E_r$ the set $E + r \mod 1$. Then the interval $[0,1]$ is the (countable) disjoint union of the sets $E_r$, and $E_r \in \mathcal{L}$ provided that $E \in \mathcal{L}$.
Assuming $E \in \mathcal{L}$, we would get from translation-invariance
\[
\sum_r \lambda(E) = \sum_r \lambda(E_r) = \lambda \left( \bigcup_r E_r \right) = \lambda([0,1]) = 1.
\]
The right-hand side is positive and finite. But on the left we have added infinitely many times the same value $\lambda(E)$. Both for $\lambda(E) = 0$ and for $\lambda(E) > 0$ we get a contradiction.

Remark 1.1 In the above situation every $\mathcal{L}$-measurable subset $F$ of $E$ must be a null set.

Proof. The sets $F_r = F + r \mod 1$ are disjoint and therefore as above
\[
\sum_r \lambda(F) = \sum_r \lambda(F_r) = \lambda \left( \bigcup_r F_r \right) \leq \lambda([0,1]) = 1
\]
and hence $\lambda(F) = 0$.

\[\text{2i.e. } E \text{ is not 'Lebesgue-measurable'}\]
Chapter 2

Construction of measures

Even though it is clear how to measure the size of simple sets by volume or similar finitely additive functionals, it is not obvious that these definitions can be extended to a measure on a whole σ-algebra of sets. The main result in this chapter is Carathéodory’s extension theorem (Theorem 2.2). As a special case this contains the construction of Lebesgue measure in Euclidean space.

There are several approaches to this theorem. Our proof is based on an approximation argument. It can easily be reframed to work also in the functional analytic context.\footnote{Carathéodory’s original proof is useful in the construction of Hausdorff measures in the theory of fractals, cf. e.g. the script \cite{fractal} “Fractal sets and Preparation to Geometric Measure Theory” (http://www.mathematik.uni-kl.de/~wwwstoch/2002w/skriptrev.pdf). A proof via approximation from inside is also possible, see the script \cite{fractal}}

2.1 Set algebras and contents

The challenge in the construction of measures is their continuity when passing to countable unions. We consider first preliminary forms of the concepts of Chapter 1 in which only finite set operations are allowed.

Definition 2.1 Let $S$ be a set. Let $\mathcal{A}$ be a collection of subsets of $S$. Then $\mathcal{A}$ is called a (set) algebra ((Mengen-)Algebra) on $S$, if the following hold:

a) $S \in \mathcal{A}$,

b) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$.

c) $\mathcal{A}$ is $\cup$-stable: $A, B \in \mathcal{A}$ imply $A \cup B \in \mathcal{A}$.

Remark: Like in the case of σ-algebras these conditions automatically imply that $\emptyset \in \mathcal{A}$ and $\mathcal{A}$ is also $\cap$-stable. Every countable union of elements of $\mathcal{A}$ can be converted to a countable disjoint union of elements of $\mathcal{A}$ via the formula

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup \bigcup_{n=2}^{\infty} (A_n \cap \bigcup_{m=1}^{n-1} A_m^c).$$

This shows that an algebra is a σ-algebra if and only if it is closed under forming countable disjoint sequences.
Example 2.1 Let $A$ be the set algebra on $\mathbb{R}$ which is generated by all intervals which are closed on the right and open on the left. Then $A$ consists precisely of all sets of the form
\[ A = \bigcup_{i=1}^{n} J_i \]  
where each $J_i$ is an interval of one of the three types $(-\infty, b]$, $(a, b]$ or $(a, \infty)$. It is called the interval algebra over $\mathbb{R}$.

Sometimes it is convenient not to require that the underlying set $S$ itself belong to the system of sets. The 'local' version of a set algebra is called a 'ring':

Definition 2.2 Let $S$ be a set and let $\mathcal{R}$ be a non-empty system of subsets of $S$. Then $\mathcal{R}$ is called a ring ((Mengen-)Ring) on $S$, if the following hold:

b') For $A, B \in \mathcal{R}$ the difference set $A \setminus B = A \cap B^c$ also belongs to $\mathcal{R}$.

c') For $A, B \in \mathcal{R}$ the union $A \cup B$ is in $\mathcal{R}$.

A ring $\mathcal{R}$ is also stable under the operation $\Delta$ (cf. 1.2). Moreover (1.3) shows that $A \cap B = (A \cup B) \setminus (A \Delta B)$ and hence a ring is stable under intersections. If one takes $\Delta$ as addition and $\cap$ as multiplication then a ring in our sense becomes a 'commutative ring' in the algebraic sense.\(^2\)

Example 2.2 An example of a ring is the interval ring $\mathcal{R}$ of all bounded sets in the interval algebra $A$ of Example 2.1.

Obviously a ring is a set algebra if and only if it has the set $S$ as an element.

Definition 2.3 Let $S$ be a set and let $\mathcal{R}$ be a ring of subsets of $S$. A map $m : A \rightarrow [0, \infty]$ is a content (Inhalt), if $m(\emptyset) = 0$ and $m$ is additive, i.e.
\[ m(A \cup B) = m(A) + m(B) \]  
holds for any pair of disjoint sets $A, B$ in $\mathcal{A}$.

Remarks: 1. In analogy to measures contents are monotone and finitely sub-additive.

2. For any two elements $A, B$ of the ring $\mathcal{R}$ the number $m(A \Delta B)$ quantifies how much the two sets differ. In the proof of the extension theorem 2.2 we shall make use of this idea.

3. The formula (2.2) has a natural extension for not necessarily disjoint sets:
\[ m(A \cup B) + m(A \cap B) = m(A) + m(B). \]  
In fact both sides are equal to $m(A) + m(B \setminus A) + m(A \cap B)$. A set function with (2.3) is called modular.

\(^2\)Here is a short way to see this: identify a set $A$ with its indicator function $1_A$ and note that $1_{A \cap B} = 1_A 1_B$ and $1_{A \setminus B} \equiv 1_A + 1_B \pmod{2}$. A set algebra becomes a 'commutative ring with 1' (the set $S$ is the unit). In older literature set algebras are called 'fields' ('Mengen-Körper'). The associated algebraic structures are called Boolean rings resp. algebras.
Example 2.3 Let $A$ (resp. $R$) be the above algebra (resp. ring) of unions of intervals. One can assume that in the representation (2.1) the intervals $J_i$ are disjoint. If $F : \mathbb{R} \to \mathbb{R}$ is a non-decreasing function then we get a content $m_F$ on $A$ resp. $R$ by the formula

$$m_F(\bigcup_{i=1}^{n} J_i) = \sum_{i=1}^{n} F(b_i) - F(a_i),$$

(2.4)

where for unbounded intervals we use the limit values $F(-\infty) = \lim_{x \to -\infty} F(x)$ and $F(\infty) = \lim_{x \to \infty} F(x)$ in the obvious way. It is called **Stieltjes content**.

For $F(x) = x$ we get the usual geometric size of a set based on the length of intervals.

Clearly if a content can be extended to a measure it must be $\sigma$-additive for all sets for which it is defined.

**Definition 2.4** A content $m$ on a ring $R$ is called $\sigma$-additive if for every disjoint sequence $(A_n)$ in $R$ whose union is in $R$, one has

$$m(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} m(A_n).$$

(2.5)

So a measure is simply a $\sigma$-additive content on a $\sigma$-algebra. In view of the extension theorem (theorem 2.2) below, sometimes a $\sigma$-additive content on a ring is also called a **pre-measure** (Prämaß).

In most cases the actual verification of the $\sigma$-additivity implicitly depends on a compactness argument like in part b) of the following criterion.

**Theorem 2.1** A content $m$ on a ring $R$ of subsets of $S$ is $\sigma$-additive if one of the following conditions hold:

a) $m$ takes only finite values and is continuous at $\emptyset$, i.e. if $R_n \in R$, and $R_n \downarrow \emptyset$ then $m(R_n) \downarrow 0$.

b) $S$ is a topological (e.g. metric) space and for each $R \in R$ and every $\varepsilon > 0$ there are two sets $B, C$ with $B \subset C \subset R$ such that $B \in R$, $C$ is compact and $m(R) \leq m(B) + \varepsilon < \infty$.

**Proof.** a) Let a disjoint sequence $(A_n)$ in $R$ be given whose union $A$ is in $R$. Let $R_n = A \setminus (\bigcup_{k=1}^{n} A_k)$. Then $R_n \downarrow \emptyset$ and hence

$$m(A) - \sum_{k=1}^{n} m(A_k) = m(A) - m(\bigcup_{k=1}^{n} A_k) = m(R_n)$$

converges to 0 as $n \to \infty$. This implies the $\sigma$-additivity (2.5).

b) Under the given assumptions we verify condition a). Let $(R_n)$ be a sequence of elements of $R$ with $R_n \downarrow \emptyset$. Let $\varepsilon > 0$. For each $n$ choose the sets $B_n \in R$, $C_n$ compact with $B_n \subset C_n \subset R_n$ and $m(R_n) \leq m(B_n) + \varepsilon 2^{-n}$. Then also $\bigcap_{n=1}^{\infty} C_n = \emptyset$ and by compactness there is already some $N \in \mathbb{N}$ with $\bigcap_{n=1}^{N} C_n = \emptyset$. In particular $\bigcap_{n=1}^{N} B_n = \emptyset$. Thus

$$R_N = R_N \cap \bigcup_{n=1}^{N} B_n \subset \bigcup_{n=1}^{N} R_n \setminus B_n$$
and hence for all $k \geq N$

$$m(R_k) \leq m(R_N) \leq \sum_{n=1}^{N} m(R_n) - \varepsilon 2^{-n} < \varepsilon.$$ 

This implies $m(R_n) \to 0$ as required.

\begin{corollary}
Let $F : \mathbb{R} \to \mathbb{R}$ be nondecreasing and right-continuous. Then the associated Stieltjes content $m_F$ on the interval ring $\mathcal{R}$ is $\sigma$-additive.
\end{corollary}

\begin{proof}
We check condition $b)$ in theorem 2.1. Let $R \in \mathcal{R}$ and $\varepsilon > 0$ be given. First let $R = (a, b]$. By right-continuity of $F$ we can choose points $c, d$ with $a < c < d < b$ and $F(a) \leq F(d) < F(a) + \varepsilon$. Let $B = (d, b]$ and $C = [c, b]$. Then $B \subset C \subset R$, $B \in \mathcal{R}$ and $C$ is compact. Moreover

$$m_F(B) = m_F((d, b]) = F(b) - F(d) > F(b) - F(a) - \varepsilon = m_F(R) - \varepsilon.$$ 

The additivity of $m_F$ allows to extend this argument to finite unions of intervals, i.e. to all elements of $\mathcal{R}$.
\end{proof}

In an analogous way one easily verifies the $\sigma$-additivity of the $n$-dimensional volume function on the ring of finite unions of rectangles of the form

$$R = (a_1, b_1] \times \cdots (a_n, b_n).$$

\begin{definition}
A content $m$ on a ring $\mathcal{R}$ over the set $S$ is called $\sigma$-finite, if $S$ is the union of an increasing sequence $(E_k)$ of elements of $\mathcal{R}$ with $m(E_k) < \infty$.
\end{definition}

This is a very mild condition. For example the volume function on the ring generated by rectangular sets in $\mathbb{R}^d$ is $\sigma$-finite since $\mathbb{R}^d = \bigcup_k (-k, k)^d$. We need this condition in order to ensure uniqueness in the following theorem which is the central result in this chapter.

\begin{theorem}
Let $m$ be a $\sigma$-additive and $\sigma$-finite content on the ring $\mathcal{R}$. Then there is a unique extension of $m$ to a measure on $\sigma(\mathcal{R})$.
\end{theorem}

The remainder of this chapter is devoted to the proof of this theorem and some additional comments. In a first round the reader may pass directly to the next chapter.

\section{Outer measures}

We saw that typically measures cannot be defined on all subsets of the underlying space. If we relax the condition of $\sigma$-additivity, this difficulty disappears.

\begin{definition}
Let $S$ be a set. A map $\mu^*$ from the power set $2^S$ to $[0, \infty]$ is called an outer measure (äußeres Maß) on $S$, if it has the following properties:

1. $\mu^*(\emptyset) = 0$,

2. $A \subset B$ implies $\mu^*(A) \leq \mu^*(B)$ (monotonicity),
\end{definition}
3. \( \mu^*(\bigcup_{n=1}^{\infty} E_n) \leq \sum_{n=1}^{\infty} \mu^*(E_n) \) (\( \sigma \)-sub-additivity).

Outer measures often are generated as follows.

**Proposition 2.1** Let \( \mathcal{R} \) be any system of subsets of \( S \) which contains \( \emptyset \), and let \( m : \mathcal{R} \to [0, \infty] \) be a set function with \( m(\emptyset) = 0 \). Let

\[
\mu^*(A) := \inf \left\{ \sum_{i=1}^{\infty} m(R_i) : R_i \in \mathcal{R}, A \subseteq \bigcup_{i=1}^{\infty} R_i \right\}
\]  

(2.6)

where \( \mu^*(A) = \infty \) if there is no covering of \( A \) by a sequence of elements of \( \mathcal{R} \). Then \( \mu^* \) is an outer measure and \( \mu^*(R) \leq m(R) \) for all \( R \in \mathcal{R} \).

**Proof.** Obviously \( \mu^*(\emptyset) = 0 \). In order to check monotonicity let \( A \subseteq B \). Then in (2.6) for \( A \) the infimum is taken over a larger class of numbers than for \( B \). Thus \( \mu^*(A) \leq \mu^*(B) \). For \( \sigma \)-sub-additivity let now \((E_n)\) be a sequence of sets and let \( \varepsilon > 0 \). For each \( n \) we choose a sequence \((R_i^n)\) of elements of \( G \) satisfying \( E_n \subseteq \bigcup_{i=1}^{\infty} R_i^n \) and

\[
\sum_{i=1}^{\infty} m(R_i^n) \leq \mu^*(E_n) + \varepsilon 2^{-n}.
\]

Then the set \( \bigcup_{n=1}^{\infty} E_n \) is contained in the set \( \bigcup_{i,n=1}^{\infty} R_i^n \) and hence

\[
\mu^*(\bigcup_{n=1}^{\infty} E_n) \leq \sum_{i,n=1}^{\infty} m(R_i^n) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \varepsilon.
\]

Thus all properties of an outer measure are verified. The final inequality in the assertion is clear since every \( R \in \mathcal{R} \) forms an admissible covering of itself on the right hand side of (2.6).

In the special case of a \( \sigma \)-additive content on a ring \( \mathcal{R} \) the restriction of the outer measure \( \mu^* \) defined in (2.6) to \( \mathcal{R} \) actually coincides with \( m \). This will be important below.

**Lemma 2.1** Let \( \mathcal{R} \) be a ring and let \( m \) be a \( \sigma \)-additive content. Then

\[
\mu^{|\mathcal{R}} = m.
\]

**Proof.** Let \( A \in \mathcal{R} \) be given. Suppose \( A \subseteq \bigcup_{i} R_i \) and \( R_i \in \mathcal{R} \). Then \( A = \bigcup_{i} A \cap R_i \). As \( m \) is \( \sigma \)-sub-additive and monotone we get

\[
m(A) \leq \sum_{i=1}^{\infty} m(A \cap R_i) \leq \sum_{i=1}^{\infty} m(R_i)\]

Taking the infimum over these coverings gives \( m(A) \leq \mu^*(A) \). The converse inequality was part of the proposition.

Finally every outer measure induces a distance (Abstand) between subsets of \( S \).
Lemma 2.2 Let $\mu^*$ be an outer measure on $S$ and define for $A, B \subset S$ the (possibly infinite) 'distance' $d(A, B)$ by

$$d(A, B) = \mu^*(A \triangle B).$$

(2.7)

Then the triangular inequality

$$d(A, C) \leq d(A, B) + d(B, C).$$

(2.8)

holds. Moreover the function $\mu^*$ and the set operations $\cup, \cap, \setminus$ are continuous for $d$: If both $d(A_n, A) \to 0$ and $d(B_n, B) \to 0$ then $\mu^*(A_n) \to \mu^*(A)$, $d(A_n \cup B_n, A \cup B) \to 0$, and similarly for $\cap$ and $\setminus$.

Proof. Obviously for three sets $A, B, C$ one has $A \triangle C \subset (A \triangle B) \cup (B \triangle C)$. Since $\mu^*$ is monotone and sub-additive we get the triangular inequality. The triangular inequality yields in particular the inequality in

$$|\mu^*(A) - \mu^*(B)| = |d(\emptyset, A) - d(\emptyset, B)| \leq d(A, B) = \mu^*(A \triangle B)$$

(2.9)

which in turn implies the continuity of $\mu^*$. Let now $\circ$ denote one of the above set operations. Let $A, B$ and $A', B'$ be two pairs of sets. If a point $x$ belongs to $A \circ B$ but not to $A' \circ B'$ or conversely, then $x$ is at least in one of the symmetric difference sets $A \triangle A'$ or $B \triangle B'$. More formally

$$(A \circ B) \triangle (A' \circ B') \subset (A \triangle A') \cup (B \triangle B').$$

Then again the fact that $\mu^*$ is monotone and sub-additive implies

$$d(A \circ B, A' \circ B') \leq d(A, A') + d(B, B').$$

This proves the continuity.

2.3 Proof of the Measure Extension Theorem

Actually the proof of theorem 2.2 yields more detailed information which sometimes is useful. The existence proof does not need $\sigma$-finiteness. It yields a complete measure with a useful approximation property. The uniqueness part of theorem 2.2 will be shown in remark 2.2 below. A much stronger uniqueness result is established in the next (optional) section.

Theorem 2.3 Let $m$ be a $\sigma$-additive content on the ring $\mathcal{R}$. Let $\mu^*$ be the associated outer measure given by proposition 2.1. Let $\mathcal{N} = \{N \subset S : \mu^*(N) = 0\}$ be the class of $\mu^*$-null-sets. Let $\mathcal{M} = \sigma(\mathcal{R} \cup \mathcal{N})$ and let $\mu = \mu^*|_{\mathcal{M}}$ be the restriction of $\mu^*$ to the $\sigma$-algebra $\mathcal{M}$. Then

a) $\mu$ is a complete measure which extends $m$ i.e. $\mu|_{\mathcal{R}} = m$.

b) For every $A \in \mathcal{M}$ with $\mu(A) < \infty$ there is a sequence $(A_n)$ in $\mathcal{R}$ with $\mu(A \triangle A_n) \to 0$.

Proof. I 1. Let $d$ denote the 'distance' on the power set of $S$ which is associated to $\mu^*$, cf. (2.7). Let $\mathcal{R}_f = \{R \in \mathcal{R} : m(R) < \infty\}$ and let $\mathcal{M}_f$ be the system of all sets $A \subset S$ which are in the closure of $\mathcal{R}_f$ with respect to $d$, i.e. $A \in \mathcal{M}_f$ iff there is a sequence $(R_n)$ in $\mathcal{R}_f$ with $d(R_n, A) \to 0$; it
is then clear that $d$ is finite on $\mathcal{M}_f$. Obviously $\mathcal{R}_f \subset \mathcal{M}_f$ and by lemma 2.2 $\mu^*(A) = \lim m(R_n) < \infty$ for each $A \in \mathcal{M}_f$. Moreover $\mathcal{N} \subset \mathcal{M}_f$ since every $\mu^*$ null-set $N$ satisfies $d(\emptyset, N) = \mu^*(N) = 0$.

2. Clearly $\mathcal{R}_f$ is a ring and as a consequence of lemma 2.2 $\mathcal{M}_f$ is a ring as well: If two sets $A$ and $B$ can be approximated by elements of the ring $\mathcal{R}_f$, then the same is true for $A \cup B$ and $A \setminus B$.

3. $\mu^*_|\mathcal{M}_f$ is a content: Let $A, B \in \mathcal{M}_f$ such that $A \cap B = \emptyset$. Choose in $\mathcal{R}$ approximating sequences $(A_n)$ for $A$ and $(B_n)$ for $B$. Then by lemmas 2.1 and 2.2 and the modularity (2.3) of $m$

$$\mu^*(A \cup B) = \lim m(A_n \cup B_n) = \lim m(A_n) + \lim m(B_n) - \lim m(A_n \cap B_n) = \mu^*(A) + \mu^*(B) + \mu^*(\emptyset) = \mu^*(A) + \mu^*(B).$$

4. We claim that $\mu^*_|\mathcal{M}_f$ is $\sigma$-additive, and moreover for every disjoint sequence $(A_n)$ in $\mathcal{M}_f$ the union $A = \bigcup A_n$ also belongs to $\mathcal{M}_f$ if and only if $\sum_{n=1}^{\infty} \mu^*(A_n) < \infty$. Clearly if $A \in \mathcal{M}_f$ (or more generally if $\mu^*(A) < \infty$) then by step 2 the partial sums of the series are all dominated by $\mu^*(A)$ and the series converges. Conversely, assume that the series converges. Then

$$d(A, \bigcup_{n=1}^{N} A_n) = \mu^*(A \triangle \bigcup_{n=1}^{N} A_n) = \mu^*(\bigcup_{n=N+1}^{\infty} A_n) \leq \sum_{n=N+1}^{\infty} \mu^*(A_n) \to 0$$

as $N \to \infty$. Therefore $A$ can be approximated by elements of $\mathcal{M}_f$. By the triangular inequality and the definition of $\mathcal{M}_f$ the set $A$ then can also be approximated by elements of $\mathcal{R}_f$ and as a result we get $A \in \mathcal{M}_f$. Moreover

$$\mu^*(A) = \lim_{N \to \infty} \mu^*(\bigcup_{n=1}^{N} A_n) = \lim_{N \to \infty} \sum_{n=1}^{N} \mu^*(A_n) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

6. We note in passing that in the case where $\mathcal{R}$ is an algebra and $m(S) < \infty$ the proof is now complete: The class $\mathcal{M}_f$ then is an algebra according to step 2. By step 3, the series in the statement of step 4 always are bounded and thus $\mathcal{M}_f$ is closed under countable disjoint unions. Hence $\mathcal{M}_f$ is a $\sigma$-algebra which contains both $\mathcal{R} = \mathcal{R}_f$ and $\mathcal{N}$, i.e. $\sigma(\mathcal{R} \cup \mathcal{N}) \subset \mathcal{M}_f$. Further $\mu^*$ is $\sigma$-additive on $\mathcal{M}_f$, in particular it induces a measure extension of $m$ on $\sigma(\mathcal{R})$ which is complete because $\mathcal{N}$ is closed under taking subsets.

II. 7. The remainder of the proof is devoted to the extension to the unbounded case. Consider the system of sets

$$\mathcal{M} = \{A \subset S : A \cap B \in \mathcal{M}_f \text{ for all } B \in \mathcal{M}_f\}.$$  

(2.10)

Then $\mathcal{M}$ contains $\mathcal{R}$. In fact let $R \in \mathcal{R}$ and let $B \in \mathcal{M}_f$. By definition of $\mathcal{M}_f$ there is an approximating sequence $(R_n)$ in $\mathcal{R}_f$ for $B$. Then $(R \cap R_n)$ is an approximating sequence for $R \cap B$, i.e. $R \cap B \in \mathcal{M}_f$. Therefore $R \in \mathcal{M}$. Moreover $\mathcal{N} \subset \mathcal{M}$ since $B \cap \mathcal{N}$ is again in $\mathcal{N}$ and hence in $\mathcal{M}_f$ for all $B$ in $\mathcal{M}_f$.\footnote{i.e. for most applications in probability}
8. We claim that $\mathcal{M}$ is a $\sigma$-algebra. Clearly $S \in \mathcal{M}$. Also $\mathcal{M}$ is a ring since $\mathcal{M}_f$ is a ring. So $\mathcal{M}$ is an algebra. Let now a disjoint sequence $(A_n)$ in $\mathcal{M}$ be given with union $A = \bigcup A_n$. For every $B \in \mathcal{M}_f$ the additivity of $\mu^*$ on $\mathcal{M}_f$ (step 3) gives

$$\sum_{n=1}^{\infty} \mu^*(A_n \cap B) = \lim_{N} \sum_{n=1}^{N} \mu^*(A_n \cap B) = \lim_{N} \mu^*(\bigcup_{n=1}^{N} A_n \cap B) \leq \mu^*(B) < \infty$$

and thus by step 4 $A \cap B = \bigcup A_n \cap B \in \mathcal{M}_f$. This holds for every $B \in \mathcal{M}_f$ and so $A \in \mathcal{M}$. As an algebra which is closed under countable disjoint unions $\mathcal{M}$ is a $\sigma$-algebra.

9. We claim that every set $A \in \mathcal{M}$ with $\mu^*(A) < \infty$ is in $\mathcal{M}_f$. By definition of $\mu^*$ there is a sequence $(B_i)$ in $\mathcal{R}$ such that $A \subseteq \bigcup B_i$ and $\sum m(B_i) \leq \mu^*(A) + 1 < \infty$. Replacing $B_i$ by $B_i \setminus \bigcup_{j<i} B_j$ if necessary we may assume that the $B_i$ are disjoint. Then again the partial sums of the series $\sum_i \mu^*(A \cap B_i)$ are bounded and hence step 4 implies that $A = \bigcup A \cap B_i \in \mathcal{M}_f$.

10. The restriction $\mu^*_|\mathcal{M}$ is a measure. Let $(A_n)$ be as in 8. In the proof of $\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n)$ we only need to check $\geq$ since $\mu^*$ is $\sigma$-sub-additive. Here we may assume $\mu^*(A) < \infty$. According to step 9 $A \in \mathcal{M}_f$ and step 4 gives the assertion.

11. Steps 7 and 8 imply that $\sigma(\mathcal{R} \cup \mathcal{N}) \subseteq \mathcal{M}$. Step 10 thus shows that the restriction $\mu$ of $\mu^*$ to $\sigma(\mathcal{R} \cup \mathcal{N})$ is a measure which extends $m$ according to lemma 2.1. It is complete since $\mathcal{N}$ is contained in the $\sigma$-algebra which is the domain of $\mu$. Step 9 implies that every element of $\sigma(\mathcal{R} \cup \mathcal{N})$ of finite measure is actually in $\mathcal{M}_f$ and hence it can be approximated in the sense of assertion $b)$. This completes the proof of the theorem.

In general measure extensions are not unique. We shall see an example in chapter 5. However it is easy to see that any measure extension is dominated by the outer measure $\mu^*$.

**Remark 2.1** If $\nu$ is a measure extension of the content $m$ to some $\sigma$-algebra $\mathcal{A}$ then $\nu(A) \leq \mu^*(A)$ for all $A \in \mathcal{A}$. Choose any covering of $A$ by a sequence $(R_i)$ of elements of $\mathcal{R}$. Then $\nu(A) \leq \sum_i \nu(R_i) = \sum_i m(R_i)$. Passing to the infimum of such sums we get the assertion.

**Remark 2.2** Let $\mu$ be the extension of theorem 2.3. Then at least on $\{ A \in \sigma(\mathcal{R} \cup \mathcal{N}) : \mu(A) < \infty \}$ all measure extensions of the content $m$ agree with $\mu$: Let $\nu$ be such a measure extension and let $A$ be in this class. According to step 9 of the above proof $A$ belongs to $\mathcal{M}_f$, i.e. there is an approximating sequence $(R_n)$ in $\mathcal{R}$ with $\mu(R_n \Delta A) \to 0$. Then the previous remark implies that a fortiori $\nu(R_n \Delta A) \to 0$ and thus by lemma 2.1 and lemma 2.2 applied both to $\nu$ and $\mu^*$ we get

$$\nu(A) = \lim_n \nu(R_n) = \lim_n m(R_n) = \mu^*(A).$$

We can now give the uniqueness argument for theorem 2.2: Let $m$ be $\sigma$-finite and let $(E_k)$ be an increasing sequence in $\mathcal{R}$ of finite content whose union is the whole set $S$. Let $A \in \sigma(\mathcal{R})$. Then

$$\nu(A) = \lim_k \nu(A \cap E_k) = \lim_k \mu(A \cap E_k) = \mu(A).$$
Example 2.4 (Lebesgue measure) Let \( m \) be the geometric volume function on the ring \( \mathcal{R} \) of finite unions of (half open) rectangular sets in \( \mathbb{R}^d \). The measure extension constructed in theorem 2.3 is called the \( d \)-dimensional Lebesgue measure. We shall denote it by \( \lambda^d \). The \( \sigma \)-algebra \( \sigma(\mathcal{R}) \) is the Borel-\( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \) according to section 1.1. The class \( \mathcal{N} \) is the class of Lebesgue null-sets. By definition a set \( N \subset \mathbb{R}^d \) belongs to \( \mathcal{N} \) iff for each \( k \) there are countably many rectangles \( R_k^i \) such that \( N \subset \bigcup_k R_k^i \) and \( \sum_k \text{vol}(R_k^i) < \frac{1}{k} \). Since \( N \subset \bigcap_{k=1}^\infty \bigcup_{i=1}^\infty R_k^i \in \sigma(\mathcal{R}) = \mathcal{B}(\mathbb{R}^d) \) the class \( \mathcal{N} \) consists precisely of the null-sets in the sense of definition 1.6 for the measure \( \lambda^d_{\mathbb{R}^d} \). The \( \sigma \)-algebra \( \mathcal{L} = \sigma(\mathcal{R} \cup \mathcal{N}) \) is then easily seen to be the completion of the Borel sets for Lebesgue measure in the sense of proposition 1.1. It is called the \( \sigma \)-algebra of Lebesgue measurable sets. Every Lebesgue measurable set differs only by a Lebesgue null-set from a Borel set.

Example 2.5 (Counting measure) Let \( \mathcal{R} \) be the family of all finite subsets of a set \( S \). Let \( m \) assign to each finite set its cardinality. The corresponding outer measure assigns to each infinite set the value \( +\infty \). The empty set \( \emptyset \) is the only null-set and \( \sigma(\mathcal{R}) = \sigma(\mathcal{R} \cup \mathcal{N}) \) consists of all countable subsets of \( S \) and their complements. Let us look into the proof of theorem 2.3 in this case. The system \( \mathcal{M}_J \) defined there agrees with the system of finite sets whereas the \( \sigma \)-algebra \( \mathcal{M} \) defined in (2.10) coincides with the full power set. We see that the proof actually sometimes produces a measure extension on a larger \( \sigma \)-algebra than just on \( \sigma(\mathcal{R} \cup \mathcal{N}) \).

2.4 Monotone classes and more on uniqueness

This section can be read independently from the rest of this chapter. The monotone class theorem below sometimes is quite convenient in order to extend a statement from a certain class of sets to the \( \sigma \)-algebra generated by this class. We shall use it e.g. in the proof of Fubini’s theorem (theorem 5.1).\(^4\)

As a consequence of theorem 2.2 two measures on some \( \sigma \)-algebra \( \mathcal{F} \) agree if they agree on a ring \( \mathcal{R} \) which generates \( \mathcal{F} \) and on which they are \( \sigma \)-finite. Even for finite measures one might ask: Is a finite measure on a \( \sigma \)-algebra \( \mathcal{F} \) already uniquely determined by its values on an arbitrary generating system \( \mathcal{E} \) of \( \mathcal{F} \)?

Example 2.6 Here is a counter-example even if the total mass of the measures is fixed. Let \( S = \{1, 2, 3, 4\} \) and let the system \( \mathcal{E} \) consist of the two sets \( \{1, 2\} \) and \( \{2, 3\} \). Then all four singletons \( \{1\}, \{2\}, \{3\}, \{4\} \) can easily be produced via algebra operations from these two sets, and hence \( \sigma(\mathcal{E}) \) is the power set of \( S \). The two weight assignments \( \{1/4, 1/4, 1/4, 1/4\} \) and \( \{1/2, 0, 1/2, 0\} \) clearly induce two different measures \( \mu \) and \( \nu \) on the power set \( S \) with \( \mu(S) = \nu(S) = 1 \) and \( \mu(A) = \nu(A) = 1/2 \) for both sets \( A \in \mathcal{E} \).

We shall see in a minute that the essential requirement is that \( \mathcal{E} \) be stable under intersection. The proof uses the following concept.

Definition 2.7 A monotone class (monotone Klasse) over the set \( S \) is a system \( \mathcal{D} \) of subsets of \( S \) with the following three properties:

\(^4\)In chapter 5 we also sketch a way how to avoid theorem 2.4.
a) \( S \in \mathcal{D} \).

b) If \( A, B \in \mathcal{D} \) and \( A \subset B \) then \( B \setminus A \in \mathcal{D} \).

c) If \((A_n)\) is an increasing sequence of elements of \( \mathcal{D} \), then their union \( \bigcup_{n=1}^{\infty} A_n \) belongs \( \mathcal{D} \).

**Remark.** Monotone classes and \( \sigma \)-algebras are built in a similar fashion. In particular again the intersection of an arbitrary number of monotone classes is again a monotone class. The essential difference to \( \sigma \)-algebras is the fact that in c) the sets are assumed to be increasing. This makes this condition much easier to verify. On the other hand we have

**Theorem 2.4** *(Monotone class theorem)* Let \( \mathcal{E} \) be a class of subsets of \( S \) which is stable under (finite) intersections. If a monotone class \( \mathcal{D} \) contains all elements of \( \mathcal{E} \) then it even contains all elements of \( \sigma(\mathcal{E}) \).

**Proof.** Let \( \mathcal{D}^* \) be the intersection of all monotone classes which contain \( \mathcal{E} \), i.e. \( \mathcal{D}^* \) is the monotone class generated by \( \mathcal{E} \). Then \( \mathcal{D}^* \subset \mathcal{D} \) and we only need to show that \( \sigma(\mathcal{E}) \subset \mathcal{D}^* \). Let us proceed step by step.

1. We claim: for every set \( E \in \mathcal{E} \) and every \( A \in \mathcal{D}^* \) we have also \( A \cap E \in \mathcal{D}^* \): Consider the class \( \mathcal{D}_E \) of all sets \( A \) satisfying \( A \cap E \in \mathcal{D}^* \). Then \( \mathcal{D}_E \) is itself a monotone class because \( \mathcal{D}^* \) is one. Moreover \( \mathcal{D}_E \) contains all elements of \( \mathcal{E} \) since \( \mathcal{E} \) is stable under intersections. By definition of \( \mathcal{D}^* \) we conclude \( \mathcal{D}^* \subset \mathcal{D}_E \) and from this the assertion.

2. We claim: \( \mathcal{D}^* \) itself is stable under intersections: Let \( A, B \in \mathcal{D}^* \). The class \( \mathcal{D}_A \) of all \( B \) with \( A \cap B \in \mathcal{D}^* \) is a monotone class which according to step 1 contains the system \( \mathcal{E} \). Again by definition of \( \mathcal{D}^* \) this shows \( \mathcal{D}^* \subset \mathcal{D}_A \) and the assertion.

3. It remains to show that every monotone class \( \mathcal{D}^* \) which is stable under intersections is a \( \sigma \)-algebra, because then \( \mathcal{D}^* \) is even a \( \sigma \)-algebra which contains \( \mathcal{E} \) and hence \( \sigma(\mathcal{E}) \). Properties a) and b) imply that \( \mathcal{D}^* \) is stable under taking complements. Then \( \mathcal{D}^* \) is not only stable under finite intersections but also under finite unions. But together with property c) we conclude that \( \mathcal{D}^* \) is stable under arbitrary countable unions. Thus \( \mathcal{D}^* \) is indeed a \( \sigma \)-algebra.

**Theorem 2.5** *(Uniqueness criterion for measures)* Let the system \( \mathcal{E} \) be stable under intersections and generate the \( \sigma \)-algebra \( \mathcal{F} \). Let \( \mu \) and \( \nu \) be two measures on \( \mathcal{F} \) which agree on \( \mathcal{E} \). If the set \( S \) is the union of an increasing sequence of sets \( E_k \in \mathcal{E} \) with \( \mu(E_k) = \nu(E_k) < \infty \) then \( \mu = \nu \).

**Proof.** First of all it is easy to verify that for each \( k \) the system of sets \( \mathcal{D}_k = \{ A \in \mathcal{F} : \mu(A \cap E_k) = \nu(A \cap E_k) \} \) is a monotone class which contains \( \mathcal{E} \). Thus by the monotone class theorem \( \mu(A \cap E_k) = \nu(A \cap E_k) \) for all \( k \) and all \( A \in \mathcal{F} \). This implies the assertion by continuity from below of both measures.

A typical application is

**Corollary 2.2** A finite measure on the Borel subsets of a metric space is uniquely determined by its values on open sets.
Chapter 3

Measurable functions

Definition 3.1 Let \((S, \mathcal{F})\) and \((Y, \mathcal{G})\) be measurable spaces. A map \(f: S \to Y\) is called \(\mathcal{F}\)-\(\mathcal{G}\)-measurable (meßbar) (or just: measurable), if

\[f^{-1}(G) \in \mathcal{F}\text{ for all } G \in \mathcal{G}.
\]

In the special case when \(Y\) is a metric space a map \(f: S \to Y\) is called \(\mathcal{F}\)-measurable or measurable, if \(f\) is \(\mathcal{F}\)-\(\mathcal{B}(Y)\)-measurable.

For functions with one-dimensional range often the values \(\pm \infty\) occur in limit procedures. Therefore it is convenient to consider also extended real valued functions. The following notation is convenient.

Notation: Let \(f, g: S \to \mathbb{R}\) be two functions and let \(a \in \mathbb{R}\). Then denote by \(\{f \leq a\}\) the set \(\{x \in S | f(x) \leq a\} = f^{-1}((-\infty; a])\), by \(\{f = g\}\) the set \(\{x \in S | f(x) = g(x)\}\) and introduce similarly the notations \(\{f < a\}, \{f \geq a\}, \{f > a\}, \{f < g\}\) etc.

Moreover we recall that the extended real line \([-\infty, \infty]\) is a metric space when endowed with the metric

\[d: [-\infty, \infty] \times [-\infty, \infty] \to [0, 1], \quad d(x, y) = \frac{|x - y|}{1 + |x - y|}.
\]

It is not hard to see that the mapping \(f: [-\infty, \infty] \to [-1, 1]\) given by \(f(x) = \frac{x}{1 + |x|}\) is homeomorphic, and hence the above metric defines the usual topology on \([-\infty, \infty]\).

For extended real valued functions measurability has a simpler description.

Definition 3.2 An extended real-valued function \(f: S \to [-\infty, +\infty]\) is called \(\mathcal{F}\)-measurable if \(\{f \leq a\} \in \mathcal{F}\) for all \(a \in \mathbb{R}\) (or alternatively \(\{f > a\} \in \mathcal{F}\) for all \(a\), etc.)

Remarks: 1. If \(f\) takes values in the extended real number system we have now two definitions of measurability. We shall see below in lemma 3.2 that the two definitions are compatible.
2. The (extended) real-valued measurable functions are - with certain restrictions to be discussed later - the natural candidates for the definition of the integral.

The verification of measurability is often much simplified by the following criterion.

**Lemma 3.1** A map \( f : S \to Y \) is \( F \)-\( G \)-measurable if \( f^{-1}(E) \in F \) for all \( E \in \mathcal{E} \) where \( \mathcal{E} \) is any system of subsets of the space \( Y \) such that \( G = \sigma(\mathcal{E}) \).

**Proof.** Consider the following class of subsets of \( Y \)

\[
\mathcal{A} = \{ A \subseteq Y | f^{-1}(A) \in F \}.
\]

According to our assumption \( E \subseteq A \). Moreover since \( f^{-1}(Y) = S \in F \) we have \( Y \in \mathcal{A} \). Let \( A \in \mathcal{A} \). Then \( f^{-1}(A^c) = S \setminus f^{-1}(A) = (f^{-1}(A))^c \in F \), and therefore \( A^c \in \mathcal{A} \). Finally let \( A_n \in \mathcal{A} \) for all \( n \in \mathbb{N} \). Then

\[
f^{-1} \left( \bigcup_{n \in \mathbb{N}} A_n \right) = \bigcup_{n \in \mathbb{N}} f^{-1}(A_n) \in F
\]

and thus \( \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A} \). So \( \mathcal{A} \) is a \( \sigma \)-algebra which contains \( \mathcal{E} \). This implies that \( G = \sigma(\mathcal{E}) \subseteq \mathcal{A} \), i.e. our assertion.

**Lemma 3.2** Let \( (S, \mathcal{F}) \) be a measurable space. For \( f : S \to [-\infty, \infty] \) the following are equivalent.

1. \( f \) is measurable (in the sense of Definition 3.1)
2. \( \{ f < a \} \in \mathcal{F} \) for all \( a \in \mathbb{R} \)
3. \( \{ f \leq a \} \in \mathcal{F} \) for all \( a \in \mathbb{R} \)
4. \( \{ f > a \} \in \mathcal{F} \) for all \( a \in \mathbb{R} \)
5. \( \{ f \geq a \} \in \mathcal{F} \) for all \( a \in \mathbb{R} \).

**Proof.** We recall that \( [-\infty, \infty] \) is homeomorphic to \( [-1, 1] \); hence by lemma 3.1 it suffices to show that

\[
\mathcal{B}([-1, 1]) = \sigma(\mathcal{E}),
\]

where \( \mathcal{E} \) denotes the class of intervals of the form \([-1, a), [-1, a], (a, 1], \) and \([a, 1] \) for \( a \in (-1, 1) \). The inclusion \( \supseteq \) is clear, and the reverse inclusion follows from the fact that

\[
(a, b) = [-1, b] \setminus [-1, a] \text{ for } -1 \leq a \leq b \leq 1,
\]

and similarly for the remaining systems of intervals.

Since the singleton sets \( \{-\infty\} \) and \( \{+\infty\} \) are Borel measurable, the previous criterion admits an obvious analogue for extended real-valued functions.

**Theorem 3.1** Let \( f, g \) be extended real valued functions. Then the sets \( \{ f \leq g \}, \{ f < g \}, \{ f = g \} \) are in \( \mathcal{F} \).

**Proof.** We have \( \{ f < g \} = \bigcup_{r \in \mathbb{Q}} \{ f \leq r \} \cap \{ g > r \} \in \mathcal{F} \), and thus also \( \{ f \leq g \} = (g < f)^c \in \mathcal{F} \) and \( \{ f = g \} = (g \leq f) \cap \{ f \leq g \} \in \mathcal{F} \).
Remark 3.1 1. If \( c \in \mathbb{R} \) and \( f(x) = c \) for all \( x \in S \) then
\[
\{ f < a \} = \begin{cases} 
\emptyset & \text{for } a \leq c \\
S & \text{for } a > c
\end{cases}
\]
So the constant functions are measurable.
2. If \( f : \mathbb{R}^m \rightarrow \mathbb{R}^n \) is continuous, Then \( f^{-1}(U) \) is open and hence Borel for all open sets \( U \). So by lemma 3.1 all continuous functions are measurable.

Lemma 3.3 Let \((S, F), (Y, \mathcal{G}), (Z, \mathcal{H})\) be measurable spaces. Let \( f : Y \rightarrow Z, g : S \rightarrow Y \) be measurable maps. Then the composition \( f \circ g : S \rightarrow Z \) is measurable.

Proof. For \( H \in \mathcal{H} \) we get \( f^{-1}(H) \in \mathcal{G} \) and \( (f \circ g)^{-1}(H) = g^{-1}(f^{-1}(H)) \in \mathcal{F} \).

Lemma 3.4 A vector-valued function \( f = (f_1, \ldots, f_n) : S \rightarrow \mathbb{R}^m \) is \( \mathcal{F} \)-measurable if and only if the components \( f_i : S \rightarrow \mathbb{R} \) are \( \mathcal{F} \)-measurable for all \( i \in \{1, \ldots, m\} \).

Proof. Let \( f \) be measurable. Then for each \( a \in \mathbb{R} \) the set \( A = [-\infty, a] \times \mathbb{R}^{m-1} \subseteq \mathbb{R}^m \) is closed, i.e. \( A \in \mathcal{B}(\mathbb{R}^m) \). Thus \( \{f_1 \leq a\} = f^{-1}(A) \in \mathcal{F} \) for all \( a \in \mathbb{R} \). Therefore \( f_1 \) is measurable by lemma 3.2. A similar argument works for the other components.

Conversely assume that \( f_i \) is measurable for all \( i \). Let a box \( Q = [a_1, b_1] \times \ldots \times [a_n, b_n] \) be given. Then
\[
f^{-1}(Q) = \bigcap_{i=1}^n f_i^{-1}([a_i, b_i]) \in \mathcal{F}
\]
This holds for all boxes and lemma 3.1 together with theorem 1.2 implies the measurability of \( f \).

Corollary 3.1 If \( f_1, f_2 : S \rightarrow \mathbb{R} \) are measurable then the same is true for \( f_1 + f_2, f_1 - f_2, f_1 \cdot f_2, \min(f_1, f_2) \) and \( \max(f_1, f_2) \).

Proof. The maps \('+', '-', '+', '-\)' are continuous. By remark 3.12 they are measurable. The map \( f = (f_1, f_2) : S \rightarrow \mathbb{R}^2 \) is measurable according to lemma 3.4. Since \( f_1 + f_2 = '+' \circ f \) the function \( f_1 + f_2 \) is measurable by lemma 3.3. The same arguments work also for difference and multiplication.

Theorem 3.2 Let \((S, \mathcal{F})\) be a measurable space and let \( \{f_n\}_{n \in \mathbb{N}} \) be a sequence of measurable extended real valued functions on \( S \). a) Then the (pointwise defined) functions \( \sup_n f_n, \inf_n f_n, \limsup_{n \rightarrow \infty} f_n, \liminf_{n \rightarrow \infty} f_n \) are measurable.

b) The set \( F = \{x : \lim f_n(x) \text{ exists}\} \) is in \( \mathcal{F} \) and \( \lim_{n} 1_{F} f_n = \text{measurable} \).

Proof. For all \( a \in \mathbb{R} \) the set \( \{sup f_n \leq a\} = \cap_{n \in \mathbb{N}} \{f_n \leq a\} \) is in \( \mathcal{F} \). So \( sup f_n \) is measurable. By symmetry \( inf f_n \) is measurable. Therefore also the functions \( \limsup_{n \rightarrow \infty} f_n = \inf_m (\sup_{n \geq m} f_n) \) and \( \limsup_{n \rightarrow \infty} f_n = \sup_m (\inf_{n \geq m} f_n) \) are measurable and the set \( F = \{\limsup_n f_n = \lim inf_n f_n\} \) is in \( \mathcal{F} \).

Simple examples of real-valued measurable functions are the indicator functions \( 1_A, A \in \mathcal{F} \). In fact the sets \( \{1_A > a\} \) are in \( \mathcal{F} \) because they are either empty (if \( a \geq 1 \), equal to \( A \) (if \( 0 \leq a < 1 \), or equal to \( S \) (if \( a < 0 \)).
Definition 3.3 Let \((S, \mathcal{F})\) be a measurable space. A function of the form 
\[ f(x) = \sum_{i=1}^{n} \alpha_i 1_{A_i}(x) \] 
with \(\alpha_i \in \mathbb{R}, A_i \in \mathcal{F}\) is called an elementary function or measurable step function (meßbare Treppenfunktion).

Remark 3.2 Clearly the measurable step functions form a vector space. A measurable step function takes only finitely many different values \(\beta_1, \ldots, \beta_m\). The sets \(B_j = \{ f = \beta_j \}\) are successively contained in each other we get 
\[ A \subseteq B_1 \subseteq B_2 \subseteq \cdots \]
and \(f\) also admits the representation 
\[ f(x) = \sum_{j=1}^{m} \beta_j 1_{B_j}(x) \] 
where the sets \(B_j\) form a (disjoint) partition of the set \(S\).

Theorem 3.3 For every measurable function \(f : S \rightarrow [0, \infty]\) there is a sequence \((f_n)\) of non-negative elementary functions with 
\[ f_n \uparrow f, \text{ i.e., } f_1 \leq f_2 \leq \cdots \] 
and \(f = \sup_n f_n\).

Proof. Let 
\[ f_n(x) = \sum_{k=0}^{n^2-1} k 2^{-n} 1_{\{k2^{-n} \leq x < (k+1)2^{-n}\}} + n1_{\{f \geq n\}}, \]  
(i.e. if \(f(x) < n\), then \(f_n(x)\) is the largest member of the grid \(2^{-n}\mathbb{N}_0\) which is situated on the left of \(f(x)\), and \(f_n(x) = n\) if \(f(x) \geq n\). Because these grids are successively contained in each other we get \(f_n(x) \leq f_{n+1}(x)\). Moreover \(f_n \leq f\). Finally \(f(x) - f_n(x) \leq 2^{-n}\), as soon as \(f(x) \leq 2^n \cdot n \cdot 2^{-n} = n\) and \(f_n(x) = n \uparrow \infty = f(x)\) if \(f(x) = \infty\). Therefore \((f_n)\) converges point-wise to \(f\).

The next result is the only place in this section where a measure appears.

Corollary 3.2 Let \(\mu\) be a measure on the \(\sigma\)-algebra \(\mathcal{F}\). Let \(f\) be extended real-valued \(\mathcal{F}_\mu\)-measurable. Then \(f\) is \(\mu\)-almost everywhere equal to an \(\mathcal{F}\)-measurable function \(g\).

Proof. Consider elementary functions 
\[ f_n = \sum_{k=1}^{k_n} \beta_{nk} 1_{B_{nk}} \] 
with \(B_{nk} \in \mathcal{F}_\mu\) such that \(f_n \uparrow f^+\). Each of the sets \(B_{nk}\) is \(\mu\)-equivalent to a set \(B_{nk}' \in \mathcal{F}\). Outside the countably many null-sets \(B_{nk} \Delta B_{nk}'\) \(f^+\) is equal to the \(\mathcal{F}\)-measurable function 
\[ g^+ = \sup_n \sum_{k=0}^{k_n} \beta_{nk} 1_{B_{nk}}. \] 
Similarly \(f^-\) is \(\mu\)-a.e. equal to an \(\mathcal{F}\)-measurable function \(g^-\). The set \(N = \{g^+ = g^- = \infty\}\) is a null-set. Let \(g = g^+ - g^-\) outside \(N\) and \(g = 0\) on \(N\). Then \(f = g\) \(\mu\)-a.e.

Sometimes functions are not defined at all points of the set \(S\). Then the following convention is helpful.

Definition 3.4 Let \(A\) be a subset of the set \(S\). Let \(f\) be an extended real-valued function which may not be defined outside of \(A\). Then \(1_A f\) or \(f 1_A\) denotes the function on \(S\) which agrees with \(f\) on \(A\) and vanishes outside of \(A\).
Chapter 4

The Integral

In this chapter \((S, \mathcal{F}, \mu)\) is a measure space. We define the integral with respect to \(\mu\) and study some of its properties.

4.1 The elementary integral

The simplest form of the integral is for elementary functions. We use the convention \(0 \cdot \infty = 0\).

**Definition 4.1** Let \((S, \mathcal{F}, \mu)\) be a measure space. Let \(f\) be a non-negative elementary function with the representation

\[
f = \sum_{i=1}^{n} \alpha_i 1_{A_i}
\]

with disjoint sets \(A_i \in \mathcal{F}\). We define the integral of \(f\) with respect to \(\mu\) by

\[
\int f \, d\mu := \int f(x) \, d\mu(x) := \sum_{i=1}^{n} \alpha_i \mu(A_i)
\]

The representation 4.1 is not unique, because the sets of the form \(\{f = \alpha\}\) can be split in many different ways into disjoint measurable parts. But

**Lemma 4.1** The integral in Definition 4.1 does not depend on the choice of the representation in (4.1).

**Proof.** Let \(f = \sum_{j=1}^{m} \beta_j 1_{B_j}\) be a second representation of \(f\) with disjoint sets \(B_j \in \mathcal{F}\). Without loss of generality we may assume \(S = \bigcup_{i} A_i = \bigcup_{j} B_j\) since the coefficient 0 is admitted. Moreover \(\alpha_i = \beta_j\) for all pairs \((i, j)\) of indices with \(A_i \cap B_j \neq \emptyset\), because both coefficients agree with the common value which the function \(f\) takes on this intersection. The additivity of \(\mu\) shows

\[
\sum_{i=1}^{n} \alpha_i \mu(A_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \mu(A_i \cap B_j) = \sum_{j=1}^{m} \sum_{i=1}^{n} \beta_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} \beta_j \mu(B_j).
\]

\[\blacksquare\]
For the sake of simplicity of this lemma we have assumed in Definition 4.1 that the sets $A_i$ in the representation (4.1) are disjoint. Part a) of the following theorem shows in particular that the formula (4.2) is also valid if the $A_i$ are not disjoint.

**Theorem 4.1** The integral as a functional on the set of non-negative elementary functions is

a) additive and positively homogeneous, i.e. $\int a f + bg \, d\mu = a \int f \, d\mu + b \int g \, d\mu$ if $a \in [0, \infty)$,

b) monotone increasing, i.e. $f \leq g$ implies $\int f \, d\mu \leq \int g \, d\mu$,

c) $\sigma$-continuous from below, i.e. if $f_n$ is $[0,\infty]$-valued and $f_n \uparrow f$ point-wise then $\int f_n \, d\mu \uparrow \int f \, d\mu$.

**Proof.** The proofs for a) and b) follow from the fact that for any two elementary functions there is a common partition of $S$ into disjoint measurable subsets on which both functions are constant.

c) Let $f = \sum_{j=1}^{m} \alpha_j 1_{A_j}$, and assume without loss of generality that $\alpha_j > 0$ for all $j = 1, \ldots, m$. We consider the sets $B_{j,k}^{n} = \{ x \in A_j : f_n(x) > \alpha_j - \frac{1}{k} \}$. For fixed $k$ we get $B_{j,k}^{n} \uparrow A_j$ as $n \to \infty$ because each $x \in A_j$ is in $B_{j,k}^{n}$ for eventually all $n$. Moreover, if $k$ is large enough to make $\alpha_j - \frac{1}{k} > 0$ for all $j = 1, \ldots, m$, we have $f_n \geq \sum_{j=1}^{m} (\alpha_j - \frac{1}{k}) 1_{B_{j,k}^{n}}$ by the positivity assumption, and thus

$$\lim_n \int f_n \, d\mu \geq \lim_n \sum_{j=1}^{m} (\alpha_j - \frac{1}{k}) \mu(B_{j,k}^{n}) = \sum_{j=1}^{m} (\alpha_j - \frac{1}{k}) \mu(A_j).$$

As $k \to \infty$ we get $\lim_n \int f_n \, d\mu \geq \sum \alpha_j \mu(A_j) = \int f \, dx$. The converse inequality follows from the monotonicity of the integral.

---

4.2 The Integral for non-negative measurable and for integrable functions

We extend the definition of the integral from elementary functions in two steps: First for all non-negative measurable functions and then for those measurable functions whose positive and negative parts have a finite integral.

**Definition 4.2** For a measurable function $f : S \to [0, \infty]$ we define $\int f \, d\mu$ by

$$\int f \, d\mu = \sup \{ \int g \, d\mu : g \leq f, g \text{ is an elementary function} \}.$$ 

Theorem 3.3 says that a measurable $f \geq 0$ can be approximated from below by an increasing sequence of elementary functions. One reaches the sup in the definition by any such sequence.

**Lemma 4.2** Let $(f_n)$ be any sequence of elementary functions such that $f_n \uparrow f$. Then $\int f_n \, d\mu \uparrow \int f \, d\mu$.

**Proof.** Obviously by definition $\int f \, d\mu \geq \int f_n \, d\mu$ for each $n$ and thus $\int f \, d\mu \geq \lim_n \int f_n \, d\mu$. For the converse inequality let $g$ be an elementary
function with \( g \leq f \). Then the functions \( \min(f_n, g) \) are also elementary and \( \min(f_n, g) \uparrow \min(f, g) = g \). By theorem 4.1 part c) we get \( \lim_n \int f_n \, d\mu \geq \lim_n \int \min(f_n, g) \, d\mu = \int g \, d\mu \). The definition of \( \int f \, d\mu \) yields the desired inequality.

From Definition 4.2 and lemma 4.2 it is straightforward to see that the integral for non-negative measurable functions is again additive, monotone and positively homogeneous.

We now admit functions which also take negative values. Every function \( f : S \to [-\infty, \infty] \) has the decomposition \( f = f^+ - f^- \), where \( f^+ = \max(f, 0) \) and \( f^- = \max(-f, 0) \). Then \( |f| = f^+ + f^- \). All these functions are measurable if \( f \) is measurable (cf. corollary 3.1).

**Definition 4.3**

a) A measurable function \( f : S \to [-\infty, \infty] \) with \( \int |f| \, d\mu < \infty \) is called \( \mu \)-integrable (\( \mu \)-integrierbar). For a \( \mu \)-integrable function the integral is defined by

\[
\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu.
\]

The set of finite-valued \( \mu \)-integrable functions is denoted by \( \mathcal{L}^1(\mu) \).

b) If \( A \in \mathcal{F} \) and \( f \) is either measurable and \( \geq 0 \) or integrable the same is true for \( 1_A f \) and one writes \( \int_A f \, d\mu \) instead of \( \int 1_A f \, d\mu \).

**Example 4.1**

As in Example 2.4 the letter \( \mathcal{L} \) is a reference to the name Lebesgue. In the special case where \( S = \mathbb{R}^d \) and \( \mathcal{F} \) is the \( \sigma \)-algebra of Lebesgue measurable sets and \( \mu = \lambda^d \) the elements of \( \mathcal{L}^1(\lambda^d) \) are called Lebesgue-integrable. One often writes \( \mathcal{L}^1(\mathbb{R}^d) \) and \( \int_{\mathbb{R}^d} f(x) \, dx \) instead of \( \mathcal{L}^1(\lambda^d) \) and \( \int f \, d\lambda^d \). In Example 4.3 below we study the relation to the Riemann integral.

We collect some elementary facts

**Proposition 4.1** Let \( f \) be measurable. Then

a) (Markov’s inequality). For all \( a > 0 \)

\[
\mu\{ |f| \geq a \} \leq \frac{1}{a} \int |f| \, d\mu.
\]  

b) \( \int |f| \, d\mu = 0 \) if and only if \( \mu\{ f \neq 0 \} = 0 \).

c) If \( f \) is integrable then \( \mu\{ |f| = \infty \} = 0 \) and \( \int \mu \, d\mu \leq \int |f| \, d\mu \).

d) \( f \) is integrable if there is an integrable function \( h \) with \( |f| \leq h \).

**Proof.**
a) Clearly \( |f| \geq a 1_{\{|f| \geq a\}} \). This implies

\[
\int |f| \, d\mu \geq \int a 1_{\{|f| \geq a\}} \, d\mu = a \mu\{ |f| \geq a \}.
\]

b) First suppose that \( \mu(N) = 0 \) for \( N = \{ f \neq 0 \} \). Then for every elementary function \( g \) with \( g \leq |f| \) we have also \( \mu\{ g \neq 0 \} = 0 \) and hence by Definition 4.1 \( \int g \, d\mu = 0 \). Then according to Definition 4.2 \( \int |f| \, d\mu = 0 \).

Conversely assume \( \int |f| \, d\mu = 0 \). One has \( \{ |f| > 1/n \} \uparrow \{ |f| > 0 \} \) and therefore by Markov’s inequality

\[
\mu\{ |f| > 0 \} = \lim_n \mu\{ |f| > 1/n \} \leq \lim_n \int |f| \, d\mu = 0.
\]
c) The first assertion follows by letting $a$ tend to $+\infty$ in (4.3). For the second assertion note that $-(\int f^+ \, d\mu + \int f^- \, d\mu) \leq \int f^+ \, d\mu - \int f^- \, d\mu \leq \int f^+ \, d\mu + \int f^- \, d\mu$.

d) Clearly if $|f| \leq h$ and $h$ is integrable then $\int |f| \, d\mu \leq \int h \, d\mu < \infty$. 

In extension to theorem 4.1 one has

**Theorem 4.2** $L^1(\mu)$ is a linear space and the integral is a linear and monotone functional on $L^1(\mu)$.

**Proof.**

1. We have $\int af \, d\mu = a \int f \, d\mu$ for all $a \in \mathbb{R}$. If $a \geq 0$ this follows from the corresponding property of the elementary integral first for $f^+$ and $f^-$ by Definition 4.2 and then for $f$. If $a < 0$ it suffices to consider $a = -1$. Then

   \[
   \int -f \, d\mu = \int (-f)^+ \, d\mu - \int (-f)^- \, d\mu = \int f^- \, d\mu - \int f^+ \, d\mu = -\int f \, d\mu.
   \]

2. If $f, g \in L^1(\mu)$, then $(f + g)^+ - (f + g)^- = f + g = f^+ + g^+ - f^- - g^-$, i.e. $(f + g)^+ + f^- + g^- = (f + g)^+ + f^+ + g^-$. The additivity of the integral for non-negative functions and subtraction give

   \[
   \int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu.
   \]

3. If $f \leq g$, then $\int g \, d\mu = \int f \, d\mu + \int g - f \, d\mu \geq \int f \, d\mu$ since $g - f \geq 0$.

**Example 4.2** Consider the counting measure $\zeta$ on the power set of $\mathbb{N}$. Then the integral is just a series:

\[
\int f \, d\zeta = \sum_{n=1}^{\infty} f(n). \tag{4.4}
\]

For elementary functions this is obvious. For nonnegative functions (sequences) this follows easily by approximation from below. A sequence $f = (f(n))_{n \in \mathbb{N}}$ is $\zeta$-integrable if and only if the series in (4.4) converges absolutely. So every theorem about integrals for measures contains a result about absolutely converging series. Like in example 4.1 however a series like in $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} = \log(2)$ which converges but does not converge absolutely cannot be treated by this method. The point is that the Lebesgue theory does not pay attention to any order structure in the set $S$. (This gives a new explanation for the elementary analysis statement that an absolutely converging series can be arbitrarily rearranged without changing its value.)

Let us now replace $\mathbb{N}$ by an arbitrary set $S$. Then the formula

\[
\int f \, d\zeta = \sum_{x \in S} f(x). \tag{4.5}
\]

allows to extend the definition of a (absolutely converging) series to functions on $S$. If $f$ is $\zeta$-integrable then the Markov inequality implies that for each $n \in \mathbb{N}$ the set $\{|f| > \frac{1}{n}\}$ is finite. In particular the set $\{f \neq 0\}$ is countable.
Let the two functions
\[ f \text{ and } g \]
be \( \mu \)-almost everywhere (fast überall), in short a.e., if the set of points which do not have this property is a \( \mu \)-null-set. For example we just proved that a \( \mu \)-integrable function \( f \) is \( \mu \)-almost everywhere finite. Similarly a measurable function \( f \) with
\[ \int |f| \, d\mu = 0 \]
vanishes \( \mu \)-a.e.. In analogy to section 1.3 we call two functions \( f, g \) \( \mu \)-equivalent if they agree \( \mu \)-a.e..

**Proposition 4.2** Let the two functions \( f, g : S \to [-\infty, \infty] \) be \( \mu \)-equivalent.

a) If \( f \) is \( F \) measurable then \( g \) is \( \mathcal{F}_F \)-measurable.

b) If \( f \) is \( \mu \)-integrable then so is \( g \) and \( \int f \, d\mu = \int g \, d\mu \).

**Proof.** a) Let \( B \in \mathcal{B}(\mathbb{R}) \). Then the set \( \{ f \in B \} \) is \( \mu \)-equivalent to the set \( \{ g \in B \} \in \mathcal{F} \) and hence an element of the completion \( \mathcal{F}_\mu \) (cf. proposition 1.1). This proves the assertion.

b) Let \( N = \{ f \neq g \} \). Part a) and lemma 3.2 show that \( N \in \mathcal{F}_\mu \) and \( \mu(N) = 0 \). Moreover \( |g| \leq |f| + 1_N \) and so by part b) of proposition 4.1 \( \int |g| \, d\mu \leq \int |f| \, d\mu + 0 < \infty \), i.e. \( g \) is integrable. The identity
\[ \int g \, d\mu = \int g1_{N^c} \, d\mu = \int f1_{N^c} \, d\mu = \int f \, d\mu \]
follows first for \( g \geq 0 \) and then for general \( g \).

**Example 4.3 (Riemann integrals).** Let \([a, b]\) be a real interval and let \( f : [a, b] \to \mathbb{R} \) be Riemann integrable. In the special case where \( f = \sum_{i=1}^{k} \alpha_i 1_{(a_i, b_i]} \) is a step function clearly the Riemann sum \( \sum_{i=1}^{k} \alpha_i (b_i - a_i) \) coincides with the elementary Lebesgue integral. In the general case there are two sequences \((\mathcal{F}_n)\) and \((\mathcal{I}_n)\) of step functions such that for \( \mathcal{F} = \inf_n \mathcal{F}_n \) and \( \mathcal{I} = \sup_n \mathcal{I}_n \) we have
\[ \mathcal{I} \leq f \leq \mathcal{F} \quad \text{and} \quad \sup_n \int_a^b \mathcal{F}_n(x) \, dx = \inf_n \int_a^b \mathcal{I}_n(x) \, dx. \]
The two functions \( \mathcal{I} \) and \( \mathcal{F} \) are -according to theorem 3.2- Borel measurable. Since they are bounded and on a bounded interval the constants are integrable they are Lebesgue-integrable by proposition 4.1 d). Monotonicity of the integral implies \( \mathcal{I} \mathcal{F} \, d\lambda^1 = \mathcal{F} \mathcal{F} \, d\lambda^1 \). Thus \( \mathcal{F} - \mathcal{I} = 0 \) and by proposition 4.1 d) \( \mathcal{F} - f = 0 \) a.e.. Therefore \( f = \mathcal{F} \) a.e. and proposition 4.2 shows that \( f \) is Lebesgue-integrable and
\[ \int f \, d\lambda^1 = \int \mathcal{F} \, d\lambda^1 = \int_a^b \mathcal{F}(x) \, dx = \int_a^b f(x) \, dx. \]
However a so-called improper Riemann integral like in the formula \( \int_0^\infty \frac{\sin x}{x} \, dx = \pi/2 \) typically is not a Lebesgue-integral, e.g. in this example \( \int_0^\infty \frac{\sin x}{x} \, dx = \infty \), i.e. the integrand is not \( \lambda^1 \)-integrable in the sense of Definition 4.3.

**Definition 4.4** We denote by \( L^1(\mu) \) the set of \( \mu \)-equivalence of integrable functions.

Part a) of the following lemma follows from proposition 4.1 c). Part b) is obvious.
Lemma 4.3  
a) Every integrable function \( f \) is \( \mu \)-equivalent to the real-valued function \( f1_{\{f < \infty\}} \in L^1(\mu) \).

b) Let \( f, f', g, g' \in L^1(\mu) \) be such that \( f = f' \) a.e. and \( g = g' \) a.e.. Then for all \( a, b \in \mathbb{R} \) one has \( af + bg = af' + bg' \) a.e..

Thus each equivalence class in \( L^1(\mu) \) has a representative in \( L^1(\mu) \). In view of theorem 4.2 the space \( L^1(\mu) \) inherits from \( L^1(\mu) \) the vector space structure and the natural ordering and the integral induces a linear monotone increasing functional on \( L^1(\mu) \). Often one does not really distinguish between an element of \( L^1(\mu) \) and its equivalence in \( L^1(\mu) \).

4.3 The Limit Theorems

The limit results of this section are a crucial feature of the Lebesgue integral. They all treat in one way or another the question whether the integral can be exchanged with a limit.

Theorem 4.3 (Theorem of monotone convergence). Let \( (f_n) \) be a non-decreasing sequence of non-negative measurable functions. Let \( f = \sup_n f_n \). Then

\[ \int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu. \]

Proof. Theorem 3.3 shows that for every \( n \in \mathbb{N} \) there exists a sequence \( (f^k_n)_{k \in \mathbb{N}} \) of elementary functions such that \( f^k_n \uparrow n \to \infty f_n \). Let us consider the elementary function \( g_k = \max(f^1_n, \ldots, f^n_k) \). For each \( k \) we have \( g_k+1 \geq g_k \) and \( f \geq g_k \). Moreover \( \sup_k g_k \geq \sup_{k \geq n} f^k_n = f_n \) for each \( n \). So \( f \geq \sup_k g_k \geq \sup_n f_n = f \), i.e. \( g_k \uparrow f \). Since the \( g_n \) are elementary lemma 4.2 gives

\[ \int f \, d\mu = \lim_{n \to \infty} \int g_n \, d\mu \leq \lim_{n \to \infty} \int f_n \, d\mu. \]

The converse inequality \( \int f \, d\mu \geq \lim_{n \to \infty} \int f_n \, d\mu \) follows from the monotonicity of the integral.

Remark 4.1 The additivity of the integral and the theorem of monotone convergence together show that every non-negative measurable function \( f \) gives rise to a new measure \( f\mu : \mathcal{F} \to [0, \infty] \), defined by

\[ f\mu(A) = \int 1_A f \, d\mu. \]

In fact for every sequence \( (A_n) \) of pairwise disjoint sets with union \( A \)

\[ f\mu(A) = \int \left( \sum_{n=1}^{\infty} 1_{A_n} \right) f \, d\mu = \sum_{n=1}^{\infty} \int 1_{A_n} f \, d\mu = \sum_{n=1}^{\infty} f\mu(A_n). \]

In chapter 8 we shall characterize those measures which appear in this way.

For non-negative sequences of functions which do not converge one has
Theorem 4.4 (Fatou’s lemma). Let \((f_n)\) be a sequence of non-negative measurable functions. Then

\[
\int \liminf_{n \to \infty} f_n \, d\mu \leq \liminf_{n \to \infty} \int f_n \, d\mu.
\]

Proof. We consider the function \(g_m = \inf_{n \geq m} f_n\). It is measurable by theorem 3.2. Then \(g_m \geq 0\) and \(g_m \uparrow \sup g_m = \liminf f_n\). On the other hand \(g_m \leq f_n\) and \(\int g_m \, d\mu \leq \int f_n \, d\mu\) for all \(n \geq m\). Together with the theorem of monotone convergence one gets

\[
\int \liminf_{n \to \infty} f_n \, d\mu = \int \sup_m g_m \, d\mu = \sup_m \int f_n \, d\mu \leq \sup_m \inf_{n \geq m} \int f_n \, d\mu.
\]

Example 4.4 In general the inequality in Fatou’s lemma is strict even if the functions converge point-wise and are in \(L^1(\mu)\). We consider the standard Lebesgue measure space \(([0, 1], B([0, 1]), \lambda|B([0, 1]))\). Let \(f = 0\) and \(f_n = n 1_{(0, \frac{1}{n}]}\). Then \(f_n \to f\) point-wise. On the other hand

\[
\lim_{n} \int_{0}^{1} f_n(x) \, dx = \lim_{n} \frac{1}{n} = 1 \neq 0 = \int_{0}^{1} f(x) \, dx.
\]

Lebesgue’s result below shows that domination by an integrable function is sufficient to avoid the effect in this example. Note that the function \(x \mapsto \frac{1}{x}\) which is a natural upper bound for our sequence fails to be integrable.

The following theorem is perhaps the most useful single result in measure theory.

Theorem 4.5 (Lebesgue’s dominated convergence theorem, Satz von der majorisierten Konvergenz) Let \((f_n)\) be a sequence of measurable functions such that there is an integrable function \(h\) with \(|f_n| \leq h\) \(\mu\)-a.e.. Let \(f\) be such that \(f(x) = \lim_{n \to \infty} f_n(x)\) \(\mu\)-a.e.. Then \(f\) is integrable as well and

\[
\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu
\]

and

\[
\lim_{n \to \infty} \int |f - f_n| \, d\mu = 0.
\]

Proof. The union \(N\) of the countably many null-sets \(\{f_n \nrightarrow f\}\), \(\{|f_n| > h\}\) and \(\{|h| = \infty\}\) is a null-set. Without changing the integrals or the integrability of the functions involved we can set all function to vanish on \(N\). Let us consider the auxiliary functions \(g_n = 2h - (f - f_n)\). Then \(\lim_n g_n = 2h\) and \(g_n \geq 0\), so by Fatou’s lemma

\[
\int 2h \, d\mu = \int \lim_n g_n \, d\mu \leq \liminf_n \int g_n \, d\mu
\]

\[
= \int 2h \, d\mu - \limsup_n (\int f \, d\mu - \int f_n \, d\mu).
\]
Therefore $\limsup_n \left( \int f \, d\mu - \int f_n \, d\mu \right) \leq 0$. Similarly we can apply Fatou’s lemma to the non-negative functions $2h + f - f_n$, and get in the same way the estimate $\liminf_n \left( \int f \, d\mu - \int f_n \, d\mu \right) \geq 0$. Together this gives the first of the two limit assertions. The second one follows if we apply the first one to the sequence $(|f - f_n|)$ which is dominated by $2h$ and converges everywhere to $0$. 

As a corollary one gets the following rule for differentiation under the integral sign.

**Theorem 4.6** Let $(a, b)$ an open interval and let $f : (a, b) \times S \to \mathbb{R}$ be a function such that

i) for fixed $t \in (a, b)$ is the function $f(t, \cdot)$ is $\mu$-integrable on $S$,

ii) for fixed $x \in S$ the function $f(\cdot, x)$ is continuously differentiable on $(a, b),

iii) there is a $\mu$-integrable $h : S \to [0, \infty]$ such that $|\frac{\partial}{\partial t} f(t, x)| \leq h(x)$ for all $x \in S$ and all $t \in (a, b)$.

Then the integral $\int_S f(t, x) \, d\mu(x)$ is continuously differentiable in $t$ and

$$\frac{\partial}{\partial t} \int_S f(t, x) \, d\mu(x) = \int_S \frac{\partial}{\partial t} f(t, x) \, d\mu(x). \tag{4.6}$$

**Proof.** Let $(t_n)$ be sequence of points which converges to some $t \in (a, b)$. Consider the measurable functions $g_n$ defined by $g_n(x) = \frac{f(t_n, x) - f(t, x)}{t_n - t}$. By the mean-value theorem for each $x$ there is a point $t_n^*(x) \in (a, b)$ such that $g_n(x) = \frac{\partial}{\partial t} f(t_n^*(x), x)$. By hypothesis we can conclude $|g_n(x)| \leq h(x)$. Furthermore $g_n(x) \to \frac{\partial}{\partial t} f(t, x)$ for all $x \in S$. The theorem of dominated convergence gives

$$\lim_{n} \int_S f(t_n, x) \, d\mu(x) - \int_S f(t, x) \, d\mu(x) = \int_S g_n(x) \, d\mu \to \int_S \frac{\partial}{\partial t} f(t, x) \, d\mu.$$ 

i.e. the relation (4.6). Similarly the continuity of the right-hand side in (4.6) follows because $iii$ implies

$$\lim_n \int \frac{\partial}{\partial t} f(t_n, x) \, d\mu(dx) = \int \lim_n \frac{\partial}{\partial t} f(t_n, x) \, d\mu(dx) = \int \frac{\partial}{\partial t} f(t, x) \, d\mu(dx).$$

$\blacksquare$
Chapter 5

Fubini’s Theorem

Fubini’s theorem is the main tool for multiple integrals.

**Theorem 5.1 (Fubini)** Let \((X, \mathcal{F}, \mu)\) and \((Y, \mathcal{G}, \nu)\) two \(\sigma\)-finite measure spaces. Let \(\mathcal{F} \otimes \mathcal{G}\) be the \(\sigma\)-algebra on the cartesian product \(X \times Y\) which is generated by the 'measurable rectangles' \(F \times G\) with \(F \in \mathcal{F}, G \in \mathcal{G}\). Then

a) There is a unique measure \(\mu \otimes \nu\) on the \(\sigma\)-algebra \(\mathcal{F} \otimes \mathcal{G}\), such that

\[
\mu \otimes \nu(F \times G) = \mu(F) \nu(G)
\]

for all \(F \in \mathcal{F}, G \in \mathcal{G}\).

b) Let \(f : X \times Y \to [0, \infty]\) be a \(\mathcal{F} \otimes \mathcal{G}\)-measurable function. For every \(x \in X\) the sectional function \(f(x, \cdot) : y \mapsto f(x, y)\) is \(\mathcal{G}\)-measurable and the function \(x \mapsto \int_Y f(x, y) \, d\nu(y)\) is \(\mathcal{F}\)-measurable. Similarly the sectional function \(f(\cdot, y)\) is \(\mathcal{F}\)-measurable for each \(y \in Y\) and the map \(y \mapsto \int_X f(x, y) \, d\mu(x)\) is \(\mathcal{G}\)-measurable.

Moreover

\[
\int_{X \times Y} f(x, y) \, d\mu \otimes \nu(x, y) = \int_X \int_Y f(x, y) \, d\nu(y) \, d\mu(x) = \int_Y \int_X f(x, y) \, d\mu(x) \, d\nu(y).
\]

(5.1)

c) An \(\mathcal{F} \otimes \mathcal{G}\)-measurable function \(f : X \times Y \to [-\infty, \infty]\) is \(\mu \otimes \nu\)-integrable if and only if

\[
\int_X \int_Y |f(x, y)| \, d\nu(y) \, d\mu(x) < \infty.
\]

In this case \(f(x, \cdot) \in L^1(\nu)\) for \(\mu\)-almost all \(x\) and \(f(\cdot, y) \in L^1(\mu)\) for \(\nu\)-almost all \(y\) and the formula in b) is valid in the sense that the inner integrals denote an arbitrary measurable function on the respective null-sets where they are not defined.

**Definition 5.1** The measure \(\mu \otimes \nu\) is called the *product (measure)* of \(\mu\) and \(\nu\). The \(\sigma\)-algebra \(\mathcal{F} \otimes \mathcal{G}\) is called the *product (\(\sigma\)-algebra)* of \(\mathcal{F}\) and \(\mathcal{G}\).

Before the proof of the theorem let us write down two important corollaries:
Corollary 5.1 (generalized Cavalieri principle) For every set \( A \in \mathcal{F} \otimes \mathcal{G} \) let \( A_x = \{ y \in Y : (x, y) \in A \} \) and \( A^y = \{ x \in X : (x, y) \in A \} \). Then
\[
\mu \otimes \nu(A) = \int \nu(A_x) \, d\mu(x) = \int \mu(A^y) \, d\nu(y).
\]

Corollary 5.2 Let \( N \subset X \times Y \). The following are equivalent:

a) \( N \) is a \( \mu \otimes \nu \)-null-set,

b) for \( \mu \)-almost all \( x \in X \) one has \( \nu(N_x) = 0 \).

c) for \( \nu \)-almost all \( y \in Y \) one has \( \mu(N^y) = 0 \).

Proof. 1. Suppose first that \( \mu(X) < \infty \) and \( \nu(Y) < \infty \). Consider the system \( \mathcal{D} \) of all subsets \( A \) of \( X \times Y \) for which i) the section sets \( A_x \) are in \( \mathcal{G} \) for all \( x \in X \) and ii) the map \( x \mapsto \nu(A_x) \) is \( \mathcal{F} \)-measurable. Clearly \( X \times Y \in \mathcal{D} \). The set operation \( A \mapsto A_x \) commutes with countable unions and the difference operator \( \setminus \), moreover \( \nu \) is a finite measure. These two facts imply that \( \mathcal{D} \) is a monotone class in the sense of section 2.4. Obviously \( \mathcal{D} \) contains the system \( \mathcal{E} \) of all \( \mathcal{G} \)-measurable rectangles \( F \times G \) with \( F \in \mathcal{F}, G \in \mathcal{G} \). The intersection of two such sets is again in \( \mathcal{E} \). So theorem 2.4 applies and gives \( \mathcal{F} \otimes \mathcal{G} \subset \mathcal{D} \). So the map \( \mu \otimes \nu : A \mapsto \int_X \nu(A_x) \, d\mu(x) \) is well-defined and the \( \sigma \)-additivity of \( \nu \) and the theorem of monotone convergence show that it is a measure on \( \mathcal{F} \otimes \mathcal{G} \).

Moreover
\[
\mu \otimes \nu(F \times G) = \int_X \nu((F \times G)_x) \, d\mu(x) = \int_F \nu(G) \, d\mu(x) = \mu(F) \nu(G).
\]

2. In a completely symmetric fashion can define a measure \( \rho \) on \( \mathcal{F} \otimes \mathcal{G} \) with \( \rho(A) = \int_Y \mu(A^y) \, d\nu \). This measure \( \rho \) satisfies as well \( \rho(F \times G) = \mu(F) \nu(G) \) for all \( F \times G \in \mathcal{E} \). So the uniqueness theorem 2.5 shows \( \rho = \mu \otimes \nu \). This proves also corollary 5.1. The second corollary follows from the first if we use the fact that a non-negative measurable function (here the functions \( x \mapsto \nu(A_x) \) and \( y \mapsto \mu(A_y) \)) vanishes a.e. if and only if its integral vanishes (proposition 4.1 b).

3. If \( f \) is an elementary \( \mathcal{F} \otimes \mathcal{G} \)-measurable function then the statement in part b) for \( f \) follows from the additivity of the integral. If \( f \) is a general non-negative measurable function then it can be approximated by a nondecreasing sequence of elementary functions (theorem 3.3) and part b) follows by a repeated application of the theorem of monotone convergence.

4. Let now \( f \) be a \( \mathcal{F} \otimes \mathcal{G} \)-measurable function. That \( f \) is \( \mu \otimes \nu \)-integrable iff the double integral in c) converges is just part b) applied to \( |f| \). Now let \( f \) be \( \mu \otimes \nu \)-integrable. Then the function \( x \mapsto \int |f(x,y)| \, d\nu(y) \) is \( \mu \)-integrable and thus there is a \( \mu \)-null-set \( N_1 \in \mathcal{F} \) such that this integral is finite and hence the value \( \int_{N_1} f(x,y) \, d\nu(y) \) is well-defined for \( x \notin N_1 \). Let \( g : X \to \mathbb{R} \) be any \( \mathcal{F} \)-measurable function. Let the symbol \( \int_{N} f(x,y) \, d\nu(y) \) denote the value \( g(x) \) for \( x \in N_1 \). Then the formula
\[
\int f(x,y) \, d\nu = \int f^+(x,y) \, d\nu - \int f^-(x,y) \, d\nu
\]
holds \( \mu \)-a.e. and therefore the identity (5.1) carries over from \( f^+ \) and \( f^- \) to \( f \). The second identity in b) follows by exchanging the roles of the variables \( x, y \).

5. Finally we get rid of the finiteness condition on the measures. Let \( (Q_n) \) and \( (R_n) \) be two sequences in \( \mathcal{F} \) resp. \( \mathcal{G} \), such that \( \mu(Q_n) < \infty \), \( \nu(R_n) < \infty \) and
We have used monotone classes. If one does not want to rely on theorem 2.5. This can be verified first on the system of these boxes and then by appealing to the form $Q \otimes B$. Because $\lambda$ only difficulty is that without monotone classes it is not so easy to prove the $F \otimes G$ the unique extension to $R$. Then note that $R$ of the above proof contains the $F$-measurable and the theorem of monotone convergence that $\mu \otimes \nu$ defined by $\mu \otimes \nu(A) = \lim \mu_n \otimes \nu_n(A)$ is a measure. One easily shows via approximation by elementary functions and the theorem of monotone convergence that

$$\int f \, d\mu \otimes \nu = \lim_n \int_{Q_n \times R_n} f \, d\mu_n \otimes \nu_n$$

for all $F \otimes G$-measurable $f \geq 0$. Then the extension of the various assertions are straightforward. The uniqueness statement in a) again is a consequence of theorem 2.5.

Remark 5.1 We have used monotone classes. If one does not want to rely on section 2.4 one can get almost the same result as follows. Let $\mathcal{R}$ be the ring of finite disjoint unions of measurable rectangles. It is easy to extend the set function $F \times G \rightarrow \mu(F)\nu(G)$ by additivity to a content on $\mathcal{R}$ and to see that the function $x \mapsto \nu(R_x)$ is $\mathcal{F}$-measurable and $m(R) = \int_x \nu(R_x) \, d\mu$ for all $R \in \mathcal{R}$. This implies that $m$ is $\sigma$-additive and $\sigma$-finite. Let $\mu \otimes \nu$ denote the unique extension to $\mathcal{F} \otimes \mathcal{G}$. Let now $f \geq 0$ be $\mathcal{F} \otimes \mathcal{G}$-measurable. The only difficulty is that without monotone classes it is not so easy to prove the $\mathcal{F}$-measurability of $x \mapsto \int f(x,y) \, d\nu(y)$. However the $\mathcal{F}_\mu$-measurability can be verified. For this let first $A \in \mathcal{F} \otimes \mathcal{G}$. Then for each $n$ there is a sequence $(R^n_i)$ of (without loss of generality disjoint) sets in $\mathcal{R}$ such that $A \subset \bigcup_i R^n_i =: A_n$ and $\mu \otimes \nu(A) \geq \sum_i m(R^n_i) - \frac{1}{n} = \mu \otimes \nu(A_n) - \frac{1}{n}$. The class $\mathcal{D}$ in part 1. of the above proof contains the $R^n_i$ and their disjoint union $A_n$. Similarly, applying the same argument to $A^c$ we find a set $A_n \in \mathcal{D}$ such that $A_n \subset A$ and $\mu \otimes \nu(A_n) \geq \mu \otimes \nu(A) - \frac{1}{n}$. Then using the same argument as for the Lebesgue-integrability of Riemann integrable functions we see that $x \mapsto \nu(A_x)$ is $\mathcal{F}_\mu$-measurable and the first identity in corollary 5.1 holds. The rest of the proof works essentially like before.

In the special case of Lebesgue-measure some additional comments are in order.

Remarks 1. If we choose $\mathcal{F} = \mathcal{B}(\mathbb{R}^m)$ and $\mathcal{G} = \mathcal{B}(\mathbb{R}^k)$ we have $\mathcal{B}(\mathbb{R}^{m+k}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^k)$, because both $\sigma$-algebras are generated by the compact boxes of the form $Q \times R$. Moreover $\lambda^m_{\mathbb{R}^m} \otimes \lambda_k^{[0,1]}(B) = \lambda^{m+k}(B)$ for all $B \in \mathcal{B}(\mathbb{R}^{m+k})$. This can be verified first on the system of these boxes and then by appealing to theorem 2.5.

2. In some texts on analysis the multidimensional integral of continuous functions of compact support is defined via successive integration. Fubini’s theorem and our remark 4.3 shows that this integral is actually the Lebegue integral of these functions.

3. Let us denote by $L^d$ the $\sigma$-algebra of Lebesgue measurable sets in $\mathbb{R}^d$. Note that $L^{m+k}$ is strictly larger than $L^m \otimes L^k$. For example let $E \subset [0,1]$ be the Vitali set from section 1.4. Then $A = \{0\} \times E$ is a $\lambda^2$-null-set since it is contained in the product measurable $\{0\} \times [0,1]$ which has $\lambda^2$-measure $0 \cdot 1 = 0$. Because $\lambda^2$ is complete we get $A \in L^2$. If $A$ were an element of $L^1 \otimes L^1$ then...
all sections $A_x$ would be in $L^1$ which is not true since $A_0 = E$. Nevertheless, $L^2$ being the $\lambda^2$-completion of the Borel sets, every element of $L^2$ differs only on a $\lambda^2$-null-set from a suitable Borel set.

4. We give an example that the uniqueness of the product measure may get lost if one of the measures is no longer $\sigma$-finite. In view of the arguments of remark 5.1 it also illustrates the importance of $\sigma$-finiteness in the uniqueness part of Carathéodory’s theorem 2.2. Let $X = Y = [0, 1]$ and $\mathcal{F} = \mathcal{G} = \mathcal{B}([0, 1])$. Let $\lambda$ be Lebesgue measure on $\mathcal{F}$ and let $\zeta$ be the counting measure on $\mathcal{G}$. As in the beginning of the proof of theorem 5 we can define two measures $\rho_1$ and $\rho_2$ on $\mathcal{F} \otimes \mathcal{G}$ by

$$\rho_1(B) = \int_0^1 \zeta(B^x) \, d\lambda(x),$$
$$\rho_2(B) = \int_0^1 \lambda(B^y) \, d\zeta(y) = \sum_y \lambda(B^y).$$

For every measurable rectangle

$$\rho_1(F \times G) = \lambda(F)\zeta(G) = \rho_2(F \times G),$$

but for the diagonal $D \subset [0, 1]^2$ we get

$$\rho_1(D) = \int_0^1 \zeta(D^x) \, d\lambda(x) = \int_0^1 1 \, dx = 1,$$

whereas

$$\rho_2(D) = \sum_y \lambda(D^y) = \sum_y 0 = 0.$$
Chapter 6

Convergence in measure

Besides point-wise convergence or a.e. convergence there are several other ways to express that a sequence of measurable functions converges. The $L^p$-spaces of the next chapter will give a full scale of such concepts. Here we discuss convergence in measure which is particularly useful in probability theory. The reader who wants to pass directly to the next chapter will need the completeness result of theorem 6.4b) which can be understood independently of the other parts of this section.

6.1 Comparison with other types of convergence

Definition 6.1 Let $(S, \mathcal{F}, \mu)$ be a measure space. Let $f, f_n (n \in \mathbb{N})$ be extended real valued measurable functions which are $\mu$-a.e. finite. One says that the sequence $(f_n)$ converges to $f$

a) $\mu$-almost everywhere (fast überall), if $\mu\{x : f_n(x) \not\rightarrow f(x)\} = 0$,

b) in measure (nach Maß) if $\mu\{x : |f_n(x) − f(x)| \geq \varepsilon\} \rightarrow 0$ for each $\varepsilon > 0$,

c) in the mean (im Mittel), if $\int |f − f_n| \, d\mu \rightarrow 0$ and all functions are integrable.

Comment. Note that in a) and b) the sets in question are actually in $\mathcal{F}$. We see that in all three parts of Definition 6.1 only the $\mu$-equivalence class of the functions matters. Therefore one usually extends these concepts also to all extended real valued functions which are $\mu$-a.e. finite, i.e. which are equivalent to a finite-valued measurable function.

For infinite measures there is a weaker form of convergence in measure which in finite measure spaces is equivalent to convergence in measure.

Definition 6.2 One speaks of stochastic convergence or local convergence in measure if $f_n 1_A \longrightarrow f 1_A$ in measure for every $A \in \mathcal{F}$ with $\mu(A) < \infty$.

Let $\lambda^1$ be Lebesgue measure on the real line. The indicator functions $f_n = 1_{[n,n+1]}$ do not converge to 0 in measure but point-wise and hence locally in measure by part b) of the following theorem.
Theorem 6.1  Between the three concepts in Definition 6.1 the following implications hold:

a) If \( f_n \rightarrow f \) in the mean then \( f_n \rightarrow f \) in measure.

b) If \( f_n \rightarrow f \mu \) a.e. then \( f_n 1_A \rightarrow f 1_A \) in measure for every \( A \in \mathcal{F} \) with \( \mu(A) < \infty \).

The converse implications are wrong even for finite measure spaces.

Proof. 1. Markov’s inequality (theorem 4.1a) implies for all \( \varepsilon > 0 \)

\[
\mu(\{|f_n - f| \geq \varepsilon\}) \leq \frac{1}{\varepsilon} \int |f_n - f| \, d\mu.
\]

So convergence in mean implies convergence in measure.

2. Assume that \( f_n \rightarrow f \mu \) a.e. Let \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) be given. Let \( A_n = \bigcup_{n \geq n} \{x \in A : |f_n(x) - f(x)| \geq \varepsilon\} \). Then \( A_1 \supset A_2 \supset \ldots \) and \( \mu(A_1) \leq \mu(A) < \infty \). So the continuity from above of \( \mu \) (theorem 1.3) shows \( \mu(A_n) \downarrow \mu(A_\infty) \) where \( A_\infty := \bigcap_{n} A_n \). On the set \( A_\infty \) one has \( \limsup_n |f_n - f| \geq \varepsilon \). Therefore \( A_\infty \) is a \( \mu \)-null-set. Together we get

\[
\mu(\{x \in A : |f_n - f| \geq \varepsilon\}) \leq \mu(A_n) \downarrow \mu(A_\infty) = 0,
\]

i.e. the convergence in measure of \( f_n 1_A \) to \( f 1_A \).

3. The sequence in example 4.4 does not converge in the mean but it converges in measure according to part b) which we just proved.

4. \( f_n \rightarrow f \) in the mean does not imply \( f_n \rightarrow f \mu \) a.e.. Again we work on the standard Lebesgue measure on \([0, 1]\). Let \( (J_n) \), \( n = 1, 2, \ldots \) be an enumeration of all intervals of the form \( \left[k \frac{1}{m}, k \frac{1}{m+1}\right) \) with \( k \leq m; k, m \in \mathbb{N} \). Then for \( f_n = 1_{J_n} \), \( f = 0 \) one gets on one side \( \int |f_n - f| \, d\lambda = \lambda(J_n) \rightarrow 0 \), and on the other hand \( (f_n) \) does not converge to \( f \) a.e. because for all \( x \)

\[
\limsup_{n \rightarrow \infty} |f_n(x) - f(x)| = \limsup_{n \rightarrow \infty} 1_{J_n}(x) = 1.
\]

This completes the proof. \( \blacksquare \)

Part b) of the theorem has a partial converse.

Theorem 6.2  Let \( \mu \) be \( \sigma \)-finite. A sequence \( (f_n) \) converges locally in measure to \( f \) if and only if for every sub-sequence \( (f_{n_k})_k \) there is a sub-sub-sequence \( (f_{n_{k_l}})_l \) which converges \( \mu \)-a.e. to \( f \).

Proof.  We may assume (for simplicity of notation) \( f = 0 \). 1. Suppose \( (f_n) \) converges locally in measure to 0 and let \( (f_{n_k})_k \) be a sub-sequence. Let \( (A_m)_m \) be an increasing sequence of sets of finite measure whose union is \( S \). Then for each \( l \in \mathbb{N} \), the numbers \( \mu(A_l \cap \{|f_{n_k}| > \frac{1}{l}\}) \) converge to 0 as \( k \rightarrow \infty \). So there is an increasing sequence of indices \( k_l \) such that \( \mu(B_l) < 2^{-l} \) where \( B_l = A_l \cap \{|f_{n_{k_l}}| > \frac{1}{l}\} \), for all \( l \).

We claim that \( f_{n_{k_l}} \rightarrow 0 \mu \)-a.e.. Let \( N \) be the exceptional set of all points \( x \) for which \( f_{n_{k_l}}(x) \) does not converge to 0. Let \( x \in N \cap A_m \). There is some \( r \in \mathbb{N} \) such that \( f_{n_{k_l}}(x) > \frac{1}{l} \) for infinitely many \( l \). In particular there is some \( l > \max(r, m) \) with this inequality. Then \( x \in B_l \). Therefore for fixed \( m_0 \) we get
Let $N \cap A_{m_0} \subset N \cap A_m \subset \bigcup_{m=m_0}^{\infty} B_1$ and $\mu(N \cap A_{m_0}) \leq \sum_{m=m_0}^{\infty} 2^{-1} = 2 \cdot 2^{-m}$ for all $m \geq m_0$. Hence $\mu(N \cap A_{m_0}) = 0$ for all $m_0$ and finally $\mu(N) = 0$ as desired.

2. Conversely suppose the subsequence condition holds. Let $A \in \mathcal{F}$ be a set of finite measure and $\varepsilon > 0$. Choose a sub-sequence such that

$$
\lim_{k \to \infty} \sup_n \mu(A \cap \{|f_n - f| > \varepsilon\}) = \lim_{k \to \infty} \mu(A \cap \{|f_{n_k} - f| > \varepsilon\}).
$$

Let $(f_{n_k})_i$ be a sub-sub-sequence such that $f_{n_k} \to f \text{ } \mu\text{-a.e.}$. Then by theorem 6.1 this sub-sub-sequence converges locally in measure to $f$, in particular the above lim sup is equal to $\lim_{k \to \infty} \mu(A \cap \{|f_{n_k} - f| > \varepsilon\}) = 0$. This proves that the full sequence $(f_n)$ also converges locally in measure to $f$.

The following corollary is a partial converse to part a) of theorem 6.1 and an extension of the theorem of dominated convergence to sequences which converge locally in measure.

**Corollary 6.1** Let $h \in L^1(\mu)$ and let $M_h$ be the set all measurable functions $g$ with $|g| \leq h$. Then for sequences in $M_h$ convergence in mean and local convergence in measure agree.

**Proof.** a). Suppose $f_n \in M_h$ for all $n$ and assume that $f_n$ converges locally in measure. Then theorem 6.2 and the theorem of dominated convergence together imply that for every sub-sequence $(f_{n_k})_k$ there is a sub-sub-sequence $(f_{n_{k_j}})_j$ such that $\int |f_{n_{k_j}} - f| \, d\mu \to 0$. But this implies even $\int |f_n - f| \, d\mu \to 0$, i.e. convergence in mean. The converse is theorem 6.1 a).

We conclude this section with a little theorem about uniform convergence on large sets.

**Theorem 6.3** (Egorov). Let $\mu(S) < \infty$ and let $f_n \to f$ $\mu$-a.e.. Then for each $\varepsilon > 0$ there is a set $A \in \mathcal{F}$ with $\mu(A^c) < \varepsilon$ such that $f_n$ converges to $f$ even uniformly on $A$.

**Proof.** Let $g_n = \sup_{k \geq n} |f_k - f|$. Then $g_n \downarrow 0 \text{ } \mu$-a.e., i.e. $\mu(\{g_n > \frac{1}{m}\}) \to 0$ for all $m \in \mathbb{N}$. Then we find a sequence of indices such that $\sum_{m=1}^{\infty} \mu(\{g_{n_m} > \frac{1}{m}\}) < \varepsilon$. Let $A = \cap_{m=1}^{\infty} \{g_{n_m} \leq \frac{1}{m}\}$. Then $\mu(A^c) < \varepsilon$. For the verification of uniform convergence let $\eta > 0$. Choose $m \geq \frac{1}{\eta}$. Then for all $x \in A$ and $n \geq n_m$ one gets $|f_n(x) - f(x)| \leq g_{n_m}(x) \leq \frac{1}{m} \leq \eta$.

### 6.2 A Metric for Local Convergence in Measure

In analysis most convergence concepts correspond to a topological structure, e.g. to a metric. At least for $\sigma$-finite measure spaces there are many ways to define a metric between (equivalence classes of) measurable functions which describes local convergence in measure. Theorem 6.5 will give one of the possibilities.

The following result shows that many metrics between equivalence classes of measurable functions are complete. It implies in particular the completeness of the $L^p$-spaces in the next chapter.\footnote{From this section only theorem 6.4 will be needed later.}
Theorem 6.4 Let $\rho$ be a functional which assigns to every non-negative $\mathcal{F}$-measurable functions $g$ a value $\rho(g) \in [0, \infty]$ with the following properties:

\begin{align*}
\rho(g) &= 0 \quad \text{if and only if} \quad g = 0 \quad \mu - a.e., \\
0 \leq f \leq g &\implies \rho(f) \leq \rho(g), \\
\rho(\sum_{i=1}^{\infty} g_i) &\leq \sum_{i=1}^{\infty} \rho(g_i), \\
\rho(g) < \infty &\implies g < \infty \quad \mu - a.e..
\end{align*}

Then the set $\mathcal{L}^\rho$ of all real valued measurable functions $f$ with $\rho(|f|) < \infty$ is complete with respect to the pseudo-metric $d(f, g) = \rho(|f - g|)$.

**Proof.** The triangular inequality $d(f, h) \leq d(f, g) + d(g, h)$ follows from $(6.2)$ and $(6.3)$. Let $(f_n)$ be a Cauchy sequence with respect to $\rho([f_n - f_m])$ converges to $0$ as $m, n$ tend to $\infty$. Choose a subsequence $(n_k)$ such that $\sum_{k=1}^{\infty} \rho(|f_{n_k+1} - f_{n_k}|) < \infty$. Let $A = \{x : \sum_{k=1}^{\infty} |f_{n_k+1}(x) - f_{n_k}(x)| < \infty\}$. Then $A \in \mathcal{F}$ and $\mu(A^c) = 0$ according to $(6.4)$. Since

$$f_{n_k}(x) = f_{k_0}(x) + \sum_{i=k_0}^{k-1} (f_{n_{i+1}}(x) - f_{n_i}(x)) \quad \text{for} \quad k > k_0,$$

the sequence $(f_{n_k}(x))$ converges for $k \to \infty$ to a finite real number $f(x)$ for all $x \in A$. Outside of $A$ let $f(x) = 0$. The measurability of the $f_n$ and of $A$ imply the measurability of $f$. Moreover $|f - f_{n_k}| \leq \sum_{l=k}^{\infty} |f_{n_{l+1}} - f_{n_l}|$ everywhere and by $(6.2)$ and $(6.3)$

$$\rho(|f - f_{n_k}|) \leq \sum_{l=k}^{\infty} \rho(|f_{n_{l+1}} - f_{n_l}|) < \infty$$

for each $k$. In particular $\rho(|f_{n_k} - f|) \to 0$ for $k \to \infty$. The triangular inequality gives $\rho(|f|) < \infty$ and finally also $\rho(|f - f_{n_k}|) \leq \rho(|f - f_{n_k}|) + \rho(|f_{n_k} - f_{n_k}|) \to 0$ for the full sequence. This proves the completeness. \hfill \Box

In order to define a metric for local convergence in measure we use a strictly positive integrable function. If $\mu(S) < \infty$ one takes the constant function $1$. If $\mu = \lambda^1$ take e.g. $h(x) = e^{-|x|}$.

**Lemma 6.1** There exists an everywhere positive function $h \in L^1(\mu)$ if and only if the measure $\mu$ is $\sigma$-finite.

**Proof.** If $h \in L^1(\mu)$ is everywhere positive, then $S = \bigcup_{n=1}^{\infty} \{ |h| > \frac{1}{n} \}$ and each of the sets in this union has finite $\mu$-measure by Markov’s inequality. So $\mu$ is $\sigma$-finite. Conversely if $\mu$ is $\sigma$-finite there is a sequence $(A_n)$ of sets in $\mathcal{F}$ with $A_n \uparrow S$ and $\mu(A_n) < \infty$ for all $n$. Choose constants $c_n > 0$ such that $\sum c_n \mu(A_n) < \infty$. Then $h = \sum_{n=1}^{\infty} c_n 1_{A_n}$ is measurable and $\int |h| \, d\mu = \sum_{n=1}^{\infty} c_n \mu(A_n) < \infty$ and of course $h(x) > 0$ for all $x \in S$. \hfill \Box

**Theorem 6.5** Let $h \in L^1(\mu)$ be everywhere positive. Define for finite measurable functions $f, g$ the distance $d(f, g)$ by

$$d(f, g) = \int_S h \wedge |f - g| \, d\mu.$$
Then $d$ is a complete pseudo-metric such that $\lim_{n \to \infty} d(f_n, f) = 0$ if and only if $(f_n)$ converges locally in measure to $f$. If moreover $\sum_n d(f_n, f) < \infty$ then $f_n \to f \mu$-a.e..

Proof. Let $\rho(g) = \int h \wedge g \, d\mu$. Then $\rho$ satisfies the conditions (6.1) - (6.4). The first two are obvious. For the third one note that $h \wedge \sum_{i=1}^{\infty} g_i \leq \sum_{i=1}^{\infty} (h \wedge g_i)$ (distinguish the two cases $g_i \leq h$ for all $i$ or $g_i > h$ for some $i$). Thus

$$
\rho\left(\sum_{i=1}^{\infty} g_i\right) = \int h \wedge \sum_{i=1}^{\infty} g_i \, d\mu \leq \int \sum_{i=1}^{\infty} (h \wedge g_i) \, d\mu = \sum_{i=1}^{\infty} \int h \wedge g_i \, d\mu = \sum_{i=1}^{\infty} \rho(g_i).
$$

For the proof of (6.4) assume that in this calculation the last series converges. Then the function $\sum_{i=1}^{\infty} (h \wedge g_i)$ is integrable and hence a.e. finite. Let $x$ be a point such that $\sum_{i=1}^{\infty} (h(x) \wedge g_i(x))$ converges. Then $g_i(x) < h(x)$ for eventually all indices because $h(x) > 0$ and therefore the series $\sum_{i=1}^{\infty} g_i(x)$ converges. Thus $\sum_{i=1}^{\infty} g_i$ is finite a.e.. Therefore $d$ is a complete pseudo-metric.

If $\sum_n d(f_n, f) < \infty$ then the series $\sum_n f_n(x) - f(x)$ converges absolutely $\mu$-a.e. because of (6.4) and in particular $f_n \to f \mu$-a.e..

Assume that $\lim_n d(f_n, f) = 0$. In order to prove local convergence in measure we verify the sub-sequence criterion of theorem 6.2. Let $(f_{n_k})$ be a sub-sequence. Then every sub-sub-sequence with $\sum_l d(f_{n_{k_l}}, f) < \infty$ converges to $f \mu$-a.e..

Conversely assume local convergence in measure of $f_n$ to $f$ or $|f_n - f|$ to $0$. Then the same is true for the functions $|f_n - f| \wedge h$. By corollary 6.1 these functions converge also in mean, i.e. by definition $d(f_n, f) = \int |f_n - f| \, d\mu \to 0$. ■
Chapter 7

The $L^p$-spaces

**Definition 7.1** Let $(S, \mathcal{F}, \mu)$ be a measure space and let $f$ be an extended real-valued $\mathcal{F}$-measurable function. We define for $p \in [1, \infty]$ the number $\|f\|_p$ by

$$
\|f\|_p := \left( \int |f|^p \, d\mu \right)^{1/p} \quad (1 \leq p < \infty)
$$

$$
\|f\|_\infty := \inf\{\alpha > 0 : \mu(|f| > \alpha) = 0\}.
$$

Further let $L^p(\mu) = \{ f : f$ is $\mathcal{F}$-measurable and $\|f\|_p < \infty\}$. $L^p(\mu)$ denotes the space of all $\mu$-equivalence classes of elements of $L^p(\mu)$.

If $p = 1$ this is in accordance with with Definition 4.3. For $p = \infty$ a measurable function is in $L^\infty(\mu)$ if and only if it is $\mu$-equivalent to a bounded function. In the case $p = 2$ the functional $\|\cdot\|_2$ is the (semi-) norm induced by the scalar product $\langle f, g \rangle = \int fg \, d\mu$ via $\|f\|_2 = \sqrt{\langle f, f \rangle}$. In particular one has the Cauchy-Schwarz(-Bunyakowsky) inequality

$$
|\int_S fg \, d\mu| = |\langle f, g \rangle| \leq \|f\|_2 \|g\|_2. \tag{7.1}
$$

Note that an extended real-valued function is $\mathcal{F}_\mu$-measurable if and only if it is $\mu$-equivalent to an $\mathcal{F}$-measurable function according to corollary 3.2 and proposition 4.2 a). Thus the equivalence classes of $\mathcal{F}_\mu$-measurable functions and of $\mathcal{F}$-measurable functions agree and each measure has the same $L^p$-spaces as its completion.

The following theorem gives an extension of (7.1) and it shows that $\|\cdot\|_p$ is really a norm on $L^p(\mu)$.

**Theorem 7.1** For all $\lambda \in \mathbb{R}$ $\|\lambda f\|_p = |\lambda|\|f\|_p$. Moreover

a) $\|f \cdot g\|_r \leq \|f\|_p \|g\|_q$ whenever $r, p, q \in [1, \infty]$ satisfy $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.

(Hölder’s inequality)

b) $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ for all $p \in [1, \infty]$.

(Minkowski’s inequality)

**Proof.** The homogeneity is obvious.
a) For $p = \infty$ the condition on $r,p,q$ implies $r = q$. Moreover $|fg| \leq \|f\|_\infty |g|$ $\mu$-a.e. and hence
\[
\|fg\|_r \leq \|\|f\|_\infty |g|\|_r = \|f\|_\infty \|g\|_r = \|f\|_\infty \|g\|_q.
\]
For finite $p,q$ we use the following

**Lemma 7.1** Let $p,q,r \in [1,\infty)$ satisfy the relation $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Then for all $\alpha,\beta \geq 0$
\[
\alpha^\frac{r}{p} \beta^\frac{r}{q} \leq \frac{r}{p} \alpha + \frac{r}{q} \beta.
\]

**Proof.** (of the lemma) The function $\varphi(x) = x^\frac{r}{p}$ is concave because of $p > r$. Using the equation of the tangent at the point $(1,1)$ we get $\varphi(x) = \frac{r}{p}(x-1)+1 = \frac{r}{p}x + \frac{r}{q}$. In this proof we may assume $\beta > 0$. For $x = \frac{q}{p}$ we get $(\frac{q}{p})^\frac{r}{p} \leq \frac{r}{p} \beta + \frac{r}{q}$. Since $\beta(\frac{q}{p})^\frac{r}{p} = \alpha^\frac{r}{p} \beta^\frac{r}{q} = \alpha^\frac{r}{p} \beta^\frac{r}{q}$ the lemma is proved. 

Now let two measurable functions $f, g$ be given. We may assume $f \geq 0, g \geq 0$ and $0 < \|f\|_p, \|g\|_q < \infty$. Choosing $\alpha = \frac{f^p(x)}{\|f\|_p}$ and $\beta = \frac{g^q(x)}{\|g\|_q}$ in the lemma and integrating we get
\[
\int \frac{(f(x)g(x))^r}{\|f\|_p^p \|g\|_q^q} \, d\mu \leq \int \frac{r f^p(x)}{p \|f\|_p^p} \, d\mu + \int \frac{g^q(x)}{\|g\|_q^q} \, d\mu = \frac{r}{p} + \frac{r}{q} = 1,
\]
i.e. Hölder’s inequality.

b) (Minkowski) For $p = 1$ and $p = \infty$ the assertion is easy. Let now $1 < p < \infty$. Let $0 < \|f\|_p + \|g\|_p < \infty$. For the proof of Minkowski’s inequality we may assume $f, g \geq 0$. Because of $\|f + g\| \leq 2^p \max(f,g)^p \leq 2^p (f^p + g^p)$ one gets $\|f + g\|_p < \infty$. Letting $r = 1, q = \frac{p}{p-1}$ we get from a)
\[
\|f + g\|_p^p = \|f(f + g)^{p-1}\|_1 + \|g(f + g)^{p-1}\|_1 \\
\leq \|f\|_p((f + g)^{p-1})_q + \|g\|_p((f + g)^{p-1})_q \\
= (\|f\|_p + \|g\|_p)(f + g)^{p-1},
\]
which implies the assertion. In the last step we have used the general relation
\[
\|h^{p-1}\|_q = \left( \int h^{(p-1)q} \, d\mu \right)^\frac{1}{p} = \left( \int h^p \, d\mu \right)^\frac{p-1}{p} = \|h\|_p^{p-1}.
\]

**Corollary 7.1** Let $\mu(S) < \infty$ and $p < q$. Then $\|f\|_p \leq \|f\|_q \mu(S)^\frac{1}{q} - \frac{1}{p}$ for all measurable functions $f$ and in particular $L^q(\mu) \subset L^p(\mu)$.

**Proof.** Let $s$ be the number for which $\frac{1}{s} = \frac{1}{p} - \frac{1}{q}$. The constant function 1 is in $L^s(\mu)$ with $\|1\|_s = \mu(S)^\frac{1}{s}$. Hölder’s inequality yields $\|f\|_p = \|f \cdot 1\|_p \leq \|f\|_q \|1\|_s$, i.e. the assertion. 

**Theorem 7.2** For each $p \in [1,\infty]$ the space $L^p(\mu)$ with the norm $\| \cdot \|_p$ is a Banach space.
Proof. It remains to prove the completeness. In the case \( p = \infty \) the proof is analogous to the better known case of the space of bounded continuous functions with the sup-norm. In the case \( p < \infty \) we let \( \rho(f) = \|f\|_p \) where \( f \) is any non-negative measurable function and \( \hat{f} \) is the equivalence class of \( f \). One can verify the conditions (6.1)-(6.4): The first two conditions are trivial. Minkowski’s inequality together with the theorem of monotone convergence imply (6.3). For the proof of (6.4) let \( f \) satisfy \( \rho(f) < \infty \). Then \( |f|^p \) is integrable and hence \( f \) is finite \( \mu \)-a.e. Thus theorem 6.4 implies the asserted completeness. □

As a preparation of the remaining results we remark that for a measurable set \( B \in \mathcal{F} \) and \( 1 \leq p < \infty \) one has

\[
\|1_B\|_p = \left( \int_B 1_B^p \, d\mu \right)^{\frac{1}{p}} = \left( \int_B 1 \, d\mu \right)^{\frac{1}{p}} = \mu(B)^{\frac{1}{p}}
\]

and hence \( \mu(B) < \infty \) is equivalent to \( 1_B \in L^p(\mu) \). Moreover the theorem 4.5 of dominated convergence holds also in \( L^p \): If a sequence of measurable functions satisfies \( |f_n| \leq h \in L^p(\mu) \) and \( f_n \rightarrow f \) point-wise, then \( \|f_n - f\|_p \rightarrow 0 \). In fact \( |f_n - f|^p \leq h^p \) and we can apply the original theorem 4.5 to these functions and conclude \( \int |f_n - f|^p \, d\mu \rightarrow 0 \).

**Proposition 7.1** Let \( \mathcal{E} \) be a \( \cap \)-stable generating system of the \( \sigma \)-algebra \( \mathcal{F} \). Assume that there is an increasing sequence \( E_k \) of sets in \( \mathcal{E} \) with \( \mu(E_k) < \infty \) and \( S = \bigcup_k E_k \). Then the linear span of all indicator functions \( 1_{E_k}, E \in \mathcal{E} \) is dense in \( L^p(\mu) \) for \( 1 \leq p < \infty \).

**Proof.** Let \( L \) be the smallest closed linear subspace of \( L^p(\mu) \) which contains the indicator functions \( 1_{E_k}, E \in \mathcal{E} \). We want to show \( L = L^p(\mu) \). Fix a number \( k \in \mathbb{N} \). We consider the class \( \mathcal{D}_k \) of all sets \( F \in \mathcal{F} \) such that \( 1_{F \cap E_k} \in L \). For every \( E \in \mathcal{E} \) we have \( E \cap E_k \in \mathcal{E} \) and hence \( 1_{E \cap E_k} \in L \) by assumption. So \( \mathcal{E} \subset \mathcal{D}_k \).

We claim that \( \mathcal{D}_k \) is a monotone class. Since \( S \cap E_k = E_k \in \mathcal{E} \) we have \( S \in \mathcal{D}_k \). Let \( F, G \in \mathcal{D}_k \) such that \( G \subset F \). Then \( 1_{(F \cap G) \cap E_k} = 1_{F \cap E_k} - 1_{G \cap E_k} \in L \) since \( L \) is a linear space. Thus \( F \setminus G \in \mathcal{D}_k \). Let \( F_n \) be an increasing sequence in \( \mathcal{D}_k \) and \( F_n \uparrow F \). Then \( F_n \cap E_k \uparrow F \cap E_k \) as \( n \rightarrow \infty \) and \( \mu(F \cap E_k) < \infty \). Thus \( \|1_{F \cap E_k} - 1_{F_n \cap E_k}\|_p = \mu((F \setminus F_n) \cap E_k) \rightarrow 0 \) as \( n \rightarrow \infty \) and hence \( 1_F \in L \) since \( L \) is closed. Therefore \( F \in \mathcal{D}_k \). The monotone class theorem 2.4 shows that \( \mathcal{F} \subset \mathcal{D}_k \) for all \( k \). Let now \( F \in \mathcal{F} \) be any measurable set with \( \mu(F) < \infty \). Then \( F \cap E_k \uparrow F \) and as above we see that \( 1_F = \lim_n 1_{F \cap E_k} \in L \).

Finally let \( f \in L^p(\mu) \). By theorem 3.3 there is a sequence of elementary functions \( (f_n)_n \) which increases to \( f^+ \). Each \( f_n \) is in \( L \) as a linear combination of indicators of sets of finite measure. Then \( f_n \rightarrow f^+ \) in \( L^p(\mu) \) by the above extended dominated convergence and hence \( f^+ \in L \) and similarly \( f^- \in L \) and \( f \in L \).

**Theorem 7.3** Let \( 1 \leq p < \infty \). Let \( \mu \) be a measure on \( \mathcal{B}(\mathbb{R}^d) \) for which bounded sets have finite measure. Then the continuous functions with compact support are dense in \( L^p(\mu) \). Similarly for a finite measure \( \mu \) on the Borel sets of a metric space the bounded continuous functions are dense in \( L^p(\mu) \).

**Proof.** In the first situation let \( L \) be the closure of \( C_c(\mathbb{R}^d) \) in \( L^p(\mu) \). Let \( U \) be an open set of finite \( \mu \)-measure. There is a sequence \( (f_n) \) in \( C_c(\mathbb{R}^d) \) such
that \( 0 \leq f_n \uparrow 1_U \) point-wise. Then by the remark above about dominated convergence in \( L^p \) we see that \( 1_U \in L \). The Borel \( \sigma \)-algebra is generated by these sets \( U \) and the result follows from part \( b) \) of the preceding proposition.

In the second situation we simply replace \( C_c(\mathbb{R}^d) \) by the space of all bounded continuous functions.

\[
\text{Corollary 7.2 (Lusin’s Theorem)} \quad \text{Let } \mu \text{ be as in the second part of the previous theorem. Let } f \text{ be measurable and real-valued. Let } A \text{ be a Borel set of finite measure. Then for each } \varepsilon > 0 \text{ there is a set } K \subset A \text{ with } \mu(K) > \mu(A) - \varepsilon \text{ such that } f \text{ is continuous on } K.
\]

**Proof.** We may assume that \( f \) vanishes outside \( A \). Since \( A = \bigcup_n A \cap \{|f| < n\} \) we may assume that \( f \) is bounded and in particular \( f \in L^1(\mu) \). There is a sequence \( (f_n) \) of continuous functions which converges in \( L^1(\mu) \) to \( f \). According to theorems 6.1 and 6.2 we can assume that \( f_n \rightharpoonup f \) \( \mu \)-a.e.. By Egorov’s theorem 6.3 for each \( \varepsilon > 0 \) there is a measurable set \( K \subset A \) with \( \mu(K) > \mu(A) - \varepsilon \) such that the \( f_n \) converge uniformly to \( f \) on \( K \). Since each \( f_n \) is continuous the same is true for the restriction of \( f \) to \( K \).
Chapter 8

The Radon-Nikodym Theorem

8.1 Absolute continuity of measures

In this section \((S, \mathcal{F})\) is a measurable space and \(\mu\) and \(\nu\) are measures on \(\mathcal{F}\). As a consequence of the theorem of monotone convergence we saw in remark 4.1 that for a measurable function \(f \geq 0\) the set function \(f\mu : \mathcal{F} \to [0, \infty]\) defined by

\[
f\mu : A \mapsto \int_A f \, d\mu
\]  

(8.1)

is a measure. The theorem of Radon-Nikodym characterizes the measures which have such a representation.

**Definition 8.1** Let \(\mu\) and \(\nu\) be two measures on the \(\sigma\)-algebra \(\mathcal{F}\). We call a function \(f\) a Radon-Nikodym-density (R-N-Dichte) or R-N-derivative (R-N-Ableitung) of \(\nu\) w.r.t. \(\mu\) if

\[
\nu(A) = \int_A f \, d\mu.
\]  

(8.2)

for all sets \(A \in \mathcal{F}\). We denote any such function \(f\) by the symbol \(\frac{d\nu}{d\mu}\). The \(\mu\)-equivalence class of \(\frac{d\nu}{d\mu}\) is uniquely determined by \(\nu\).

**Proposition 8.1** Let \(\nu, \nu'\) be two measures with the Radon-Nikodym densities \(f = \frac{d\nu}{d\mu}\) and \(f' = \frac{d\nu'}{d\mu}\). Then \(\nu \leq \nu'\) if and only if \(f \leq f'\) \(\mu\)-a.e.. Two measurable functions are Radon-Nikodym densities of the same measure \(\nu\) if and only if they are \(\mu\)-equivalent.

**Proof.** Suppose that \(\nu \leq \nu', \) yet \(\mu(\{f > f'\}) > 0\). Then proposition 4.1 part b) applied to the function \(1_{\{f > f'\}}(f - f')\) would give the contradiction

\[
\nu(\{f' > f\}) = \int_{\{f' > f\}} f \, d\mu > \int_{\{f' > f\}} f' \, d\mu = \nu'(\{f' > f\}).
\]

Therefore \(f \leq f'\) \(\mu\)-a.e.. Conversely clearly \(f \leq f'\) \(\mu\)-a.e. implies \(\nu \leq \nu'\). The second statement easily follows from the first. \(\blacksquare\)
Definition 8.2 Let $\mu$ and $\nu$ be two measures on the $\sigma$-algebra $\mathcal{F}$. Then $\nu$ is called absolutely continuous (absolut stetig) with respect to $\mu$, if every $\mu$-null-set is a $\nu$-null-set. In this case we write $\nu \ll \mu$.

Proposition 8.2 Let $\mu$ and $\nu$ be measures on the $\sigma$-algebra $\mathcal{F}$, and suppose that $\nu$ is finite. Then the following are equivalent.

a) $\nu \ll \mu$, i.e. $\nu$ is absolutely continuous w.r.t. $\mu$,

b) For every $\varepsilon > 0$ there is a $\delta > 0$, such that for each set $A \in \mathcal{F}$ with $\mu(A) < \delta$ one has $\nu(A) < \varepsilon$.

Proof. b) $\rightarrow$ a) is trivial.

a) $\rightarrow$ b). If b) does not hold then there are some $\varepsilon > 0$, a sequence $(\delta_n)$ of positive reals converging to $0$, and a sequence $(\lambda_n)$ of measurable sets with $\mu(\lambda_n) < \delta_n$ and $\nu(\lambda_n) \geq \varepsilon$ for all $n$. Without loss of generality we may assume $\sum \delta_n < \infty$. Let $A = \bigcap_n \bigcup_{m \geq n} \lambda_m$. Then $\mu(A) = 0$ because of $\sum \mu(\lambda_n) < \infty$. On the other hand $\infty > \nu(\bigcup_{m \geq n} \lambda_m) \geq \varepsilon$ for all $n$ and thus $\nu(A) \geq \varepsilon$ since $\bigcup_{m \geq n} \lambda_m \Downarrow A$ as $n \rightarrow \infty$. Thus a) fails as well. \hfill \blacksquare

Note that under condition a) every $\mu$-a.e. converging sequence of measurable functions converges also $\nu$-a.e.. If both measures are finite theorem 6.2 implies that then every sequence $(f_n)$ which converges in measure to $0$ with respect to $\mu$ does the same with respect to $\nu$. If we apply this to indicator functions this is just condition b). This gives an alternative proof of the proposition if both measures are finite.

Theorem 8.1 (Radon-Nikodym) Let $\nu$ and $\mu$ be $\sigma$-finite and suppose $\nu \ll \mu$.

a) There is a Radon-Nikodym density $\frac{d\nu}{d\mu}$.

b) If either $g \geq 0$ is measurable or $g \in L^1(\nu)$ then
\[ \int g \, d\nu = \int g \frac{d\nu}{d\mu} \, d\mu. \] (8.3)

Proof. 1. Note that b) holds for all pairs $\mu, \nu$ for which a) holds. First we get the equality in b) if $g$ is a non-negative elementary function. Then by monotone convergence one extends the equality to all non-negative functions $g$ and finally to all $g \in L^1(\nu)$.

2. For the existence proof assume first that $\nu$ and $\mu$ are finite measures with $0 \leq \nu \leq \mu$. We use, following an idea of J. v. Neumann, the following basic tool from functional analysis (theorem of Fischer-Riesz): Let $\mathcal{H}$ be a real Hilbert space with the scalar product $\langle \cdot, \cdot \rangle$. Let $\ell : \mathcal{H} \rightarrow \mathbb{R}$ be a linear functional for which there is some positive constant $c$ such that $\ell(h) \leq c \cdot \|h\|$ for all $h \in \mathcal{H}$. Then there is an element $f$ of $\mathcal{H}$ such that $\ell(h) = \langle h, f \rangle$ for all $h \in \mathcal{H}$.

To apply this we work in the space $L^2(\mu)$ and in the associated Hilbert space $L^2(\mu)$. The norm is induced by the scalar product $\langle f, h \rangle = \int fh \, d\mu$. Then by corollary 7.1 or by Cauchy-Schwarz we get
\[ | \int h \, d\nu | \leq \nu(S)^{\frac{1}{2}} \left( \int h^2 \, d\nu \right)^{\frac{1}{2}} \leq \nu(S)^{\frac{1}{2}} \left( \int h^2 \, d\mu \right)^{\frac{1}{2}} = \text{const} \cdot \|h\|_{2,\mu} \]
for all $h \in L^2(\mu)$. Moreover two $\mu$-equivalent functions are also $\nu$-equivalent and hence the value of the integral $\int h \, d\nu$ depends only on the $\mu$-equivalence
class of \( h \). Thus by \( \ell(h) = \int h \, d\nu \) a linear functional on \( L^2(\mu) \) is defined which fulfills the assumption of the theorem of Fischer-Riesz. So there is a function \( f \in L^2(\mu) \) with \( \int h \, d\nu = \ell(h) = \int hf \, d\mu \) for all \( h \in L^2(\mu) \). Let \( A \in \mathcal{F} \). Since \( \mu \) is finite we have \( 1_A \in L^2(\mu) \) and so

\[
\nu(A) = \ell(1_A) = \int_A f \, d\mu,
\]

i.e. \( a) \) is proved under the above additional assumption.

3. Let now \( \nu \) and \( \mu \) be finite measures with \( \nu \ll \mu \). Set \( \rho = \mu + \nu \). Then \( \rho \) is a finite measure and both \( \mu \leq \rho \) and \( \nu \leq \rho \). Let \( f = \frac{d\nu}{d\rho} \) and \( g = \frac{d\mu}{d\rho} \). Then \( \mu(\{g = 0\}) = 0 \) and hence \( \nu(\{g = 0\}) = 0 \) and \( \rho(\{g = 0\}) = 0 \), i.e. \( g > 0 \) \( \rho \)-a.e.

Part b) applied to the pair \( \mu, \rho \) gives

\[
\nu(A) = \int_A f \, d\rho = \int_A \frac{f}{g} \, d\rho = \int_A \frac{f}{g} \, d\mu,
\]

i.e. the RN-density \( \frac{d\nu}{d\rho} \) is the quotient \( \frac{f}{g} \) of the RN-densities with respect to \( \rho \).

4. Finally let \( \mu \) and \( \nu \) be \( \sigma \)-finite with \( \nu \ll \mu \). According to lemma 6.1 there are two strictly positive measurable functions \( g \in L^1(\mu) \) and \( h \in L^1(\nu) \). Then the two measures \( g\mu \) and \( h\nu \) (cf. (8.1)) are finite and it is easy to check that they have the same null-sets as the measures \( \mu \) and \( \nu \), respectively. Thus \( h\nu \ll g\mu \) and there is a RN-density \( f = \frac{d\nu}{d\mu} \). Then using (8.3) again we get for all \( A \in \mathcal{F} \)

\[
\nu(A) = \int 1_A \frac{1}{h} \, d\nu = \int 1_A \frac{1}{h} \, d\nu = \int 1_A \frac{1}{h} \, d\mu = \int \frac{g}{h} f \, d\mu,
\]

i.e. \( \frac{g}{h} f \) is the desired Radon-Nikodym derivative of \( \nu \) w.r.t. \( \mu \).

In order to illustrate the role of \( \sigma \)-finiteness we consider the Lebesgue measure \( \lambda \) on the unit interval and the measure \( \mu \) with the same null-sets but \( \mu(A) = \infty \) if \( \lambda(A) > 0 \). Then \( \lambda \ll \mu \) but \( \frac{d\lambda}{d\mu} \) does not exist.

### 8.2 Lebesgue and Hahn-Jordan decomposition

**Definition 8.3** Two measures \( \mu \) and \( \nu \) on \( \mathcal{F} \) are called singular (singulär) or orthogonal to each other if they live on disjoint sets, i.e. there is a measurable set \( D \) such that \( \mu(D) = 0 \) and \( \nu(D^c) = 0 \). We write in this case \( \mu \perp \nu \).

**Theorem 8.2** (Lebesgue decomposition) Let \( \mu \) and \( \nu \) be \( \sigma \)-finite measures. Then \( \nu \) has a unique decomposition

\[
\nu = \nu_{ac} + \nu_s \tag{8.4}
\]

where \( \nu_{ac} \ll \mu \) and \( \nu_s \perp \mu \).

**Proof.** Again consider \( \rho = \mu + \nu \). Then \( \rho \) is also \( \sigma \)-finite: If \( (E_n) \) and \( (E'_n) \) are two increasing sequence with union \( S \) and \( \mu(E_n) < \infty \) and \( \nu(E'_n) < \infty \) then \( (E_n \cap E'_n) \) is an increasing sequence with union \( S \) and \( \rho(E_n \cap E'_n) = \mu(E_n \cap E'_n) + \nu(E_n \cap E'_n) \leq \mu(E_n) + \nu(E'_n) < \infty \). Thus there are RN-densities \( f = \frac{d\nu}{d\rho} \) and \( g = \frac{d\mu}{d\rho} \). Let \( D = \{ f > 0 \} \). Then \( \nu = \nu_{ac} + \nu_s \), where \( \nu_{ac}(A) = \nu(A \cap D) \) and
\[ \nu_s(A) = \nu(A \cap D^c). \] Then \( \nu_s \perp \mu \) since \( \mu(D^c) = \int_{\{f=0\}} f \, d\rho = 0 \) and \( \nu_s(D) = 0 \).

Finally

\[ \nu_{ac}(A) = \nu(A \cap D) = \int_{A \cap D} g \, d\rho = \int_A 1_{D^c} g f \, d\rho = \int_A 1_{D^c} g \, d\mu. \]

Therefore \( \nu_{ac} \) has a Radon-Nikodym density with respect to \( \mu \), i.e. \( \nu_{ac} \ll \mu \).

Sometimes it is useful to admit ‘measures’ which can have also negative values.

**Definition 8.4** Let \((S, \mathcal{F})\) be a measurable space. A function \( \nu : \mathcal{F} \to \mathbb{R} \) is called a signed measure (signiertes Maß), if there are two finite measures \( \rho_1, \rho_2 \) with \( \nu = \rho_1 - \rho_2 \).

**Remark.** Actually, every real valued set function on \( \mathcal{F} \) which is \( \sigma \)-additive in the sense of equation (1.1) is a signed measure. This fact is interesting but rarely used. So we omit the proof. Note also that contrary to common use of language a signed measure in general is not a measure. The representation of a signed measures as a difference of two (non-negative) measures is not unique. The following result provides a ’minimal’ decomposition of this type.

**Theorem 8.3** (Hahn-Jordan decomposition) Every signed measure \( \nu \) on the \( \sigma \)-algebra \( \mathcal{F} \) has a unique decomposition \( \nu = \nu^+ - \nu^- \) where \( \nu^+ \) and \( \nu^- \) are two orthogonal measures. This decomposition is ’minimal’: If \( \nu = \kappa_1 - \kappa_2 \) for two measures then \( \nu^+ \leq \kappa_1 \) and \( \nu^- \leq \kappa_2 \).

**Proof.** By assumption \( \nu = \rho_1 - \rho_2 \) for some finite measures \( \rho_i \). Consider the measure \( \rho = \rho_1 + \rho_2 \). Let \( f_i = \frac{d\rho_i}{d\rho} \) for \( i = 1, 2 \). The function \( f = f_1 - f_2 \) and the set \( D = \{ f \geq 0 \} \) satisfy for all \( A \in \mathcal{F} \)

\[
\nu(A) = \int_A f_1 \, d\rho - \int_A f_2 \, d\rho = \int_A f \, d\rho = \int_{A \cap \{ f \geq 0 \}} f \, d\rho + \int_{A \cap \{ f < 0 \}} f \, d\rho = \nu(A \cap D) + \nu(A \cap D^c)
\]

where clearly in the last two sums the first term is non-negative and the second term is non-positive for all \( A \). Thus \( \nu^+ = \nu(D \cap \cdot) \) and \( \nu^- = -\nu(D^c \cap \cdot) \) provide the indicated decomposition.

Every representation \( \nu = \nu^+ - \nu^- \) as difference of two singular measures is minimal in the above sense: Suppose \( \nu = \kappa_1 - \kappa_2 \) and let \( D \) be a measurable set such that \( \nu^+(D^c) = 0 = \nu^-(D) \). Then

\[
\nu^+(A) = \nu(A \cap D) = \kappa_1(A \cap D) - \kappa_2(A \cap D) \leq \kappa_1(A \cap D) \leq \kappa_1(A)
\]

for all \( A \in \mathcal{F} \), i.e. \( \nu^+ \leq \kappa_1 \) and hence also \( \kappa_2 - \nu^- = (\kappa_1 - \nu) - (\nu^+ - \nu) = (\kappa_1 - \nu^+) \geq 0 \). If moreover \( \kappa_1 \perp \kappa_2 \) we can argue by symmetry that \( \kappa_1 \leq \nu^+ \) and \( \kappa_2 \leq \nu^- \). This proves the uniqueness.

**Definition 8.5** The space of signed measures on the measurable space \((S, \mathcal{F})\) is denoted by \( \mathcal{M}(S, \mathcal{F}) \). For \( \nu \in \mathcal{M}(S, \mathcal{F}) \) the measure \( |\nu| = \nu^+ + \nu^- \) is called the
modulus (Betrag) or total variation measure of $\nu$. The space $\mathcal{M}(S, \mathcal{F})$ is equipped with the norm of total variation (Totalvariation) $\| \cdot \|$ defined by

$$\|\nu\| = |\nu(S)| = \nu^+(S) + \nu^-(S).$$  \hfill (8.5)

The Hahn-Jordan decomposition allows to extend the Radon-Nikodym theorem to signed measures $\nu$.

**Corollary 8.1** (‘signed’ Radon-Nikodym) Let $\mu$ be a measure. Let the signed measure $\nu$ have the property that every $\mu$-null-set is a $\nu$-null-set. Then $\nu^+$ and $\nu^-$ have the same property. There is a function $\frac{d\nu}{d\mu} \in L^1(\mu)$, uniquely determined up to $\mu$-equivalence, such that $\nu(A) = \int_A \frac{d\nu}{d\mu} \, d\mu$ for all $A \in \mathcal{F}$. Moreover

$$\left( \frac{d\nu}{d\mu} \right)^+ = \frac{d\nu^+}{d\mu}, \quad \left( \frac{d\nu}{d\mu} \right)^- = \frac{d\nu^-}{d\mu}, \quad |\frac{d\nu}{d\mu}| = \frac{d|\nu|}{d\mu}. \hfill (8.6)$$

**Proof.** For the uniqueness of $\frac{d\nu}{d\mu}$ note that the proof of proposition 8.1 carries over to our situation. In order to show $\nu^+ \ll \mu$ and $\nu^- \ll \mu$ let $\mu(N) = 0$. Then $\mu(N \cap D) = \mu(N \cap D^c) = 0$ and hence $\nu^+(N) = \nu(N \cap D) = 0$ and $\nu^-(N) = -\nu(N \cap D^c) = 0$. The function $f = \frac{d\nu}{d\mu}$ defined by $\frac{d\nu}{d\mu} = \frac{d\nu^+}{d\mu} - \frac{d\nu^-}{d\mu}$ has the desired properties since according to the proof of the Hahn-Jordan decomposition $f \geq 0$ on $D$ and $f < 0$ on $D^c$. □

**Remark 8.1** The last identity in (8.6) implies

$$\|\nu\| = |\nu|(S) = \int_S |\frac{d\nu}{d\mu}| \, d\mu = \left\| \frac{d\nu}{d\mu} \right\|_1.$$  

In other words the map $f \mapsto f\mu$ induces an isometric embedding of the space $(L^1(\mu), \| \cdot \|_1)$ into the space $(\mathcal{M}(S, \mathcal{F}), \| \cdot \|)$. 